



**Cochrane**  
**Library**

Cochrane Database of Systematic Reviews

## Acupuncture for peripheral joint osteoarthritis (Review)

Manheimer E, Cheng K, Linde K, Lao L, Yoo J, Wieland S, van der Windt DAWM, Berman BM, Bouter LM

Manheimer E, Cheng K, Linde K, Lao L, Yoo J, Wieland S, van der Windt DAWM, Berman BM, Bouter LM.  
Acupuncture for peripheral joint osteoarthritis.  
*Cochrane Database of Systematic Reviews* 2010, Issue 1. Art. No.: CD001977.  
DOI: [10.1002/14651858.CD001977.pub2](https://doi.org/10.1002/14651858.CD001977.pub2).

[www.cochranelibrary.com](http://www.cochranelibrary.com)

## TABLE OF CONTENTS

ABSTRACT .....	1
PLAIN LANGUAGE SUMMARY .....	2
SUMMARY OF FINDINGS .....	4
BACKGROUND .....	9
OBJECTIVES .....	10
METHODS .....	10
RESULTS .....	13
Figure 1. ....	15
Figure 2. ....	15
Figure 3. ....	16
Figure 4. ....	20
Figure 5. ....	21
DISCUSSION .....	21
AUTHORS' CONCLUSIONS .....	24
ACKNOWLEDGEMENTS .....	25
REFERENCES .....	26
CHARACTERISTICS OF STUDIES .....	33
DATA AND ANALYSES .....	82
Analysis 1.1. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	82
Analysis 1.2. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	83
Analysis 1.3. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	83
Analysis 1.4. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 4 Pain (26 weeks after baseline). ....	84
Analysis 1.5. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 5 Function (26 weeks after baseline). ....	84
Analysis 1.6. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 6 Total score (26 weeks after baseline). ....	84
Analysis 2.1. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	85
Analysis 2.2. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	85
Analysis 2.3. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	86
Analysis 2.4. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 4 Pain (26 weeks after baseline). ....	86
Analysis 2.5. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 5 Function (26 weeks after baseline). ....	87
Analysis 2.6. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 6 Total score (26 weeks after baseline). ....	87
Analysis 3.1. Comparison 3 Acupuncture vs. sham acupuncture for hip OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	88
Analysis 3.2. Comparison 3 Acupuncture vs. sham acupuncture for hip OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	88
Analysis 4.1. Comparison 4 Acupuncture vs. sham acupuncture for hip OA post-treatment scores post-analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	88
Analysis 4.2. Comparison 4 Acupuncture vs. sham acupuncture for hip OA post-treatment scores post-analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	89
Analysis 5.1. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	89
Analysis 5.2. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	90

Analysis 5.3. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	90
Analysis 5.4. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 4 Pain (26 weeks after baseline). .....	91
Analysis 5.5. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 5 Function (26 weeks after baseline). .....	91
Analysis 5.6. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 6 Total score (26 weeks after baseline). .....	91
Analysis 6.1. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	92
Analysis 6.2. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	93
Analysis 6.3. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	93
Analysis 6.4. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 4 Pain (26 weeks after baseline). .....	94
Analysis 6.5. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 5 Function (26 weeks after baseline). .....	94
Analysis 6.6. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 6 Total score (26 weeks after baseline). .....	94
Analysis 7.1. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	97
Analysis 7.2. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). .	98
Analysis 7.3. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	99
Analysis 7.4. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 4 Pain (26 weeks after baseline). .....	100
Analysis 7.5. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 5 Function (26 weeks after baseline). .....	101
Analysis 7.6. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 6 Total score (26 weeks after baseline). .....	101
Analysis 8.1. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	105
Analysis 8.2. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). .	106
Analysis 8.3. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	107
Analysis 8.4. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 4 Pain (26 weeks after baseline). .....	108
Analysis 8.5. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 5 Function (26 weeks after baseline). .....	108
Analysis 8.6. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 6 Total score (26 weeks after baseline). .....	109
Analysis 9.1. Comparison 9 Acupuncture vs. waiting list or other active treatment controls for hip OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). ....	110
Analysis 9.2. Comparison 9 Acupuncture vs. waiting list or other active treatment controls for hip OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). .	111
Analysis 9.3. Comparison 9 Acupuncture vs. waiting list or other active treatment controls for hip OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	111
Analysis 10.1. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	114

Analysis 10.2. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	115
Analysis 10.3. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	116
Analysis 10.4. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 4 Pain (26 weeks after baseline). .....	117
Analysis 10.5. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 5 Function (26 weeks after baseline). .....	117
Analysis 10.6. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 6 Total score (26 weeks after baseline). .....	118
Analysis 11.1. Comparison 11 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	120
Analysis 11.2. Comparison 11 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	121
Analysis 11.3. Comparison 11 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization). .....	122
ADDITIONAL TABLES .....	123
WHAT'S NEW .....	137
CONTRIBUTIONS OF AUTHORS .....	137
DECLARATIONS OF INTEREST .....	137
SOURCES OF SUPPORT .....	137
DIFFERENCES BETWEEN PROTOCOL AND REVIEW .....	137
INDEX TERMS .....	139

[Intervention Review]

# Acupuncture for peripheral joint osteoarthritis

Eric Manheimer<sup>1</sup>, Ke Cheng<sup>2</sup>, Klaus Linde<sup>3</sup>, Lixing Lao<sup>4</sup>, Junghee Yoo<sup>5</sup>, Susan Wieland<sup>6</sup>, Daniëlle AWM van der Windt<sup>7</sup>, Brian M Berman<sup>1</sup>, Lex M Bouter<sup>8</sup>

<sup>1</sup>Center for Integrative Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA. <sup>2</sup>Acupuncture-Moxibustion and Tuina, Shanghai University of TCM, Shanghai Research Center of Acupuncture and Meridians, Shanghai, China. <sup>3</sup>Institut für Allgemeinmedizin / Institute of General Practice, Technische Universität München / Klinikum rechts der Isar, München, Germany. <sup>4</sup>Complementary Medicine Program, University of Maryland School of Medicine, Baltimore, Maryland, USA. <sup>5</sup>College of Oriental Medicine, Kyunghee University, Seoul, Korea, South. <sup>6</sup>Providence, Rhode Island, USA. <sup>7</sup>Department of Primary Care & Health Sciences, Keele University, Keele, UK. <sup>8</sup>Executive Board of VU University Amsterdam, Amsterdam, Netherlands

**Contact:** Eric Manheimer, Center for Integrative Medicine, University of Maryland School of Medicine, Kernan Hospital Mansion, 2200 Kernan Drive, Baltimore, Maryland, 21207-6697, USA. [emanheimer@compmed.umm.edu](mailto:emanheimer@compmed.umm.edu).

**Editorial group:** Cochrane Musculoskeletal Group.

**Publication status and date:** New, published in Issue 1, 2010.

**Citation:** Manheimer E, Cheng K, Linde K, Lao L, Yoo J, Wieland S, van der Windt DAWM, Berman BM, Bouter LM. Acupuncture for peripheral joint osteoarthritis. *Cochrane Database of Systematic Reviews* 2010, Issue 1. Art. No.: CD001977. DOI: [10.1002/14651858.CD001977.pub2](https://doi.org/10.1002/14651858.CD001977.pub2).

Copyright © 2010 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

## ABSTRACT

### Background

Peripheral joint osteoarthritis is a major cause of pain and functional limitation. Few treatments are safe and effective.

### Objectives

To assess the effects of acupuncture for treating peripheral joint osteoarthritis.

### Search methods

We searched the Cochrane Central Register of Controlled Trials (*The Cochrane Library* 2008, Issue 1), MEDLINE, and EMBASE (both through December 2007), and scanned reference lists of articles.

### Selection criteria

Randomized controlled trials (RCTs) comparing needle acupuncture with a sham, another active treatment, or a waiting list control group in people with osteoarthritis of the knee, hip, or hand.

### Data collection and analysis

Two authors independently assessed trial quality and extracted data. We contacted study authors for additional information. We calculated standardized mean differences using the differences in improvements between groups.

### Main results

Sixteen trials involving 3498 people were included. Twelve of the RCTs included only people with OA of the knee, 3 only OA of the hip, and 1 a mix of people with OA of the hip and/or knee. In comparison with a sham control, acupuncture showed statistically significant, short-term improvements in osteoarthritis pain (standardized mean difference -0.28, 95% confidence interval -0.45 to -0.11; 0.9 point greater improvement than sham on 20 point scale; absolute percent change 4.59%; relative percent change 10.32%; 9 trials; 1835 participants) and function (-0.28, -0.46 to -0.09; 2.7 point greater improvement on 68 point scale; absolute percent change 3.97%; relative percent change 8.63%); however, these pooled short-term benefits did not meet our predefined thresholds for clinical relevance (i.e. 1.3 points for pain;

3.57 points for function) and there was substantial statistical heterogeneity. Additionally, restriction to sham-controlled trials using shams judged most likely to adequately blind participants to treatment assignment (which were also the same shams judged most likely to have physiological activity), reduced heterogeneity and resulted in pooled short-term benefits of acupuncture that were smaller and non-significant. In comparison with sham acupuncture at the six-month follow-up, acupuncture showed borderline statistically significant, clinically irrelevant improvements in osteoarthritis pain (-0.10, -0.21 to 0.01; 0.4 point greater improvement than sham on 20 point scale; absolute percent change 1.81%; relative percent change 4.06%; 4 trials; 1399 participants) and function (-0.11, -0.22 to 0.00; 1.2 point greater improvement than sham on 68 point scale; absolute percent change 1.79%; relative percent change 3.89%). In a secondary analysis versus a waiting list control, acupuncture was associated with statistically significant, clinically relevant short-term improvements in osteoarthritis pain (-0.96, -1.19 to -0.72; 14.5 point greater improvement than sham on 100 point scale; absolute percent change 14.5%; relative percent change 29.14%; 4 trials; 884 participants) and function (-0.89, -1.18 to -0.60; 13.0 point greater improvement than sham on 100 point scale; absolute percent change 13.0%; relative percent change 25.21%). In the head-on comparisons of acupuncture with the 'supervised osteoarthritis education' and the 'physician consultation' control groups, acupuncture was associated with clinically relevant short- and long-term improvements in pain and function. In the head on comparisons of acupuncture with 'home exercises/advice leaflet' and 'supervised exercise', acupuncture was associated with similar treatment effects as the controls. Acupuncture as an adjuvant to an exercise based physiotherapy program did not result in any greater improvements than the exercise program alone. Information on safety was reported in only 8 trials and even in these trials there was limited reporting and heterogeneous methods.

### Authors' conclusions

Sham-controlled trials show statistically significant benefits; however, these benefits are small, do not meet our pre-defined thresholds for clinical relevance, and are probably due at least partially to placebo effects from incomplete blinding. Waiting list-controlled trials of acupuncture for peripheral joint osteoarthritis suggest statistically significant and clinically relevant benefits, much of which may be due to expectation or placebo effects.

## PLAIN LANGUAGE SUMMARY

### Acupuncture for osteoarthritis

This summary of a Cochrane review presents what we know from research about the effect of acupuncture on osteoarthritis.

#### The review shows that in people with osteoarthritis,

- Acupuncture may lead to small improvements in pain and physical function after 8 weeks.
- Acupuncture may lead to small improvements in pain and physical function after 26 weeks.

We often do not have precise information about side effects and complications. This is particularly true for rare but serious side effects. Possible side effects of acupuncture treatment include minor bruising and bleeding at the site of needle insertion.

#### What is osteoarthritis and what is acupuncture?

Osteoarthritis (OA) is a disease of the joints, such as your knee or hip. When the joint loses cartilage, the bone grows to try and repair the damage. Instead of making things better, however, the bone grows abnormally and makes things worse. For example, the bone can become misshapen and make the joint painful and unstable. This can affect your physical function or ability to use your knee.

According to the philosophy of traditional acupuncture, energy circulates in 'meridians' located throughout the body. Pain or ill health happens when something occurs to cause this meridian energy circulation to be blocked. The way to restore health is to stimulate the appropriate combination of acupuncture points in the body by inserting very thin needles. Sometimes in painful conditions, electrical stimulation along with the needles is also used. According to acupuncture theory, one way you can tell that acupuncture is relieving pain is that you may feel numbness or tingling, called *de qi*, where the needle is inserted.

#### Best estimate of what happens to people with osteoarthritis who have acupuncture:

##### Pain after 8 weeks:

- People who had acupuncture rated their pain to be improved by about 4 points on a scale of 0 to 20.
- People who received sham acupuncture rated their pain to be improved by about 3 points on a scale of 0 to 20.
- People who received acupuncture had a 1 point greater improvement on a scale of 0-20. (5% absolute improvement).

##### Pain after 26 weeks:

- People who had acupuncture rated their pain to be improved by slightly more than 3 points on a scale of 0 to 20.
- People who received sham acupuncture rated their pain to be improved by slightly less than 3 points on a scale of 0 to 20.
- People who received acupuncture had under a 1 point greater improvement on a scale of 0-20. (2% absolute improvement).

### Acupuncture for peripheral joint osteoarthritis (Review)

**Physical function after 8 weeks :**

- People who had acupuncture rated their function to be improved by about 11 points on a scale of 0 to 68.
- People who received sham acupuncture rated their function to be improved by about 8 points on a scale of 0 to 68.
- People who received acupuncture had about a 3 point greater improvement on a scale of 0-68. (4% absolute improvement)

**Physical function after 26 weeks :**

- People who had acupuncture rated their function to be improved by about 11 points on a scale of 0 to 68.
- People who received sham acupuncture rated their function to be improved by about 10 points on a scale of 0 to 68.
- People who received acupuncture had about a 1 point greater improvement on a scale of 0-68. (2% absolute improvement)

## SUMMARY OF FINDINGS

### Summary of findings for the main comparison.

#### Acupuncture compared with sham acupuncture for peripheral joint osteoarthritis

**Patient or population:** Patients with peripheral joint osteoarthritis

**Settings:**

**Intervention:** Acupuncture

**Comparison:** Sham acupuncture

Outcomes	Illustrative comparative risks* (95% CI)		Relative per- cent change	No of Parti- pants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control	Acupuncture				
<p><b>Pain (short term)</b></p> <p>WOMAC scale from 0 to 20 points (higher is worse pain).</p> <p>Follow up: 8 weeks</p>	<p>The mean pain (short term) in the control groups was</p> <p><b>-2.66 points</b><sup>1</sup></p>	<p>The mean pain (short term) in the intervention groups was</p> <p><b>0.92 lower</b></p> <p>(1.48 to 0.36 lower)<sup>2</sup></p>	-10.32% <sup>3</sup>	1835 (9 studies)	++OO low(4,5)	<p>SMD -0.28</p> <p>(-0.45 to -0.11)</p> <p>Absolute percent difference:</p> <p>-4.59% (0.92 point lower on a 0-20 point scale)<sup>6</sup></p>
<p><b>Function (short term)</b></p> <p>WOMAC scale from 0 to 68 points (higher is worse function).</p> <p>Follow up: 8 weeks</p>	<p>The mean function (short term) in the control groups was</p> <p><b>-7.86 points</b><sup>1</sup></p>	<p>The mean function (short term) in the intervention groups was</p> <p><b>2.70 lower</b></p> <p>(4.44 to 0.87 lower)<sup>2</sup></p>	-8.63% <sup>3</sup>	1767 (8 studies)	++OO low(4)	<p>SMD -0.28</p> <p>(-0.46 to -0.09)</p> <p>Absolute percent difference:</p> <p>-3.97% (2.70 points lower on a 0-68 point scale)<sup>6</sup></p>
<p><b>Pain (long term)</b></p>	<p>The mean pain (long term) in the control groups was</p>	<p>The mean pain (long term) in the intervention groups was</p>	-4.06% <sup>3</sup>	1399 (4 studies)	++++ high	<p>SMD -0.10</p> <p>(-0.21 to 0.01)</p>



WOMAC scale from 0 to 20 points (higher is worse pain). Follow up: 26 weeks	<b>-2.92 points</b> <sup>1</sup>	<b>0.36 lower</b> (0.75 lower to 0.04 higher) <sup>2</sup>				Absolute percent difference: -1.81% (0.36 point lower on a 0-20 point scale) <sup>6</sup>
<b>Function (long term)</b> WOMAC scale from 0 to 68 points (higher is worse function). Follow up: 26 weeks	The mean function (long term) in the control groups was <b>-9.94 points</b> <sup>1</sup>	The mean function (long term) in the intervention groups was <b>1.21 lower</b> (2.43 lower to 0 higher) <sup>2</sup>	-3.89% <sup>3</sup>	1398 (4 studies)	++++ high	SMD -0.11 (-0.22 to 0)  Absolute percent difference: -1.79% (1.22 points lower on a 0-68 point scale) <sup>6</sup>
<b>Adverse events</b>	See comment	See comment	Not estimable	-	See comment	Eight RCTs described adverse events across groups, and they found that the frequency of adverse events was similar between the acupuncture and control groups. The frequency of adverse events in the acupuncture group ranged from 0% (Sangdee 2002) to 7% (Berman 2004). Pooling of adverse events across these RCTs was not possible because of limited reporting and heterogeneous methods. No serious adverse events were reported to be associated with acupuncture.
<b>Side effects of acupuncture-Bruising and bleeding at injection site</b>	See comment	See comment	Not estimable	-	See comment	The frequency of minor side effects of acupuncture, primarily minor bruising and bleeding at needle insertion sites, ranged from 0% (Foster 2007) to 45% (Sangdee 2002). These frequencies varied widely because of heterogeneous and scanty reporting and different definitions of what constitutes a side effect of acupuncture versus what is an inherent part of treatment (for example, occasional bruising at needle insertion site).

\*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval

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.

<sup>1</sup> The representative trial selected for calculating the percent changes from baseline was the [Berman 2004](#) trial because this trial was sufficiently large, and because the patient characteristics and the baseline mean and SD of the control group for this trial was most similar to, and thus most representative of, the other trials.

<sup>2</sup> We calculated the main difference by choosing the [Berman 2004](#) trial as a representative study, and then calculating the difference by multiplying the SMD by the SD (standard deviation) of the mean change in the control group in this study.

<sup>3</sup> We calculated the relative percent change by multiplying the SMD by the standard deviation of change in the control group of the [Berman 2004](#) trial, dividing the result by the baseline mean in the control group of the [Berman 2004](#) trial, and multiplying by 100 to obtain the percent.

<sup>4</sup> We could not be certain that the shams used in three of the sham-controlled trials ([Sangdee 2002\\*](#); [Vas 2004](#); [Berman 2004](#)) were sufficiently credible in fully blinding participants to the treatment being evaluated.

<sup>5</sup> There was statistically significant heterogeneity of effect estimates between the two substrata for the following four variables for the pain outcome: success of blinding (Yes/Not sure); likely physiological activity of sham control (Yes/No); use of electrical stimulation of needles (Yes/No); and adequate number of acupuncture sessions (Yes/No).

<sup>6</sup> We calculated the absolute percent change by multiplying the SMD by the standard deviation of change in the control group of the [Berman 2004](#) trial, dividing the result by the number of units in the scale, and multiplying by 100 to obtain the percent.

<sup>7</sup> There was statistically significant heterogeneity of effect estimates between the trials ( $I^2 = 73\%$ ).

## Summary of findings 2.

### Acupuncture compared with sham acupuncture for knee osteoarthritis

**Patient or population:** Patients with knee osteoarthritis

**Settings:**

**Intervention:** Acupuncture

**Comparison:** Sham acupuncture

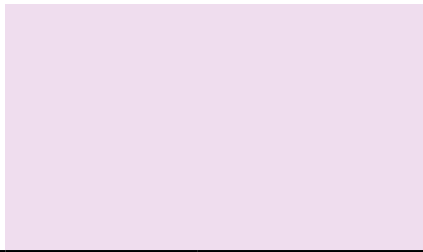
Outcomes	Illustrative comparative risks* (95% CI)		Relative percent change	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control	Acupuncture				
<b>Pain (short term)</b> WOMAC scale from 0 to 20 points (higher is worse pain). Follow up: 8 weeks	The mean pain (short term) in the control groups was <b>-2.66 points</b> <sup>1</sup>	The mean pain (short term) in the intervention groups was <b>0.96 lower</b>	-10.69% <sup>3</sup>	1773 (8 studies)	++OO low <sup>(4,5)</sup>	SMD -0.29 (-0.48 to -0.1) Absolute percent difference: -4.76% (0.95 point lower on a 0-20 point scale) <sup>6</sup>



		(1.57 to 0.33 lower) <sup>2</sup>				
<b>Function (short term)</b> WOMAC scale from 0 to 68 points (higher is worse function). Follow up: 8 weeks	The mean function (short term) in the control groups was <b>-7.86 points</b> <sup>1</sup>	The mean function (short term) in the intervention groups was <b>2.80 lower</b> (4.73 to 0.77 lower) <sup>2</sup>	-8.93% <sup>3</sup>	1767 (8 studies)	++OO low(4,7)	SMD -0.29 (-0.49 to -0.08)  Absolute percent difference: -4.11% (2.80 points lower on a 0-68 point scale) <sup>6</sup>
<b>Pain (long term)</b> WOMAC scale from 0 to 20 points (higher is worse pain). Follow up: 26 weeks	The mean pain (long term) in the control groups was <b>-2.92 points</b> <sup>1</sup>	The mean pain (long term) in the intervention groups was <b>0.36 lower</b> (0.76 lower to 0.04 higher) <sup>2</sup>	-4.06% <sup>3</sup>	1399 (4 studies)	++++ high	SMD -0.10 (-0.21 to 0.01)  Absolute percent difference: -1.81% (0.36 point lower on a 0-20 point scale) <sup>6</sup>
<b>Function (long term)</b> WOMAC scale from 0 to 68 points (higher is worse function). Follow up: 26 weeks	The mean function (long term) in the control groups was <b>-9.94 points</b> <sup>1</sup>	The mean function (long term) in the intervention groups was <b>1.22 lower</b> (2.44 lower to 0 higher) <sup>2</sup>	-3.89% <sup>3</sup>	1398 (4 studies)	++++ high	SMD -0.11 (-0.22 to 0)  Absolute percent difference: -1.79% (1.22 points lower on a 0-68 point scale) <sup>6</sup>
<b>Adverse events</b>	See comment	See comment	Not estimable	-	See comment	Seven RCTs described adverse events across groups, and they found that the frequency of adverse events was similar between the acupuncture and control groups. The frequency of adverse events in the acupuncture group ranged from 0% (Sangdee 2002) to 7% (Berman 2004). Pooling of adverse events across these RCTs was not possible because of limited reporting and heterogeneous methods. No serious adverse events were reported to be associated with acupuncture.
<b>Side effects of acupuncture-Bruis-</b>	See comment	See comment	Not estimable	-	See comment	The frequency of minor side effects of acupuncture, primarily minor bruising and bleeding at

needle insertion sites, ranged from 0% (Foster 2007) to 45% (Sangdee 2002). These frequencies varied widely because of heterogeneous and scanty reporting and different definitions of what constitutes a side effect of acupuncture versus what is an inherent part of treatment (for example, occasional bruising at needle insertion site).

**ing and bleeding at injection site**



\*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk Ratio; [other abbreviations, eg. OR, etc]

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.

1 The representative trial selected for calculating the percent changes from baseline was the [Berman 2004](#) trial because this trial was sufficiently large, and because the patient characteristics and the baseline mean and standard deviation of the control group for this trial was most similar to, and thus most representative of, the other trials.

2 We calculated the main difference by choosing the [Berman 2004](#) trial as a representative study, and then calculating the difference by multiplying the SMD by the standard deviation of the mean change in the control group in this study.

3 We calculated the relative percent change by multiplying the SMD by the standard deviation of change in the control group of the [Berman 2004](#) trial, dividing the result by the baseline mean in the control group of the [Berman 2004](#) trial, and multiplying by 100 to obtain the percent.

4 We could not be certain that the shams used in three of the sham-controlled trials ([Sangdee 2002\\*](#); [Vas 2004](#); [Berman 2004](#)) were sufficiently credible in fully blinding participants to the treatment being evaluated.

5 There was statistically significant heterogeneity of effect estimates between the trials ( $I^2 = 68\%$ ).

6 We calculated the absolute percent change by multiplying the SMD by the standard deviation of change in the control group of the [Berman 2004](#) trial, dividing the result by the number of units in the scale, and multiplying by 100 to obtain the percent.

7 There was statistically significant heterogeneity of effect estimates between the trials ( $I^2 = 73\%$ ).

## BACKGROUND

Osteoarthritis (OA), the most common form of arthritis (Lawrence 1998; Felson 2000), is the leading cause of disability among older adults (MMWR 2001; Peat 2001). Non-steroidal anti-inflammatory drugs (NSAIDs) and acetaminophen are the most commonly used pharmacological agents for treating osteoarthritis (Ausiello 2002; Wegman 2004). However, according to a recent systematic review, NSAIDs are only slightly better than placebo in providing short-term pain relief, and their effects are probably too small to be meaningful to people with OA (Bjordal 2004). Acetaminophen is often considered a safer alternative to NSAIDs (Wegman 2004), and a recent Cochrane review (Towheed 2006) indeed found that acetaminophen was associated with fewer adverse effects than NSAIDs in short-term randomized controlled trials (RCTs) (Zhang 2008). However, this review also found that acetaminophen is modestly less effective than NSAIDs (Towheed 2006), and that the clinical significance of acetaminophen is questionable (Towheed 2006; Zhang 2008), as it results in only a 5% larger improvement from baseline in pain compared to placebo in the short-term (Towheed 2006). In addition, acetaminophen may be the leading cause of acute liver failure in the United States (Larson 2005).

The most recent evidence-based treatment guidelines from the UK National Institutes of Clinical Excellence (NICE Guideline 2008) and the Osteoarthritis Research International (Zhang 2008) suggest that OA treatment should be multidisciplinary, with non-pharmacological treatments such as education, aerobic and resistance exercises, and weight loss as the "cornerstone" (NICE Guideline 2008) or "initial focus" (Zhang 2008) of patient management, and with consideration also given to pharmacological options such as acetaminophen when further treatment is required. In a recent systematic review of OA guidelines (Zhang 2008), five of the eight guidelines that considered acupuncture recommended it as an osteoarthritis treatment modality. A very recent, authoritative, guideline for knee or hip OA (Zhang 2008) was developed by an international, multidisciplinary group of experts using a Delphi process. With this approach, consensus recommendations among the committee members were developed by systematically reviewing and critically appraising both meta-analyses of efficacy as well as existing guidelines. This approach was described as "evidence-driven and clinically supported". This guideline committee recommended acupuncture as one of 12 possible non-pharmacological modalities for treating osteoarthritis; however, this recommendation achieved only a 69% consensus among the guideline committee members.

Many people with OA seek out complementary and alternative medicine (CAM) therapies (Rao 1999; Herman 2004; Quandt 2005). For example, according to a recent US national survey (Quandt 2005), 41% of a nationally representative sample of people with arthritis have used a CAM therapy in the past year. Another recent survey found that 67% of people with OA in primary care clinics were currently using at least one type of CAM therapy (Herman 2004), with glucosamine and chondroitin being by far the most commonly used therapies (Herman 2004). Though glucosamine is used by 25% of people with OA (Herman 2004), the evidence to support its use is inconsistent (Towheed 2005; Zhang 2008), with some recent large trials showing no benefit over a placebo (Cibere 2004; McAlindon 2004; Clegg 2006; Rozendaal 2008). Chondroitin is used by 18% of people with OA (Herman 2004), but a recent systematic review of large, methodologically sound

trials concluded that the symptomatic benefit is "minimal or non-existent" (Reichenbach 2007). All other CAM therapies for treating OA are used far less frequently than glucosamine and chondroitin (Herman 2004). For example, acupuncture is used by only about 1% (Quandt 2005) to 2% (Herman 2004) of people with arthritis. Even among people with OA who do use acupuncture, most do not use it specifically for treating their OA (Quandt 2005).

According to the philosophy of traditional acupuncture, energy circulates in 'meridians' located throughout the body. When something occurs to cause this meridian energy circulation to be blocked, pain or ill health will result. The way to restore energy circulation, health, and balance, is to stimulate the appropriate combination of the estimated 400 traditional meridian acupuncture points in the body (WHO 1991). Additional non-meridian tender points may also be used, and electrical stimulation of the points is also common in modern acupuncture, especially for pain-related conditions. According to acupuncture theory, one indication that acupuncture is exerting its analgesic effects is that a patient may experience a sensation of numbness or tingling, called *de qi*, at the needle insertion point.

Laboratory evidence has documented a biological basis of acupuncture analgesia. For example, animal studies provide evidence that acupuncture may simply be a particular method of stimulating the nervous system to release a range of neurotransmitters - particularly opioid peptides - which are involved in the body's own pain-suppressing mechanisms (Lewith 1984). Other research suggests that the noxious stimulation of the acupuncture needles may act to suppress the nervous system pathways that are involved in the sensory and affective components of pain (Han 2003; Hui 2005). Because pain often limits a patient's activity, decreased pain may improve function. Also, basic science studies show that acupuncture suppresses inflammation (Li 2008); any decreases in inflammation may improve physical function.

Acupuncture has been demonstrated to be a safe therapy with a very low risk of serious side effects (MacPherson 2001; White 2001; Cherkin 2003; Lao 2003; Melchart 2004). A systematic review of 12 prospective studies which surveyed more than one million acupuncture treatments found that the risk of a serious adverse event from acupuncture is estimated to be 0.05 per 10 000 treatments, and 0.55 per 10 000 individual patients. Treatment guidelines (Am Coll Rheum 2000; Jordan 2003; Wegman 2004) emphasize that treatment safety is an important consideration, especially because people with knee OA are often older, use concurrent medications, and have co-morbidities. Given its safety, the question of whether or not acupuncture is effective for treating OA of the knee therefore is highly relevant.

In everyday practice, acupuncture treatment is often individualized, continually modified to take into account changes in the patient's condition, and combined with other treatments, such as herbal medicine and mind-body exercises. In contrast, in RCTs of acupuncture, a prescribed formula of acupuncture points is generally evaluated as a sole treatment. While evaluating acupuncture as a sole, isolated treatment may not reflect everyday clinical practice, it does allow for the best estimation of the specific effects of acupuncture.

Three recent systematic reviews of RCTs have evaluated the effects of acupuncture on OA. Two of these reviews included only

participants with knee OA (Manheimer 2007; White 2007), while the third (Kwon 2006) included participants with OA of any peripheral joint. Each of these reviews included a meta-analysis which showed that acupuncture was statistically significantly superior to sham acupuncture in the short-term. This review is a substantial update of the most recently published earlier review (Manheimer 2007). This current review has been largely rewritten from the earlier version (Manheimer 2007), and includes RCTs with OA of all peripheral joints, not just the knee, as well as two large, recent, knee OA RCTs not included in any previous systematic review.

## OBJECTIVES

The objective of this review was to compare the effects of traditional needle acupuncture with a sham, another active treatment, or with a waiting list control, for people with OA of the knee, hip, or hand.

## METHODS

### Criteria for considering studies for this review

#### Types of studies

We included RCTs in any language. We included only RCTs with at least six weeks of observation because trials with a shorter duration were considered irrelevant for the question of whether acupuncture is helpful for people with a chronic disease like OA. In principle, a longer trial duration would seem even more desirable; however, given the limited number of trials available we considered a minimum duration of six weeks a reasonable compromise.

#### Types of participants

We included only studies that concerned exclusively participants with osteoarthritis of one or more of the peripheral joints (i.e. knee, hip, and hand). Studies including participants with only OA of the spine were not included. Studies that included a mix of participants with OA of the spine and OA of the peripheral joints were included only if the results for the participants with OA of the peripheral joints were reported separately from the results of the participants with OA of the spine.

#### Types of interventions

We included only studies evaluating traditional acupuncture. Traditional acupuncture involves inserting needles into traditional meridian points, usually with the intention of influencing energy flow in the meridian. In traditional acupuncture, needles may also be inserted at additional tender points and electrical stimulation of the needles may be used. We excluded trials of dry needling/trigger point therapy, a therapy which rejects traditional concepts of energy and meridians, and which involves inserting needles only at unnamed tender or trigger points to stimulate nerves or muscles. We also excluded RCTs of laser acupuncture and electro-acupuncture without needle insertion because most authorities believe acupuncture involves needle insertion (Birch 2001).

The control interventions were a sham intervention, a waiting list, and another active treatment. We also included trials that compared acupuncture plus another active treatment versus that other active treatment alone. Thus, we included all pragmatic trials that compared acupuncture with any other treatments (e.g. exercise, education, medication, etc.). Because our objective was to evaluate the effects of acupuncture compared to non-acupuncture

controls, we excluded RCTs in which one form of acupuncture was compared only with another form.

#### Types of outcome measures

At least one of the following outcome measures had to be reported: pain, function, or symptom severity.

We compiled a table of any excluded studies about which it was plausible to expect that a reader would question why the study was not included. This includes all studies that appeared on first glance to meet the selection criteria but did not, as well as studies that did not meet all of the criteria but are well known, in the same general area as the review, and may be thought relevant by some readers (Higgins 2008 (Section 7)).

#### Search methods for identification of studies

To identify RCTs, we searched the Cochrane Central Register of Controlled Trials (CENTRAL, *The Cochrane Library* 2008, Issue 1), MEDLINE (through December 2007), and EMBASE (through December 2007) (see Additional Table 1). For our MEDLINE and CENTRAL searches, we searched the following terms as both free-text terms as well as MeSH terms (except where indicated): (acupuncture; acupuncture therapy; auriculotherapy (free text only); electroacupuncture; moxibustion; Medicine, Oriental Traditional; Medicine, Chinese Traditional) AND (arthritis; osteoarthritis; arthralgia; joint diseases; joint pain (free-text only); chronic joint symptoms (free-text only); gonarthrosis (free-text only); osteoarthritis (free-text only); ostoarthrosis (free-text only); degenerative arthritis (free-text only)). We combined this search strategy with a methods filter for clinical trials (Glanville 2006). For our EMBASE search, we used a modified version of the MEDLINE strategy (see Additional Table 1).

All RCTs included in previous systematic reviews of acupuncture for OA (Ernst 1997; Ezzo 2001; Kwon 2006; White 2007; Manheimer 2007) were also reconsidered for inclusion in this review. We scanned bibliographies of retrieved articles for further references. Finally, we also searched databases of ongoing trials to identify details of trials that may be relevant for future updates of this review.

Two authors (EM with either KC or KL) independently considered articles for inclusion, with disagreements resolved by discussion.

#### Data collection and analysis

##### Data extraction

One author (EM) extracted the data for all trials except the one German language trial (Molsberger 1994). A second author (either KC or KL) completed an independent second extraction for all trials: KC completed the second extraction for five trials (Fink 2001; Haslam 2001; Stener-Victorin 2004; Foster 2007; Williamson 2007) and KL completed the extractions for the remaining 11 trials. Consensus was generally achieved by discussion; in the few instances where disagreements persisted, a third author made the final decision on the extraction of the data item. We independently extracted information pertaining to quality of the methods, participants, acupuncture and control interventions, and treatment outcomes (including adverse effects). We e-mailed all RCT corresponding authors and requested that they review the information about their RCT extracted into our Characteristics of included studies table, as well as review our quality assessments of

their RCT. When data reported in RCT publications was incomplete or ambiguous, we requested additional information or clarification from the corresponding authors.

We extracted the outcomes of pain, function, and symptom severity for all time points reported. When a given study reported more than one pain, function or symptom severity measure, we gave preference to the WOMAC pain, function, and total (i.e. sum of WOMAC pain, function, and stiffness) measures, respectively, because the WOMAC has been extensively and repeatedly validated in its original English version (Bellamy 1988; Angst 2005; Bellamy 2005) and also in its Spanish (Escobar 2002) and German (Stucki 1996) adaptations, each of which was used in one of the trials in our meta-analyses (Vas 2004; Scharf 2006). In addition, the WOMAC is the most comparable between studies because most trials in this review measured outcomes using the WOMAC scale.

The method of selecting acupuncture points was categorized as individual, fixed formula, or flexible formula. For the individual style, the practitioner is free to choose any points. For the fixed formula, the same fixed points are used for all participants. For the flexible formula, a fixed formula is used and some additional points are chosen according to the symptoms or tenderness of the patient.

For cross-over trials, we included only the data before the cross-over occurred because we considered the risk for carryover effects to be prohibitive.

### Risk of bias assessment

For the risk of bias assessment, we used the new tool recommended by the Cochrane Reviewer's Handbook. As recommended, we used the following six separate criteria:

- Adequate sequence generation
- Allocation concealment
- Blinding
- Incomplete outcome data addressed (up to 3 months after randomization)
- Incomplete follow-up outcome data addressed (4 to 12 months after randomization)
- Free of selective reporting

For the "other potential threats to validity" item on the risk of bias scale, we assessed the following items:

groups similar at baseline regarding the most important prognostic indicators; co-interventions avoided or similar; compliance acceptable in all groups; timing of the outcome assessment in all groups similar; intention-to-treat analysis.

As a first step in evaluating risk of bias, we copied information relevant for making a judgment on a criterion from the original publication into a table. If available, we also entered any additional information from the study authors into this table. Two reviewers independently made a judgment whether the risk of bias for each criterion was considered low, high or unclear. A third reviewer arbitrated any disagreements.

For the blinding item on the risk of bias scale, we assigned sham-controlled trials as "Unclear" rather than "Yes" because we could not be certain that all shams were sufficiently credible in fully blinding participants to the treatment being evaluated.

However, we assigned the "Yes" score to sham-controlled trials that either 1) evaluated the credibility of the sham and found the sham to be indistinguishable from true acupuncture, or 2) used needle acupuncture as the sham *and also* informed participants that two different types of acupuncture were being compared (i.e., did not inform participants that a sham treatment was involved). Some trials had both blinded sham control groups and unblinded comparison groups (i.e. waiting list or other active treatment control). In the risk of bias tables the column judgment always relates to the comparison with sham interventions. In the column description we also include the assessment for the other comparison group. As the risk of bias table does not include a "not applicable" option, the item "incomplete follow-up outcome data addressed (time point greater than three months and closest to six months after randomization)" was rated as "unclear" for trials which did not follow patients longer than three months.

For osteoarthritis trials, investigators typically measure a number of outcomes at multiple time points using various outcome measurement instruments. For the selective reporting item, we considered those trials as having a low risk of bias if they reported the results of the most relevant osteoarthritis outcomes measured (typically a measure of pain and function using the WOMAC scale) for the most relevant time points (end of treatment and, if done, follow-up), and if these results made it unlikely that authors had picked them out because they were particularly favorable or unfavorable. Trials which met all criteria, or all but one criterion, were considered to be of higher quality.

For ranking the strength and quality of the evidence for a given comparison, we used the GRADE and Summary of Findings tables recommended by The Cochrane Collaboration.

### Quantitative data synthesis

We placed studies into one of four categories according to which of the following comparisons were evaluated:

- 1) Acupuncture versus a sham intervention;
- 2) Acupuncture versus a waiting list;
- 3) Acupuncture versus another active treatment;
- 4) Acupuncture plus another active treatment versus that other active treatment alone.

Trials using different active treatment comparators were analyzed separately.

Trials of acupuncture for different peripheral joints were each analyzed in a separate joint-specific meta-analysis. An additional analysis that evaluated acupuncture for OA of any peripheral joint combined trials of OA of the knee with trials of OA of other peripheral joints (i.e. hip or hand). If any trials included both knee OA and other peripheral joint OA participants and reported the outcomes separately for the participants with OA of the knee and participants with OA of the other peripheral joints, the separate outcomes from the knee, hip, or hand joint participants were included in the relevant joint-specific analysis, while the results from all included participants were included in the peripheral joint OA analysis.

The outcomes of the review were the standardized mean differences of acupuncture, as compared with each comparison group, on pain, function, and symptom severity, at both the short- and longer-term follow-up time points. For our meta-analyses,

we defined the short-term outcome as the measurement point closest to eight weeks, and less than or equal to three months, following randomization. We defined the longer-term outcome as the measurement point closest to six months, and more than three months, following randomization.

Standardized mean differences were calculated using the differences in improvements between groups. We used standardized mean differences as the principal measure of effect size because the trials assessed the same outcomes but measured them in various ways (e.g., WOMAC VAS and Likert scales).

For the five RCTs that used a waiting list control group (Christensen 1992; Berman 1999; Tukmachi 2004; Witt 2005; Witt 2006), we excluded all outcome measurements after participants on the waiting list began acupuncture.

We pooled data using the random-effects model to account for expected heterogeneity. To evaluate heterogeneity within our categories of trials, we used  $I^2$  tests on all outcomes meta-analyzed. If there was "considerable heterogeneity", which is defined by the Cochrane Handbook for Systematic Reviews of Interventions as an  $I^2$  value between 75% and 100% (Higgins 2008 (Section 9.5.2)), the data were not pooled; otherwise, data were pooled. However, when studies showed "substantial heterogeneity" (i.e.  $I^2 \geq 50\%$ , (Higgins 2008 (Section 9.5.2))), the pooled results were interpreted with caution. We also tried to determine the cause of statistically heterogeneous study results, using subgroup analyses (Ioannidis 2008), as described below. We also conducted a sensitivity analysis using a fixed-effect analysis, to assess the robustness of our findings. Although the random-effects analysis is the preferred approach because of the large heterogeneity, we also conducted a fixed-effect as a sensitivity analysis because the larger studies are more valid and thus these studies will remain most influential in a fixed-effect analysis.

To allow for a more clinically relevant interpretation, we also evaluated whether the pooled effects of acupuncture met the threshold for minimal clinically important differences, defined as the smallest differences in scores that people with OA would perceive to be beneficial (Angst 2002). The clinically relevant effects for knee osteoarthritis have been estimated to be standardized mean differences of 0.39 for WOMAC pain and 0.37 for WOMAC function (Angst 2002).

To further aid clinical interpretation, we also converted the pooled standardized mean differences to the percent changes from baseline of acupuncture relative to the comparison control, using the approach recommended by the Cochrane Musculoskeletal Group. For these calculations, we first calculated the absolute change by multiplying the pooled standardized mean difference by the standard deviation of the control group of the trial that has a large weighting in the overall result in RevMan, and is most representative, in terms of the patient characteristics and the baseline mean and standard deviation of the control group. This absolute change was then divided by the baseline value of this trial's control group to calculate a relative percent change from baseline. If there is "considerable or substantial heterogeneity" (Higgins 2008 (Section 9.5.2)), it may be misleading to quote an average clinical value for the intervention's effect, and in such cases, we did not compute pooled absolute and relative percent changes.

If any studies within any category reported insufficient data for pooling even after asking authors for more details, we excluded such studies from the meta-analyses and described their results narratively.

If at least six studies were available for a meta-analysis, we assessed the likelihood of small study bias by constructing funnel plots (Sutton 2000).

### Acupuncture adequacy assessments

Two acupuncturists (LL, Marcos Hsu), who have a combined acupuncture clinical experience of nearly forty years in treating knee OA, and who have both previously worked on RCTs and systematic reviews of acupuncture, independently assessed the adequacy of the acupuncture administered in the trials. Consensus was achieved by discussion. The adequacy of four aspects of the acupuncture (Furlan 2005) were assessed: choice of acupuncture points; number of sessions; needling technique; and acupuncturist's experience. The adequacy of the sham intervention was also assessed, using an open-ended question. The acupuncturists' assessments of adequacy were based only on the description of the study population and the acupuncture procedure; the assessors were blinded to the results of the study and the publication (author and journal).

The assessors had previously used this adequacy assessment instrument for the earlier systematic review (Manheimer 2007), of which this is an update. However, the acupuncturists decided that their previous assessments may have overestimated the adequacy of the acupuncture in some of the trials, and therefore, for this update, the acupuncturists assessed again the 11 trials previously assessed for the earlier version and also assessed the five new trials. For this update, the acupuncturists decided that the trials needed to include both an adequate number of treatments and also an adequate overall duration of treatment to earn a score of adequate on the 'number of sessions' item, whereas in the earlier review, only the number of treatments was considered, without regard for whether the duration of treatment (i.e. number of weeks of treatment) was of an adequate length. We asked the acupuncturists to guess the identity of each study being assessed to test the success of the blinding. The results of the tests of blinding to the results of the studies that are reported in Additional Table 2 were based on the guesses of the studies made during the second adequacy assessment.

### Subgroup and sensitivity analyses

Of the 11 elements of the Cochrane Back Review Group scale (see Table 3, Table 4) we conducted subgroup analyses only on the elements that relate specifically to randomization, blinding, and follow-up because we believed that these elements would be most likely to affect the effect of acupuncture for osteoarthritis. These six elements were the following: generation of random sequence; allocation concealment; baseline comparability; blinding success; acceptability of drop-out rate; and intention-to-treat analysis. For these subgroup tests, which were conducted among the sham-controlled trials only, which by design were all intended to have participants blinded, we used blinding success rather than patient blinding as the variable for the testing. We conducted another subgroup analysis on whether or not the sham used in the trial was likely to have analgesic/physiological activity, according to the acupuncturists' assessments. We also conducted a subgroup analysis on high ( $\geq 6$  out of 11) versus low quality trials,



according to the Cochrane Back Review Group scale. Additionally, we conducted subgroup tests on the following 7 clinical variables: the 4 variables related to acupuncture adequacy (see section 'Acupuncture adequacy assessments' above), 1 variable related to the method of selecting acupuncture points (i.e. fixed formula or flexible formula), 1 variable related to electrical stimulation of acupuncture needles (Y/N), and 1 variable related to length of follow-up (i.e. greater than or equal to versus less than 3 months from randomization).

For these subgroup analyses, we used a significance test, as described by Deeks et al (Deeks 2001; Deeks 2005) to investigate whether differences in effects of acupuncture between any two subgroups for any variable were statistically significant for either the pain or function outcome. For these subgroup analyses, we subgrouped all peripheral joint OA sham-controlled trials. Among the 15 potential variables for subgroup analyses described above, we conducted a statistical test for interaction only if there were more than one study in each subgroup, as is required (Deeks 2001). For all subgroup comparisons, the P value for the interaction test was reported, as were the pooled estimates and I<sup>2</sup> values from each of the two relevant subgroups (Higgins 2008 (Section 9.5.2)).

We conducted two sensitivity analyses. First, for the acupuncture versus sham comparison, for which the effect sizes were small and therefore potentially sensitive to use of different analytic methods, we calculated the standardized mean differences using both changes from baseline and post-treatment values. Namely, we used a comparison of means and standard deviations (SDs) of changes from baseline for each group for the primary analysis and we used a comparison of means and SDs of post-treatment scores for each group for a sensitivity analysis (Higgins 2008 (Section 9.4.5.2)). We conducted separate analyses using both changes from baseline and post treatment scores in order to test whether the statistical significance of the pooled standardized mean differences differed depending on the sets of values used. We recorded any assumptions or imputations necessary to calculate either SDs of changes from baseline or SDs of post-treatment scores.

Second, when both crude values and adjusted values were reported in a publication, we used crude values preferentially for the primary analysis, and we used adjusted changes from baseline, when reported, for an additional sensitivity analysis, for the acupuncture versus sham comparison only.

## RESULTS

### Description of studies

#### Excluded studies

Among the 20 excluded studies (for details see [Characteristics of excluded studies](#)), 6 had observation periods less than 6 weeks (Ammer 1988; Petrou 1988; Dickens 1989; McIndoe 1994; Yurtkuran 1999; Ng 2003), 6 included people with OA of the neck rather than the peripheral joints (Coan 1982; Loy 1983; Petrie 1983; Petrie 1986; Lundeberg 1991; Thomas 1991); 1 included people with OA of a variety of joints, including the spine, and the results for the people with OA of the peripheral joints were not reported separately from the results of the people with OA of the spine (Gaw 1975), 3 were not explicitly randomized (Junnila 1982; Zhrebkin 1998; Tillu 2002), 2 compared two types of active acupuncture (Kwon 2001; Tillu 2001),

1 paper was not obtainable (Milligan 1981), and 1 used trigger point therapy, not Chinese style acupuncture (Weiner 2007).

#### Included studies

##### **General study characteristics: Numbers of studies and participants; languages of publication; peripheral joints included; unpublished information obtained from authors:**

We included 16 RCTs (Christensen 1992; Molsberger 1994; Takeda 1994; Berman 1999; Fink 2001; Haslam 2001; Sangdee 2002 \*; Berman 2004; Stener-Victorin 2004; Tukmachi 2004; Vas 2004; Witt 2005; Scharf 2006; Witt 2006; Foster 2007; Williamson 2007) representing 3498 randomized OA participants. Twelve of the RCTs included only people with OA of the knee, 3 included only people with OA of the hip (Fink 2001; Haslam 2001; Stener-Victorin 2004), and 1 included a mix of people with OA of the hip and/or OA of the knee (Witt 2006). All studies but 1 (Molsberger 1994) were published in English. We obtained unpublished data from 9 authors, including both outcome data (Berman 2004; Vas 2004; Witt 2005; Scharf 2006; Foster 2007; Williamson 2007) and methodological information (Christensen 1992; Takeda 1994; Berman 1999; Vas 2004; Foster 2007).

##### **Characteristics of participants and details on outcome measures**

The [Characteristics of included studies](#) table shows the most important characteristics of the 16 included RCTs. All RCTs included primarily older participants, with a mean age of 60 or greater, and a mean duration of osteoarthritis knee pain of 5 years or more. For all RCTs, participants needed to be diagnosed with OA to be eligible, and all but two RCTs (Sangdee 2002 \*; Foster 2007) required radiological evidence of OA. For all but two RCTs (Fink 2001; Stener-Victorin 2004), the WOMAC instrument was used to measure outcomes. (One other RCT (Haslam 2001) used a modified version of the WOMAC.) No RCTs reported that the OA diagnosis was made according to the principles of traditional Chinese medicine.

##### **Method of selecting acupuncture points and stimulating needles**

Five RCTs (Stener-Victorin 2004; Witt 2005; Scharf 2006; Foster 2007; Williamson 2007) used a flexible formula for point selection, and ten RCTs (Christensen 1992; Molsberger 1994; Takeda 1994; Berman 1999; Fink 2001; Haslam 2001; Sangdee 2002 \*; Berman 2004; Tukmachi 2004; Vas 2004) used a set formula. For the one remaining pragmatic trial (Witt 2006), the point selection and needling technique were entirely at the discretion of the treating physicians. Superficial needling alone was used in one trial (Sangdee 2002 \*), whereas 13 trials (Christensen 1992; Molsberger 1994; Takeda 1994; Berman 1999; Fink 2001; Berman 2004; Stener-Victorin 2004; Tukmachi 2004; Vas 2004; Witt 2005; Scharf 2006; Foster 2007; Williamson 2007) used sufficiently deep needle stimulation to elicit the de qi needling sensation. (Two trials did not report on de qi: the pragmatic trial (Witt 2006) and one other small trial (Haslam 2001).) Electrical stimulation of the needles was used in five trials (Berman 1999; Sangdee 2002 \*; Berman 2004; Stener-Victorin 2004; Vas 2004), and for three of these trials (Sangdee 2002 \*; Stener-Victorin 2004; Vas 2004), all needles were electrically stimulated.

##### **Assumptions used to derive the meta-analysis study data**

For five of the RCTs (Takeda 1994; Berman 1999; Fink 2001; Haslam 2001; Tukmachi 2004) included in the meta-analyses, there

was incomplete reporting of SDs of change, and we needed to make some conservative assumptions to calculate standardized mean differences (SMDs) for the changes from baseline analysis. Namely, for the [Berman 1999](#) and [Tukmachi 2004](#) trials, estimated SDs of changes from baseline were calculated from the reported within group change P values extracted from publications, using conservative assumptions when exact P values were not reported. For these calculations, the reported P values and the mean changes in each group were used to calculate the standard errors of change for each group, which were then converted to SDs of change for each group. For another trial ([Haslam 2001](#)), the P value of the difference in changes between the two groups was used to calculate the standardized mean difference and standard error, which were then entered into RevMan using the generic inverse variance. (The generic inverse variance was used because this trial ([Haslam 2001](#)) was the only study that compared acupuncture versus supervised exercise for hip OA so could not be pooled with any other studies in a meta-analysis.) For two trials ([Takeda 1994](#); [Fink 2001](#)), there were no reported SDs of changes or any statistics (e.g. P values) that would allow us to directly calculate these, for any outcomes. For these two trials, we therefore used the pre- and post-treatment means and SDs for each group, and assumed a conservative within-subject pretest-post-test correlation of 0.5, to calculate the SDs of change for each group. For the [Sangdee 2002 \\*](#) study we did not need to make any assumptions to calculate SMDs for the changes from baseline analysis; however, for the post-treatment scores analysis, we needed to use baseline SDs as estimates for post-treatment SDs, because post-treatment SDs were not reported.

For the seven remaining RCTs included in our meta-analyses ([Berman 2004](#); [Vas 2004](#); [Witt 2005](#); [Scharf 2006](#); [Witt 2006](#); [Foster 2007](#); [Williamson 2007](#)), we obtained from the investigators the unadjusted means and SDs necessary to calculate SMDs as both differences in improvements and differences in post-treatment scores, without needing to make any assumptions. The publications for three ([Berman 2004](#); [Witt 2005](#); [Scharf 2006](#)) of these seven RCTs also reported means and SDs of change from an adjusted changes from baseline analysis. Namely, [Berman 2004](#) reported adjusted results (as P values in their [Table 4](#) ([Berman 2004](#))), from a mixed model analysis, with adjustment for baseline values of outcome variable, within-participant correlation, and clinical site as a random effect; [Witt 2005](#) reported adjusted results, for the 8 week outcome only (as means and standard errors in their [Table 1](#) ([Witt 2005](#))), from an ANCOVA analysis which adjusted for baseline values of the outcome variable; and [Scharf 2006](#) reported adjusted results (as means and standard errors in their [Table 1](#) and [Appendix Table 4](#) ([Scharf 2006](#))), based on a linear mixed-effects model, which adjusted for number of affected knees and regional cluster. As noted above, for an additional sensitivity analysis, we calculated SMDs of change for the acupuncture versus sham comparison, substituting the adjusted values reported in these 3 publications for the unadjusted values used in our primary analysis.

### Risk of bias in included studies

Among the ten RCTs that included a sham control (counting the [Sangdee 2002 \\*](#) twice because it included two different comparisons for our meta-analysis, as described below, in the sub-section 'Results: effects of interventions: acupuncture versus sham'), we considered the five ([Berman 2004](#); [Vas 2004](#); [Witt 2005](#); [Scharf 2006](#); [Foster 2007](#)) with the highest quality ratings on the Cochrane Back Review Group scale ([van Tulder 2003](#))

(see Additional [Table 4](#)) to comprise the bulk of the evidence for this review. These five RCTs have been published in leading international medical journals, all since 2004. Four of the five also included a waiting list or other active treatment control group ([Berman 2004](#); [Witt 2005](#); [Scharf 2006](#); [Foster 2007](#)). Only two of the five had any obvious methodological flaws, which in both cases were due to higher dropout rates, in the sham group for one ([Vas 2004](#)) and in the education control group for another ([Berman 2004](#)). Four of these trials ([Berman 2004](#); [Witt 2005](#); [Scharf 2006](#); [Foster 2007](#)) had a six month outcome assessment, but for only one ([Berman 2004](#)) was a treatment schedule maintained up until the final six month measurement point.

For all sham-controlled RCTs, the schedule for the sham acupuncture procedure was the same as that for the true acupuncture procedure. We could not be certain that the shams used in three of the sham-controlled trials were sufficiently credible in fully blinding participants to the treatment being evaluated ([Sangdee 2002 \\*](#); [Berman 2004](#); [Vas 2004](#)). For all waiting list-controlled trials, participants on the waiting lists were allowed to receive the current level of oral NSAID or analgesic therapy.

### Assessments of acupuncture adequacy

All of the trials included in this review were judged adequate on "Choice of acupoints" and "Needling technique", but only two of the trials ([Berman 1999](#); [Berman 2004](#)) were judged adequate in terms of the acupuncturist's experience. For five of the trials ([Christensen 1992](#); [Takeda 1994](#); [Haslam 2001](#); [Foster 2007](#); [Williamson 2007](#)), the number of acupuncture sessions was judged inadequate. Also, for six trials ([Molsberger 1994](#); [Takeda 1994](#); [Fink 2001](#); [Witt 2005](#); [Scharf 2006](#); [Foster 2007](#)), the acupuncture adequacy assessors noted that the sham needling may have had physiologic activity (see Additional [Table 2](#)).

### Funding sources

Five RCTs did not report funding sources ([Christensen 1992](#); [Molsberger 1994](#); [Takeda 1994](#); [Haslam 2001](#); [Tukmachi 2004](#)). Five ([Berman 1999](#); [Berman 2004](#); [Stener-Victorin 2004](#); [Vas 2004](#); [Foster 2007](#)) were funded by government grants, 1 ([Sangdee 2002 \\*](#)) by a university, 1 ([Williamson 2007](#)) by a hospital, 1 ([Fink 2001](#)) by a pharmaceutical company, and 3 ([Witt 2005](#); [Scharf 2006](#); [Witt 2006](#)) by social health insurance companies.

### Effects of interventions

See: [Summary of findings for the main comparison](#); [Summary of findings 2](#)

Of the sixteen RCTs that met the selection criteria, all except three ([Christensen 1992](#); [Molsberger 1994](#); [Stener-Victorin 2004](#)) reported extractable outcome data. Three trials included people with only hip OA ([Fink 2001](#); [Haslam 2001](#); [Stener-Victorin 2004](#)), but we did not meta-analyze these three trials together in a hip-specific analysis because there was heterogeneity of controls and outcome measures, and the outcomes were poorly reported or nonstandard. For knee OA, eleven trials reported extractable outcome data.

### Acupuncture versus sham

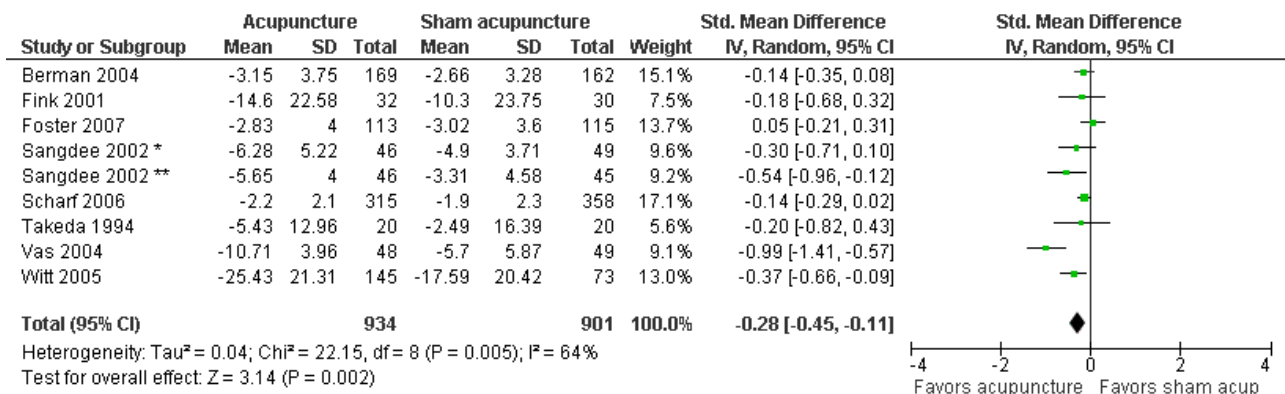
Ten trials included a sham control, nine in people with knee OA and one in people with hip OA. Data for all but one ([Molsberger 1994](#)) sham-controlled knee OA trial could be used for our quantitative analysis.

One sham-controlled trial (Sangdee 2002 \*; Sangdee 2002 \*\*) randomized participants to the following four groups: 'placebo tablet plus sham electroacupuncture', 'diclofenac tablet plus sham electroacupuncture', 'placebo tablet plus true electroacupuncture', and 'diclofenac tablet plus true electroacupuncture'. Because our meta-analysis addressed the question of whether acupuncture is more effective than sham, we included in our meta-analysis the following two comparisons from this trial as if they were from different studies, as recommended (Higgins 2008 (Section 16.5.4)): 'acupuncture versus sham with placebo tablet co-intervention' (Sangdee 2002 \*\*) and 'acupuncture versus sham with diclofenac tablet co-intervention' (Sangdee 2002 \*). With this approach, intervention groups from this trial were only entered once in the meta-analysis.

In comparison with a sham control at the short-term follow-up, acupuncture showed improvements in OA pain (standardized mean difference -0.28, 95% confidence interval -0.45 to -0.11; 9 trials; 1835 participants;  $I^2 = 64%$ ) (see Figure 1 directly below), function (-0.28, -0.46 to -0.09; 9 trials; 1829 participants;  $I^2 = 69%$ ) (Analysis 5.2) and symptom severity (-0.29, -0.50 to -0.09; 9 trials; 1767 participants;  $I^2 = 74%$ ) (Analysis 5.3), but the results were heterogeneous. The

range of SMDs for the pain outcome was from -0.99 in the trial showing the greatest benefit to +0.05 in the trial showing no benefit (see Figure 1). This corresponds to an absolute and relative percent improvement relative to a sham control of -29.06% and -48.03% in the trial showing the greatest benefit of acupuncture (Vas 2004) to 0.90% and 2.02% in the trial showing no benefit of acupuncture (Foster 2007). Two out of the nine trials showed effect estimates higher than the pre-defined clinical relevance thresholds of 0.37 and 0.39 for pain and function respectively, but the pooled estimates were lower than the threshold. Results of the short-term meta-analyses of acupuncture versus sham for peripheral joint OA described above were unchanged when we restricted to only the knee-only trials (i.e. when we removed the Fink 2001 study, the one relevant hip osteoarthritis sham-controlled trial with extractable short-term data) (Analysis 1.1; Analysis 1.2; Analysis 1.3). In comparison with a sham acupuncture control at the six-month follow-up, acupuncture showed borderline statistically significant, clinically irrelevant improvements in knee OA pain (-0.10, -0.21 to 0.01; 4 trials; 1399 participants,  $I^2 = 0%$ ) (see Figure 2 below), function (-0.11, -0.22 to 0.00, 4 trials; 1398 participants;  $I^2 = 6%$ ) (Analysis 5.5) and symptom severity (-0.11, -0.22 to 0.00, 4 trials; 1398 participants;  $I^2 = 2%$ ) (Analysis 5.6), and there was low heterogeneity.

**Figure 1. Effects of acupuncture versus a sham control group on the pain outcome at the short-term measurement point \*Comparison of electroacupuncture with sham acupuncture using a diclofenac co-intervention. \*\*Comparison of electroacupuncture with sham acupuncture using a placebo diclofenac co-intervention.**



**Figure 2. Effects of acupuncture versus a sham acupuncture control group on the pain outcome at the long-term measurement point**

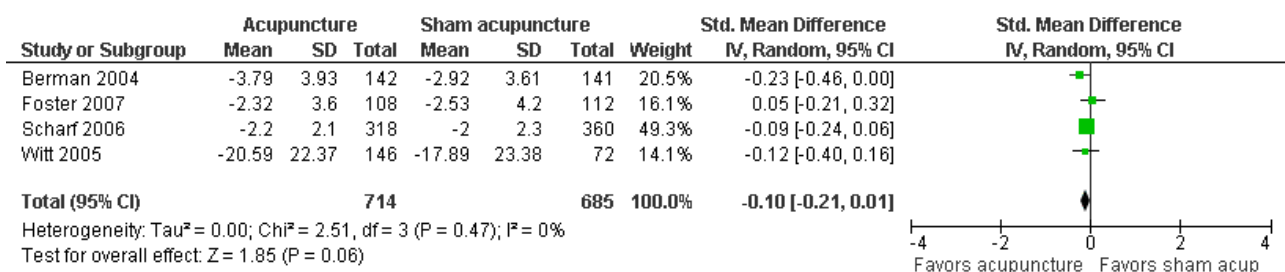


Figure 1; Figure 2

**Acupuncture versus waiting list**

In comparison with a waiting list control, acupuncture was associated with clinically relevant short-term improvements in

OA pain (-0.96, -1.19 to -0.72; 4 trials; 884 participants;  $I^2 = 41%$ ) (see [Figure 3](#) below), function (-0.89, -1.18 to -0.60; 3 trials; 864 participants;  $I^2 = 64%$ ) ([Analysis 10.2](#)), and symptom severity (-0.92, -1.16 to -0.67; 3 trials; 864 participants;  $I^2 = 52%$ ) ([Analysis 10.3](#)). The pooled effect estimate for the pain outcome corresponds to an absolute and relative percent improvement of

-14.54% and -29.14% relative to a waiting list. There was moderate heterogeneity ([Higgins 2008 \(Section 9.5.2\)](#)) but the benefits of acupuncture in each individual trial, as well as the pooled benefits, were much larger than our predefined thresholds for clinical relevance.

**Figure 3. Effects of acupuncture compared with a waiting list or other active treatment control group at the short-term measurement point**

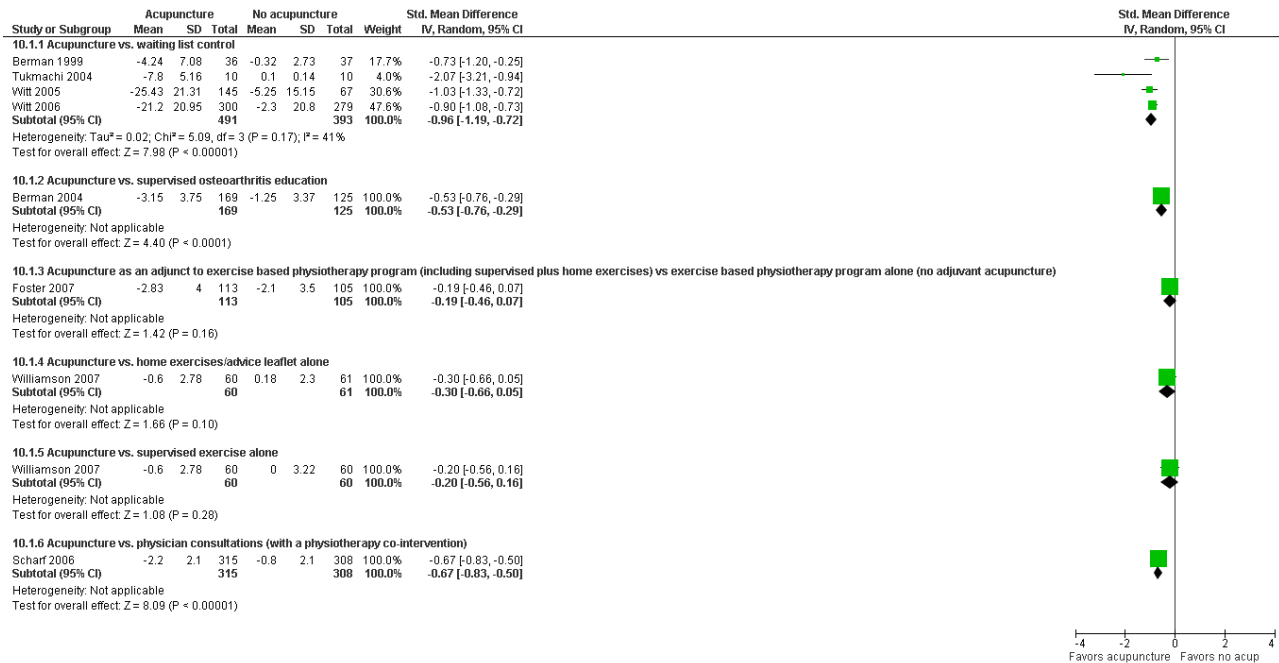


Figure 3

**Acupuncture versus other active treatments**

The effects of acupuncture were heterogeneous across trials that involved 'head-on' comparisons of acupuncture versus different active comparator controls (see [Figure 3](#) above, and [Analysis 10.1](#) to [Analysis 10.6](#)). These head-on comparisons each included only a single trial so no meta-analysis could be performed. In the head-on comparisons of acupuncture with the 'supervised osteoarthritis education' control ([Berman 2004](#)) and the 'physician consultations (with a physiotherapy co-intervention)' control ([Scharf 2006](#)), acupuncture was associated with short- and long-term improvements in pain and function. In both cases, the benefits of acupuncture exceeded our thresholds for clinical relevance. In the head-on comparisons of acupuncture with the 'home exercises/advice leaflet alone' ([Williamson 2007](#)), and 'supervised exercise alone' ([Williamson 2007](#)) controls, there was also evidence that acupuncture was associated with similar treatment effects as the controls.

**Acupuncture plus another active treatment versus other active treatment alone**

The [Foster 2007](#) trial include 3 treatment arms: an exercise based physiotherapy program (including supervised plus home exercises), exercise based physiotherapy program plus true acupuncture, and exercise based physiotherapy program plus non-penetrating acupuncture. For this trial ([Foster 2007](#)), which

was the only trial with extractable outcome data that evaluated acupuncture as strictly an adjuvant to existing care, acupuncture as an adjuvant to an exercise based physiotherapy program (including supervised plus home exercises) did not result in any greater improvements than the exercise based physiotherapy program alone (see [Figure 3](#) and [Analysis 10.1](#) to [Analysis 10.6](#)).

**Trials not included in meta-analyses**

Three RCTs ([Christensen 1992](#); [Molsberger 1994](#); [Stener-Victorin 2004](#)), one each using a sham, waiting list, and active treatment control, did not have extractable outcome data either because their results were reported in a way that could not be entered in the meta-analyses or because their results were difficult to interpret. The sham-controlled trial ([Molsberger 1994](#)) found a statistically significant reduction in pain but no improvement in function, at both the end of treatment, and three months later. The waiting list controlled trial ([Christensen 1992](#)) found a statistically significant reduction in pain, analgesic consumption, and most objective measures in the acupuncture group relative to the waiting list group. The active intervention controlled trial ([Stener-Victorin 2004](#)) which compared patient education alone, electroacupuncture and hydrotherapy (the last two were evaluated as adjuvant treatments to patient education) found a statistically significant reduction in pain and improvement in function for both the electroacupuncture and hydrotherapy groups relative to the patient education alone group.

The three hip OA RCTs (Fink 2001; Haslam 2001; Stener-Victorin 2004) each had outcome data collected during the follow-up time relevant for our predefined long-term time point analysis. However, because for each of these trials, the attrition rate was so high (almost 50% for at least one of the comparison groups), a post hoc decision was made to exclude these follow-up data from our meta-analyses. However, these outcomes are briefly described here. Namely, in the Fink 2001 study, only 41 out of 67 randomized participants (and only 17 out of 33 randomized to the acupuncture group) were available at the six month follow-up point, at which time there were no differences in pain or function between the acupuncture and sham groups. In the Haslam 2001 study, the dropout rate in the control group was very high (almost 44%) at the 14 week follow-up point, at which time there was a "significant improvement in group A [acupuncture] compared with group B [supervised exercise]  $P = .03$ ." In the Stener-Victorin 2004 study, only 25 out of 45 randomized were available for follow-up at the 6 month point, at which time there were no differences between the groups on the pain outcome but greater improvements on the function outcome in the group that received adjuvant electroacupuncture compared with the group that received patient education alone.

### Safety of acupuncture

Eight RCTs (Fink 2001; Haslam 2001; Sangdee 2002 \*; Berman 2004; Witt 2005; Scharf 2006; Foster 2007; Williamson 2007) described adverse events across groups, and they found that the frequency of adverse events was similar between the acupuncture and control groups (Table 5). Pooling of adverse events across these RCTs was not possible because of limited reporting and heterogeneous methods. No serious adverse events were reported to be associated with acupuncture. The frequency of minor side effects of acupuncture, primarily minor bruising and bleeding at needle insertion sites, ranged from 0% (36) to 45% (44). These frequencies varied widely because of heterogeneous and scanty reporting and different definitions of what constitutes a side effect of acupuncture versus what is an inherent part of treatment (for example, occasional bruising at needle insertion site).

### Subgroup and sensitivity analyses

We planned subgroup analyses for 15 variables, but only seven had sufficient data available for a statistical test for interaction. Of the seven subgroup analyses on clinical and methodological variables, there was statistically significant heterogeneity of effect estimates between the two substrata for the following four variables for the pain outcome: success of blinding (Yes/Not sure); likely physiological activity of sham control (Yes/No); use of electrical stimulation of needles (Yes/No); and adequate number of acupuncture sessions (Yes/No) (see Table directly below, which is the same as Additional Table 6). The function outcome subgroup meta-analyses results were similar (see Additional Table 7). The six trials that were judged to have been successfully blinded were the same six trials for which the acupuncture adequacy assessors noted that the sham needling may have had physiologic effects. When restricting to these successfully blinded trials with potentially physiologically active shams, the pooled results were smaller and only borderline statistically significant for the pain outcome and no longer statistically significant for the function outcome. Also, dividing trials into the two subgroups defined by this variable of blinding/sham type substantially reduced the overall heterogeneity (see Table below and Additional Table 6), suggesting that the type of sham used and the likelihood of

whether or not it successfully blinds participants to treatment assignment may explain some of the overall heterogeneity of the sham controlled trials. For the 'sufficient number of sessions delivered over an adequate treatment duration (yes/no)' and 'electrical stimulation (yes/no)' criteria, the pooled estimates were statistically significant only in the subgroups that met either of the criteria (see Table below and Additional Table 7), but neither of these variables explained the heterogeneity.

**Table.** Results of the subgroup meta-analyses for the pain outcome\*

Variable	Total RCTs, n	Pts randomly assigned, n	Effect size (95% CI)	I <sup>2</sup> , %	P value for interaction†	Not met‡
<b>All trials</b>	9	1835	-0.28 (-0.45,-0.11)	63.9	-	-
<b>Methodological variables</b>						
<b>Generation of random sequence</b>					0.141	SS
Adequate	7	1649	-0.25 (-0.45,-0.05)	69		
Unclear or no	2	186	-0.42 (-0.71,-0.13)	0		
<b>Allocation concealment</b>					0.215	SSFi
Adequate	6	1587	-0.26 (-0.48,-0.04)	74.2		
Unclear	3	248	-0.36 (-0.61,-0.11)	0		
<b>Blinding success</b>					0.042	BSSV
Yes	5	1221	-0.15 (-0.28,-0.01)	14.5		
Uncertain	4	614	-0.47 (-0.84,-0.10)	77.5		
<b>Intention-to-treat analysis</b>					0.532	FiFoSST
Yes	4	1319	-0.35 (-0.63,-0.07)	80.7		
Unclear or no	5	516	-0.20 (-0.43,0.03)	35.2		
<b>Clinical variables</b>					0.047	FoT
<b>Sufficient number of sessions delivered over an adequate treatment duration</b>						
Yes	7	1567	-0.34 (-0.54, -0.15)	66.1		
No	2	268	0.01 (-0.23, 0.25)	0		

<b>Electrical stimulation was used with the acupuncture</b>					0.042	BSSV met
Yes	4	614	-0.50 (-0.81,-0.20)	66		
No	5	1215	-0.11 (-0.29,0.07)	42.3		
<b>Physiological activity of sham, as judged by acupuncturist</b>					0.042	BSSV
Likely	5	1221	-0.15 (-0.28,-0.01)	14.5		
Not likely	4	614	-0.47 (-0.84,-0.10)	77.5		
<b>Formula versus flexible formula for point selection</b>					0.057	BFISSV used formula
Formula	6	716	-0.39 (-0.66,-0.12)	72.8		
Flexible formula	3	1119	-0.14 (-0.34,0.05)	77		

\* Subgroup analyses for the sham controlled peripheral joint OA trials, based on changes of pain scores from baseline at the short-term (i.e. closest to 8 wks) follow up. The structure and organization of this table was largely modeled after a similar table in a recent review (Reichenbach 2007).

†A statistical test for interaction could not be performed for the following prespecified subgroups because there were one or fewer studies in one of the strata of each of these subgroups: comparability of baseline, acceptability of drop-out rate, CBRG score  $\geq 6$ , follow-up length  $\geq 3$  months, choice of acupoints, needling technique, acupuncturists' experience.

‡"Not met" column lists the first (or first two) letters of the author of the studies that do not meet the criterion.

Subgroup analyses restricting to only the high quality trials were uninformative because almost all the sham-controlled trials meta-analyzed were classified as high quality according to the Cochrane Back Review Group quality evaluation scale (see Additional Table 4).

Sensitivity analyses showed that the pooled standardized mean differences calculated using differences in post-treatment scores between groups (see Comparisons 2, 4, 6, 8, 9, and 11 for post-treatment scores analyses) were slightly larger than the pooled standardized mean differences calculated using differences in improvements between groups, for the acupuncture versus sham comparison; this was attributed to the slightly better baseline

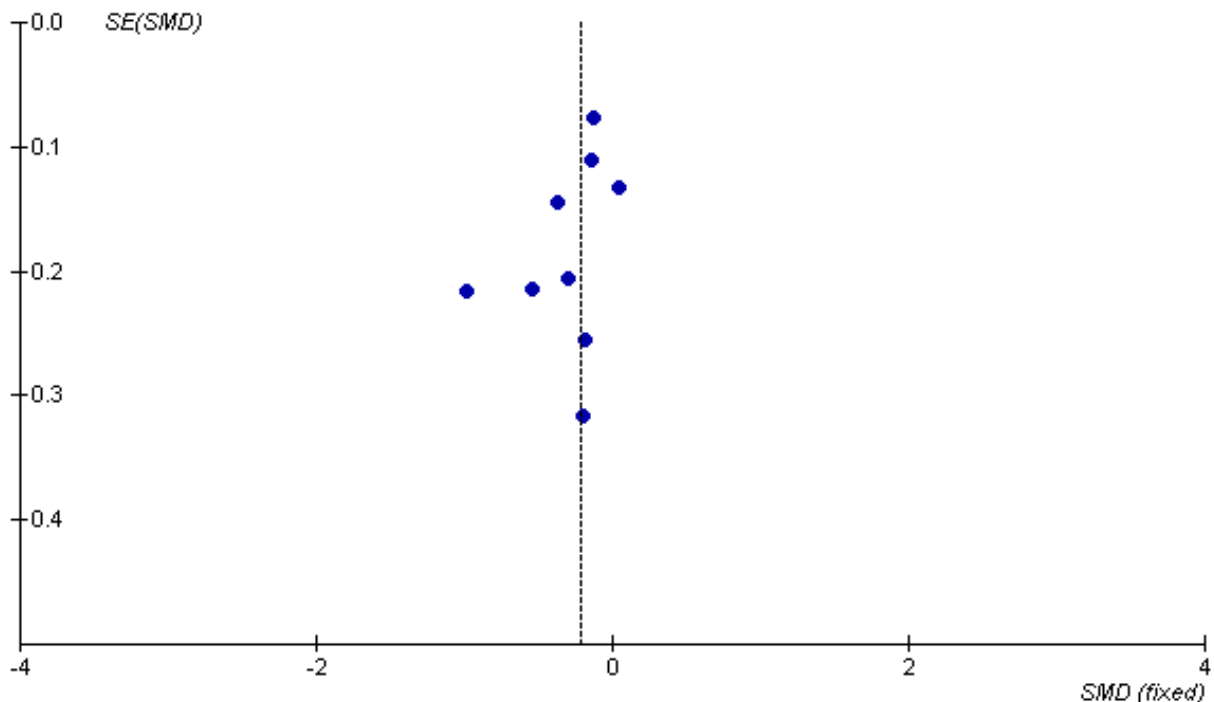
WOMAC scores in the acupuncture groups than the sham groups, for three RCTs (Vas 2004; Witt 2005; Scharf 2006). However, there were no outcomes at any time points for which the statistical significance of the pooled result changed depending on whether post-treatment scores or differences in improvements were used. The pooled standardized mean differences calculated using the fixed-effect analysis were slightly smaller than the pooled standardized mean differences using the random effects analysis, for the acupuncture versus sham short-term comparison. However, the statistical significance of the pooled result did not change depending on whether the random-effects or fixed-effect analyses were used.

Standardized mean differences of change calculated by substituting the adjusted values for the unadjusted values for the three publications (Berman 2004; Witt 2005; Scharf 2006) that reported both resulted in slightly larger results when pooling all sham-controlled trials, for the short- and long-term pain and function outcomes. These slightly larger benefits using pooled adjusted versus unadjusted standardized mean differences was attributed to the fact that, for two of these three RCTs (Berman 2004; Witt 2005) that reported adjusted values, the adjusted standardized mean differences were larger than the corresponding unadjusted standardized mean differences.

The interpretation of our funnel plots (Figure 4; Figure 5) was hindered by the small number of very small trials and the small number of trials overall.

**Figure 4. Funnel plot of acupuncture vs. sham short term change for peripheral OA pain**

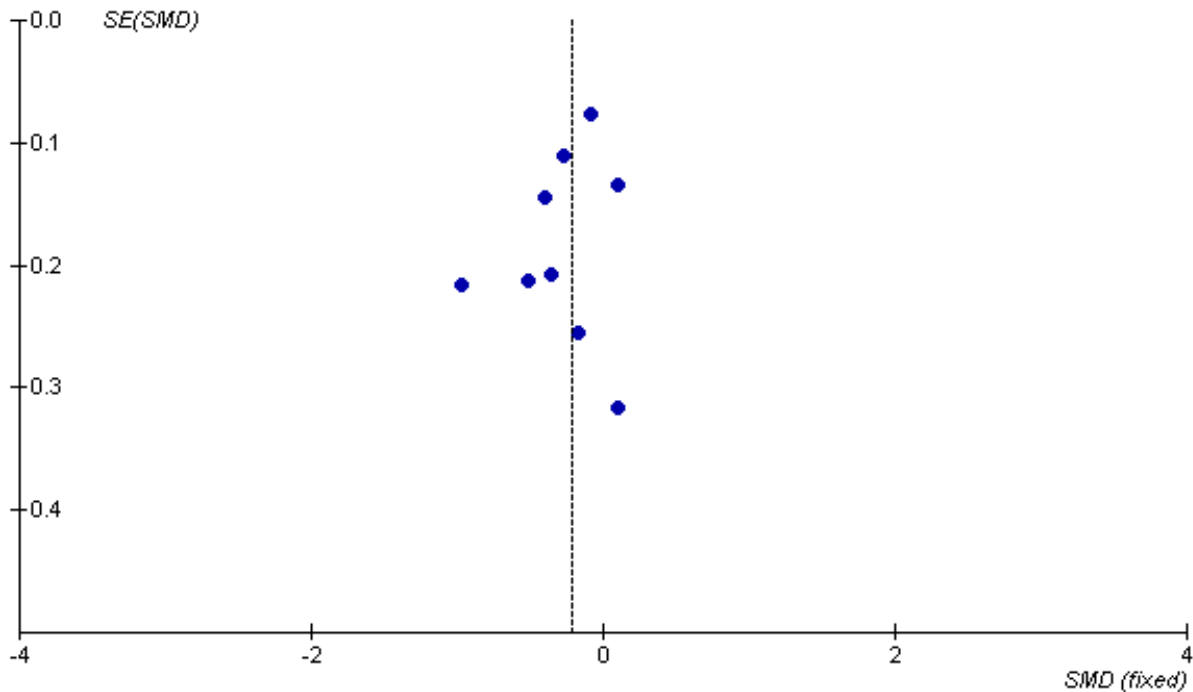
Review: Acupuncture for Osteoarthritis (February 25 13PM 2008)  
 Comparison: 05 Acupuncture vs sham acupuncture for peripheral joint OA change from baseline analysis  
 Outcome: 02 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)





**Figure 5. Funnel plot of acupuncture vs. sham short term change for peripheral OA function**

Review: Acupuncture for Osteoarthritis (February 25 13PM 2008)  
 Comparison: 05 Acupuncture vs sham acupuncture for peripheral joint OA change from baseline analysis  
 Outcome: 04 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)



The only reported adverse events attributable to the acupuncture were slight bruising or bleeding at acupuncture points (see Additional Table 5).

**DISCUSSION**

**Summary of main results**

Sham-controlled trials are designed to minimize placebo effects, and thereby measure the true biological effects of acupuncture. Currently available sham-controlled trials show statistically significant pooled benefits of acupuncture relative to sham; however, these benefits are small and do not meet our pre-defined criteria for clinical relevance. Also, some sham-controlled trials showed no benefit of acupuncture over sham and there was substantial heterogeneity of results, which may be due to differences in sham interventions, differences in acupuncture protocols, differences in settings, and varying proficiencies of the acupuncturists. One trial that evaluated acupuncture as an adjuvant to a standardized exercise/advice program found no additional benefit of acupuncture (Foster 2007). Head to head trials of acupuncture versus other active treatments have had heterogeneous results: two trials showed no statistically significant differences between acupuncture and exercise intervention controls (Williamson 2007 - two different exercise controls used), while two other trials showed a statistically significant benefit of acupuncture compared to the active intervention controls of 'supervised osteoarthritis education' (Berman 2004) and 'physician consultation (with a physiotherapy co-intervention)' (Scharf 2006).

However, these two latter trials are difficult to interpret because both the acupuncture and the sham arms showed benefits over the active treatment controls, suggesting placebo effects may have played a role. Currently available trials of acupuncture for osteoarthritis suggest statistically significant and clinically relevant benefits of acupuncture compared to a waiting list control. We considered patient blinding to be the most critical factor related to the applicability of the evidence and the risk of bias of the included studies. Therefore, the remainder of the Discussion focuses primarily on the design of the included trials, and particularly the different types of controls used, and how this variability of controls may explain the variability in the findings.

**Limitations and challenges of the sham control design**

Using a sham acupuncture comparison is intended to control for patient expectations and placebo effects, and thereby estimate the effects of acupuncture due solely to the point specific placement of the acupuncture needles. However, there are two considerations that must be borne in mind in designing a sham control. First, it is important to design a sham that is physiologically inert. Yet, there are no standards by which to determine which 'sham' point locations, depths or directions of sham needle insertion, or durations of sham needle placement will truly result in no physiological activity, and which are therefore truly inert. However, many authorities agree that in order to be inert, sham needles should not penetrate the skin (Birch 2006). The reason that this is important is because one commonly proposed mechanism suggests that acupuncture (as well as sham acupuncture),

may work through a diffuse noxious inhibitory control (DNIC) mechanism. That is, the DNIC theory suggests that noxious stimuli (e.g. needles) applied to any part of the body can produce analgesic effects, even at distant sites (Le Bars 1979; Lewith 1983). Because inserting needles, even at non-acupoints, may produce an analgesic effect, needle insertive sham acupuncture may not be an appropriate control, and may be better understood as a "poor form of acupuncture treatment" (Leewith 1983)). Indeed, physiological activity of needle penetration, even if superficial and at non-acupuncture points, is suggested by several lines of research (Birch 2006), including RCTs showing larger effects of a superficially-penetrating needle acupuncture than a non-penetrating sham control (Macdonald 1983; Sangdee 2002 \*); positron emission tomography research indicating sham acupuncture can stimulate regions of the brain associated with natural opiate production (Pariente 2005); and animal studies showing sham needle-insertion can have non-specific analgesic effects through the postulated DNIC mechanism (Le Bars 1979). One possible explanation for our subgroup analysis which showed smaller benefits of acupuncture when compared with a sham involving needle penetration (versus a sham with no needle penetration) is that needle penetrating shams may have physiological activity, even if the needles are inserted only superficially. Indeed, superficial needling is a common technique used in many authentic traditional Japanese acupuncture styles (Birch 1998).

The second important consideration in designing an appropriate sham is somewhat in tension with the first: the sham must be sufficiently believable to participants as an authentic treatment to fully allow for the control of placebo effects. Developing a sham that is believable to trial participants and yet that does not involve needles that penetrate the skin is a challenge. One way to increase participants' beliefs in the authenticity of the sham may be to inform participants of the sham and true acupuncture treatment arms in a manner that suggests two different real, active treatments are compared, without mentioning terms such as placebo or sham (Witt 2005). However, this disclosure practice may not be permitted by some internal review boards, and some bioethicists (Miller 2007) have argued that this practice violates the ethical requirements of informed consent.

To test whether a sham was believable to the participants, it has been proposed that investigators should ask participants to guess their treatment assignment and/or guess whether or not they received an authentic treatment or a sham (Fergusson 2004). However, the validity of testing credibility of the sham is unclear and deserves further study because correctly guessing treatment assignment could be highly correlated with a treatment's effectiveness, or lack thereof (Schulz 2002; Altman 2004; Berman 2005; Sackett 2007). That is, if only true acupuncture and not sham has a true physiological benefit, then one might expect that participants assigned to true acupuncture would be more likely to guess that they were receiving an authentic treatment than would participants assigned to the sham acupuncture because only the true acupuncture participants would be experiencing the beneficial effects of the treatment. Or as Altman and colleagues describe it, "end of trial tests of blindness might be tests of hunches for adverse effects or efficacy" (Altman 2004). Indeed, both Altman and colleagues (Altman 2004) as well as David Sackett (Sackett 2004; Sackett 2007) have argued that trial investigators should vigorously test for blinding before trials start, but that tests of blinding at the end of trials can be difficult to interpret, and therefore cannot be

recommended in all circumstances. Nevertheless, in this review, we have considered the results of blinding tests after the trial has begun because acupuncture is a particularly difficult intervention for which to develop a believable sham, and it is possible that the trial participants would continue to try to figure out whether they were receiving the true or sham treatment while the trial is underway. While we have considered the results of such post-treatment blinding tests, we recognize that these results can be uninterpretable.

### Possible explanations for the findings of the sham-controlled trials

For the two sham-controlled RCTs that found clinically relevant benefits of acupuncture (Sangdee 2002 \*; Vas 2004), the credibility of the sham was not tested and the informed consent procedure was not described. In one of these RCTs (Sangdee 2002 \*) the sham involved patch electrodes attached to the same 4 knee points used in the true acupuncture group, with mock electrical stimulation. Because no needles were used, participants randomized to sham probably did not believe they were receiving traditional acupuncture. Nevertheless, these participants may have believed they were receiving an authentic treatment because the non-needle sham used in this trial is comparable to a sham that was shown to be similarly credible to authentic acupuncture in a previous RCT (White 2004). In the other sham-controlled RCT with highly positive results (Vas 2004), some unblinding may have been possible: the sham needles did not penetrate the skin and therefore may have been less believable to participants than a sham that involves needle penetration of the skin. While a non-penetrating sham needle similar to the one used in this trial (Vas 2004) has been shown to be indistinguishable from the real penetration of a needle among acupuncture-naïve participants in an earlier validation study (Streitberger 1998), the credibility of the specific sham needle used in this trial (Vas 2004) was not tested among the participants included (who were also acupuncture naïve) so we cannot be certain that all the participants in this trial believed this sham to be an authentic treatment. Although some unblinding may have been possible in these two sham-controlled RCTs, a possible alternative explanation for the fact that these were the only two sham-controlled trials that clearly showed clinically relevant benefits is that only these two sham-controlled trials used intensive electrical stimulation at all local knee points; electrical stimulation of needles is not always used because of logistical and cost constraints, but it may produce stronger analgesic effects than manual stimulation of needles (Ulett 1998).

A third large trial (Berman 2004) found small, statistically significant benefits of acupuncture relative to sham but these benefits did not meet our threshold for clinical relevance. This trial used an innovative sham that involved a combination of penetrating and non-penetrating needles. Most participants believed they received true acupuncture at the 4 week credibility test, but at the 26 week test, the sham group participants were more likely than the acupuncture group participants to guess that they received a sham. At the 26 week test, this may be partly explained by the fact that the participants receiving sham were experiencing no benefit, as further described above.

Two other large sham-controlled RCTs (Witt 2005; Scharf 2006) used fully needle-penetrating shams, which involved an average of 10 (Scharf 2006) to 13 (Witt 2005) needles, inserted superficially at non-acupuncture points. These two RCTs (Witt 2005; Scharf

2006) found clinically irrelevant (Scharf 2006) or minimally clinically relevant (Witt 2005) differences between acupuncture and sham, but clinically relevant differences between both acupuncture groups (i.e. both true and sham acupuncture) and the non-acupuncture comparison group. While the fully needle-penetrating shams used in these two RCTs were sufficiently similar to true acupuncture to ensure adequate blinding, at the same time, these shams were so similar to true acupuncture that they may have had weak physiological activity and not been true placebos (Birch 2006). Indeed, in another recent systematic review on acupuncture for knee osteoarthritis (White 2007), the reviewers judged that the shams used in these two RCTs were probably physiologically active and inappropriate as controls.

In these two RCTs (Witt 2005; Scharf 2006), the trial participants may have had a positive attitude towards acupuncture and high expectations of a benefit (Brinkhaus 2006), but less enthusiasm for the control treatments, which are standard care therapies that the trial participants could have easily obtained on their own. Also, these two RCTs were funded by German insurance agencies to determine whether acupuncture should be reimbursable (Linde 2004). Participating people with OA and their physicians probably knew that patients would not need to pay for their acupuncture treatments out of pocket in the future if these RCTs had positive results (Ernst 2006). As a result, participants may have overstated benefits of the treatment they believed to be acupuncture, which in these well-blinded trials (Witt 2005; Scharf 2006) were both the true and sham acupuncture treatments, and understated assessments of non-acupuncture controls. Any such biased assessments may have contributed to the large differences between the acupuncture (i.e. both true or sham) and the non-acupuncture control and the small differences between the acupuncture and sham groups.

The most recent sham-controlled trial (Foster 2007) used a sham that involved non-penetrating acupuncture at the same points as the true acupuncture, with both true and sham acupuncture administered over a duration of three weeks. In this methodologically sound trial, the sham was found to be highly credible and believable as an authentic treatment. However, because the non-insertive sham needles were placed at the true acupuncture points, these sham needles may have had some physiological activity due to a massaging effect on the acupuncture points. Indeed, the investigators found that a considerable proportion of participants in the sham group experienced the *de qi* sensation, and they noted that "we cannot consider this [sham] intervention as inert."

As noted above, for only two trials (Sangdee 2002 \*; Vas 2004) did the acupuncture meet the thresholds for clinical relevance relative to the sham; however, these thresholds should be interpreted with caution because the sham comparator may be an active treatment in itself rather than an inert placebo. Also, these clinical relevance thresholds, while helpful as general guidelines for estimating whether OA treatment effects are meaningful, cannot be considered fixed across all different clinical situations, and may vary depending on the characteristics of the population, the condition studied, the types of interventions, and the types of controls. Furthermore, data on clinical relevance of benefits need to be considered together with data on costs and adverse reactions. Clinicians must weigh all of this information to decide whether estimates of benefit are important for their patient population, and in their setting.

## Limitations of the waiting list and other active treatment control designs

The major limitation of the waiting list and the other active treatment control designs is the lack of blinding. The clinically relevant benefits of acupuncture in comparison to the waiting list and, in some cases, the other active treatment controls, might be partly attributable to either non-specific effects associated with the patient-acupuncturist relationship (Paterson 2005; Kaptchuk 2008) (e.g. attention, compassionate care (Kaptchuk 2002) or to expectations of a benefit by participants (Linde 2007). That is, study participants know they are getting acupuncture and might expect acupuncture to benefit them more than other active therapies more commonly used, perhaps because of the novelty of acupuncture, its ritualistic associations, or its ancient history of use (Kaptchuk 2006; Kaptchuk 2008). Therefore study participants receiving acupuncture may report feeling better, regardless of whether the acupuncture worked. In fact, the placebo effects of acupuncture are expected to be even stronger than the placebo effects of conventional active treatments (Hrobjartsson 2001; Kaptchuk 2006; Kaptchuk 2008), which is a limitation of head-on comparisons. However, despite the limitations of the head-on and waiting list comparator designs, these designs may still best approximate the average likely response to acupuncture in clinical practice, in which treatment effects and placebo factors, expectation effects, and patient preferences may all operate (Roland 1998).

## Possible interpretations of subgroup analyses findings

Our subgroup analyses on blinding success suggests that the benefits of acupuncture relative to sham are smaller and borderline to non-significant in the sham-controlled trials that are judged most likely to adequately blind participants to treatment assignment. However, this finding can be interpreted in at least two different ways. The first interpretation is that acupuncture is mostly a placebo, and that when acupuncture is compared with a credible sham that adequately controls for the placebo effect, there is little if any remaining benefit of acupuncture. However, a second interpretation is that the effects of acupuncture are underestimated when compared with such believable shams because such truly believable shams are often not inert placebos, but have physiological effects of their own, due either to insertion of needles, even if superficially inserted at non-acupuncture points, or placement of needle-like sham devices at true acupoints.

Furthermore, for this subgroup analysis, we considered the success of blinding to be uncertain for the trials that either did not test blinding success (Sangdee 2002 \*; Vas 2004) or that showed differential awareness of group membership among participants (Berman 2004). We might instead have assumed that all the shams used in these trials were successful at blinding participants to treatment assignment. Indeed, few systematic reviews and RCTs even consider tests of blinding success, and instead it is typically assumed that placebo controlled trials are adequately blinded. For example, in a sample of 97 placebo-controlled RCTs from five leading general medical journals, only seven RCTs provided evidence on the success of blinding, and five of these seven reported that the success of blinding was imperfect (Fergusson 2004).

Our subgroup analyses also showed that trials judged not to have an adequate number of treatment sessions delivered over

a sufficient duration showed smaller benefits compared with trials with an adequate number of sessions. This finding agrees with that of an earlier meta-analysis (Ezzo 2000), which found that fewer than six acupuncture treatments was significantly associated with smaller benefits, even after adjusting for study quality (Ezzo 2000). Finally, our subgroup analyses found that trials using electrical stimulation of needles showed a larger benefit than trials using only manual stimulation. This finding agrees with mechanistic studies of acupuncture, which have similarly suggested that electrical stimulation produces stronger analgesic effects than manual stimulation alone (Ulett 1998).

All of our subgroup analyses should be considered hypothesis generating only and not confirmatory, both because of the small number of trials relative to the large number of subgroup comparisons tested, and also because of the possibility of confounding. For example, the subgroup finding of a smaller benefit in the trials judged most likely to adequately blind participants might be confounded by the fact that the trials that tested success of blinding were the same trials that were conducted at multiple centers and that involved a large number of treating physicians or physiotherapists trained in acupuncture. A potential limitation of such large scale multi-center acupuncture trials is that, with an increasing number of study sites and practitioners, the capacity to implement and monitor adherence to the acupuncture and sham protocols could be reduced, potentially increasing the risk of a Type II error.

## AUTHORS' CONCLUSIONS

### Implications for practice

The effects of true acupuncture relative to sham do not meet our pre-specified thresholds for clinical relevance (Angst 2002). Thus, the effects of acupuncture relative to sham acupuncture are too small to be perceived by participants as beneficial (Angst 2002); however, few if any other commonly used treatments for osteoarthritis meet these thresholds (Angst 2002) for minimal clinically important differences (Bjordal 2007). For example, NSAIDs (relative to an inert placebo) do not meet these thresholds (Angst 2002; Bjordal 2004), yet NSAIDs are used regularly by half of all people with painful osteoarthritis (Bjordal 2004). Acupuncture, in contrast, is used by only about 1% (Quandt 2005) of people with osteoarthritis, and most of these people do not use it specifically for treating their osteoarthritis (Quandt 2005).

The effects of true acupuncture relative to a waiting list control and some of the other active treatment control groups do exceed our thresholds for clinical relevance (Angst 2002). The only other non-pharmacological treatment for osteoarthritis with benefits close to or exceeding the thresholds for clinical relevance is exercise, with standardized mean differences of .39 for pain and .31 for function, relative to a non-exercise control group (Fransen 2003). However, in both cases, sham treatments were not used as controls, so some of the benefits measured may be attributable to expectation or placebo effects. Although exercise cannot be compared with sham exercise, acupuncture can be compared with sham acupuncture, although sham acupuncture may not be an inert placebo. While the comparison of acupuncture with sham, which shows very small benefits of acupuncture at best, is useful for estimating the specific biological effects of acupuncture, it may be less relevant for clinical applications. Rather, the evaluation of the whole package of acupuncture, including both its specific

and non-specific components (as is the case with exercise and education), may be of equal or greater clinical relevance. Overall, the studies suggest that people with osteoarthritis find meaningful benefits through acupuncture, although these benefits may be largely mediated through placebo effects.

The fact that few if any OA treatments have specific effects that meet the threshold for clinically relevant benefits (Bjordal 2007) should not be interpreted to mean that we simply have no effective treatment for osteoarthritis. Rather, it may be that the threshold for clinical relevance is too high for any individual treatment alone, and that a multidisciplinary approach to OA patient management, with a focus on combining several nonpharmacological therapies (Jordan 2003) is necessary. Some clinicians and patients may consider acupuncture as one treatment option in such a multidisciplinary integrative approach (Jordan 2003) to treating knee osteoarthritis.

The relative benefits of acupuncture compared with other treatments cannot be reliably assessed because there is a scarcity of direct comparisons. Comparing different OA treatments by using indirect comparisons of effect sizes from different meta-analyses can be misleading because of differences in the numbers of studies, comparators used, and characteristics of participants (Zhang 2007; Zhang 2008). Indeed, the recent Osteoarthritis Research International recommendations state that at best we can only examine whether there is no overlap of the 95% confidence intervals between the meta-analytic effect sizes of different treatments to see whether there may be differential benefits (Zhang 2008). And yet because for most OA treatments there are small effect sizes with wide confidence intervals, and differences in point estimates across different meta-analyses evaluating the same treatments, it is unreliable to estimate the relative effects of acupuncture compared to other active treatments using indirect comparisons.

Safety and costs are other considerations. Safety is best determined with large prospective surveys of practitioners, and 3 such surveys (MacPherson 2001; White 2001; Melchart 2004) show that serious adverse events after acupuncture are rare. There were no adverse events associated with acupuncture in this review, although heterogeneous reporting and relatively small sample sizes limit the usefulness of this safety data. In addition to efficacy and safety, people with OA and their clinicians will also need to consider costs because acupuncture treatment often needs to be paid for out of pocket, at least in part.

### Implications for research

Considering the prevalence of knee osteoarthritis and its burden on the health system and society in general and the dearth of safe and effective treatments, it seems warranted to conduct additional RCTs evaluating the cost-effectiveness of acupuncture, as well as its short- and long-term effects relative to other active treatments and shams. Pragmatic comparisons (including cost-effectiveness studies) are now of particular clinical relevance, and some future trials should perhaps shift from sham controls to active controls. Also, future trials might shift focus from the knee to other peripheral joints, for which the current evidence is scarce.

The results of this systematic review may help inform the design of future trials in several ways. First, current RCT results (Berman 2004; Witt 2005) suggest that benefits may attenuate over time, and

therefore for future trials that assess long-term outcomes, it may be important to maintain monthly acupuncture treatments in the months prior to the long-term assessment. Second, our sensitivity analysis suggests that electrical stimulation may be associated with better outcomes, and the two sham-controlled trials in this review that used intensive electrical stimulation of all local knee points showed the greatest benefits (Sangdee 2002 \*; Vas 2004). While these findings might indicate a superiority of electroacupuncture over needle acupuncture without electrical stimulation, the finding may also be explained by the fact that electroacupuncture is probably more difficult to blind than needle acupuncture and some of the extra benefit seen with electroacupuncture may be due to incomplete blinding or placebo effects. Third, our sensitivity analysis suggests that an adequate number of treatments delivered over a time period of a sufficient duration may be associated with better outcomes. Fourth, acupuncture may elicit a greater placebo effect or meaning response (Moerman 2002) than usual care therapies, particularly among participants who have a preference for acupuncture, and therefore investigators conducting future pragmatic trials that compare acupuncture with other active therapies might consider asking participants about their preferences and expectations (before and after the intervention), and studying the potential effects of pre-treatment preferences on study outcomes. Furthermore, to minimize the recruitment of participants with a preference for acupuncture, advertisements to recruit participants should ideally not specify that acupuncture is one of the treatments being investigated. Fifth and last, our review suggests that skin-penetrating needle shams may be best at insuring blinding success, but that such penetrating shams may also have physiological activity. Future trials should therefore consider the use of non-insertive shams; however, because such non-insertive shams may be less believable to participants, if they are used, their credibility should be tested, certainly before the trial starts, and perhaps also during the trial.

There are at least three large and rigorous ongoing sham-controlled trials, all of which should be published within the next couple of years. The largest of these three trials, which was recently presented at a conference (Suarez-Almazor 2007), found no difference between true and sham acupuncture, but

found significant differences between both the true and sham acupuncture groups and the waiting list control group (Suarez-Almazor 2007). The results of the two other sham-controlled trials currently ongoing will be unlikely to shift the currently very small pooled benefits of acupuncture relative to sham towards the threshold for clinical relevance; however, these ongoing trials, likely to be successfully blinded, will be important to further assess how much of the currently observed benefit of acupuncture relative to sham is due to expectation or placebo effects and how much is due to specific effects of the needle placement. However, the truth about acupuncture effects will always be difficult to assess, even through a systematic review of well-designed and well-reported RCTs. The complexities and potential biases inherent to both the non-acupuncture and sham acupuncture control designs makes it difficult to evaluate the subjective, patient-reported outcomes in peripheral joint osteoarthritis.

## ACKNOWLEDGEMENTS

The authors thank Nancy Min, PhD, and Kevin Chen, PhD, for their support in calculating estimated SDs of change from baseline for publications that did not directly report these but did report other data that could be used to estimate SDs of change; Byungmook Lim, MD, MPH, PhD, for performing data extraction for Witt and colleagues' 2005 RCT (Witt 2005); Marcos Hsu, LAc, for assessing the adequacy of the acupuncture administered in the RCTs; Becky Gray for support with RevMan 5 questions; and Elizabeth Kimbrough Pradhan, PhD, and Jeanette Ezzo, PhD, for providing helpful comments on the manuscript. The authors also thank Claudia Witt, PhD, for providing unpublished data specific to people with only knee osteoarthritis from Witt and colleagues' 2006 RCT (Witt 2006), and Birgitte Christensen, MD; Nadine Foster, PhD; Wenlin Lee, PhD; Jorge Vas, MD; Lynn Williamson, MD; and Steffen Witte, PhD, all coauthors of included RCTs, for confirming and providing data related to their respective RCTs. Ke Cheng has been supported by Innovation Program of Shanghai Municipal Education Commission (China, Grant No. 08YZ55), Chenguang Program of Shanghai Education Committee (China, Grant No. 2008CG52) and National Natural Science Foundation of China (30572306).

## REFERENCES

### References to studies included in this review

#### Berman 1999 *{published data only}*

\* Berman BM, Singh BB, Lao L, Langenberg P, Li H, Hadhazy V, et al. A randomized trial of acupuncture as an adjunctive therapy in osteoarthritis of the knee. *Rheumatology (Oxford)* 1999;**38**(4):346-54. [MEDLINE: 10378713]

Singh BB, Berman BM, Hadhazy V, Baretta J, Lao L, Zarow FM, et al. Clinical decisions in the use of acupuncture as an adjunctive therapy for osteoarthritis of the knee. *Alternative therapies in health and medicine* 2001;**7**(4):58-65. [MEDLINE: 11452568]

#### Berman 2004 *{published and unpublished data}*

\* Berman BM, Lao L, Langenberg P, Lee WL, Gilpin AM, Hochberg MC. Effectiveness of acupuncture as adjunctive therapy in osteoarthritis of the knee: a randomized, controlled trial. *Annals of internal medicine* 2004;**141**(12):901-10. [MEDLINE: 15611487]

#### Christensen 1992 *{published data only}*

\* Christensen BV, luhl IU, Vilbek H, Bulow HH, Dreijer NC, Rasmussen HF. Acupuncture treatment of severe knee osteoarthritis. A long-term study. *Acta anaesthesiologica Scandinavica* 1992;**36**(6):519-25. [MEDLINE: 1514335]

#### Fink 2001 *{published data only}*

Fink MG, Kunsebeck HW, Wippermann B, Fink MG, Kunsebeck HW, Wippermann B, et al. [Effect of needle acupuncture on pain perception and functional impairment of patients with coxarthrosis]. *Zeitschrift fur Rheumatologie* 2000;**59**(3):191-9. [MEDLINE: 10929448]

Fink MG, Wippermann B, Gehrke A. Non-specific effects of traditional Chinese acupuncture in osteoarthritis of the hip. *Complementary therapies in medicine* 2001;**9**(2):82-9. [MEDLINE: 11444887]

Fink MG, de Groot M, Kunsebeck HW, Gutenbrunner CH, Lambrecht F, Gehrke A. Standardized acupuncture vs placebo-acupuncture in coxarthrosis patient [Abstract]. *Forschende Komplementarmedizin* 1996;**3**:316.

#### Foster 2007 *{published data only}*

Foster NE, Thomas E, Barlas P, Hill JC, Young J, Mason E, et al. Acupuncture as an adjunct to exercise based physiotherapy for osteoarthritis of the knee: randomised controlled trial. *BMJ* 2007;**335**(7617):436.

#### Haslam 2001 *{published data only}*

Haslam R. A comparison of acupuncture with advice and exercises on the symptomatic treatment of osteoarthritis of the hip--a randomised controlled trial. *Acupuncture in medicine* 2001;**19**(1):19-26. [MEDLINE: 11471578]

#### Molsberger 1994 *{published data only}*

\* Molsberger A, BÖwing G, Jensen KU, Lorek M. Acupuncture treatment for the relief of gonarthrosis pain-a controlled clinical trial [Schmerztherapie mit Akupunktur bei Gonarthrose Eine

kontrollierte Studie zur analgetischen Wirkung der Akupunktur bei Gonarthrose]. *Der Schmerz* 1994;**8**:37-42.

#### Sangdee 2002 \* *{published data only}*

\* Sangdee C, Teekachunhatean S, Sananpanich K, Sugandhavesa N, Chiewchantanakit S, Pojchamarnwiputh S, et al. Electroacupuncture versus diclofenac in symptomatic treatment of osteoarthritis of the knee: a randomized controlled trial. *BMC complementary and alternative medicine* 2002;**2**:3. [MEDLINE: 11914160]

#### Sangdee 2002 \*\* *{published data only}*

Sangdee C, Teekachunhatean S, Sananpanich K, Sugandhavesa N, Chiewchantanakit S, Pojchamarnwiputh S, et al. Electroacupuncture versus diclofenac in symptomatic treatment of osteoarthritis of the knee: a randomized controlled trial. *BMC complementary and alternative medicine* 2002;**2**:3. [MEDLINE: 11914160]

#### Scharf 2006 *{published and unpublished data}*

\* Scharf HP, Mansmann U, Streitberger K, Witte S, Kramer J, Maier C, et al. Acupuncture and knee osteoarthritis: a three-armed randomized trial. *Annals of internal medicine* 2006;**145**(1):12-20.

Streitberger K, Witte S, Mansmann U, Knauer C, Kramer J, Scharf HP, et al. Efficacy and safety of acupuncture for chronic pain caused by gonarthrosis: a study protocol of an ongoing multi-centre randomised controlled clinical trial. *BMC complementary and alternative medicine* 2004;**4**(1):6. [MEDLINE: 15040805]

#### Stener-Victorin 2004 *{published data only}*

Stener-Victorin E, Kruse-Smidje C, Jung K. Comparison between electro-acupuncture and hydrotherapy, both in combination with patient education and patient education alone, on the symptomatic treatment of osteoarthritis of the hip. *The Clinical journal of pain* 2004;**20**(3):179-85. [MEDLINE: 15100594]

#### Takeda 1994 *{published data only}*

\* Takeda W, Wessel J. Acupuncture for the treatment of pain of osteoarthritic knees. *Arthritis care and research* 1994;**7**(3):118-22. [MEDLINE: 7727550]

#### Tukmachi 2004 *{published data only}*

\* Tukmachi E, Jubb R, Dempsey E, Jones P. The effect of acupuncture on the symptoms of knee osteoarthritis-- an open randomised controlled study. *Acupuncture in medicine* 2004;**22**(1):14-22. [MEDLINE: 15077933]

#### Vas 2004 *{published and unpublished data}*

\* Vas J, Mendez C, Perea-Milla E, Vega E, Panadero MD, Leon JM, et al. Acupuncture as a complementary therapy to the pharmacological treatment of osteoarthritis of the knee: randomised controlled trial. *BMJ* 2004;**329**(7476):1216. [MEDLINE: 15494348]

**Williamson 2007** {published data only}

Williamson L, Wyatt MR, Yein K, Melton JT. Severe knee osteoarthritis: a randomized controlled trial of acupuncture, physiotherapy (supervised exercise) and standard management for patients awaiting knee replacement. *Rheumatology (Oxford)* 2007;**46**(9):1445-9.

**Witt 2005** {published and unpublished data}

Brinkhaus B, Becker-Witt C, Jena S, Linde K, Streng A, Wagenpfeil S, et al. Acupuncture Randomized Trials (ART) in patients with chronic low back pain and osteoarthritis of the knee - design and protocols. *Forschende Komplementarmedizin und Klassische Naturheilkunde* 2003;**10**(4):185-91. [MEDLINE: 12972723]

\* Witt C, Brinkhaus B, Jena S, Linde K, Streng A, Wagenpfeil S, et al. Acupuncture in patients with osteoarthritis of the knee: a randomised trial. *Lancet* 2005;**366**(9480):136-43. [MEDLINE: 16005336]

**Witt 2006** {published and unpublished data}

Witt CM, Jena S, Brinkhaus B, Liecker B, Wegscheider K, Willich SN. Acupuncture in patients with osteoarthritis of the knee or hip: a randomized, controlled trial with an additional nonrandomized arm. *Arthritis and rheumatism* 2006;**54**(11):3485-93. [MEDLINE: 17075849]

**References to studies excluded from this review**
**Ammer 1988** {published data only}

Ammer K, Petschnig R. [Comparison of the effectiveness of acupuncture and physical therapy in ambulatory patients with gonarthrosis]. *Wiener medizinische Wochenschrift* 1988;**138**(22):566-9. [MEDLINE: 3213064]

**Coan 1982** {published data only}

Coan RM, Wong G, Coan PL. The acupuncture treatment of neck pain: a randomized controlled study. *The American journal of Chinese medicine* 1982;**9**(4):326-32.

**Dickens 1989** {published data only}

Dickens W, Lewith GT. A single-blind controlled and randomized clinical trial to evaluate the effect of acupuncture in the treatment of trapezio-metacarpal osteoarthritis. *Complementary Medicine Research* 1989;**3**(2):5-8.

**Gaw 1975** {published data only}

Gaw AC, Chang LW, Shaw LC. Efficacy of acupuncture on osteoarthritic pain. A controlled, double-blind study. *The New England journal of medicine* 1975;**293**(8):375-8. [MEDLINE: 1097921]

**Junnila 1982** {published data only}

Junnila SYT. Acupuncture superior to piroxicam in the treatment of osteoarthrosis. *American journal of acupuncture* 1982;**10**:341-6.

**Kwon 2001** {published data only}

Kwon YB, Kim JH, Yoon JH, Lee JD, Han HJ, Mar WC, et al. The analgesic efficacy of bee venom acupuncture for knee osteoarthritis: a comparative study with needle acupuncture.

*The American journal of Chinese medicine* 2001;**29**(2):187-99. [MEDLINE: 11527062]

**Loy 1983** {published data only}

Loy TT. Treatment of cervical spondylosis. Electroacupuncture versus physiotherapy. *The Medical journal of Australia* 1983;**2**(1):32-4. [MEDLINE: 6346029]

**Lundeberg 1991** {published data only}

Lundeberg T, Eriksson SV, Lundeberg S, Thomas M. Effect of acupuncture and naloxone in patients with osteoarthritis pain. A sham acupuncture controlled study. *The Pain Clinic* 1991;**4**:155-61.

**McIndoe 1994** {published data only}

McIndoe AK, Young K, Bone ME. A comparison of acupuncture with intra-articular steroid injection as analgesia for osteoarthritis of the hip. British Medical Acupuncture Society. Paper presented in Cheltenham, 1994.

**Milligan 1981** {published data only}

Milligan JL, Glennie-Smith K, Dowson DI, Harris J. Comparison of acupuncture with physiotherapy in the treatment of osteoarthritis of the knees [abstract]. 15th International Congress of Rheumatology. Paris, 1981.

**Ng 2003** {published data only}

Ng MM, Leung MC, Poon DM. The effects of electro-acupuncture and transcutaneous electrical nerve stimulation on patients with painful osteoarthritic knees: a randomized controlled trial with follow-up evaluation. *Journal of alternative and complementary medicine* 2003;**9**(5):641-9. [MEDLINE: 14629842]

**Petrie 1983** {published data only}

Petrie JP, Langley GB. Acupuncture in the treatment of chronic cervical pain. A pilot study. *Clinical and experimental rheumatology* 1983;**1**(4):333-6. [MEDLINE: 6335859]

**Petrie 1986** {published data only}

Petrie JP, Hazleman BL. A controlled study of acupuncture in neck pain. A controlled study of acupuncture in neck pain. *British journal of rheumatology* 1986;**25**(3):271-5. [MEDLINE: 3524727]

**Petrou 1988** {published data only}

Petrou P, Winkler V, Genti G, Balint G. Double-blind trial to evaluate the effect of acupuncture treatment on knee osteoarthrosis. *Scandinavian Journal of Acupuncture* 1988;**3**:112-5.

**Thomas 1991** {published data only}

Thomas M, Eriksson SV, Lundeberg T. A comparative study of diazepam and acupuncture in patients with osteoarthritis pain: a placebo controlled study. *The American journal of Chinese medicine* 1991;**19**(2):95-100. [MEDLINE: 1816730]

**Tillu 2001** {published data only}

Tillu A, Roberts C, Tillu S. Unilateral versus bilateral acupuncture on knee function in advanced osteoarthritis of the knee--a prospective randomised trial. *Acupuncture in medicine* 2001;**19**(1):15-8. [MEDLINE: 11471577]

**Tillu 2002** {published data only}

Tillu A, Tillu S, Vowler S. Effect of acupuncture on knee function in advanced osteoarthritis of the knee: a prospective, non-randomised controlled study. *Acupuncture in medicine* 2002;**20**(1):19-21. [MEDLINE: 11926599]

**Weiner 2007** {published data only}

Weiner DK, Rudy TE, Morone N, Glick R, Kwok CK. Efficacy of periosteal stimulation therapy for the treatment of chronic knee pain: an initial controlled clinical trial osteoarthritis-associated chronic knee pain: an initial controlled clinical trial. *Journal of the American Geriatrics Society* 2007;**55**(10):1541-7. [MEDLINE: 17908057]

**Yurtkuran 1999** {published data only}

Yurtkuran M, Kocagil T. TENS, electroacupuncture and ice massage: comparison of treatment for osteoarthritis of the knee. *American journal of acupuncture* 1999;**27**(3-4):133-40. [MEDLINE: 10729968]

**Zherebkin 1998** {published data only}

Zherebkin VV. [The use of acupuncture reflex therapy in the combined treatment of osteoarthritis patients] [Primenenie iglorefleksoterapii v kompleksnom lechenii bol'nykh osteoartrozom.]. *Likars'ka sprava / Ministerstvo okhorony zdorov'ia Ukrainy* 1998;**8**:149-51. [MEDLINE: 10204379]

**References to studies awaiting assessment**
**Itoh 2008** {published data only}

Itoh K, Hirota S, Katsumi Y, Ochi H, Kitakoji H. Trigger point acupuncture for treatment of knee osteoarthritis--a preliminary RCT for a pragmatic trial. *Acupuncture in Medicine* 2008;**26**(1):17-26.

**Itoh 2008\*** {published data only}

Itoh K, Hirota S, Katsumi Y, Ochi H, Kitakoji H. A pilot study on using acupuncture and transcutaneous electrical nerve stimulation (TENS) to treat knee osteoarthritis (OA). *Chinese Medicine* 2008;**3**:2.

**Jubb 2008** {published data only}

Jubb RW, Tukmachi ES, Jones PW, Dempsey E, Waterhouse L, Brailsford S. A blinded randomised trial of acupuncture (manual and electroacupuncture) compared with a non-penetrating sham for the symptoms of osteoarthritis of the knee. *Acupuncture in Medicine* 2008;**26**(2):69-78.

**References to ongoing studies**
**Bower** {unpublished data only}

Bower S. A randomised, controlled, single-blind trial to test the efficacy of acupuncture on pain relief in osteoarthritis of the knee. Susanne.Bower@ncl.ac.uk.

**Farrar** {unpublished data only}

Farrar JT. Efficacy of acupuncture with physical therapy for knee osteoarthritis. Contact details: jfarrar@cceb.med.upenn.edu. [Grant number: 5R01AT000304-04]

**McNally** {unpublished data only}

McNally, J. A study of acupuncture, physiotherapy or non-intervention in management of painful (OA) knee. Dr Jeremy Mc Nally, Department of Rheumatology; Battle Hospital; Oxford Road; Reading RG30 1AG; United Kingdom Telephone: 0118-958-3666.

**Suarez-Almazor** {unpublished data only}

\* Suarez-Almazor, M. Interaction Between Patient and Healthcare Provider: Response to Acupuncture in Knee Osteoarthritis. Contact details: mes@bcm.tmc.edu. [Grant number: 5R01AR049999-02]

**van Arendok** {unpublished data only}

van Arendok, N. Does the inclusion of acupuncture improve pain and outcome measures following total knee arthroscopy?. Ms Norma van Arendok, Physiotherapist; St. Mary's Hospital; Newport; Isle of Wight; PO30 5TG, United Kingdom.

**White** {published data only}

White, Peter. The process of acupuncture: a randomised controlled trial and qualitative study to evaluate the relative contributions of specific and non-specific effects. Contact details: pjw1@soton.ac.uk. [ISRCTN78434638]

**Additional references**
**Altman 2004**

Altman DG, Schulz KF, Moher D. Turning a blind eye: testing the success of blinding and the CONSORT statement. *BMJ* 2004;**328**(7448):1135; author reply 1136. [MEDLINE: 15130995]

**Am Coll Rheum 2000**

American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. *Arthritis and rheumatism* 2000;**43**(9):1905-15. [MEDLINE: 11014340]

**Angst 2002**

Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *The Journal of rheumatology* 2002;**29**(1):131-8. [MEDLINE: 11824949]

**Angst 2005**

Angst F, Ewert T, Lehmann S, Aeschlimann A, Stucki G. The factor subdimensions of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) help to specify hip and knee osteoarthritis. a prospective evaluation and validation study. *The Journal of rheumatology* 2005;**32**(7):1324-30. [MEDLINE: 15996072]

**Ausiello 2002**

Ausiello JC, Stafford RS. Trends in medication use for osteoarthritis treatment. *The Journal of rheumatology* 2002;**29**(5):999-1005. [MEDLINE: 12022364]



**Bellamy 1988**

Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *The Journal of rheumatology* 1988;**15**(12):1833-40. [MEDLINE: 3068365]

**Bellamy 2005**

Bellamy N. The WOMAC Knee and Hip Osteoarthritis Indices: development, validation, globalization and influence on the development of the AUSCAN Hand Osteoarthritis Indices. *Clinical and experimental rheumatology* 2005;**23**(5 Suppl 39):S148-53. [MEDLINE: 16273799]

**Berman 2005**

Berman B, Gilpin A, Lao L. Acupuncture and Knee Osteoarthritis. *Annals of internal medicine* 2005;**142**(10):872-3.

**Birch 1998**

Birch S, Ida J. Japanese Acupuncture. Brookline: Paradigm Publications, 1998.

**Birch 2001**

Birch S. Systematic reviews of acupuncture -- are there problems with these?. *Clinical Acupuncture and Oriental Medicine* 2001;**2**:17-22.

**Birch 2006**

Birch S. A review and analysis of placebo treatments, placebo effects, and placebo controls in trials of medical procedures when sham is not inert. *Journal of alternative and complementary medicine* 2006;**12**(3):303-10. [MEDLINE: 16646730]

**Bjordal 2004**

Bjordal JM, Ljunggren AE, Klovning A, Slordal L. Non-steroidal anti-inflammatory drugs, including cyclo-oxygenase-2 inhibitors, in osteoarthritic knee pain: meta-analysis of randomised placebo controlled trials. *BMJ* 2004;**329**(7478):1317. [MEDLINE: 15561731]

**Bjordal 2007**

Bjordal JM, Klovning A, Ljunggren AE, Slordal L. Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: A meta-analysis of randomised placebo-controlled trials. *European journal of pain* 2007;**11**(2):125-38. [MEDLINE: 16682240]

**Brinkhaus 2006**

Brinkhaus B, Witt CM, Jena S, Linde K, Streng A, Wagenpfeil S, et al. Acupuncture in patients with chronic low back pain: a randomized controlled trial. *Archives of internal medicine* 2006;**166**(4):450-7.

**Cherkin 2003**

Cherkin DC, Sherman KJ, Deyo RA, Shekelle PG. A review of the evidence for the effectiveness, safety, and cost of acupuncture, massage therapy, and spinal manipulation for back pain. *Annals of internal medicine* 2003;**138**(11):898-906. [MEDLINE: 12779300]

**Cibere 2004**

Cibere J, Kopec JA, Thorne A, Singer J, Canvin J, Robinson DB, et al. Randomized, double-blind, placebo-controlled glucosamine discontinuation trial in knee osteoarthritis. *Arthritis and rheumatism* 2004;**51**(5):738-45. [MEDLINE: 15478160]

**Clegg 2006**

Clegg DO, Reda DJ, Harris CL, Klein MA, O'Dell JR, Hooper MM, et al. Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. *The New England journal of medicine* 2006;**354**(8):795-808. [MEDLINE: 16495392]

**Deeks 2001**

Deeks JJ ADBM. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG editor(s). *Systematic Reviews in Health Care: Meta-Analysis in Context*. 2nd Edition. London: BMJ Publishing Group, 2001.

**Deeks 2005**

Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. Section 8. In: Higgins JPT, Green S editor(s). *Cochrane Handbook for Systematic Reviews of Interventions*. 4.2.4. The Cochrane Collaboration, 2005.

**Ernst 1997**

Ernst. Acupuncture as a symptomatic treatment of osteoarthritis. A systematic review. *Scandinavian journal of rheumatology* 1997;**26**(6):444-7. [MEDLINE: 9433405]

**Ernst 2006**

Ernst E. Testing acupuncture for osteoarthritis: pragmatic trials or efficacy studies?. *Rheumatology (Oxford)* 2006;**45**(2):125.

**Escobar 2002**

Escobar A, Quintana JM, Bilbao A, Azkarate J, Guenaga JI. Validation of the Spanish version of the WOMAC questionnaire for patients with hip or knee osteoarthritis. Western Ontario and McMaster Universities Osteoarthritis Index. *Clinical rheumatology* 2002;**21**(6):466-71. [MEDLINE: 12447629]

**Ezzo 2000**

Ezzo J, Berman B, Hadhazy VA, Jadad AR, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000;**86**(3):217-25. [MEDLINE: 10812251]

**Ezzo 2001**

Ezzo J, Hadhazy V, Birch S, Lao L, Kaplan G, Hochberg M, et al. Acupuncture for osteoarthritis of the knee: a systematic review. *Arthritis and rheumatism* 2001;**44**(4):819-25. [MEDLINE: 11315921]

**Felson 1998**

Felson DT, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis and rheumatism* 1998;**41**(8):1343-55. [MEDLINE: 9704632]

**Felson 2000**

Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, et al. Osteoarthritis: new insights. Part 1: the

disease and its risk factors. *Annals of internal medicine* 2000;**133**(8):635-46. [MEDLINE: 11033593]

#### Fergusson 2004

Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004;**328**(7437):432. [MEDLINE: 14761905]

#### Fransen 2003

Fransen M, McConnell S, Bell M. Exercise for osteoarthritis of the hip or knee. *Cochrane Database of Systematic Reviews* 2003, Issue 3. [DOI: [10.1002/14651858.CD004286](https://doi.org/10.1002/14651858.CD004286)]

#### Furlan 2005

Furlan AD, van Tulder MW, Cherkin DC, Tsukayama H, Lao L, Koes BW, et al. Acupuncture and dry-needling for low back pain. *Cochrane Database of Systematic Reviews* 2005, Issue 1. [DOI: [10.1002/14651858.CD001351.pub2](https://doi.org/10.1002/14651858.CD001351.pub2); MEDLINE: 15674876]

#### Glanville 2006

Glanville JM, Lefebvre C, Miles JN, Camosso-Stepinovic J. How to identify randomized controlled trials in MEDLINE: ten years on. *Journal of the Medical Library Association : JMLA* 2006;**94**(2):130-6. [MEDLINE: 16636704]

#### Guccione 1994

Guccione AA, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. *American journal of public health* 1994;**84**(3):351-8. [MEDLINE: 16636704]

#### Han 2003

Han JS. Acupuncture: neuropeptide release produced by electrical stimulation of different frequencies. *Trends in neurosciences* 2003;**26**(1):17-22. [MEDLINE: 12495858]

#### Herman 2004

Herman CJ, Allen P, Hunt WC, Prasad A, Brady TJ. Use of complementary therapies among primary care clinic patients with arthritis. *Preventing chronic disease* 2004;**1**(4):A12. [MEDLINE: 15670444]

#### Higgins 2008 (Section 16.5.4)

Higgins JPT, Green S (editors). 16.5.4 How to include multiple groups from one study. In: Higgins JPT, Green S editor(s). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, 2008.

#### Higgins 2008 (Section 7)

Higgins JPT, Deeks JJ (editors). Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S editor(s). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, 2008.

#### Higgins 2008 (Section 9.4.5.2)

Higgins JPT, Green S (editors). 9.4.5.2 Meta-analysis of change scores. In: Higgins JPT, Green S (editors) editor(s). *Cochrane*

*Handbook for Systematic Reviews of Interventions* Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, 2008.

#### Higgins 2008 (Section 9.5.2)

Higgins JPT, Green S (editors). 9.5.2 Identifying and measuring heterogeneity. In: Higgins JPT, Green S (editors) editor(s). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, 2008.

#### Hrobjartsson 2001

Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *The New England journal of medicine* 2001;**344**(21):1594-602.

#### Hui 2005

Hui KK, Liu J, Marina O, Napadow V, Haselgrove C, Kwong KK, et al. The integrated response of the human cerebro-cerebellar and limbic systems to acupuncture stimulation at ST 36 as evidenced by fMRI. *NeuroImage* 2005;**27**(3):479-96. [MEDLINE: 16046146]

#### Ioannidis 2008

Ioannidis JP, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 2008;**336**(7658):1413-5.

#### Jadad 1996

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary?. *Controlled clinical trials* 1996;**17**(1):1-12. [MEDLINE: 8721797]

#### Jinks 2004

Jinks C, Jordan K, Ong BN, Croft P. A brief screening tool for knee pain in primary care (KNEST). 2. Results from a survey in the general population aged 50 and over. *Rheumatology (Oxford)* 2004;**43**(1):55-61. [MEDLINE: 12923283]

#### Jordan 2003

Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCSIT). *Annals of the rheumatic diseases* 2003;**62**(12):1145-55. [MEDLINE: 14644851]

#### Juni 1999

Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;**282**(11):1054-60. [MEDLINE: 10493204]

#### Kaptchuk 2002

Kaptchuk TJ. The placebo effect in alternative medicine: can the performance of a healing ritual have clinical significance?. *Annals of internal medicine* 2002;**136**(11):817-25.

#### Kaptchuk 2006

Kaptchuk TJ, Stason WB, Davis RB, Legedza AR, Schnyer RN, Kerr CE, et al. Sham device v inert pill: randomised controlled

trial of two placebo treatments. *BMJ* 2006;**332**(7538):391-7. [MEDLINE: 16452103]

#### Kaptchuk 2008

Kaptchuk TJ, Kelley JM, Conboy LA, Davis RB, Kerr CE, Jacobson EE, et al. Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. *BMJ* Epub 2008;**336**(7651):999-1003. [MEDLINE: 18390493]

#### Kearney 2006

Kearney PM, Baigent C, Godwin J, Halls H, Emberson JR, Patrono C. Do selective cyclo-oxygenase-2 inhibitors and traditional non-steroidal anti-inflammatory drugs increase the risk of atherothrombosis? Meta-analysis of randomised trials. *BMJ* 2006;**3**(332):1302-8. [MEDLINE: 16740558]

#### Kwon 2006

Kwon YD, Pittler MH, Ernst E. Acupuncture for peripheral joint osteoarthritis: A systematic review and meta-analysis. *Rheumatology (Oxford)* Epub 2006;**45**(11):1331-7. [MEDLINE: 16936326]

#### Lao 2003

Lao L, Hamilton GR, Fu J, Berman BM. Is acupuncture safe? A systematic review of case reports. *Alternative therapies in health and medicine* 2003;**9**(1):72-83. [MEDLINE: 12564354]

#### Larson 2005

Larson AM, Polson J, Fontana RJ, Davern TJ, Lalani E, Hynan LS, et al. Acetaminophen-induced acute liver failure: results of a United States multicenter, prospective study. *Hepatology* 2005;**42**(6):1364-72. [MEDLINE: 16317692]

#### Lawrence 1998

Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis and rheumatism* 1998;**41**(5):778-99. [MEDLINE: 9588729]

#### Lawrence 2008

Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part II. *Arthritis and rheumatism* 2008;**58**(1):26-35. [MEDLINE: 9588729]

#### Le Bars 1979

Le Bars D, Dickenson AH, Besson JM. Diffuse noxious inhibitory controls (DNIC). I. Effects on dorsal horn convergent neurones in the rat. *Pain* 1979;**6**(3):283-304. [MEDLINE: 460935]

#### LethbridgeCejku 2004

Lethbridge-Cejku M, Schiller JS, Bernadel L. Summary health statistics for U.S. adults: National Health Interview Survey, 2002. *Vital and health statistics. Series 10, Data from the National Health Survey* 2004;**Not available**(222):1-151. [MEDLINE: 15791763]

#### Lewith 1983

Lewith GT, Machin D. On the evaluation of the clinical effects of acupuncture. *Pain* 1983;**16**(2):111-27. [MEDLINE: 6348651]

#### Lewith 1984

Lewith GT, Kenyon JN. Physiological and psychological explanations for the mechanism of acupuncture as a treatment for chronic pain. *Social science & medicine* 1984;**19**(12):1367-78. [MEDLINE: 6085191]

#### Li 2008

Li A, Lao L, Wang Y, Xin J, Ren K, Berman BM, et al. Electroacupuncture activates corticotrophin-releasing hormone-containing neurons in the paraventricular nucleus of the hypothalamus to alleviate edema in a rat model of inflammation. *BMC Complementary and Alternative Medicine* 2008;**8**:20.

#### Linde 2004

Linde K. [Should acupuncture be reimbursed? Three acupuncture programmes of German Statutory Health Insurance Funds (SHI) and their scientific evaluation]. *Zeitschrift fur arztliche Fortbildung und Qualitätssicherung* 2004;**98**(6):467-8. [MEDLINE: 15527186]

#### Linde 2007

Linde K, Witt CM, Streng A, Weidenhammer W, Wagenpfeil S, Brinkhaus B, et al. The impact of patient expectations on outcomes in four randomized controlled trials of acupuncture in patients with chronic pain. *Pain* 2007;**3**:264-71. [MEDLINE: 17257756]

#### Macdonald 1983

Macdonald AJ, Macrae KD, Master BR, Rubin AP. Superficial acupuncture in the relief of chronic low back pain. *Annals of the Royal College of Surgeons of England* 1983;**65**(1):44-6. [MEDLINE: 6218776]

#### MacPherson 2001

MacPherson H, Thomas K, Walters S, Fitter M. The York acupuncture safety study: prospective survey of 34 000 treatments by traditional acupuncturists. *BMJ* 2001;**323**:486-7. [MEDLINE: 11532841]

#### Manheimer 2005

Manheimer E, White A, Berman B, Forsys K, Ernst E. Meta-analysis: acupuncture for low back pain. *Annals of internal medicine* 2005;**142**(8):651-63. [MEDLINE: 15838072]

#### Manheimer 2007

Manheimer E, Linde K, Lao L, Bouter LM, Berman BM. Meta-analysis: acupuncture for osteoarthritis of the knee. *Annals of internal medicine* 2007;**146**(12):868-77. [MEDLINE: 17577006]

#### McAlindon 2004

McAlindon T, Formica M, LaValley M, Lehmer M, Kabbara K. Effectiveness of glucosamine for symptoms of knee osteoarthritis: results from an internet-based randomized double-blind controlled trial. *The American journal of medicine* 2004;**117**(9):643-9. [MEDLINE: 15501201]

#### Melchart 2004

Melchart D, Weidenhammer W, Streng A, Reitmayr S, Hoppe A, Ernst E, et al. Prospective investigation of adverse effects of

acupuncture in 97 733 patients. *Archives of internal medicine* 2004;**164**(1):104-5. [MEDLINE: 14718331]

#### Miller 2007

Miller FG, Kaptchuk TJ. Acupuncture trials and informed consent. *Journal of medical ethics* 2007;**33**(1):43-4. [MEDLINE: 17209110]

#### Mitchell 2006

Mitchell HL, Carr AJ, Scott DL. The management of knee pain in primary care: factors associated with consulting the GP and referrals to secondary care. *Rheumatology (Oxford)* 2006;**45**(6):771-6. [MEDLINE: 16461443]

#### MMWR 2001

Centers for Disease Control and Prevention. Prevalence of disabilities and associated health conditions among adults--United States, 1999. *MMWR. Morbidity and mortality weekly report* 2001;**50**(7):120-5. [MEDLINE: 11393491]

#### Moerman 2002

Moerman DE, Jonas WB. Deconstructing the placebo effect and finding the meaning response. *Annals of internal medicine* 2002;**136**(6):471-6. [MEDLINE: 11900500]

#### NICE Guideline 2008

National Collaborating Centre for Chronic Conditions. Osteoarthritis: national clinical guideline for care and management in adults. London: Royal College of Physicians, 2008.

#### Oliveria 1995

Oliveria SA, Felson DT, Reed JI, Cirillo PA, Walker AM. Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. *Arthritis and rheumatism* 1995;**38**(8):1134-41. [MEDLINE: 7639811]

#### Pariente 2005

Pariente J, White P, Frackowiak RS, Lewith G. Expectancy and belief modulate the neuronal substrates of pain treated by acupuncture. *Neuroimage* 2005;**25**(4):1161-7. [MEDLINE: 15850733]

#### Paterson 2005

Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ* 2005;**330**(7501):1202-5. [MEDLINE: 15905259]

#### Peat 2001

Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. *Annals of the rheumatic diseases* 2001;**60**(2):91-7. [MEDLINE: 11156538]

#### Quandt 2005

Quandt SA, Chen H, Grzywacz JG, Bell RA, Lang W, Arcury TA. Use of complementary and alternative medicine by persons with arthritis: results of the National Health Interview Survey. *Arthritis and rheumatism* 2005;**53**(5):748-55. [MEDLINE: 16208669]

#### Rao 1999

Rao JK, Mihaliak K, Kroenke K, Bradley J, Tierney WM, Weinberger M. Use of complementary therapies for arthritis among patients of rheumatologists. *Annals of internal medicine* 1999;**131**(6):409-16. [MEDLINE: 10498556]

#### Reichenbach 2007

Reichenbach S, Sterchi R, Scherer M, Trelle S, Burgi E, Burgi U, et al. Meta-analysis: chondroitin for osteoarthritis of the knee or hip. *Annals of internal medicine* 2007;**146**(8):580-90. [MEDLINE: 17438317]

#### Roland 1998

Roland M, Torgerson DJ. What are pragmatic trials?. *BMJ* 1998;**316**(7127):285.

#### Rozendaal 2008

Rozendaal RM, Koes BW, van Osch GJ, Uitterlinden EJ, Garling EH, Willemsen SP, et al. Effect of glucosamine sulfate on hip osteoarthritis: a randomized trial. *Annals of internal medicine* 2008;**148**(4):268-77. [MEDLINE: 18283204]

#### Sackett 2004

Sackett DL. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ* 2004;**328**(7448):1136. [MEDLINE: 15130997]

#### Sackett 2007

Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't?. *International journal of epidemiology* 2007;**36**(3):664-5. [MEDLINE: 17675306]

#### Schulz 2002

Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet* 2002;**359**(9307):696-700. [MEDLINE: 11879884]

#### Streitberger 1998

Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. *Lancet* 1998;**352**(9125):364-5. [MEDLINE: 9717924]

#### Stucki 1996

Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, et al. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index]. *Zeitschrift fur Rheumatologie* 1996;**55**(1):40-9. [MEDLINE: 8868149]

#### Suarez-Almazor 2007

Suarez-Almazor M, Looney C, Street Jr R, Liu Y, Cox V, Pietz K, et al. A randomized controlled trial of acupuncture for osteoarthritis of the knee: effects of provider communication style. American College of Rheumatology Annual Conference Proceedings 2007; Vol. 007-2. Osteoarthritis: Clinical Aspects I: Abstract No. 739.

#### Sutton 2000

Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000;**320**(7249):1574-7. [MEDLINE: 10845965]

**Towheed 2005**

Towheed TE, Maxwell L, Anastassiades TP, Shea B, Houpt J, Robinson V, et al. Glucosamine therapy for treating osteoarthritis. *Cochrane Database of Systematic Reviews* 2005, Issue 2. [DOI: [10.1002/14651858.CD002946.pub2](https://doi.org/10.1002/14651858.CD002946.pub2); MEDLINE: 15846645]

**Towheed 2006**

Towheed TE, Maxwell L, Judd MG, Catton M, Hochberg MC, Wells G. Acetaminophen for osteoarthritis. *Cochrane Database of Systematic Reviews* 2006, Issue 1. [DOI: [10.1002/14651858.CD004257.pub2](https://doi.org/10.1002/14651858.CD004257.pub2); MEDLINE: 16437479]

**Ulett 1998**

Ulett GA, Han S, Han JS. Electroacupuncture: mechanisms and clinical application. *Biological psychiatry* 1998;**44**(2):129-38. [MEDLINE: 9646895]

**van Tulder 2003**

van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine* 2003;**28**(12):1290-9. [MEDLINE: 12811274]

**Wegman 2004**

Wegman A, van der Windt D, van Tulder M, Stalman W, de Vries T. Nonsteroidal antiinflammatory drugs or acetaminophen for osteoarthritis of the hip or knee? A systematic review of evidence and guidelines. *The Journal of rheumatology* 2004;**31**(2):344-54. [MEDLINE: 14760807]

**White 2001**

White A, Hayhoe S, Hart A, Ernst E. Adverse events following acupuncture: prospective survey of 32 000 consultations with doctors and physiotherapists. *BMJ* 2001;**323**:485-6.

**White 2004**

White P, Lewith G, Prescott P, Conway J. Acupuncture versus placebo for the treatment of chronic mechanical neck pain:

a randomized, controlled trial. *Annals of internal medicine* 2004;**141**(12):911-9. [MEDLINE: 15611488]

**White 2007**

White A, Foster NE, Cummings M, Barlas P. Acupuncture treatment for chronic knee pain: a systematic review. *Rheumatology (Oxford)* 2007;**46**(3):384-90. [MEDLINE: 17215263]

**WHO 1991**

World Health Organization. A proposed standard international acupuncture nomenclature: report of a WHO scientific group. WHO. Geneva: World Health Organization, 1991.

**Wilcox 2006**

Wilcox CM, Allison J, Benzuly K, Borum M, Cryer B, Grosser T, et al. Consensus development conference on the use of nonsteroidal anti-inflammatory agents, including cyclooxygenase-2 enzyme inhibitors and aspirin. *Clinical gastroenterology and hepatology* 2006;**4**(9):1082-9. [MEDLINE: 16877048]

**Zhang 2007**

Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. *Osteoarthritis and cartilage* 2007;**15**(9):981-1000. [MEDLINE: 17719803]

**Zhang 2008**

Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis and cartilage* 2008;**16**(2):137-62. [MEDLINE: 18279766]

\* Indicates the major publication for the study

**CHARACTERISTICS OF STUDIES**
**Characteristics of included studies [ordered by study ID]**
**Berman 1999**

Methods	Design: Parallel Blinding: No Attempt to confirm patient blinding for sham control?: Not applicable (no sham control) Drop-outs/withdrawals: At 12 week follow-up, 8 drop-outs in acupuncture group and 7 in control group, according to <a href="#">Figure 1</a> flow chart; however, according to text, 7 drop-outs in acupuncture group and 8 in control group CBRG score: 1-1-1-0-0-0-1-1-1-1-1 Duration: 8 weeks treatment followed by 4 weeks follow-up Type of analysis reported: intention-to-treat analysis; per protocol/treatment received analysis also done but not reported
Participants	Setting: 1 outpatient clinic, Baltimore Mean age (+/-SD or Range): ~65 (8.5) Men/Women (n/n): 29/44 Recruitment method: advertisements on print and radio

**Acupuncture for peripheral joint osteoarthritis (Review)**

**Berman 1999** (Continued)

Mean pain duration (SD) years: 7(6)  
 For bilateral OA diagnosis, which knee treated/evaluated?: Not reported  
 Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA):  
 ACR criteria met  
 Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Kellgren 2 or  
 more  
 Minimum duration (and extent) of knee pain required to be eligible: 6 months with at least moderate  
 pain in knee for most days in last month  
 Hospital inpatients? (Y/N; if Y list number inpatients): Not reported  
 Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Not reported  
 Were people with a history of acupuncture treatment excluded: No  
 Other important inclusion criteria: Age 50 or above; taking analgesics or anti-inflammatories for pain  
 control for at least 1 month  
 Important exclusion criteria: Severe chronic or uncontrolled concomitant illness; corticosteroid injec-  
 tion into knee within four weeks preceding study entry

**Interventions**

**TEST GROUP INTERVENTION: Acupuncture**  
 N allocated to acupuncture: 36  
 Style of acupuncture: Chinese  
 Point selection: Formula  
 Points stimulated: GB34, 39, Sp6, 9, St35, 36, UB60, KI3, Ex32 (Xiyan); electrostimulation at St35 and  
 Ex32  
 Total length of treatment period (weeks): 8  
 Number of sessions target (mean): 16 (?mean)  
 Times per week: 2  
 Number of points used: 9 or 18 (depending if one or both knees affected)  
 Insertion depth: .4 to .6 inches  
 Was De qi reportedly sought?: Yes  
 Duration (mins): 20  
 Method of stimulation: Electrical 2.5 - 4 Hz at St 35 and EX32

**CONTROL GROUP A (sham, if used): standard of care - patients asked to remain on current level of oral  
 therapy**  
 N allocated to control group A: 37  
 Total length of treatment period: NA  
 Number of sessions target (mean): NA  
 Times per week: NA  
 (If relevant) Number of points used: NA  
 (If relevant) Insertion depth: NA  
 Was De qi sought?: NA  
 Duration (mins): NA  
 (If relevant) Method of stimulation: NA

Any co-interventions in all groups? Acupuncture group was asked to maintain previous analgesic/an-  
 ti-inflammatory regimen and control group asked to remain on current level of oral therapy

**Outcomes**

**PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLI-  
 CATIONS: MEASUREMENT TIME POINTS**

Pain: WOMAC Pain: Baseline, 4 weeks after baseline, 8 weeks after baseline, 4 weeks after cessation of  
 treatment (i.e., 12 weeks after baseline)  
 Function: WOMAC function: Baseline, 4 weeks after baseline, 8 weeks after baseline, 4 weeks after ces-  
 sation of treatment (i.e., 12 weeks after baseline)  
 Overall index of symptom severity: WOMAC total: Baseline, 4 weeks after baseline, 8 weeks after base-  
 line, 4 weeks after cessation of treatment (i.e., 12 weeks after baseline)  
 Type of outcome data reported (i.e., post treatment/change from baseline/both): Post-treatment  
 means and standard deviations for each group, and P values of within group changes  
 Additional outcomes reported in the trial but not abstracted: Lequesne function  
 Adverse effects: Reported as none for acupuncture group

**Berman 1999** (Continued)

Notes

Comments: Carefully reported study with waiting list control (waiting list patients received acupuncture after 12 weeks). Blinding of the patients not possible; follow-up only 4 weeks (after the 8 weeks of treatment)

Weak points: Not blinded

We excluded all outcome measurements that occurred after the patients on the waiting list began acupuncture treatment.

There was no standard deviations of changes from baseline directly reported for this trial. Therefore, in order to include this trial in the between group changes analysis, we converted the reported P values for the within group comparisons for 'baseline - to week 4', 'baseline - week 8', and 'baseline - week 12', into standard deviations of changes for each group at each measurement point. These recalculated standard deviations of changes were entered into RevMan.

Source of support: "This work was supported by the Maurice Laing Foundation and National Institutes of Health-National Center of Complementary and Alternative Medicine and the National Institutes of Arthritis/Musculoskeletal/Skin Diseases (Grant no. 1 R21-RR09327-01)."

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"Seventy-three patients with symptomatic knee OA were screened .....and randomized using computer-generated assignment."
Allocation concealment?	Low risk	"Block randomization to ensure balance within groups was carried out using opaque sealed envelopes."
Blinding? Versus sham	High risk	Acupuncture vs. standard treatment, no sham control.
Incomplete outcome data addressed? Short term	High risk	At 12 week follow-up, 8(37) drop-outs in acupuncture group and 7(36) in control group, according to <a href="#">Figure 1</a> flow chart; (however, according to text, 7 drop-outs in acupuncture group and 8 in control group.) One of the percentages of withdrawals and drop-outs exceeds 20%.
Free of selective reporting?	Low risk	"Patients were then asked to record their responses to the WOMAC (pain, disability and total) and Lequesne scales. ....Patient scores were ascertained at weeks 0, 4, 8 and 12 during the randomized trial." Findings reported in <a href="#">Table 4</a> , <a href="#">3</a> and <a href="#">Figure 3</a> .
Free of other bias?	Low risk	

**Berman 2004**

Methods

Design: Parallel

Blinding: Patient and outcomes assessor blinding for sham control comparison only. (Since the assessments were done by the patients themselves using questionnaires, we considered the sham arm of this trial to have been both patient and outcomes assessor blinded.) Neither patient nor outcomes assessor blinding for education control.

Attempt to confirm patient blinding for sham control?: Yes, at both 4 and 26 weeks after randomization. Most people in both groups believed that they were receiving acupuncture at both times. However, at both time points, more in acupuncture group knew they were receiving acupuncture, and it was a significant difference at 26 weeks.

Drop-outs/withdrawals: 33 in acupuncture group, 25 in sham acupuncture group, and 70 in education group voluntarily withdrew sometime during the 6 month trial.

CBRG score: 1-1-1-.5/0-0-.5/0-1-1/0-1/0-1-1 (number preceding / is sham control group score and number following / is education control group score)

**Berman 2004** (Continued)

Duration: 26 weeks of tapering treatment

Type of analysis reported: available case analysis with intention-to-treat (multiple imputation) as sensitivity analysis. Intention-to-treat results not presented but reported to be very similar to those using non-imputed data.

Participants	<p>Setting: 2 outpatient clinics and 1 clinical trials private research firm, Baltimore</p> <p>Mean age (+/-SD or Range): 65.5 (8.6)</p> <p>Men/Women (n/n): 205/365</p> <p>Recruitment method: print and radio advertisements</p> <p>Mean pain duration (SD) years: ~50% more than 5 years</p> <p>For bilateral OA diagnosis, which knee treated/evaluated?: If both knees were affected, 9 needles were inserted in each leg. Outcome measures not targeted to whether the patient had OA in 1 or both knees and no differential effects on basis of number of knees treated were observed.</p> <p>Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): Yes, rheumatologic exam</p> <p>Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Yes, Kellgren 2 or more</p> <p>Minimum duration (and extent) of knee pain required to be eligible: Moderate or greater clinically significant pain in knee on most days of preceding month</p> <p>Hospital inpatients? (Y/N; if Y list number inpatients): No</p> <p>Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Not reported</p> <p>Were people with a history of acupuncture treatment excluded?: Yes</p> <p>Other important inclusion criteria: age 50 or older</p> <p>Important exclusion criteria: serious medical conditions that precluded participation in study; bleeding disorders that might contraindicate acupuncture</p>
--------------	---

Interventions	<p>TEST GROUP INTERVENTION: Acupuncture</p> <p>N allocated to acupuncture: 190</p> <p>Style of acupuncture: Chinese</p> <p>Point selection: Formula</p> <p>Points stimulated: GB34, 39, SP6, 9, ST35, 36, EX32 (Xiyian), UB60, KI 3; electrostimulation at EX32 (Xiyian)</p> <p>Total length of treatment period (weeks): 26 weeks</p> <p>Number of sessions target (mean): 23 (22.7) (mean number averaged for true and sham groups; adherence to acupuncture appointments did not statistically significantly differ between the true and sham groups)</p> <p>Times per week: 8w:2/w + 2w:1/w + 4w: 0.5/w, 12w:1/m</p> <p>Number of points used: 9 or 18 (depending if one or both knees affected) plus two sham points (non-penetrating)</p> <p>Insertion depth: .3 to 1 inch, depending on point location</p> <p>Was De qi reportedly sought?: Yes</p> <p>Duration (mins): 20 mins</p> <p>Method of stimulation: electrostimulation at EX32</p> <p>CONTROL GROUP A (sham, if used): Combined insertion/non-insertion procedure (penetrating needles at 2 non-points, 2 tapes, tube at true points plus one non-inserted needle)</p> <p>N allocated to control group A: 191</p> <p>Total length of treatment period: 26 weeks</p> <p>Number of sessions target (mean): 23 (22.7) (mean number averaged for true and sham groups; adherence to acupuncture appointments did not statistically significantly differ between the true and sham groups)</p> <p>Times per week: 8w:2/w + 2w:1/w + 4w:0.5/w + 12w:1/m</p> <p>(If relevant) Number of points used: 9 or 18 non-penetrating sham plus two penetrating needles</p> <p>(If relevant) Insertion depth: Not stated</p> <p>Was De qi sought?: Not stated</p> <p>Duration (mins): 20</p> <p>(If relevant) Method of stimulation: Mock electrostimulation at non-penetrating knee points</p> <p>CONTROL GROUP B: Education - 6 two-hour group sessions based on the Arthritis Self-Management Program plus mailed educational materials</p> <p>N allocated to control group B: 189</p>
---------------	--



**Berman 2004** (Continued)

Total length of treatment period: Not applicable/not stated  
 Number of sessions target (mean): 6 (?mean)  
 Times per week: Not applicable/not stated  
 Duration (minutes): 120

Any co-interventions in all groups? Pain medication

**Outcomes**

PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS

Pain: WOMAC Pain: Baseline, 4 weeks after baseline, 8 weeks after baseline, 14 weeks after baseline (i.e., start of once/month treatment schedule), 26 weeks after baseline (i.e., end of 1x/month treatment schedule and end of study)

Function: WOMAC function: Baseline, 4 weeks after baseline, 8 weeks after baseline, 14 weeks after baseline (i.e., start of once/month treatment schedule), 26 weeks after baseline (i.e., end of 1x/month treatment schedule and end of study)

Overall index of symptom severity: WOMAC total: Baseline, 4 weeks after baseline, 8 weeks after baseline, 14 weeks after baseline (i.e., start of once/month treatment schedule), 26 weeks after baseline (i.e., end of 1x/month treatment schedule and end of study)

Type of outcome data reported (i.e., post treatment/change from baseline/both): means and standard errors of changes from baseline for each group reported in publication

Additional outcomes reported in the trial but not abstracted: physical health scores of SF-36 and 6 minute walk distance

Adverse effects: 26 in all: 14 in acupuncture group (none interpreted as treatment related); 5 in sham group; 7 in education group. The differences among groups did not reach statistical significance.

**Notes**

Comments: Very fastidious and careful study. Unusual treatment: 23 meetings with 16 treatments during the first 8 weeks, followed by a tapering treatment schedule.

Weak points: not blinded for education control arm; high drop out rates, in particular in the education group; after 26 weeks, the patients randomized to real acupuncture were more likely to believe they were receiving real acupuncture.

-----  
 14 week data counted as end of treatment for this trial, since after this point the treatments were reduced to once every month.

The SAS programmer for this trial (Wenlin Lee), provided the post-treatment means and standard deviations, as well as the mean changes and standard deviations of changes from baseline for the acupuncture and placebo acupuncture groups. These unadjusted means and standard deviations provided by the statistician were directly entered into RevMan, and used for the analyses (both post-treatment and between group changes analyses).

Source of support: "The National Center for Complementary and Alternative Medicine and the National Institute of Arthritis and Musculoskeletal and Skin Diseases provided funding for this study."

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"[A] cohort of 12 to 21 patients was formed, at which point each cohort at each site was randomly assigned to 1 of 3 groups by a computer-generated process using randomly selected blocks of 3, 6, and 9."
Allocation concealment?	Low risk	"We assured allocation concealment by using disguised letter codes that were generated and sent to the site coordinators by a central statistical core." "The research assistants who collected assessments from participants, the participants themselves (in the true acupuncture and sham acupuncture groups), and the statistician were blinded to group assignment."

**Berman 2004** (Continued)

Blinding? Versus sham	Unclear risk	"At 4 weeks, 67% in the true acupuncture group and 58% in the sham group believed that they were receiving true acupuncture (P = 0.06), and 25% and 33% were unsure, respectively. By the end of the trial, more individuals in the true group (75%) than in the sham group (58%) held this belief (P = 0.003), and 23% and 32% were unsure, respectively."
Incomplete outcome data addressed? Short term	Low risk	Observations available for analysis at 14 weeks were 158(190) in acupuncture group, 157(191) in sham acupuncture group, and 113(189) in education group.  Observations available for analysis at 26 weeks were 142(190) in acupuncture group, 141(191) in sham acupuncture group, and 108(189) in education group. There were 33(190) in acupuncture group, 25(191) in sham acupuncture group, and 70(189) in education group voluntarily withdrew sometime during the 6 month trial.
Free of selective reporting?	Low risk	WOMAC pain, function, stiffness, total and physical health scores of SF-36 and 6 minute walk distance were reported at baseline and at 4, 8 14 and 26 week. Serious adverse events during trial were reported. Findings reported in Table1, Table2.
Free of other bias?	Low risk	

**Christensen 1992**

Methods	<p>Design: Parallel</p> <p>Blinding: Outcomes assessor blinding for only the objective outcome measures. However, since the outcomes that were extracted were primarily assessed by the unblinded patients, we did not consider this trial to have used blinding.</p> <p>Attempt to confirm patient blinding for sham control?: Not applicable (no sham control)</p> <p>Drop-outs/withdrawals: 3 in first eight weeks</p> <p>CBRG score: ?-0-1-0-0-1/0-?-?-1-1-0 (for outcomes assessor blinded item, number preceding / is patient assessed outcome and number following / is objective outcome assessed by blinded observer)</p> <p>Duration: 3 weeks of acupuncture followed by six weeks of follow-up for randomized part of study</p> <p>Type of analysis reported: Available case analysis (probably)</p>
Participants	<p>Setting: Hospital outpatients, Denmark</p> <p>Mean age (+/-SD or Range): 69.2 (median) (range: 48-75)</p> <p>Men/Women (n/n): 9/20</p> <p>Recruitment method: 58 patients on county waiting list for knee replacement due to osteoarthritis were invited to participate</p> <p>Mean pain duration (SD) years: 4.3 years median (range: 6 months to 44 years)</p> <p>For bilateral OA diagnosis, which knee treated/evaluated?: Both knees</p> <p>Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): All patients waiting for knee replacement due to osteoarthritis of the knee</p> <p>Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): All knees x-rayed and classified according to Ahlback classification</p> <p>Minimum duration (and extent) of knee pain required to be eligible: Not reported</p> <p>Hospital inpatients? (Y/N; if Y list number inpatients): Not reported</p> <p>Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Not reported</p> <p>Were people with a history of acupuncture treatment excluded? No, but acupuncture treatment within past year was an exclusion criterion</p> <p>Other important inclusion criteria: Not stated</p> <p>Important exclusion criteria: Neurological diseases, psychiatric disorders, connective tissue disease affecting knee, patients who experienced infection of knee or changed analgesic during course of study were excluded</p>

**Christensen 1992** (Continued)

Interventions	<p>TEST GROUP INTERVENTION: acupuncture          N allocated to acupuncture: 17?          Style of acupuncture: Chinese          Point selection: Formula          Points stimulated: ST34, 35, 36, SP10, EX32 (Xiyian), LI4          Total length of treatment period (weeks): 3          Number of sessions target (mean): 6 (?mean)          Times per week: 2          Number of points used: 6 or 12 depending if one or both knees affected          Insertion depth: 10-15 mm          Was De qi reportedly sought?: Yes          Duration (mins): 20          Method of stimulation: Manual stimulation of needles</p> <p>CONTROL GROUP A (sham, if used): Waiting list          N allocated to control group A: 15?          Total length of treatment period: Not applicable          Number of sessions target (mean): Not applicable          Times per week: Not applicable          (If relevant) Number of points used: Not applicable          (If relevant) Insertion depth: Not applicable          Was De qi sought?: Not applicable          Duration (mins): Not applicable          (If relevant) Method of stimulation: Not applicable</p> <p>Any co-interventions in all groups? Analgesics</p>
Outcomes	<p>PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS</p> <p>(Outcome data not included in meta-analysis because of confusing presentation and difficulties in data interpretation.)          Pain: VAS pain intensity scale: Baseline, end of treatment, and 2, 3, and 4 weeks after end of treatment          Function: HSS knee function scale and walking distance: Baseline, end of treatment and 4 weeks after end of treatment          Total: WOMAC total          Type of outcome data reported (i.e., post treatment/change from baseline/both): (Median values, ranges, and 95% confidence limits of median difference at the post-treatment measurement points)          Additional outcomes reported in the trial but not abstracted: Analgesic use          Adverse effects: Only reported for total group</p>
Notes	<p>Comments: Carefully carried out, but somewhat confusing presentation of data makes study difficult to interpret, which is the reason that this trial did not contribute to effect size estimates. The patients in the waiting list group received acupuncture after 9 weeks and all results after 9 weeks were pooled for the two treated groups, and therefore, post-9-week results were not extracted.</p> <p>Weak points: Not blinded, poor report, small sample size</p> <p>Source of support: Not reported.</p> <p>For this trial, the method of generating the randomization sequence was not described in the publication. According to the corresponding author "The patients were randomized at week 0 by drawing sealed opaque envelopes to group A or B." Based on the author's description, the method of randomization was adequate and the treatment allocation was concealed. Including this information from the RCT author, the Jadad scale score for this trial would be increased from 1 (based on the publication alone) to 2, and the Cochrane Back Review Group score would be increased from 4/3 (based on the publication alone) to 6/5.</p> <p>Overall conclusions: This trial compared acupuncture to usual care and found a statistically significant reduction in pain, analgesic consumption, and most objective measures in the acupuncture group relative to the control.</p>

**Christensen 1992** (Continued)

The overall conclusions of the RCT authors, as reported in the abstract, were the following: "Comparing group A [acupuncture treatment group] to B [no-treatment control group] there was a significant reduction in pain, analgesic consumption and in most objective measures."..."Conclusions: Acupuncture can ease the discomfort while waiting for an operation and perhaps even serve as an alternative to surgery."

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"At the start of the short-term study, the patients were randomly assigned to Group A or B by drawing lots."  The method of generating the randomization sequence was not described in the publication. According to the corresponding author "The patients were randomized at week 0 by drawing sealed opaque envelopes to group A or B." Based on the author's description, the method of randomization was adequate.
Allocation concealment?	Low risk	The method of generating the randomization sequence was not described in the publication. According to the corresponding author "The patients were randomized at week 0 by drawing sealed opaque envelopes to group A or B." Based on the author's description, the treatment allocation was concealed.
Blinding? Versus sham	High risk	Acupuncture vs. waiting list. no sham control.
Incomplete outcome data addressed? Short term	Low risk	"Three out of the 32 patients did not participate from week 0. They received acupuncture treatment with Group B, and they have been excluded from the rest of the results of the short-term study, leaving 10 women and four men in Group A and 10 women and 5 men in Group B."
Free of selective reporting?	Low risk	WOMAC pain, function, stiffness, total and physical health scores of SF-36 and 6 minute walk distance were reported at baseline and at 4, 8 14 and 26 week. Serious adverse events during trial were reported. Findings reported in <a href="#">Table 3</a> , <a href="#">Table 4</a> .
Free of other bias?	High risk	Co-interventions were not reported.  Compliance acceptability condition was not reported.  ITT analysis was not applied.

**Fink 2001**

Methods	Design: Parallel Blinding: successful blinding of patients. Physicians carrying out the clinical evaluations were also blinded. Attempt to confirm patient blinding for sham control?: Yes, credibility and expectation assessed using a "modified Borkovec Expectancy Scale." Patients rated credibility of acupuncture and sham much as the same. The authors noted that "After the first treatment session, a modified Credibility Scale showed no difference between the groups concerning the expectation, acceptance and credibility for the ongoing treatment. Therefore, coping and expectation should not influence treatment results." Drop-outs/withdrawals: Out of 67 patients randomized (33 to acupuncture and 34 to sham), 65 were available at first follow-up (33 in acup group; 32 in sham group) and 62 were available at second follow-up (32 in acup group; 30 in sham group). Only 41/67 were available at the 6 months follow-up (i.e. third follow-up) of which only 17/34 were available at the third follow-up for the acupuncture group. Because of the high drop-outs at the third follow-up time point, the RCT authors did not analyze the da-
---------	---

**Fink 2001** (Continued)

ta from this time point in their statistical analysis. Similarly, because at the third follow-up time point, 48% of the acupuncture group was unavailable for analysis, an ad hoc decision was made to exclude this third follow-up time point data from this meta-analysis. This trial did not earn the point for acceptability of drop-out rate on the CBRG scale because of the high numbers of drop-outs at the long-term time point. However, our subgroup analyses used the short-term time point, at which time this trial had low drop-outs, and therefore this trial was classified as having met the "acceptability of drop-outs" criteria for our subgroup analyses.

CBRG score: 1-?-1-1-0-1-?-0-0-1-0

Duration: 3 weeks of acupuncture, patients followed up for 6 months

Type of analysis reported: Not explicitly stated but appears to be a per protocol analysis, since, for example, [Table 6](#) indicates that patients who had a total hip replacement between follow-up 2 and follow-up 3 did not appear to be included in the follow-up 3 analysis. Thus, it appears that only those participants who completed the trial and complied with their allocated treatments were included in the analysis.

**Participants**

Setting: a university department for physical medicine and rehabilitation, Hannover, Germany

Mean age (+/-SD or Range): ~62 (9)

Men/Women (n/n): 22/43

Recruitment method: through advertisements in a local newspaper

Mean pain duration (SD) years: ~5.2(3.8) is mean "duration of complaints"

For bilateral OA diagnosis, which hip treated/evaluated?: Not reported

Diagnosis of hip OA required to be eligible? (if yes, describe how patients were verified to have OA): Yes, hip arthritis with reduced range of motion and significant radiographic changes, 2° minimum as per Kellgren-score on x-ray of pelvis from the last year

Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Yes (Kellgren 2° minimum)

Minimum duration (and extent) of hip pain required to be eligible: Complaints for at least 6 months and pain on most days of previous month

Hospital inpatients? (Y/N; if Y list number inpatients): Not reported (assumed none)

Previous hip surgery? (Y/N; if Y list number with previous knee surgery): Not reported

Were people with a history of acupuncture treatment excluded?: Not reported

Other important inclusion criteria: None other than those described above

Important exclusion criteria: "Exclusion criteria were mainly supposed to eliminate all factors influencing the symptoms of hip arthritis during the intervention, i.e. physical therapy or medical treatment"

**Interventions**

TEST GROUP INTERVENTION: acupuncture

N allocated to acupuncture: 23

Style of acupuncture: Chinese

Point selection: Formula

Points stimulated: "Within the hip area, six pressure sensitive locations ('Ah-Shi'-points) were used.

In addition, the regional meridian points 'GB-30', 'GB-31', 'BL-37' and the distal meridian points 'ST-40' and 'BL-54' were chosen, as well as the master point for tendons and muscles 'GB-34'."

Total length of treatment period (weeks): 3

Number of sessions target (mean): 10 (?mean)

Times per week: assumed about 3 because 10 sessions total and duration is 3 weeks

Number of points used: 12

Insertion depth: Not reported (but needles were inserted deeper until subjects experienced Te chi)

Was De qi reportedly sought?: Yes

Duration (mins): 20

Method of stimulation: Manual manipulation (not reported when during the treatment session or how many times during the treatment session)

CONTROL GROUP A (sham, if used): needle acupuncture "but the selected puncture sites were at least 5 cm away from the classical acupuncture points and their interconnecting lines (meridians) and also clear of painful pressure points (Ah-Shi or trigger points)."

N allocated to control group A: 34

Total length of treatment period: 3 weeks

Number of sessions target (mean): 10

Times per week: about 3

(If relevant) Number of points used: same number as true acupuncture groups

**Fink 2001** (Continued)

(If relevant) Insertion depth: same depth as true acupuncture group  
 Was De qi sought?: No  
 Duration (mins): same as true acupuncture group  
 (If relevant) Method of stimulation: None

**Outcomes**

PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS

Pain: VAS scale: Baseline (one week before the beginning of treatment but assumedly after randomization), 2 weeks after the end of treatment (i.e. ~6 weeks after randomization), 2 months after end of treatment (i.e. ~3 months (or slightly less) after randomization), and 6 months after end of treatment (i.e. ~7 months after randomization)

Function: Lequesne: same time points as for pain measurement

Overall index of symptom severity: None (although there was an 'overall assessment of patient satisfaction' measure, this measure was not considered to be comparable to a measure of overall index of symptom severity (e.g. WOMAC total) and therefore we did not extract data for this 'overall assessment' outcome)

Type of outcome data reported (i.e., post treatment/change from baseline/both): Post-treatment means and standard deviations only (data extracted from charts)

Additional outcomes reported in the trial but not abstracted: 'overall assessment of patient satisfaction' measure and quality of life measure

Adverse effects: "During the study course, no side effects occurred."

**Notes**

Because the changes from baseline values were not reported, they were imputed for the changes from baseline analysis, primarily so that the data from this study could be combined with changes from baseline data for the knee studies.

The time points of the F1 and F2 follow-ups were described differently in the text and in the figures 3 and 4. The F1 was described as 2 weeks after the end of treatment in the text and as 1 week after the end of treatment in the figures. The F2 was described as 2 months after the end of treatment in the text and 6 weeks after end of treatment in the figures. The choice of either F1 or F2 for the closest to 8 weeks short-term analysis would depend on whether the description in the figure or the manuscript was correct. I assumed that the time point listed in the figure was correct, and therefore used the F2 time point for the primary short-term meta-analysis time point closest to 8 weeks and less than 3 months after randomization. (The F2 time point was chosen to be more conservative, as the effects were smaller at this time point and also because it was a longer follow-up time than the F1 time point.). For the sensitivity analysis, in which the 'short-term' follow-up was classified as the 'end of treatment', the F1 time point was used.

A questionnaire was used to assess the expectation and credibility. However, the results data from this questionnaire were not reported in the Results section. Only in the Discussion section was the following reported: "After the first treatment session, a modified Credibility Scale<sup>19</sup> showed no difference between the groups concerning the expectation, acceptance and credibility for the ongoing treatment. Therefore, coping and expectation should not influence treatment results." Based on this text in the Discussions section, for our risk of bias scoring, we assigned 1 point to this trial to indicate that this trial evaluated the credibility of the sham and found it to be indistinguishable from true acupuncture. However, there is some uncertainty because the credibility was only assessed after the first treatment, and the numerical results of this assessment are not reported.

Source of support: "This study was supported in part by a grant from the PharmaMED Foundation Germany." PharmaMED is the IDA subsidiary in Malta, and is a pharmaceuticals company.

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"Written informed consent was obtained and patients were then randomly assigned to the treatment groups using computer generated random codes."
Allocation concealment?	Unclear risk	The method of concealment is not described.

**Fink 2001** (Continued)

Blinding? Versus sham	Low risk	<p>“The physician carrying out clinical evaluation of the study parameters was not informed about the acupuncture treatment, so he could not decide to which treatment group each patient belonged.”</p> <p>“Needle acupuncture was also performed for the control group, but the selected puncture sites were at least 5 cm away from the classical acupuncture points and their interconnecting lines (meridians) and also clear of painful pressure points (Ah-Shi or trigger points).”</p> <p>“A modified Borkovec Expectancy Scale<sup>19</sup> was used to assess credibility and expectation. The items in this questionnaire are listed in <a href="#">Table 2</a>.”</p> <p>“After the first treatment session, a modified Credibility Scale showed no difference between the groups concerning the expectation, acceptance and credibility for the ongoing treatment. Therefore, coping and expectation should not influence treatment results.”</p>
Incomplete outcome data addressed? Short term	Low risk	<p>According to <a href="#">Table 6</a>, out of 67 patients randomized (33 to acupuncture and 34 to sham), 32(33) in acup group and 30(34) in sham group were available at second follow-up (6 weeks after end of treatment).</p> <p>According to <a href="#">Table 6</a>, only 41(67) were available at the 6 months follow-up (i.e. third follow-up) of which only 17(34) in acup group and 24(33) in sham group were available at the third follow-up.</p>
Free of selective reporting?	Low risk	<p>“Carlsson’s ‘Comparative Scale’, which evaluates overall patient satisfaction, was used after 2 months (F2). To assess alterations in the quality of life, Bullinger’s ‘Everyday Life’ questionnaire was used?. Hip index according to Lequesne<sup>14</sup> determines the impairment caused by pain and restriction of range of movement. For pain measurement, the Visual analogue Scale (VAS) was used.” The above four scales were reported at baseline and at 1 and 6 weeks, 6 months after end of treatment. Findings reported in Fig 1-4.</p> <p>Adverse events during trial were reported. “During the study course, no side effects occurred.”</p>
Free of other bias?	High risk	<p>Co-interventions were not reported.</p> <p>No ITT analysis was applied.</p> <p>It appears to be a per protocol analysis, since, for example, <a href="#">Table 6</a> indicates that patients who had a total hip replacement between follow-up 2 and follow-up 3 did not appear to be included in the follow-up 3 analysis. Thus, it appears that only those participants who completed the trial and complied with their allocated treatments were included in the analysis.</p>

**Foster 2007**

Methods	<p>Design: Parallel</p> <p>Blinding: "good credibility and blinding of participants receiving the nonpenetrating acupuncture intervention, and successful blinding of the research team"</p> <p>Attempt to confirm patient blinding for sham control?: Yes, credibility assessed two weeks (and 6 weeks) after treatment started. Patients rated credibility of acupuncture and sham much as the same and very high. "Participants receiving either acupuncture intervention were significantly more confident that treatment could help their knee problem than those receiving advice and exercise alone."</p> <p>Drop-outs/withdrawals: "Follow-up rate at six months was 94%."</p> <p>CBRG score: 1-1-1-1/0-0-1/0-1/0-1-1-1-0 (number preceding / is sham control group score and number following / is advice and exercise control group score)</p>
---------	--

**Acupuncture for peripheral joint osteoarthritis (Review)**

**Foster 2007** (Continued)

Duration: 3 weeks of acupuncture, patients followed up for 12 months

Type of analysis reported: Methods section states that "Analysis was by intention to treat."; however, the Table of Results (i.e. [Table 2](#)) presents the "No analyzed" for each time point, and this is the available number of cases at that time point, rather than the number randomized. Dr. Foster, the principle investigator provided the following clarification: "We analysed people according to the treatment they were randomised to rather than the treatment they received. Although in the end we had very few protocol violators. The reasons we gave "numbers analysed" was because some people did not complete FU questionnaires and we did not impute any FU data as we had such high response rates." Therefore although no imputations were made for a full intention-to-treat analysis that includes all participants regardless of whether their outcomes were collected, the available case analysis used in this trial is very unlikely to result in bias because the missing outcome data was balanced in numbers across interventions groups, with similar reasons for missing data across groups, and the effect sizes among the missing outcomes were not enough to impact to any clinically relevant extent on the observed effect sizes.

Participants	<p>Setting: "37 physiotherapy centres accepting primary care patients referred from general practitioners in the Midlands, United Kingdom."</p> <p>Mean age (+/-SD or Range): ~63.5(9)</p> <p>Men/Women (n/n): 136/216</p> <p>Recruitment method: "Participants were recruited from 37 NHS physiotherapy centres providing services for general practices within the Midlands and Cheshire regions of the United Kingdom between November 2003 and October 2005..." "They were referred to physiotherapy by their general practitioner"</p> <p>Mean pain duration (SD) years: ~40% less than 1 year</p> <p>For bilateral OA diagnosis, which knee treated/evaluated?: Not reported [if bilateral problems - both knees were treated]</p> <p>Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): Clinical diagnosis</p> <p>Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): No, clinical diagnosis only</p> <p>Minimum duration (and extent) of knee pain required to be eligible: Not reported [there was no minimum but all patients were consultants to their family doctor because of their knee problem]</p> <p>Hospital inpatients? (Y/N; if Y list number inpatients): No</p> <p>Previous knee surgery? (Y/N; if Y list number with previous knee surgery): No</p> <p>Were people with a history of acupuncture treatment excluded?: Yes</p> <p>Other important inclusion criteria: age 50 or older</p> <p>Important exclusion criteria: Patients with potentially serious pathology (e.g. inflammatory arthritis, malignancy etc) on the basis of general practice or physiotherapy diagnosis or from past medical history, those who have had a knee or hip replacement on the affected side(s), are already on a surgical waiting list for total knee replacement, or for whom the trial interventions are contraindicated are excluded from the trial. Those who have received an exercise programme, from a physiotherapist, for their knee problem within the last 3 months or an intra-articular injection to the knee in the last 6 months are also excluded.</p>
Interventions	<p>TEST GROUP INTERVENTION: Acupuncture</p> <p>N allocated to acupuncture: 117</p> <p>Style of acupuncture: Chinese</p> <p>Point selection: Flexible formula</p> <p>Points stimulated: "For each individualised treatment session between six and 10 acupuncture points from 16 commonly used local and distal points were selected. Local points were Sp 9, Sp 10, St 34, St 35, St 36, Xiyan, Gb 34, and trigger points. Distal points were LI 4, TH5, Sp 6, Liv 3, St 44, Ki 3, Bl 60, and Gb 41."</p> <p>Total length of treatment period (weeks): 3 weeks</p> <p>Number of sessions target (mean): 6 (median: 6)</p> <p>Times per week: assumedly 2x/week since 6 treatments were delivered over 3 weeks [correct]</p> <p>Number of points used: between 6 and 10</p> <p>Insertion depth: "the depth of insertion was between 5 mm and 25 mm, depending on the points selected"</p> <p>Was De qi reportedly sought?: Yes</p> <p>Duration (mins): 25-35 mins</p>



**Foster 2007** (Continued)

Method of stimulation: manual stimulation only

CONTROL GROUP A (sham, if used): "...non-penetrating acupuncture at the same points as the true acupuncture" "Patients randomized to receive...non-penetrating acupuncture had acupuncture delivered through needles with a blunt tip. The shaft of these needles collapses into the handle, creating an illusion of insertion. They meet the recommendations for acceptable controls for acupuncture research<sup>18</sup> and have been used in previous trials.<sup>22 31</sup> Some authors believe that a small acupressure effect may be induced by this method.<sup>31</sup> The same protocol was used as for true acupuncture, thus all criteria for harnessing nonspecific effects were included..."

N allocated to control group A: 119

Total length of treatment period: 3 weeks

Number of sessions target (mean): 6 (median = 6)

Times per week: assumedly 2x/week since 6 treatments were delivered over 3 weeks [correct]

(If relevant) Number of points used: same as true acupuncture group

(If relevant) Insertion depth: Non-insertive needles

Was De qi sought?: De qi was not sought, but patients were asked to report the sensations when the needles were placed and "a considerable proportion of participants in our non-penetrating acupuncture group reported sensations fitting the normal descriptions of de qi"

Duration (mins): same as true acupuncture

(If relevant) Method of stimulation: None

CONTROL GROUP B: Advice and exercise -"The package consisted of up to six sessions of 30 minutes (including the prerandomisation session) over six weeks."

N allocated to control group B: 116

Total length of treatment period: 6 weeks

Number of sessions target (mean): 6 (median = 6)

Times per week: 1x/week

Duration (minutes): 30 minutes

Any co-interventions in all groups? Advice and exercise, although the advice and exercise co-intervention was delivered over six weeks in the 'advice and exercise' group and delivered over 3 weeks in the acupuncture and sham acupuncture groups. In a personal e-mail, Dr. Foster also reported the following: "We allowed co-interventions in all 3 groups as per usual primary care - such as medications and gathered self-report data on these (grouped as analgesics or NSAIDS) at each visit. Other co-interventions were avoided during the treatment period."

**Outcomes**

PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS

Pain: WOMAC Pain: Baseline, 6 weeks, 6 months, and 12 months after randomization

Function: WOMAC function: Baseline, 6 weeks, 6 months, and 12 months after randomization

Total: WOMAC total: Not reported [Provided by Dr. Foster]

Type of outcome data reported (i.e., post treatment/change from baseline/both): Post-treatment means and standard deviations for each group as well as change scores (i.e. reported in [Table 2](#) as both crude changes from baseline as well as changes adjusted for recruitment age, sex, duration of pain, and scores for pain or function)

Additional outcomes reported in the trial but not abstracted: "participants' global assessment of change compared with baseline, pain severity and unpleasantness, severity of patient nominated main functional problem, arthritis self efficacy, satisfaction with care, and adherence with exercises, and all measures at six weeks and 12 months. Side effects of treatment, adverse events, and use of co-interventions were also recorded. We identified the proportion of each group who achieved a clinically significant response according to criteria from the outcome measures in Rheumatology and Osteoarthritis Research Society international initiative (OMERACT-OARSI)"

Adverse effects: "No adverse events occurred in the advice and exercise group or in the advice and exercise plus nonpenetrating acupuncture group. Five adverse events were reported for participants receiving true acupuncture (pain, sleepiness, fainting, nausea, and swelling around the treated knee)."

**Notes**

Nadine Foster provided the WOMAC totals for all time points. These were not reported in the BMJ publication.

**Foster 2007** (Continued)

Source of support: "This study was supported by a project grant from the Arthritis Research Campaign, UK (grant H0640) and Support for Science funding secured by the North Staffordshire Primary Care Research Consortium for NHS service support costs." Assume it's funded by government grant.

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	<p>"[T]he physiotherapist telephoned an administrator at the research centre to record the selected acupuncture points and to receive a computed generated randomisation group for the patient."</p> <p>"Using random permuted blocks of 12 (blocked by treatment centre) the participants were randomised to advice and exercise, advice and exercise plus true acupuncture, or advice and exercise plus nonpenetrating acupuncture."</p>
Allocation concealment?	Low risk	<p>"[T]he physiotherapist telephoned an administrator at the research centre to record the selected acupuncture points and to receive a computed generated randomisation group for the patient."---Central allocation</p> <p>"Researchers who collected, entered, and analysed data were unaware of treatment allocation. By necessity the physiotherapists delivering the interventions were not blind to allocation."</p>
Blinding? Versus sham	Low risk	<p>"Participants were not told they may receive a sham intervention (rather they were told they may receive one of two forms of acupuncture)."</p> <p>" [We] achieved good credibility and blinding of participants receiving the non-penetrating acupuncture intervention, and successful blinding of the research team"</p> <p>"Acceptability and credibility of the intervention groups were evaluated during a telephone call two weeks after the start of treatment and at six weeks by questionnaire, using items adapted from previous literature.<sup>36</sup>"</p> <p>"Table 4 summarises treatment credibility two weeks after treatment started. Participants receiving either acupuncture intervention were significantly more confident that treatment could help their knee problem than those receiving advice and exercise alone."</p> <p>" Satisfaction with care was significantly greater for participants receiving advice and exercise plus non-penetrating acupuncture than for those receiving advice and exercise alone (table 8). No statistically significant differences were seen between advice and exercise plus true acupuncture and advice and exercise plus nonpenetrating acupuncture for any other of the outcomes measured (data not shown)."</p> <p>--Patients rated credibility of acupuncture and sham much as the same and very high.</p> <p>"Participants receiving either acupuncture intervention were significantly more confident that treatment could help their knee problem than those receiving advice and exercise alone."</p> <p>See also allocation concealment</p>
Incomplete outcome data addressed? Short term	Low risk	<p>Returned questionnaire at 6 weeks is 105(116) in advice and exercise group, 113(116) in advice, exercise and true acupuncture group, 115(119) in advice, exercise, and non-penetrating acupuncture group. Finding reported in Fig.1</p> <p>"Follow-up rate at six months was 94%."</p>

**Foster 2007** (Continued)

Returned questionnaire at 6 weeks is 105(116) in advice and exercise group, 109(116) in advice, exercise and true acupuncture group, 113(119) in advice, exercise, and non-penetrating acupuncture group. Finding reported in Fig.1

Free of selective reporting?	Low risk	"Follow-up was at two weeks (by telephone), six weeks, and six and 12 months (by postal questionnaire) after randomisation. The primary outcome measure was change at six months in the pain subscale score of the Western Ontario and McMaster Universities osteoarthritis index (Likert 3.0). Secondary outcomes at six months included the function subscale score of the Western Ontario and McMaster Universities osteoarthritis index, participants' global assessment of change compared with baseline, pain severity and unpleasantness, severity of patient nominated main functional problem, arthritis self efficacy, satisfaction with care, and adherence with exercises, and all measures at six weeks and 12 months. Side effects of treatment, adverse events, and use of co-interventions were also recorded. We identified the proportion of each group who achieved a clinically significant response according to criteria from the outcome measures in Rheumatology and Osteoarthritis Research Society international initiative (OMERACT-OARSI)." Findings reported in <a href="#">Table 3</a> , 4-8.
------------------------------	----------	---

Free of other bias?	Low risk
---------------------	----------

**Haslam 2001**

Methods	Design: Parallel Blinding: None Attempt to confirm patient blinding for sham control?: NA Drop-outs/withdrawals: All acupuncture patients available for post-treatment assessment and 1 acupuncture patient dropped out later and was unavailable for the eight week follow-up assessment. In the advice and exercise group, 4 patients were unavailable at the post-treatment assessment and an additional 3 patients dropped out between the post-treatment assessment and the eight-week follow-up.  CBRG score: 1-?-1-0-0-?-?-0-1-0 Duration: 6 weeks of acupuncture, patients followed up for 8 additional weeks, for a total duration of 14 weeks Type of analysis reported: Unclear -- assume available case. The 4 dropouts in the control group during the intervention period appear to have been excluded from the analysis, as suggested in <a href="#">Table 3</a> .
---------	--

Participants	Setting: Physiotherapy department in a District General Hospital, Princess Margaret hospital, Swindon Mean age (+/-SD or Range): ~67 Men/Women (n/n): 7/21 Recruitment method: Participants were recruited from those on a waiting list for a total hip arthroplasty Mean pain duration (SD) years: ~8 years For bilateral OA diagnosis, which hip treated/evaluated?: Not reported Diagnosis of hip OA required to be eligible? (if yes, describe how patients were verified to have OA): Yes, radiographic changes consistent with OA Radiologic evidence of hip OA required to be eligible? (if yes, describe requirement): Yes Minimum duration (and extent) of knee pain required to be eligible: Not reported Hospital inpatients? (Y/N; if Y list number inpatients): No Previous hip surgery? (Y/N; if Y list number with previous knee surgery): No Were people with a history of acupuncture treatment excluded?: Not reported Other important inclusion criteria: None reported Important exclusion criteria: Patients with rheumatoid arthritis, previous hip injection or surgery, intra-articular steroid injection in last three months
--------------	---

Interventions	TEST GROUP INTERVENTION: acupuncture N allocated to acupuncture: 16
---------------	--

**Haslam 2001** (Continued)

Style of acupuncture: Chinese  
 Point selection: Formula  
 Points stimulated: "GB29, GB30, GB34, GB43, ST44, LI4 bilaterally, and four 'ah shi' points around the great trochanter, in a north, south, east, west formation."  
 Total length of treatment period (weeks): 6  
 Number of sessions target (mean): 6 (?mean)  
 Times per week: 1  
 Number of points used: 16  
 Insertion depth: Not reported  
 Was De qi reportedly sought?: Not reported  
 Duration (mins): 10 minutes for first session, and 25 minutes for subsequent sessions  
 Method of stimulation: For first session, "Each needle was manually stimulated for approximately ten seconds half way through the treatment." For subsequent sessions, there was "manual stimulation of each needle, for approximately ten seconds, every five minutes."

CONTROL GROUP A (sham, if used): advice and exercise "Group B attended the department three times over the same time period. Each session lasted 30 minutes. The first session involved giving each patient an advice sheet on OA, and on self-help. The advice sheet also contained a set of five exercises that were demonstrated to the patient, who then practiced them to the author's satisfaction. Three weeks later the patients came back for a review to check that they were doing their exercises correctly, and they were advised, if appropriate, on how to gently progress. Three weeks later the patients came back for their final review and were encouraged to continue with the exercises and follow the advice given."

N allocated to control group A: 16  
 Total length of treatment period: 6 weeks  
 Number of sessions target (mean): 3 (? mean)  
 Times per week: 1x/3 weeks (i.e. baseline session; 3 weeks post-baseline session; 6 week post-baseline session)  
 (If relevant) Number of points used: NA  
 (If relevant) Insertion depth: NA  
 Was De qi sought?: NA  
 Duration (mins): NA  
 (If relevant) Method of stimulation: NA

Any co-interventions in all groups?: None reported

**Outcomes**

A slightly modified version of the WOMAC (visual analogue scale version) was used. Their modified version had 16 questions rather than the full 24 questions. The major modification was that some of the function questions from the original WOMAC were dropped so that the questionnaire could be made shorter. Authors only reported WOMAC totals and not the WOMAC pain and function sub-scores. For the post-treatment scores analysis, the WOMAC total post-treatment means and SDs were extracted from [Table 4](#). For the changes from baseline analysis, the authors only reported the Ns for each group and the P value of the difference in changes between the two groups (i.e.  $P = .02$ , reported only in the abstract) for the short-term follow-up (i.e. 6 weeks after randomization which was the end of treatment time point), which were then entered into the Comprehensive Meta-analysis program to calculate the SMD and standard error. The SMDs and SEs were then in turn entered into RevMan using the Generic Inverse Variance data entry screen. The SMDs calculated using the between group change P values were much higher than the SMDs calculated using the reported post-treatment means and SDs.

**Notes**

Patients had severe OA considering they were on waiting list for a total hip arthroplasty. The dropout rate in the control group was very high (almost 44%) at the 14 week follow-up point. Although the 14 week outcome data was reported (i.e. in [Table 4](#) for post-treatment means and SDs, and in abstract as between group change p-value), a post hoc decision was made to exclude this 14 week follow-up data from the long-term effects meta-analysis because of the high attrition.

Source of support: Not reported.

**Risk of bias**
**Bias**

**Authors' judgement    Support for judgement**

**Haslam 2001** (Continued)

Adequate sequence generation?	Low risk	“The patients were randomly allocated, using a random number table, to group A or B, with 16 in each group.”
Allocation concealment?	Unclear risk	The method of concealment is not described.
Blinding? Versus sham	High risk	Acupuncture vs. advice and exercise, no sham control.
Incomplete outcome data addressed? Short term	High risk	“Group A remained intact but in group B, four patients withdrew from the study during the intervention period. One patient was admitted for arthroplasty, another withdrew after the first session after having a fall at home, another withdrew after the first session due to difficulty in getting to the department because of poor mobility. The fourth patient attended the first session then did not attend the second and could not be re-contacted. Three further patients dropped out of group B between the post-treatment assessment and the eight-week follow-up. One reported having had an arthroplasty and the other two were not contactable. One patient dropped out of group A during this period to have a total hip replacement.” 1(16) acupuncture patient was unavailable for the eight week follow-up assessment. In the advice and exercise group, 7(16) patients were unavailable at the eight-week follow-up.
Free of selective reporting?	High risk	<p>“A slightly modified version of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index was used as the outcome measure (Appendix 1)” at baseline, post-treatment and 8 weeks follow-up.</p> <p>From author: Their modified version had 16 questions rather than the full 24 questions. The major modification was that some of the function questions from the original WOMAC were dropped so that the questionnaire could be made shorter. Authors only reported WOMAC totals and not the WOMAC pain and function sub-scores. (unpublished information).</p> <p>Findings reported in <a href="#">Table 4</a>, <a href="#">Figure 1</a>.</p>
Free of other bias?	High risk	<p>Co-interventions were not reported.</p> <p>Patients compliance condition was not reported.</p> <p>No ITT analysis was applied.</p>

**Molsberger 1994**

Methods	Design: Parallel Blinding: Patient and outcomes assessor blinded Attempt to confirm patient blinding for sham control?: No Drop-outs/withdrawals: 6 out of 103  CBRG score: ?-0-1-.5-0-.5-1-?-0-1-0 Duration: 5 weeks treatment and 3 months follow-up Type of analysis reported: Analysis of only the 97 'completers'
Participants	Setting: Setting unclear, Germany Mean age (+/-SD or Range): 59.7 Men/Women (n/n): 36/61 Recruitment method: Not reported Mean pain duration (SD) years: 7.8 years mean duration of disease For bilateral OA diagnosis, which knee treated/evaluated?: Not reported Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): Diagnosis confirmed for at least a year

**Acupuncture for peripheral joint osteoarthritis (Review)**

**Molsberger 1994** (Continued)

Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Wirth criteria I-IV  
 Minimum duration (and extent) of knee pain required to be eligible: Chronic knee pain for at least two months  
 Hospital inpatients? (Y/N; if Y list number inpatients): Not reported  
 Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Excluded  
 Were people with a history of acupuncture treatment excluded?: Not reported  
 Other important inclusion criteria: Not medicated during the study and no other therapy during the study  
 Important exclusion criteria: No other symptomatic diseases (e.g., RA) or obvious psychiatric problems; patients with other joint pain on the knee

**Interventions**

TEST GROUP INTERVENTION: acupuncture  
 N allocated to acupuncture: unclear (71 analyzed)  
 Style of acupuncture: Chinese  
 Point selection: Formula  
 Points stimulated: on ST34, 35, 36, SP9, 10, UB40, GB 34, EX31, EX32  
 Total length of treatment period (weeks): 5 weeks  
 Number of sessions target (mean): 10 (?mean)  
 Times per week: 2  
 Number of points used: 9  
 Insertion depth: .5 to 1.5 cm  
 Was De qi reportedly sought?: Yes  
 Duration (mins): 20  
 Method of stimulation: Manual

CONTROL GROUP A (sham, if used): insertion at non-points (0.5 cm) at the homo-lateral tibia front line  
 N allocated to control group A: unclear (26 analyzed)  
 Total length of treatment period: 5 weeks  
 Number of sessions target (mean): 10  
 Times per week: 2  
 (If relevant) Number of points used: Not reported  
 (If relevant) Insertion depth: .5cm  
 Was De qi sought?: unclear  
 Duration (mins): 20 mins  
 (If relevant) Method of stimulation: Not reported

Any co-interventions in all groups?: Not reported

**Outcomes**

PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS

Pain: VAS pain at baseline and average pain reductions on the VAS at end of treatment (acupuncture = -2.227, sham = -1.164), and 3 months after end of treatment; proportion with at least 50% pain relief (acupuncture = 36 patients or 50.7%, sham = 8 patients or 30.8%); percent pain reduction on VAS (end of treatment: acupuncture = 47.5%, sham = 26.1%; 3 months after end of treatment: acupuncture = 48%, sham = 6.1%)  
 Function: Lysholm at baseline and at end of treatment (acupuncture = 47.0, sham = 50.8); average percent reduction in Lysholm scores (acupuncture = 16.8%, sham = 8.2% (P = .22))  
 Overall index of symptom severity: WOMAC total  
 Type of outcome data reported (i.e., post treatment/change from baseline/both): See description above  
 Additional outcomes reported in the trial but not abstracted: All outcomes presented listed above  
 Adverse effects: Not reported

**Notes**

Comments: Presentation lacks detail. No clear mention of drop outs. Overall, this trial found a statistically significant reduction in pain but no improvement in function, at both the end of treatment and three months later.  
 Weak points: Results presentation  
 Only one reviewer (KL) extracted the data from this German language publication

**Molsberger 1994** (Continued)

Source of support: Not reported.

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Unclear risk	Randomization method was not described.
Allocation concealment?	High risk	
Blinding? Versus sham	Unclear risk	In sham acupuncture group, needles were inserted at non-points (0.5 cm) at the homo-lateral tibia front line.  No attempt to confirm patient blinding for sham control.
Incomplete outcome data addressed? Short term	Low risk	Drop-outs were 6 out of 103. See <a href="#">Figure 1</a> .
Free of selective reporting?	Low risk	Pain: VAS pain at baseline and average pain reductions on the VAS at end of treatment (acupuncture = -2.227, sham = -1.164), and 3 months after end of treatment; proportion with at least 50% pain relief (acupuncture = 36 patients or 50.7%, sham = 8 patients or 30.8%); percent pain reduction on VAS (end of treatment: acupuncture = 47.5%, sham = 26.1%; 3 months after end of treatment: acupuncture = 48%, sham = 6.1%) Function: Lysholm at baseline and at end of treatment (acupuncture = 47.0, sham = 50.8); average percent reduction in Lysholm scores (acupuncture = 16.8%, sham = 8.2% (P = .22)) Overall index of symptom severity: WOMAC total.  Findings reported in <a href="#">Table 4</a> and <a href="#">Figure 3</a> .
Free of other bias?	High risk	Patients compliance condition was not reported.  No ITT analysis was applied.

**Sangdee 2002 \***

Methods	Design: Parallel (2x2) Blinding: Clinical outcomes assessors were blinded for all treatment groups and patients were blinded for acupuncture/placebo treatment groups only. Since patients assessed their own outcomes, we considered this trial to have used blinding only for the acupuncture/placebo acupuncture groups for which patients were blinded. Attempt to confirm patient blinding for sham control?: No Drop-outs/withdrawals: 7 withdrawn from the trial for various medical reasons, none of which were attributable to acupuncture treatment  CBRG score: ?-?-1-.5-0-.5-1-1-1-1-0 (number preceding / is placebo control group score and number following / is medication alone control group score) Duration: 4 weeks of treatment and two additional months of follow-up for responders Type of analysis reported: only analysis of the 186 completers (intention-to-treat analysis also done as sensitivity analysis, but not reported because the results were reportedly not affected by type of analysis due to small number of withdrawals)
Participants	Setting: University outpatient center in Thailand Mean age (+/-SD or Range): ~63(7) Men/Women (n/n): 43/150

**Sangdee 2002** \* (Continued)

Recruitment method: Not reported  
 Mean pain duration (SD) years: ~5(3) years duration of OA  
 For bilateral OA diagnosis, which knee treated/evaluated?: Not reported  
 Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA):  
 ACR criteria  
 Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): No  
 Minimum duration (and extent) of knee pain required to be eligible: more than 3 months as suffering  
 from OA of knee  
 Hospital inpatients? (Y/N; if Y list number inpatients): N  
 Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Not reported  
 Were people with a history of acupuncture treatment excluded?: No, but EA within the last 3 months  
 was an exclusion criteria  
 Other important inclusion criteria: age>40; able to walk; Lequesne functional index greater than or  
 equal to 6 points at baseline; informed consent  
 Important exclusion criteria: underlying inflammatory arthropathy; surgery in future; injury in area af-  
 fected by OA knee, intraarticular corticosteroid injections or EA in the last 3 months

**Interventions**

**TEST GROUP INTERVENTION: Electroacupuncture plus placebo diclofenac**  
 N allocated to acupuncture: 48  
 Style of acupuncture: Chinese  
 Point selection: Formula  
 Points stimulated: ST35, LR8, EX32 (Xiyan) + 1 Triggerpoint  
 Total length of treatment period (weeks): 4  
 Number of sessions target (mean): 12 (?mean)  
 Times per week: 3  
 Number of points used: 4  
 Insertion depth: not more than .5 inches  
 Was De qi reportedly sought?: No  
 Duration (mins): 20  
 Method of stimulation: Electrical stimulation (2 Hz) at maximum tolerable level to patient to each pair  
 of needles

**CONTROL GROUP A (sham, if used): Sham procedure plus placebo diclofenac at 1 tablet 3x/day for 4  
 weeks. Sham involving patch electrodes attached to the same acupuncture points used in acupuncture  
 group and other end attached to sound producing dummy mode of the same EA machine**  
 N allocated to control group A: 47  
 Total length of treatment period: 4 weeks  
 Number of sessions target (mean): 12 (not reported)  
 Times per week: 3  
 (If relevant) Number of points used: 4  
 (If relevant) Insertion depth: Not applicable  
 Was De qi sought?: Not applicable  
 Duration (mins): 20  
 (If relevant) Method of stimulation: Mock electrostimulation

**CONTROL GROUP B: Sham procedure plus diclofenac 1 table 3x/day for 4 wks**  
 N allocated to control group B: 49  
 Total length of treatment period: 4 weeks  
 Number of sessions target (mean): Not applicable (3 tablets/day)  
 Times per week: Not applicable (daily)  
 Duration (minutes): Not applicable

**CONTROL GROUP C: Electroacupuncture plus diclofenac 1 table 3x/day for 4 wks**  
 N allocated to control group C: 49  
 Total length of treatment period: 4 weeks  
 Number of sessions target (mean): 12 plus diclofenac 3x/day  
 Times per week: 3  
 Duration (minutes): 20 minutes

Any co-interventions in all groups? During the study period, all additional therapies for OA (e.g., oral or  
 topical NSAIDS, intraarticular corticosteroid injection, other analgesics, chondro-protective, agents,



**Sangdee 2002** \* (Continued)

surgical procedures on the knee joint, etc.) were not allowed. However, 2 tablets of 500 mg Paracetamol/day were still prescribed as a rescue analgesic during the study. Other treatments for concomitant diseases could be continued if treatments administered were documented.

**Outcomes**

**PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS**

Pain: WOMAC Pain: Baseline, after 4 week treatment period, and, for responders only, 2 months later  
Function: WOMAC function: Baseline, after 4 week treatment period, and, for responders only, 2 months later

Overall index of symptom severity: WOMAC total: Baseline, after 4 week treatment period, and, for responders only, 2 months later

Type of outcome data reported (i.e., post treatment/change from baseline/both): means and standard errors of changes for each group reported in publication

Additional outcomes reported in the trial but not abstracted: amount of paracetamol tablets taken/week; 50 feet-walk time; a patient's global pain as 100 mm VAS; Lesquesne's functional index; and clinician's and patient's overall opinion of change

Adverse effects: reported that the "percentage of patients who experienced adverse effects...during the study did not differ between the four groups (data not shown), whereas, local contusions around the knee were more common in the EA and combined group (approximately 45%). However, the contusions usually disappeared within 5-7 days."

**Notes**

Comments: Well reported study with 2x2 factorial design.

Weak points: Only responders were followed-up beyond the initial four week treatment period so no randomized long-term data available

This trial did not report standard deviations of post-treatment scores (because the authors of this trial used a comparison of between group changes for their analysis). For this trial, we used the baseline standard deviations, provided in Table 4, as an estimate for the post-treatment standard deviation scores. This was the only trial that reported pre-treatment standard deviations but not post-treatment standard deviations, and was, therefore, the only trial for which we used pre-treatment standard deviations to calculate variability. We discussed the use of pre-treatment standard deviations as estimates for post-treatment standard deviations with two biostatisticians, who both considered this a reasonable assumption.

This trial had a factorial design, and compared electro-acupuncture versus placebo acupuncture, using either a diclofenac or a placebo diclofenac co-intervention. This study ID compares acupuncture versus sham using the true diclofenac co-intervention.

The corresponding author was e-mailed (at csangdee@mail.med.cmu.ac.th) to request review of information extracted from the trial, to obtain further details about the randomization, and to request the SDs of the post-treatment scores; however, no response was received.

Source of support: "This work was supported by the Faculty of Medicine, Chiang Mai University, Thailand."

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Unclear risk	<p>"The patients who had persistent pain and a Lequesne's functional index of at least 6 points at the end of the run-in period were randomized into the four groups mentioned above." No more detailed description about randomization.</p> <p>The corresponding author was e-mailed to obtain further details about the randomization; however, no response was received.</p>
Allocation concealment?	Unclear risk	The method of concealment is not described.

**Sangdee 2002 \*** (Continued)

Blinding? Versus sham	Unclear risk	<p>“Clinical assessments in each patient were evaluated by the same physician who was blinded to the treatment.”</p> <p>“The placebo EA was performed by attaching patch electrodes to the selected acupuncture points. Each electrode was connected to the sound producing dummy mode of the same apparatus, as in the true EA treatment.”</p> <p>No attempt to confirm patient blinding for sham control was reported.</p>
Incomplete outcome data addressed? Short term	Low risk	<p>“Of the 193 study patients, 186 (96.37%) completed the study. The remaining 7 patients were withdrawn from the trial due to flare of pain with joint swelling (2 in the placebo and 1 in the EA group), severe GI side effects (3 in the combined group), and flare of pain from an accidental fall not related to treatment (1 in the EA group).”</p> <p>45(47), 49(49), 46(48) and 46(49) patients were available at week 4 according to <a href="#">Table 2</a> and above description.</p>
Free of selective reporting?	Low risk	<p>“Clinical assessments were evaluated for base-line data at the end of the run-in period (week 0) and again at the end of the study (week 4). These assessments included the amount of paracetamol tablets taken/week, 50 feet-walk time, a patient's global pain as 100 mm visual analog scale (VAS) over the previous 3 days, the Western Ontario and McMaster Universities OA Index (WOM-AC: score ranging from 0-96) [21], and Lequesne's functional index (score ranging from 0-24) [22]. Complete physical examination and non-directive questioning for adverse events were also performed weekly for 4 weeks in order to acquire a safety assessment.”</p> <p>Findings reported in <a href="#">Table 4</a>, 4-6 and <a href="#">Figure 2-4</a>.</p>
Free of other bias?	High risk	No ITT analysis was applied.

**Sangdee 2002 \*\***

Methods	The Sangdee trial used a factorial design, comparing electroacupuncture versus placebo electroacupuncture, with one comparison including a diclofenac co-intervention and the other comparison using a placebo diclofenac co-intervention. This study ID was created so that the comparison arms with the placebo diclofenac co-intervention could be entered in the meta-analysis.	
Participants		
Interventions		
Outcomes		
Notes		
<b>Risk of bias</b>		
<b>Bias</b>	<b>Authors' judgement</b>	<b>Support for judgement</b>
Adequate sequence generation?	Unclear risk	<p>“The patients who had persistent pain and a Lequesne's functional index of at least 6 points at the end of the run-in period were randomized into the four groups mentioned above.” No more detailed description about randomization.</p> <p>The corresponding author was e-mailed (at csangdee@mail.med.cmu.ac.th) to request review of information extracted from the trial, to obtain further details</p>

**Sangdee 2002 \*\*** (Continued)

		about the randomization, and to request the SDs of the post-treatment scores; however, no response was received.
Allocation concealment?	Unclear risk	The method of concealment is not described.
Blinding? Versus sham	Unclear risk	<p>“Clinical assessments in each patient were evaluated by the same physician who was blinded to the treatment.”</p> <p>“The placebo EA was performed by attaching patch electrodes to the selected acupuncture points. Each electrode was connected to the sound producing dummy mode of the same apparatus, as in the true EA treatment.”</p> <p>No attempt to confirm patient blinding for sham control was reported.</p>
Incomplete outcome data addressed? Short term	Low risk	<p>“Of the 193 study patients, 186 (96.37%) completed the study. The remaining 7 patients were withdrawn from the trial due to flare of pain with joint swelling (2 in the placebo and 1 in the EA group), severe GI side effects (3 in the combined group), and flare of pain from an accidental fall not related to treatment (1 in the EA group).”</p> <p>45(47), 49(49), 46(48) and 46(49) patients were available at week 4 according to <a href="#">Table 2</a> and above description.</p>
Free of selective reporting?	Low risk	<p>“Clinical assessments were evaluated for base-line data at the end of the run-in period (week 0) and again at the end of the study (week 4). These assessments included the amount of paracetamol tablets taken/week, 50 feet-walk time, a patient's global pain as 100 mm visual analog scale (VAS) over the previous 3 days, the Western Ontario and McMaster Universities OA Index (WOMAC: score ranging from 0-96) [21], and Lequesne's functional index (score ranging from 0-24) [22]. Complete physical examination and non-directive questioning for adverse events were also performed weekly for 4 weeks in order to acquire a safety assessment.”</p> <p>Findings reported in <a href="#">Table 4</a>, 4-6 and <a href="#">Figure 2-4</a>.</p>
Free of other bias?	High risk	No ITT analysis was applied.

**Scharf 2006**

Methods	<p>Design: Parallel</p> <p>Blinding: All outcomes assessments were documented by telephone interviewers who were blinded for all three treatments. However, for this trial, we did not consider the usual care arm to have been outcomes assessor blinding because we always considered the patient to be the outcomes assessor for the WOMAC questionnaire, a self-assessed outcome; the patients in the 'no acupuncture' usual care arm obviously could not be blinded to whether or not they received acupuncture.</p> <p>Attempt to confirm patient blinding for sham control?: The quality of blinding between the true and sham acupuncture groups was assessed in the final interview. Only 33 patients (16 in acupuncture and 17 in sham group) stated that they had been unblinded by the treating physician. Almost one-half (49%) of the 675 patients who received true or sham acupuncture were unaware which treatment they received. Of the 345 who said they knew their treatment, 45% guessed incorrectly. Therefore, blinding between true and sham acupuncture groups was successful.</p> <p>Drop-outs/withdrawals: 22 patients were lost to follow-up: 8 in true acupuncture group; 5 in sham group; and 9 in standard care group. These lost to follow-up patients were included in the intention-to-treat analysis. (Of the total of n = 1039 randomly assigned patients, thirty-two additional patients dropped out immediately after randomization (i.e., 4 in acupuncture; 2 in sham; 26 in standard care) and were excluded from the intention-to-treat analysis.)</p>
---------	---

**Scharf 2006** (Continued)

CBRG score: 1-1-1-1/0-0-1/0-0-1-1-1-1 (number preceding / is sham control group score and number following / is standard care control group score)  
 Duration: 6 weeks of treatment, followed by an evaluation at week 7; during weeks 7-13, there were further treatments for 'partial responders' only (i.e., non-responders and full responders did not get additional sessions); follow-up evaluation at weeks 13 and 26  
 Type of analysis reported: Intention-to-treat for the primary analysis (excluding the 32 patients who declined further participation immediately after randomization and for whom no measurements were available); per protocol for a sensitivity analysis.

**Participants**

Setting: 315 primary care practices staffed by 320 practitioners, Germany  
 Mean age (+/-SD or Range): ~63(10)  
 Men/Women (n/n): 314/693  
 Recruitment method: Patients of 320 physicians, selected from a group of experienced primary care practitioners participating in a large cohort study on acupuncture  
 Mean pain duration (SD) years: ~5.4 (6)  
 For bilateral OA diagnosis, which knee treated/evaluated?: Both knees were treated if affected. In patients with 2 affected knees, 1 knee was randomly chosen for evaluation during the initial telephone interview.  
 Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): ACR criteria  
 Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Yes, Kellgren of 2 or 3  
 Minimum duration (and extent) of knee pain required to be eligible: Chronic pain for at least 6 months due to OA of the knee  
 Hospital inpatients? (Y/N; if Y list number inpatients): N  
 Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Patients excluded if surgery of the afflicted extremity during the last six months or planned surgery  
 Were people with a history of acupuncture treatment excluded?: Patients with previous acupuncture treatment for gonarthrosis (ever) or previous acupuncture treatment for any condition (in the past 12 months) were excluded  
 Other important inclusion criteria: WOMAC score of at least 3 points (0-10 scale); chronic pain score of at least 1, according to the criteria of von Korff and colleagues  
 Important exclusion criteria: Patients with other diseases affecting the knee, neurologic and psychiatric diseases, severe coagulopathy, pregnancy, or previous acupuncture treatment for osteoarthritis of the knee were excluded.

**Interventions**

**TEST GROUP INTERVENTION: acupuncture**  
 N allocated to acupuncture: 330 (4 withdrew consent immediately after randomization, leaving 326 for the full analysis set)  
 Style of acupuncture: Chinese  
 Point selection: Flexible formula  
 Points stimulated: Obligatory Unilateral: St34, 36, Xiyan, Sp9, 10, GB34; Optional Unilateral: 1-4 Ahshi points; bilateral according to Traditional Chinese Medicine: 1-2 of 16 defined distant points  
 Total length of treatment period (weeks): 6 weeks (Patients could request up to 5 additional sessions or visits in weeks 7 to 13 if the initial treatment was viewed as being partially successful.)  
 Number of sessions target (mean): 10 [+5] (12.5) "+5" indicates 5 additional sessions or visits if treatment was graded as partially successful after 6 weeks.  
 Times per week: almost 2x/wk (i.e., 10 sessions over 6 weeks)  
 Number of points used (mean): 7-15 (or 14-30 if bilateral)  
 Insertion depth: 0.5-3.5 until de qi  
 Was De qi reportedly sought?: Yes  
 Duration (mins): 20-30 [mean 30.3]  
 Method of stimulation: Manual, twice

**CONTROL GROUP A (sham, if used): sham acupuncture with 10 points total, four on each leg and one on each arm, with a minimal depth of needling (not more than 5mm) and avoiding real acupoints**  
 N allocated to control group A: 367 (2 withdrew consent immediately after randomization, leaving 365 for the full analysis set)  
 Total length of treatment period: 6 weeks (Patients could request up to 5 additional sessions or visits in weeks 7 to 13 if the initial treatment was viewed as being partially successful.)

**Scharf 2006** (Continued)

Number of sessions target (mean): 10 [+5] (12.5) “+5” indicates 5 additional sessions or visits if treatment was graded as partially successful after 6 weeks.  
 Times per week: almost 2x/wk (i.e., 10 sessions over 6 weeks)  
 (If relevant) Number of points used (mean): 10 (12.5)  
 (If relevant) Insertion depth: Up to 0.5 without de qi  
 Was De qi sought?: No  
 Duration (mins): 20–30 [mean 29.9]  
 (If relevant) Method of stimulation: None

CONTROL GROUP B: Conservative therapy involved 10 visits to physician practitioners [if possible] with consultation and a prescription for diclofenac, up to 150 mg/d, or rofecoxib, 25 mg/d, as needed, until week 23.

N allocated to control group B: 342 (26 withdrew consent immediately after randomization, leaving 316 for the full analysis set)

Total length of treatment period: 6 weeks (Patients could request up to 5 additional sessions or visits in weeks 7 to 13 if the initial treatment was viewed as being partially successful.)

Number of sessions target (mean): 10 [+5] [8.4] “+5” indicates 5 additional sessions or visits if treatment was graded as partially successful after 6 weeks.

Times per week: almost 2x/wk (i.e., 10 visits over 6 weeks)

Duration (minutes): Duration of consultation visits with physicians not stated

Any co-interventions in all groups? The acupuncture and sham groups could take up to 150 mg of diclofenac per day during the first 2 treatment weeks and up to a total of 1 g until week 23. The standard care group could take Diclofenac, 150 mg/d, or rofecoxib, 25 mg/d up to week 23 (with no total limit). Each of the 3 treatment groups had up to 6 physiotherapy sessions.

**Outcomes**

PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS

Pain: WOMAC Pain: Baseline, 13 weeks after randomization/start of treatment, and at follow-up 26 weeks after randomization/start of treatment

Function: WOMAC function: Baseline, 13 weeks after randomization/start of treatment, and at follow-up 26 weeks after randomization/start of treatment

Total: WOMAC total: Baseline, 13 weeks after randomization/start of treatment, and at follow-up 26 weeks after randomization/start of treatment

Type of outcome data reported (i.e., post treatment/change from baseline/both): Post-treatment means and standard deviations for each group as well as means and standard deviations of changes from baseline for each groups

Additional outcomes reported in the trial but not abstracted: Success rates calculated according to a change of at least 36% from baseline WOMAC scores at 13 and 26 weeks after the start of treatment; SF-12; global patient assessment

Adverse effects: Number of total adverse effects and serious adverse effects were comparable for the three groups. Haematoma or minor bleeding was reported more often in the acupuncture and sham acupuncture groups than in the standard care group.

**Notes**

Comments: Very well reported study. Some strengths are the successful blinding between the true and sham acupuncture groups, the low number of withdrawals, and homogeneous treatment groups at baseline. The primary outcome measures used by the RCT authors was the success rate, as defined by at least 36% improvement in WOMAC score at 26 weeks, which is different from the primary outcome used in our meta-analysis (i.e., the WOMAC total, pain, and function scores).

Weak points: Two weak points were acknowledged by the RCT authors in their discussion: 1) Large number of treatment sites and participating physicians might make it difficult to monitor adherence to the predefined acupuncture schemes, and 2) The recruited patients assumedly had an interest in acupuncture, possibly introducing a selection effect. Another possible weak point is that a very intensive sham procedure was used.

The intention-to-treat Ns (using imputation for missing values) were used for the success rate outcome (which the authors considered the primary outcome), but for the WOMAC pain, function and total mean and SD outcomes (which were not considered primary), no imputations were made, and the analyses were based on the numbers for whom there were available data.

**Scharf 2006** (Continued)

Steffen Witte provided the WOMAC total post-treatment SDs for the 3 month and 6 month time points. In addition, he provided additional information, not reported in the publication, about the participants and interventions. We have enclosed this additional information in brackets.

Source of support: "A consortium of German social health care insurance organizations (Allgemeine Ortskrankenkassen, Betriebskrankenkassen, Innungskrankenkassen, Bundesknappschaft, Landwirtschaftliche Krankenkassen, and See-Krankenkasse) provided funding for this study at the request of German health authorities."

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"The 1:1:1 block randomization with block size of 6, stratified by center, was computer-generated by an independent statistician and was transferred to the randomization center."
Allocation concealment?	Low risk	"For each eligible patient, the practitioner called the randomization center and was immediately informed by fax of the patient's treatment group."
Blinding? Versus sham	Low risk	"The quality of blinding between the TCA and sham acupuncture groups was assessed in the final interview. Only 33 patients (16 in the TCA group and 17 in the sham acupuncture group) stated that they had been unblinded by the treating physician. Almost half (n = 330 [49%]) of 675 patients with TCA or sham acupuncture were unaware of which treatment they received. Data on this variable were missing in 16 patients. Of the 345 patients who said they knew their treatment, 154 (45%) guessed incorrectly. The 33 patients who were unblinded by the treating physician are among the 345 patients increasing the rate of correct presumptions. Therefore, the estimates were roughly random, and blinding between the TCA and sham acupuncture groups was successful (Appendix Table 6)."
Incomplete outcome data addressed? Short term	Unclear risk	<p>"During the 26-weeks assessment, 22 patients were lost to follow-up: 8 in true acupuncture group; 5 in sham group; and 9 in standard care group. These lost to follow-up patients were included in the intention-to-treat analysis. (Of the total of n = 1039 randomly assigned patients, thirty-two additional patients dropped out immediately after randomization (i.e., 4 in acupuncture; 2 in sham; 26 in standard care) and were excluded from the intention-to-treat analysis.)"</p> <p>See study flow diagram.</p> <p>Observations available for analysis are 326(330) in acupuncture group, 365(367) in sham acupuncture group, and 316(342) in conservative therapy group at 13 and 26 weeks.</p> <p>"For the primary analyses, an intention-to-treat sample was used, including all randomly assigned patients with at least 1 treatment."</p> <p>See Table 4-4.</p>
Free of selective reporting?	Low risk	"The effect on pain and function was measured with the WOMAC score (total score and the subscales standardized to 0 to 10). In patients with 2 affected knees, 1 knee was randomly chosen for evaluation during the initial telephone interview. According to the recommendations of the OsteoArthritis Research Society International (21), success rates were calculated according to a change of at least 36% from baseline WOMAC scores at 13 and 26 weeks after the start of treatment (22). Patients with missing data were considered to have had treatment failure."

**Scharf 2006** (Continued)

“Blinded central telephone interviews were conducted in weeks 13 and 26 to record the main and secondary outcome measures (SF-12 [14] and global patient assessment [23]). At each visit, the investigator documented adverse events since the last visit, and a medical doctor performed coding using the Medical Dictionary for Regulatory Activities (MedDRA).”

Findings reported in table 3, 4 and appendix table 2, 3, 6.

Steffen Witte provided the WOMAC total post-treatment SDs for the 3 month and 6 month time points. In addition, he provided additional information, not reported in the publication, about the participants and interventions.

Free of other bias?	Low risk
---------------------	----------

**Stener-Victorin 2004**

Methods	Design: Parallel Blinding: No blinding Attempt to confirm patient blinding for sham control?: NA Drop-outs/withdrawals: At the end of treatment assessment (i.e. at end of 5 weeks of treatment) there were 13/15 available in the EA group; 13/15 in the hydrotherapy group; and 8/15 in the patient education group. For the short-term follow-up as defined for this review (i.e. 1 month after the end of treatment or ~9 weeks after baseline), the numbers available for analysis were not provided in the paper. However, the authors provided these data as follows: EA = 13; Water = 13; Control = 12. At the six months post-end of treatment assessment (i.e. the long-term follow-up as defined for this review), there were 9/15 available in both the EA and hydrotherapy groups, and 7/15 available in the patient education group.  CBRG score: [1]-[1]-0-0-0-0-1-?-0-0-0 Duration: 5 weeks of treatment with EA or hydrotherapy followed by an observation period of an additional six months Type of analysis reported: not explicitly stated but assumedly per protocol analysis
Participants	Setting: outpatient physiotherapy clinic, Molndal, Sweden Mean age (+/-SD or Range): ~67 Men/Women (n/n): 18/27 Recruitment method: Recruited from patients on a waiting list for total hip arthroplasty Mean pain duration (SD) years: Not reported For bilateral OA diagnosis, which hip treated/evaluated?: Not reported Diagnosis of hip OA required to be eligible? (if yes, describe how patients were verified to have OA): Yes, diagnosis was based on radiographic changes consistent with hip OA and pain related to motion and/or pain on load or rest Radiologic evidence of hip OA required to be eligible? (if yes, describe requirement): Yes, only states that radiographic changes consistent with hip OA required to be eligible Minimum duration (and extent) of hip pain required to be eligible: Not reported Hospital inpatients? (Y/N; if Y list number inpatients): No Previous hip surgery? (Y/N; if Y list number with previous knee surgery): Not reported Were people with a history of acupuncture treatment excluded? Not reported (assume no, since no sham group) Other important inclusion criteria: None Important exclusion criteria: patients with other rheumatoid diseases
Interventions	TEST GROUP INTERVENTION: acupuncture N allocated to acupuncture: 15 Style of acupuncture: Chinese Point selection: Flexible formula Points stimulated: Electroacupuncture locally at 4 of following points: BL 54, 36, GB 29, 30, 31, and ST 31. Distal points were always the same: GB34 and BL 60 ipsilateral

**Stener-Victorin 2004** (Continued)

Total length of treatment period (weeks): 5 weeks  
 Number of sessions target (mean): 10 (?mean)  
 Times per week: 2  
 Number of points used: 6  
 Insertion depth: 15-35mm  
 Was De qi reportedly sought?: Yes  
 Duration (mins): 30  
 Method of stimulation: Electrical stimulation at all points and needles were also rotated manually 4 times during each treatment to evoke needle sensation

CONTROL GROUP A (sham, if used): hydrotherapy -- consisted of warming up, mobility and strengthening and stretching exercises for area around the hip, in small groups of 1-3, in warm water

N allocated to control group A: 15  
 Total length of treatment period: 5 weeks  
 Number of sessions target (mean): 10 (?mean)  
 Times per week: 2  
 (If relevant) Number of points used: NA  
 (If relevant) Insertion depth: NA  
 Was De qi sought?: NA  
 Duration (mins): 30  
 (If relevant) Method of stimulation: NA

CONTROL GROUP B: Patient education about anatomy and physiology of hip, pain relief, and total hip arthroplasty. They were also given home exercise instructions. They were taught to train once per day with intensity below pain.

N allocated to control group B: 15  
 Total length of treatment period: Not reported  
 Number of sessions target (mean): 2 (?mean)  
 Times per week: Not clear  
 Duration (minutes): 120

Any co-interventions in all groups?: All groups received the patient education co-intervention.

Outcomes	The pain outcome was pain intensity using the VAS scale and the function outcome was the disease-related handicap as assessed by the (disability rating index) DRI. Other outcomes that were not extracted were the overall assessment of patient satisfaction on the Comparative Scale (Carlsson) and the Quality of life measurement, using the 'Everyday Life' questionnaire (Bullinger). The outcomes were reported as medians and interquartile ranges rather than as means and SDs.
----------	---

Notes	Randomization stated to be done using "sealed, unlabeled envelopes". In response to an e-mail request for further information about this, the author stated that the generation of the randomization sequence was computerized, the envelopes were opaque and placed in a dark box and then selected by the trial participants, and that the process was administered by an independent person not responsible for determining the eligibility of the patients.
-------	---

This study reported median results rather than means, which is the primary reason that this trial did not contribute to effect size estimates. In addition, because of the high attrition rate at 6 months, a post hoc decision was made to not enter the 6 months follow-up median data as 'Other data' in RevMan. The short-term follow-up median data (i.e. the data at 1 month after last treatment = 9 weeks after baseline) was also not entered as 'Other data' in RevMan because the numbers available for follow-up were not reported at that time point.

Overall conclusions of RCT author: "The main outcome of the study was that EA and hydrotherapy, both in combination with patient education, offer clear advantages for patients with hip pain caused by osteoarthritis over patient education alone, as shown by reduced pain, increased function, and increased quality of life."

Source of support: This study was "supported by Research and Development Unit, Vastra Goteborg, Sweden". Assumed it's supported by government grant.

**Risk of bias**



**Stener-Victorin 2004** (Continued)

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	<p>“Participants were randomly allocated to EA in combination with patient education (n = 15) or hydrotherapy in combination with patient education (n = 15) as a control by using sealed, unlabeled envelopes.”</p> <p>In response to an e-mail request for further information about this, the author stated that the generation of the randomization sequence was computerized, and that the process was administered by an independent person not responsible for determining the eligibility of the patients.</p>
Allocation concealment?	Low risk	<p>“Participants were randomly allocated to EA in combination with patient education (n = 15) or hydrotherapy in combination with patient education (n = 15) as a control by using sealed, unlabeled envelopes.”</p> <p>In response to an e-mail request for further information about this, the author stated that the envelopes were opaque and placed in a dark box and then selected by the trial participants, and that the process was administered by an independent person not responsible for determining the eligibility of the patients.</p>
Blinding? Versus sham	High risk	EA vs. hydrotherapy and patient education, no sham control.
Incomplete outcome data addressed? Short term	Low risk	<p>For the short-term follow-up as defined for this review (i.e. 1 month after the end of treatment or ~9 weeks after baseline), the numbers available for analysis were not provided in the paper. However, the authors provided these data as follows: EA = 13(15); Water = 13(15); Control = 12(15).</p> <p>At the six months post-end of treatment assessment (i.e. the long-term follow-up as defined for this review), there were 9(15) available in both the EA and hydrotherapy groups, and 7(15) available in the patient education group.</p> <p>See <a href="#">Figure 1</a>.</p>
Free of selective reporting?	Low risk	<p>“The EA and hydrotherapy groups were assessed before the intervention and immediately after the last treatment. Follow-up included 1 assessment at 1, 3 and 6 months after the last treatment. The control group was assessed at the same point of time except from the time immediately after the last treatment.”</p> <p>“Outcome measures were determined by a functional index, called the disability rating index (DRI), a quality-of-life status called the global self-rating index (GSI), and for pain, the visual analogue scale (VAS).“</p> <p>Findings reported in <a href="#">Table 4, 3</a>.</p>
Free of other bias?	High risk	<p>Baseline is not similar among groups.</p> <p>“The control group (patient education group) was assessed at the same point of time except from the time immediately after the last treatment.” Timings of assessment were not similar among groups.</p> <p>Patients compliance condition was not reported.</p> <p>No ITT analysis was applied.</p>

**Takeda 1994**

Methods

Design: Parallel

**Acupuncture for peripheral joint osteoarthritis (Review)**

Copyright © 2010 The Cochrane Collaboration. Published by John Wiley &amp; Sons, Ltd.

**Takeda 1994** (Continued)

Blinding: Patient and outcomes assessor blinding  
 Attempt to confirm patient blinding for sham control?: No  
 Drop-outs/withdrawals: 2 (one from each group dropped out for personal/work commitments and each was replaced by a new recruit)

CBRG score: [1]-[1]-0-.5-0-.5-?-?-1-1-0  
 Duration: 3 weeks of treatment and four weeks of follow-up  
 Type of analysis reported: Only analysis of the 40 completers

---

**Participants**

Setting: University physical therapy department (probably), Canada  
 Mean age (+/-SD or Range): 62 (9)  
 Men/Women (n/n): 20/20  
 Recruitment method: Not reported (participants were 'volunteers')  
 Mean pain duration (SD) years: Not reported  
 For bilateral OA diagnosis, which knee treated/evaluated?: Only the most painful knee  
 Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): Yes, radiological evidence of OA was inclusion criteria  
 Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Yes (Kellgren I through IV)  
 Minimum duration (and extent) of knee pain required to be eligible: Not reported  
 Hospital inpatients? (Y/N; if Y list number inpatients): Not reported  
 Previous knee surgery? (Y/N; if Y list number with previous knee surgery): People who had reconstructive surgery of the affected knee were excluded  
 Were people with a history of acupuncture treatment excluded?: Yes (No previous experience with acupuncture of the knee was an inclusion criteria.)  
 Other important inclusion criteria: pain in one or both knees; no change in medications for arthritis and other conditions in the last 3 weeks  
 Important exclusion criteria: serious systematic conditions; hemophilia; receiving any treatment other than medication for their arthritis

---

**Interventions**

TEST GROUP INTERVENTION: acupuncture  
 N allocated to acupuncture: 21  
 Style of acupuncture: Chinese  
 Point selection: Formula  
 Points stimulated: GB34, ST35, SP9, EX31 (Heding), EX32 (Xiyan)  
 Total length of treatment period (weeks): 3  
 Number of sessions target (mean): 9 (?mean)  
 Times per week: 3  
 Number of points used: 5  
 Insertion depth: Not reported (but needles were inserted deeper until subjects experienced Te chi)  
 Was De qi reportedly sought?: Yes  
 Duration (mins): 30  
 Method of stimulation: Rotated manually for 5 minutes

CONTROL GROUP A (sham, if used): superficial needling one inch from the acupuncture points  
 N allocated to control group A: 21  
 Total length of treatment period: 3 weeks  
 Number of sessions target (mean): 9  
 Times per week: 3  
 (If relevant) Number of points used: 5?  
 (If relevant) Insertion depth: superficial  
 Was De qi sought?: No, however, 'te chi was experienced regularly during treatment by 25 subjects (14 experimental, 11 control...'  
 Duration (mins): 30  
 (If relevant) Method of stimulation: None; however the needles were "only touched periodically to give the impression that movement of the needles was taking place."

---

**Outcomes**

PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS

Pain: WOMAC Pain: Baseline, after 3 weeks of treatment, and at follow-up 4 weeks later

**Takeda 1994** (Continued)

Function: WOMAC function: Baseline, after 3 weeks of treatment, and at follow-up 4 weeks later  
 Overall index of symptom severity: WOMAC total: Baseline, after 3 weeks of treatment, and at follow-up 4 weeks later  
 Type of outcome data reported (i.e., post treatment/change from baseline/both): Post-treatment means and standard deviations  
 Additional outcomes reported in the trial but not abstracted: McGill Pain Questionnaire, a pain threshold dolorimeter  
 Adverse effects: Not reported

**Notes**

Comments: Small yet carefully conducted study.

Weak points: the large baseline differences, with the acupuncture group having better scores, especially function scores, at baseline, complicates the interpretation of the results; only 7 weeks duration; small sample size

For this trial, the method of generating the randomization sequence and the allocation concealment were not described in the publication, but were both adequate according to the results of a survey conducted to determine whether the description of methods in the published reports is an accurate and complete reflection of study procedures used. (survey reference: Manheimer E, Ezzo J, Hadhazy V, Berman B. Published reports of acupuncture trials showed important limitations. *Journal of Clinical Epidemiology* 59 (2006) 107–113).

The text states “each was rotated back and forth manually for 5 minutes.” However, considering there were five needles, this would mean that the manual manipulation would have been 25 minutes total (5 minutes for each of five needles). This is highly unlikely and we assume that the authors meant that the needles were rotated every five minutes. (There was no response to an e-mail contact to Jean Wessel, the corresponding author for this study.) LL was asked about this and agreed that we should interpret the text to mean that the needles were rotated every five minutes.

This trial did not report WOMAC totals. We calculated means of WOMAC totals by adding the reported means of the three WOMAC sub-scale (i.e., pain, function and stiffness) scores. We calculated SDs of WOMAC totals by taking the square root of the sum of the variances of the sub-scale scores, making the assumption that these sub-scale scores were independent.

For WOMAC pain, function, and totals, the trial did not report standard deviations of changes from baseline or any statistics that would allow us to calculate these, for any outcomes. For the between-group changes analyses, we assumed a conservative within-subject pretest–post-test correlation of 0.5. We entered the pre- and post- means and standard deviations and the pre-test-post-test correlation of .5 in Comprehensive Meta-analysis Version 2, from which we obtained the standard deviations of change for each group. We then entered these standard deviations of change scores into RevMan.

Source of support: Not reported.

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	“Subjects were stratified by gender and randomly allocated to groups in blocks of four.”  For this trial, the method of generating the randomization sequence and the allocation concealment were not described in the publication, but were both adequate according to the results of a survey conducted to determine whether the description of methods in the published reports is an accurate and complete reflection of study procedures used. (survey reference: Manheimer E, Ezzo J, Hadhazy V, Berman B. Published reports of acupuncture trials showed important limitations. <i>Journal of Clinical Epidemiology</i> 59 (2006) 107?113).
Allocation concealment?	Low risk	For this trial, the method of generating the randomization sequence and the allocation concealment were not described in the publication, but were both adequate according to the results of a survey conducted to determine whether

**Takeda 1994** (Continued)

the description of methods in the published reports is an accurate and complete reflection of study procedures used. (survey reference: Manheimer E, Ezzo J, Hadhazy V, Berman B. Published reports of acupuncture trials showed important limitations. *Journal of Clinical Epidemiology* 59 (2006) 107-113).

Blinding? Versus sham	Unclear risk	"For the placebo treatment, the same type of needles were inserted superficially (just enough to puncture the skin) approximately 1 inch from the acupuncture points (Figure 1), in areas not considered active acupuncture points."  "All measurements were performed by a professional assistant who was blind to group assignment of the subjects."  But no attempt to confirm patient blinding for sham control.
Incomplete outcome data addressed? Short term	Low risk	"Subjects were 40 volunteers (20 men, 20 women)..." Each group included 20 subjects originally (see Table 3).  "One subject from each group dropped out of the study because of personal/work commitments unrelated to the arthritis or treatment. These subjects were replaced by new recruits."
Free of selective reporting?	Low risk	"Measurements of pain, stiffness, and function were taken before (pretest) and after (midtest) 3 weeks of treatment, and at follow-up 4 weeks later (posttest)."  "At each test session, the subjects were tested with the McGill Pain Questionnaire [MPQ], the Western Ontario and McMaster Universities (WOMAC) OA Index, and a pain threshold dolorimeter."  WOMAC, PRI and pain threshold were reported in Table 4, 3.
Free of other bias?	High risk	Baselines were not similar. (see Table 3).  Co-intervention was not reported.  Patients compliance condition was not reported.  No ITT analysis was applied.

**Tukmachi 2004**

Methods	Design: Parallel Blinding: No Attempt to confirm patient blinding for sham control?: Not applicable (no sham control) Drop-outs/withdrawals: One participant excluded from the analysis due to deviation from protocol  CBRG score: 1-?-?-0-0-?-?-1-1-0 Duration: 5 weeks of treatment followed by outcomes assessment, followed by an additional five weeks during which acupuncture was included as an add-on treatment for all patients in the three arms of this study. We did not extract outcomes data after the 5 week assessment time point because 'acupuncture versus acupuncture' comparisons are not included in our review. Type of analysis reported: One patient was excluded, in all remaining patients missing values were replaced using last value carried forward (therefore, not strictly an intention-to-treat analysis).
Participants	Setting: University outpatient clinic (probably, based on context), Birmingham, Great Britain Mean age (+/-SD or Range): 62 (Range: 42-77) Men/Women (n/n): 5/24 Recruitment method: Recruited from hospital department of rheumatology Mean pain duration (SD) years: 10 (?SD)

**Tukmachi 2004** (Continued)

For bilateral OA diagnosis, which knee treated/evaluated?: More painful knee  
 Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA):  
 Yes, diagnosis was based on clinical and radiological findings  
 Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Yes, Kellgren II-III. (Kellgren IV was mentioned in Box 1, but this is probably an error as it contradicts Table 3, which indicates that all patients were either Kellgren II or III)  
 Minimum duration (and extent) of knee pain required to be eligible: Duration of knee osteoarthritis of six months or more required  
 Hospital inpatients? (Y/N; if Y list number inpatients): Not reported  
 Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Not reported  
 Were people with a history of acupuncture treatment excluded? Yes  
 Other important inclusion criteria: Previous non-response to inpatient or outpatient treatment  
 Important exclusion criteria: Other type of arthritis (e.g., RA), hemophilia, use of anticoagulants, cortisone or oral corticosteroid medication

**Interventions**

TEST GROUP INTERVENTION: acupuncture without medication (patients agreed not to take any NSAIDS or analgesics throughout the treatment period, stopping one week before treatment began)  
 N allocated to acupuncture: 10  
 Style of acupuncture: Chinese  
 Point selection: Formula  
 Points stimulated: Electroacupuncture at SP9, GB34, BL40, 57, two Xiyao points; manual stimulation at GB34; needle insertion only at ST36, LR3, LI4  
 Total length of treatment period (weeks): 5  
 Number of sessions target (mean): 10 (?mean)  
 Times per week: 2  
 Number of points used: 9  
 Insertion depth: 1-1.5cm  
 Was De qi reportedly sought?: Yes  
 Duration (mins): 20-30  
 Method of stimulation: Manual at GB34 and electrostimulation at SP9, GB34, BL40, BL57, and two Xiyao points (for the other points, there was no method of stimulation)

CONTROL GROUP A (sham, if used): symptomatic medication (with acupuncture course added after five weeks)  
 N allocated to control group A: 10  
 Total length of treatment period: 5 weeks  
 Number of sessions target (mean): NA  
 Times per week: NA  
 (If relevant) Number of points used: NA  
 (If relevant) Insertion depth: NA  
 Was De qi sought?: NA  
 Duration (mins): NA  
 (If relevant) Method of stimulation: NA

CONTROL GROUP B: Acupuncture plus continued on medication (same acupuncture procedure as that used for 'acupuncture without medication group described above)  
 N allocated to control group B: 10  
 Total length of treatment period: 5 weeks  
 Number of sessions target (mean): 10 (?mean)  
 Times per week: 2  
 Duration (minutes): 20-30

Any co-interventions in all groups?: All groups continued with any medications unrelated to their osteoarthritis.

**Outcomes**

PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS

Pain: WOMAC Pain: Baseline, after 5 weeks of treatment, and at follow-up five weeks later (at five weeks all patients were started on acupuncture so post-five week follow-up data not included in review)  
 Function: WOMAC function: Not reported

**Tukmachi 2004** (Continued)

Overall index of symptom severity: WOMAC total: Not reported  
 Type of outcome data reported (i.e., post treatment/change from baseline/both): Post-treatment means and standard deviations for each group, and P values of within group changes  
 Additional outcomes reported in the trial but not abstracted: VAS pain score; patient and practitioner global assessment using the VAS scale  
 Adverse effects: Reported as none

**Notes**

Comments: Small trial which is in general carefully described. We excluded all outcome measurements that occurred after the patients on the waiting list began acupuncture treatment. All patients received acupuncture after five-weeks, and therefore, no usable randomized data is available beyond five weeks.

For this trial, two arms received acupuncture and one arm was a waiting list control. We used the 'acupuncture plus continued on medication arm' rather than the 'acupuncture without medication arm' for our 'acupuncture versus waiting list comparison' because for all other included trials, the patients in the acupuncture arm were not restricted from using NSAIDs or analgesic medications.

Weak points: small sample size; medication regimen of patients not clearly described; minor differences in baseline pain (baseline VAS = 6 for groups A and B and 7 for group C), although there were no statistical tests reported to indicate whether these differences were statistically significant; WOMAC function scale is not presented - only the pain and stiffness subscales.

In response to a request for the means and standard deviations of the WOMAC function and WOMAC total scores, the authors provided 34 pages of computer printout of the analyses for the WOMAC scores. However, it was difficult to interpret some of the computer printout, and there were minor discrepancies between the publication and the computer print-outs for the WOMAC pain data. There was no response to a follow-up request for clarification, and therefore, we did not include the WOMAC function or WOMAC total scores for this trial.

There were no standard deviations of changes from baseline directly reported for the WOMAC pain scores. Therefore, in order to include this trial in the between group changes analysis, we converted the reported P values for the within group comparisons for baseline - week 5 into standard deviations of changes for each group. These recalculated standard deviations of changes for each group were entered into RevMan. We made the following conservative assumptions in extracting these P values for the between group changes analysis: For the 'acupuncture and medication group', the P value was reported only as 'P < .001' so we used P = .001 (i.e., the maximum possible P value) to calculate standard deviation of changes; for the 'medication only' group, the P value was reported only as 'not significant', so we used P = .05 (i.e., the minimum possible P value) to calculate standard deviation of changes.

Third author consulted for final decision of data extraction item: For the Jadad scale randomization item, EM scored as 2 and KL scored as 1. EM scored it as 2 because a point was scored for both 'study described as randomized' and 'method of generating randomization sequence appropriate' components. BB agreed with EM and thought we should score this study as 'yes' for item 'method of generating randomization sequence appropriate'.

Source of support: Not reported.

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"The randomisation process used block randomisation, in groups of 10, with sealed envelopes containing cards designating the treatment group prepared by a research nurse unconnected with the study."
Allocation concealment?	Low risk	"The randomisation process used block randomisation, in groups of 10, with sealed envelopes containing cards designating the treatment group prepared by a research nurse unconnected with the study."
Blinding? Versus sham	High risk	Acupuncture vs. symptomatic medication and Acupuncture plus continued on medication. No sham control.

**Acupuncture for peripheral joint osteoarthritis (Review)**

**Tukmachi 2004** (Continued)

Incomplete outcome data addressed? Short term	Low risk	“One patient in group A (10) continued taking concomitant analgesic and anti-inflammatory medication, contrary to the protocol; this patient's data were excluded from the analysis.”
Free of selective reporting?	High risk	<p>“The primary outcome measure was the change in pain as assessed by a 10cm visual analogue scale (VAS). Secondary end points included the WOMAC self-assessment questionnaire, ... Likert scales, give total scores in the ranges of 0-25 (pain) and 0-10 (stiffness). Additionally, patient and practitioner were asked to make a global assessment of the effect at week five and at the final visit, by marking a 10cm visual analogue scale rating labelled 'useless' and 'excellent' at opposite ends.” It appears that the entire WOMAC scale was used, but the WOMAC function results are not reported.</p> <p>Findings reported in Tables 2-4 and <a href="#">Figure 3, 4</a>.</p>
Free of other bias?	High risk	<p>Baseline of WOMAC and VAS information was not reported.</p> <p>Co-intervention was not reported.</p> <p>Compliance acceptability was not reported.</p> <p>No ITT analysis was applied.</p>

**Vas 2004**

Methods	<p>Design: Parallel</p> <p>Blinding: Patient and outcomes assessor blinding</p> <p>Attempt to confirm patient blinding for sham control?: No, however report stated that no patients lost to the study made any reference to belonging to one or the other of the two groups</p> <p>Drop-outs/withdrawals: acupuncture 2 and sham 8</p> <p>CBRG score: 1-1-1-.5-0-.5-1-?-0-1-1</p> <p>Duration: 12 weeks of treatment with final evaluation one week after end of treatment</p> <p>Type of analysis reported: Intention-to-treat analysis, with conservative replacement strategy</p>
Participants	<p>Setting: Outpatient pain clinic, Spain</p> <p>Mean age (+/-SD or Range): ~67(10)</p> <p>Men/Women (n/n): 16/81</p> <p>Recruitment method: referral by doctors at three health centers in study area</p> <p>Mean pain duration (SD) years: 7.5(8.5)</p> <p>For bilateral OA diagnosis, which knee treated/evaluated?: unclear for treatment; for evaluation, the knee with the worst results [In the case of bilateral arthritis, treatment was effected on both knees, although the assessment was limited to the one presenting the worse results]</p> <p>Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): ACR criteria</p> <p>Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Yes, at least grade 1 according to Ahlback classification</p> <p>Minimum duration (and extent) of knee pain required to be eligible: Pain in one or both knees for preceding 3 months or longer; illness had to be symptomatic at moment of selection</p> <p>Hospital inpatients? (Y/N; if Y list number inpatients): No</p> <p>Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Not reported [Patients with a previous history of knee surgery were implicitly excluded, although this was not, in fact, stated as so in the text]</p> <p>Were people with a history of acupuncture treatment excluded? Yes</p> <p>Other important inclusion criteria: 45 years or older; willing and able to complete study questionnaire</p> <p>Important exclusion criteria: severe concomitant illness; inflammatory, metabolic, or neuropathic arthropathies; existing treatment with antineoplastic, corticoid, or immunosuppressive drugs</p>

**Vas 2004** (Continued)

Interventions	<p>TEST GROUP INTERVENTION: acupuncture          N allocated to acupuncture: 48          Style of acupuncture: Chinese          Point selection: Formula          Points stimulated: Electroacupuncture at GB34, Sp6, 9, St36, 40, KI3, LI4, EX32 (Xiyan)          Total length of treatment period (weeks): 12 weeks          Number of sessions target (mean): 12 (?mean) [All the participants in the study received 12 sessions]          Times per week: 1          Number of points used: 8          Insertion depth: Not reported [The puncture was carried out in accordance with the classical doctrines of traditional Chinese medicine, until Deqi was achieved]          Was De qi reportedly sought?: Yes          Duration (mins): Not reported [The duration of each of the sessions was 20 minutes]          Method of stimulation: Electrical with WQ-10D1</p> <p>CONTROL GROUP A (sham, if used): retractable needles went into cylinders (ie non-penetrating Streitberger needle) and placed at true points, and mock electrostimulation          N allocated to control group A: 49          Total length of treatment period: Not reported [12 weeks]          Number of sessions target (mean): 12          Times per week: 1          (If relevant) Number of points used: 8          (If relevant) Insertion depth: Not applicable          Was De qi sought?: Very unlikely          Duration (mins): Not reported [The duration of each of the sessions was 20 minutes]          (If relevant) Method of stimulation: None (Mock electrostimulation)</p> <p>Any co-interventions in all groups?: In all treatment groups, patients were allowed to treat osteoarthritis knee pain with oral NSAIDs if necessary. The use of other pain treatments, such as drugs acting through the central nervous system, or corticosteroids, was not allowed.</p>
Outcomes	<p>PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS</p> <p>Pain: WOMAC Pain: Baseline, and one week after the end of the 12 week treatment period          Function: WOMAC function: Baseline, and one week after the end of the 12 week treatment period          Overall index of symptom severity: WOMAC total: Baseline, and one week after the end of the 12 week treatment period          Type of outcome data reported (i.e., post treatment/change from baseline/both): Post-treatment means and standard deviations for each group          Additional outcomes reported in the trial but not abstracted: Pain intensity on VAS scale; profile of quality of life in the chronically ill instrument; dosage of diclofenac taken during treatment          Adverse effects: reported that adverse effects were limited to 3 patients who reported bruising at one of the acupuncture points (SP9)</p>
Notes	<p>Comments: Well reported, methodologically high-quality study. Extremely conservative intention-to-treat analysis: For patients who withdrew, the score used was the worst of the scores obtained for the intervention group and the best obtained for the control group.</p> <p>Weak points: extremely good results; more participants discontinued the intervention in the sham group than the acupuncture group (8 versus 1).</p> <p>Jorge Vas provided the mean changes and standard deviations of changes for the acupuncture and placebo acupuncture groups, which were entered into RevMan for the between group changes analysis. In addition, he provided additional information, not reported in the publication, about the participants and interventions. We have enclosed this additional information in brackets.</p> <p>Third author consulted for final decision of data extraction item: Based on the publication text, EM had scored the 'compliance' item (i.e., item I of the Cochrane Back Review Group scale) as Yes and KL had scored this item as Don't know; BB was asked to make the final determination, and BB agreed with KL that it should be scored as Don't know.</p>



**Vas 2004** (Continued)

For this trial, the acceptability of the compliance in the two groups was determined to be not adequately described in the publication and, therefore, we scored the compliance item in the Cochrane Back Review Group scale (i.e., Item I) as 'Don't know'. Dr. Vas informed us that "I believe that, with the comments made regarding the other sections, we have made it clear that the two treatments were indistinguishable, and presented a similar duration and periodicity of treatment." Based on the additional information the author provided us about the duration and number of sessions for both the acupuncture and sham groups, the compliance would be considered acceptable. Including this additional information about compliance from the RCT author, the Cochrane Back Review Group score would be increased from 7 (based on the publication alone) to 8.

Source of support: "This study was partly financed by Servicio Andaluz de Salud (Grant No 192/99)."

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"We used a computer program to assign the patients randomly to one or the other group."
Allocation concealment?	Low risk	"We used a simple random allocation method. We sent out sealed opaque envelopes. Only the doctor applying the treatment was aware which group each patient had been assigned to, and he did not participate in any phase of the subsequent evaluation. We took precautions to maintain the confidentiality of the data concerning the participating patients."
Blinding? Versus sham	Unclear risk	<p>"The same specialist carried out the placebo acupuncture, at the same frequency and for the same duration as for the group receiving the true intervention. Retractable needles went into small adhesive cylinders, such that the needle was supported but did not perforate the skin. The acupuncturist then placed the needles over the same points as were used for the true acupuncture group. He connected the same pairs of electrodes and simulated the electrical connection."</p> <p>The credibility of the specific sham needle used in this trial was not tested among the participants included.</p> <p>However report stated that "none of the patients lost to the study made any reference to belonging to one or the other of the two groups."</p>
Incomplete outcome data addressed? Short term	High risk	<p>"Although the standard treatment was given for 12 weeks, participants who dropped out of the study were mainly from the control group, and only six people were lost to the study because their condition did not improve.</p> <p>Observations available for analysis is 47(48) in acupuncture group, 41(49) in sham acupuncture group (see Figure2). Drop-outs/ withdrawals: acupuncture 2 and sham 8.</p> <p>Even though the drop-outs were low, the differential drop-out in the sham and acupuncture groups may lead to bias.</p>
Free of selective reporting?	Low risk	<p>"We used as the primary efficacy end point the WOMAC index and its three subscales (pain (0-20), stiffness (0-8), and physical function (0-68)), pain in the knee on a visual analogue scale from 0 to 100, the dosage of diclofenac accumulated, and the profile of quality of life in the chronically ill (PQLC) instrument." WOMAC pain, function, stiffness, total and Pain intensity on VAS scale; profile of quality of life in the chronically ill instrument; dosage of diclofenac taken during treatment were reported at baseline and one week after the end of the 12 week treatment period.</p> <p>Findings reported in Table2, Table3.</p>

**Vas 2004** (Continued)

Free of other bias?                      Low risk

**Williamson 2007**

Methods	Design: Parallel Blinding: No Drop-outs/withdrawals: "89% of patients could be assessed at 7 weeks and 66% at 12 weeks."  CBRG score: 1-1-1-0-0-0-?-?-0-1-1 Duration: 6 weeks of acupuncture, patients followed up for 12 weeks (and also assessed again later, 3 months after their knee replacement surgery) Type of analysis reported: "intention-to-treat with the baseline values used in place of any missing follow-up values."
Participants	Setting: NHS outpatient group setting, in the Swindon, Wiltshire, UK Mean age (+/-SD or Range): ~71(8) Men/Women (n/n): 84/97 Recruitment method: "We invited 559 patients on the waiting list for knee replacement surgery (total or unicompartmental, unilateral or bilateral) to take part in the study" Mean pain duration (SD) years: Not reported For bilateral OA diagnosis, which knee treated/evaluated?: Not reported Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): Assumedly so, since inclusion criteria was that "patients listed for knee arthroplasty due to OA" Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Not reported Minimum duration (and extent) of knee pain required to be eligible: pain lasting more than 3 months Hospital inpatients? (Y/N; if Y list number inpatients): No, outpatient setting Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Not reported Were people with a history of acupuncture treatment excluded?: No, but no acupuncture allowed within the last year Other important inclusion criteria: None reported Important exclusion criteria: "taking anticoagulants; within 2 months after receiving an intra-articular steroid injection; experiencing back pain associated with referred leg pain; suffering from ipsilateral OA of the hip; suffering psoriasis or other skin disease in the region of the knee; suffering from rheumatoid arthritis; and if they had received acupuncture or physiotherapy treatment in the last year."
Interventions	TEST GROUP INTERVENTION: Acupuncture N allocated to acupuncture: 60 Style of acupuncture: Chinese Point selection: Flexible formula Points stimulated: Following points used for all patients: SP10, ST35, ST36, GB34, Xiyuan, SP9, LIV3. Also, up to 3 needles were used at trigger or traditional points, at the discretion of the physiotherapist. Total length of treatment period (weeks): 6 weeks Number of sessions target (mean): 6 (mean not reported) Times per week: 1x/week Number of points used: formula of 7 points, plus 3 additional points possible Insertion depth: Not reported Was De qi reportedly sought?: Yes Duration (mins): 20 mins Method of stimulation: manual stimulation only  CONTROL GROUP A (sham, if used): physiotherapy N allocated to control group A: 60 Total length of treatment period: 6 weeks Number of sessions target (mean): 6 Times per week: 1x/week (If relevant) Number of points used: NA (If relevant) Insertion depth: NA Was De qi sought?: NA

**Williamson 2007** (Continued)

Duration (mins): 60  
 (If relevant) Method of stimulation: NA

CONTROL GROUP B: Advice and exercise -"received an exercise and advice leaflet"  
 N allocated to control group B: 61  
 Total length of treatment period: NA -- they only received the leaflet at enrollment  
 Number of sessions target (mean): 0  
 Times per week:0  
 Duration (minutes):0

Any co-interventions in all groups?

**Outcomes** Wrote to the author to request the WOMAC pain and function data at 7 and 12 weeks -- if she cannot provide it, we may have to consider using the VAS outcome for our pain measure. Also the changes from baseline data were requested from the authors, since only the post-treatment data were reported in the article.

**Notes** The dropout rate was very high at the 12 week follow-up point; however, this point was not included in the meta-analysis (i.e. because it was not longer than 3 months and hence not eligible as the long-term follow-up, and was also not less than three months and closest to 8 weeks and hence not eligible as the short-term follow-up). The WOMAC pain and function sub-scores were not included in the publication and were provided by the author (as both post-treatment values and changes from baseline). The change from baseline values of WOMAC total scores were not included in the publication and were not provided by the author so are not included in the meta-analysis.

The WOMAC subscore data source came from "Data for Cochrane" sent by Matthew Wyatt on January 29, 2008. We calculated SDs of WOMAC totals by taking the square root of the sum of the variances of the sub-scale scores, making the assumption that these sub-scale scores were independent.

Source of support: The study was funded by "Research and Development Grant, The Great Western Hospital, Swindon".

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"An orthopaedic consultant using computerized block randomization provided 180 sealed opaque envelopes. The computer was asked to block randomize all 180 allocations at once. No stratification was used. The 181st patient was offered a choice of three sealed opaque envelopes each containing one group allocation."
Allocation concealment?	Low risk	"An orthopaedic consultant using computerized block randomization provided 180 sealed opaque envelopes. The computer was asked to block randomize all 180 allocations at once. No stratification was used. The 181st patient was offered a choice of three sealed opaque envelopes each containing one group allocation. The envelopes were opened by the patient in the presence of the study physiotherapist immediately after recruitment to the study."
Blinding? Versus sham	High risk	Acupuncture vs. physiotherapy and advice and exercise. No sham control group.
Incomplete outcome data addressed? Short term	High risk	"89% of patients could be assessed at 7 weeks and 66% at 12 weeks."  At 7 weeks follow-up, there were 56(61) in acupuncture groups, 43(60) in physiotherapy group, 59(61) in control (home exercise group) group. Imbalance in losses to follow-up among the treatment groups is not addressed.  See Fig 2.

**Williamson 2007** (Continued)

Free of selective reporting?	Low risk	<p>“A patient completed a questionnaire containing: Oxford Knee Score (OKS) [19] (primary outcome measure); Western Ontario MacMaster (WOMAC) Score [20], 10-cm pain visual analogue scale (VAS); Hospital Anxiety and Depression score (HAD) [21]. Patients also undertook a 50-m timed walk, and were weighed at each assessment. Assessments took place at baseline, week 7, week 12 and 3 months post-operatively. Duration of hospital stay was determined from medical records.”</p> <p>Findings reported in <a href="#">Table 4, 3.</a></p>
Free of other bias?	High risk	<p>Co-intervention was not reported.</p> <p>Compliance acceptability was not reported.</p> <p>Unclear whether ITT analysis applied.</p>

**Witt 2005**

Methods	<p>Design: Parallel</p> <p>Blinding: Patient and outcomes assessor blinding for sham control comparison only. (Since the assessments were done by the patients themselves using questionnaires, we considered the sham arm of this trial to have been both patient and outcomes assessor blinded.) Neither patient nor outcomes assessor blinding for waiting list control.</p> <p>Attempt to confirm patient blinding for sham control?: Yes, credibility questionnaire after the 3rd acupuncture session; also patients asked at end of study which type of acupuncture they thought they had received. After three sessions, patients rated credibility of acupuncture and sham much as the same and very high, and at the end of study, most patients believed they received real acupuncture</p> <p>Drop-outs/withdrawals: At 8 weeks, acupuncture = 4, sham = 3, waiting list = 7; At 52 weeks after baseline, acupuncture = 4 (no additional drop-outs) and sham = 5 (2 additional drop-outs)</p> <p>CBRG score: 1-1-1-1/0-0-1/0-1-1-1-1-1 (number preceding / is sham control group score and number following / is waiting list control group score)</p> <p>Duration: 8 weeks of treatment and additional follow-up for one year after baseline, with measurements made at end of treatment and 26 and 52 weeks after baseline</p> <p>Type of analysis reported: Available cases analysis was the main analysis; additionally intention-to-treat and per protocol as sensitivity analyses</p>
Participants	<p>Setting: Outpatient clinics in Germany (28 centers)</p> <p>Mean age (+/-SD or Range): 64(7)</p> <p>Men/Women (n/n): 99/195</p> <p>Recruitment method: Mostly reports in local newspapers</p> <p>Mean pain duration (SD) years: 9.2(7.9)</p> <p>For bilateral OA diagnosis, which knee treated/evaluated?: Both knees were treated, and knee defined at baseline as most painful was evaluated throughout the study</p> <p>Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): ACR criteria</p> <p>Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Yes, Kellgren at least 2</p> <p>Minimum duration (and extent) of knee pain required to be eligible: Average pain intensity of 40 or more on a 100 mm VAS in 7 days before baseline assessment</p> <p>Hospital inpatients? (Y/N; if Y list number inpatients): No</p> <p>Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Patients with previous knee surgery or arthroscopy of the affected knee in the past year excluded.</p> <p>Were people with a history of acupuncture treatment excluded?: No, but people with acupuncture in past 12 months were excluded</p> <p>Other important inclusion criteria: written informed consent</p>

**Witt 2005** (Continued)

Important exclusion criteria: malignant, inflammatory, or autoimmune disease; arthroscopy of knee; systemic corticosteroid treatment

**Interventions**

TEST GROUP INTERVENTION: acupuncture  
 N allocated to acupuncture: 150  
 Style of acupuncture: Chinese  
 Point selection: Flexible formula  
 Points stimulated: At least 6 out of a possible 12 local points (i.e., St34, 35, 36, Sp9, 10, BL40, KI10, GB33, 34, LR8, EX31(Heding), and EX32 (Xiyian)) and at least 2 from possible 10 distant points (i.e., Sp4, 5, 6, St6, BL20, 57, 58, 60, 62, KI3), with points chosen according to principles of traditional Chinese medicine  
 Total length of treatment period (weeks): 8  
 Number of sessions target (mean): 12  
 Times per week: Twice a week for first four weeks and once a week for second four weeks  
 Number of points used: 17 mean (minimum 8 for unilateral and 16 for bilateral)  
 Insertion depth: Not standardized (Insertion depth based on instruction in common TCM textbook)  
 Was De qi reportedly sought?: Yes  
 Duration (mins): 30  
 Method of stimulation: Manual stimulation (at least once during each session)

CONTROL GROUP A (sham, if used): minimal (sham) acupuncture involved superficial insertion of fine needles (20-40 mm) at predefined distant non-acupuncture points (i.e., not in the area of the knee)  
 N allocated to control group A: 76  
 Total length of treatment period: 8 weeks  
 Number of sessions target (mean): 12  
 Times per week: Twice a week for first four weeks and one time a week for second four weeks  
 (If relevant) Number of points used: 13 mean (SD = 3)  
 (If relevant) Insertion depth: superficial  
 Was De qi sought?: No  
 Duration (mins): [30]  
 (If relevant) Method of stimulation: None

CONTROL GROUP B: Waiting list  
 N allocated to control group B: 74  
 Total length of treatment period: NA / 8 week waiting list  
 Number of sessions target (mean): NA  
 Times per week: NA  
 Duration (minutes): NA

Any co-interventions in all groups? Oral NSAIDs for treating knee osteoarthritis allowed in all three groups

**Outcomes**

PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS

Pain: WOMAC Pain: Baseline, after 8 week treatment period, and at follow-up 26 and 52 weeks after baseline  
 Function: WOMAC function: Baseline, after 8 week treatment period, and at follow-up 26 and 52 weeks after baseline  
 Total: WOMAC total: Baseline, after 8 week treatment period, and at follow-up 26 and 52 weeks after baseline  
 Type of outcome data reported (i.e., post treatment/change from baseline/both): Post-treatment means and standard deviations for each group (adjusted for baseline values) as well as post-treatment differences between groups (also adjusted for baseline values)  
 Additional outcomes reported in the trial but not abstracted: German version of the pain disability index; scale for assessing emotional aspects of pain; depression scale; German version of the SF-36; numerical rating scales for pain intensity; workdays lost; global assessment; number of days with pain and medication.  
 Adverse effects: acupuncture (n = 20); minimal acupuncture (n = 13); most often haematoma or minor bleeding

**Witt 2005** (Continued)

## Notes

Comments: At the end of the 8 weeks of treatment, acupuncture was better than sham acupuncture and waiting list groups for nearly all outcomes; at follow-up, there was usually only a trend favoring acupuncture

At the end of 8 weeks, the waiting list group was administered acupuncture. Therefore, data from the waiting list group was extracted only for the end of treatment time point, and there is no data comparing acupuncture versus waiting list at the 26 or 52 week time points.

Klaus Linde, a co-author of this RCT and a co-author of this systematic review, provided the unadjusted mean changes and standard deviations of changes for the acupuncture and sham groups, which were entered into RevMan for the between group changes analysis. He also provided the unadjusted post-treatment values, which were entered into RevMan for the post-treatment analysis, (instead of the baseline adjusted post-treatment values reported in the publication). Finally, he provided additional information about the interventions, which we have enclosed in brackets.

The data source came from "full raw analyses for the 0-100 scale for the WOMAC " sent by Dr. Klaus Linde on March 14, 2007.

Source of support: "The trial was initiated after a request from German health authorities (Federal Committee of Physicians and Social Health Insurance Companies, German Federal Social Insurance Authority) and sponsored by German Social Health Insurance Companies."

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"Patients were randomly assigned to a treatment group stratified by centre in a 2: 1: 1 ratio (acupuncture: minimal acupuncture: waiting list) with a centralised telephone randomisation procedure (random list generated with Sample Size 2.0)."
Allocation concealment?	Low risk	"Patients were randomly assigned to a treatment group stratified by centre in a 2: 1: 1 ratio (acupuncture: minimal acupuncture: waiting list) with a centralised telephone randomisation procedure (random list generated with Sample Size 2.0)."
Blinding? Versus sham	Low risk	<p>Patient and outcomes assessor blinding for sham control comparison only. (Since the assessments were done by the patients themselves using questionnaires, we considered the sham arm of this trial to have been both patient and outcomes assessor blinded.) Neither patient nor outcomes assessor blinding for waiting list control.</p> <p>"After three treatment sessions, patients rated the credibility of acupuncture and minimal acupuncture much the same and as very high, and at the end of the study most patients believed that they had received acupuncture following the principles of Chinese medicine (table 2)."</p> <p>At end of week 52, 66% in the true acupuncture group and 56% in the sham group believed that they were receiving true acupuncture (P = 0.06), and 6% and 6% believed that they were receiving the other type of acupuncture, respectively. 28% and 38% didn't know which type, respectively. (P = 0.332)" (See Table 4)</p>
Incomplete outcome data addressed? Short term	Low risk	<p>Observations available for analysis is 146(150) in acupuncture group, 73(76) in sham acupuncture group, and 67(74) in waiting list group at 8 weeks. See Figure 2.</p> <p>Observations available for analysis is 146(150) in acupuncture group, 71(76) in sham acupuncture group, and 67(74) in waiting list group at 52 weeks. See Figure 2.</p>

**Witt 2005** (Continued)

Free of selective reporting?	Low risk	<p>“The primary outcome measure was the Western Ontario and McMaster Universities Osteoarthritis Index.<sup>10,11</sup> In cases of bilateral osteoarthritis, the knee defined at baseline as most painful was the one assessed throughout the entire study. Furthermore, the patient questionnaire included a modified version of the German Society for the Study of Pain survey,<sup>12</sup> which uses the German version of the pain disability index;<sup>13</sup> a scale for assessing emotional aspects of pain (Schmerzempfindungs-Skala [SES]);<sup>14</sup> the depression scale (Allgemeine Depressionsskala [ADS]);<sup>15</sup> and the German version of the SF-3616 (MOS36-item short form quality-of-life questionnaire) to assess health-related quality of life. Additionally, several questions on sociodemographic characteristics, numerical rating scales for pain intensity, questions about workdays lost, and global assessments were asked. The number of days with pain and medication were documented in a diary by the patients.”</p> <p>“Blinding to treatment and the credibility of the treatment method were assessed by the patients with a credibility questionnaire<sup>17</sup> after the third acupuncture session. At the end of the study, patients were asked whether they thought they had received acupuncture following the principles of Chinese medicine or the other type of acupuncture. Physicians documented medical history, acupuncture treatment, serious adverse events, and side-effects for each session. Patients also reported side-effects at the end of week 8.”</p> <p>Findings reported in Table 1-4, Figure 3, 4.</p>
Free of other bias?	Low risk	

**Witt 2006**

Methods	Design: Parallel Blinding: No Attempt to confirm patient blinding for sham control?: NA Drop-outs/withdrawals: "After 3 months, data were available on 93.2% of the patients (308 in the randomized acupuncture group, 289 in the control group...)"  CBRG score: 1-1-1-0-0-0-?-?-1-1-1 Duration: up to 15 sessions of acupuncture over a 3 month period, and additional follow-up for 6 months after baseline. Measurements made at end of treatment, 3, and 6 months after baseline Type of analysis reported: Available cases analysis was the main analysis; additionally intention-to-treat as sensitivity analyses
Participants	Setting: Outpatient clinics in Germany Mean age (+/-SD or Range): ~61(10) Men/Women (n/n): 135/207 Recruitment method: Patients of selected from a group of experienced primary care practitioners participating in a large research initiative on acupuncture Mean pain duration (SD) years: 5.4(7) For bilateral OA diagnosis, which knee treated/evaluated?: Knee defined at baseline as most painful was evaluated throughout the study (treatment was at discretion of physician) Diagnosis of knee OA required to be eligible? (if yes, describe how patients were verified to have OA): Yes, clinical diagnosis of OA-associated pain in the knee with disease duration of >6 months Radiologic evidence of knee OA required to be eligible? (if yes, describe requirement): Yes, radiologic evidence of OA (osteophyte formation) Minimum duration (and extent) of knee pain required to be eligible: At least 15 days with pain in the preceding 30 days Hospital inpatients? (Y/N; if Y list number inpatients): No Previous knee surgery? (Y/N; if Y list number with previous knee surgery): Not reported

**Witt 2006** (Continued)

Were people with a history of acupuncture treatment excluded?: No. About 31% of patients listed "previous successful acupuncture" as their reason for participating in this study

Other important inclusion criteria: written informed consent

Important exclusion criteria: Knee or hip pain due to inflammation or malignancy was a criterion for exclusion.

**Interventions**

TEST GROUP INTERVENTION: acupuncture

N allocated to acupuncture: 175

Style of acupuncture: Because the specifics of acupuncture treatment were left to the discretion of the physician, acupuncture treatment regimens varied among patients in the study.

Point selection: Individualized (determined by treating physician)

Points stimulated: Number of needles and acupuncture points were chosen at the physician's discretion

Total length of treatment period (weeks): 13

Number of sessions target (mean): target: up to 15; mean: ~11(2.5)

Times per week: Not specifically reported, but assumedly at physician's discretion

Number of points used: At physician's discretion

Insertion depth: Not specifically reported, but assumedly at physician's discretion

Was De qi reportedly sought?: Not specifically reported, but assumedly at physician's discretion

Duration (mins): 30

Method of stimulation: only manual stimulation was allowed (administered at physician's discretion)

CONTROL GROUP A: Waiting list

N allocated to control group B: 167

Total length of treatment period: NA / 13 week waiting list

Number of sessions target (mean): NA

Times per week: NA

Duration (minutes): NA

Any co-interventions in all groups? Patients were permitted to receive any additional conventional treatments as needed.

**Outcomes**

PAIN, FUNCTION, AND OVERALL INDEX OF SYMPTOM SEVERITY OUTCOMES EXTRACTED FROM PUBLICATIONS: MEASUREMENT TIME POINTS

Pain: WOMAC Pain: Baseline, after 13 week treatment period, and at follow-up 26 weeks after baseline

Function: WOMAC function: Baseline, after 13 week treatment period, and at follow-up 26 weeks after baseline

Total: WOMAC total: Baseline, after 13 week treatment period, and at follow-up 26 weeks after baseline

Type of outcome data reported (i.e., post treatment/change from baseline/both): Post-treatment means and standard deviations for each group as well as post-treatment differences between groups

Additional outcomes reported in the trial but not abstracted: Percent reduction in the WOMAC index; German version of the SF-36 physical and mental component scores;

Adverse effects: No life threatening side effects were reported. Only side effects resulting from acupuncture were reported, for the acupuncture patients only; data specific to the knee OA patients could not be obtained from authors. For knee plus hip OA patients: in 5.2% (n = 184), a total of 219 side effects were reported after the patients had acupuncture (66% minor local bleeding or hematoma, 5% pain at the site of needle insertion, 4% vegetative symptoms, and 25% other)

**Notes**

Comments: Need to confirm that co-interventions were indeed either avoided in the trial design or similar between the index and control groups. For example, could the acupuncturists also offer herbs? If they could, I need to change the co-intervention score for the CBRG scale item.

For the patients for whom the site of OA was only the knee, the corresponding author provided the crude means and SDs, both post-treatment and changes from baseline, of the WOMAC totals and sub-scores. For the patients with OA of the hip or knee (i.e. "All" patients), the between group changes were calculated using the pre- and post- means and SDs, and assuming a within-subject pretest-post-test correlation of 0.5. (For the N's the numbers with the WOMAC data complete at 3 months were used, as shown in [Figure 1](#).) For the patients for whom the site of OA was only the hip, no data were available to allow the SDs of change to be calculated and therefore only the post-treatment analysis is included for



**Witt 2006** (Continued)

the hip OA only patients. For the hip OA patients, the number available at the 3 month time point was not reported, only the numbers with hip OA randomized and the numbers with WOMAC data complete at 3 months for all patients. 84.0% of acupuncture patients and 78.6% of all control patients had WOMAC data complete at 3 months. Therefore these percentages were multiplied by the numbers of hip OA only patients randomized to estimate number of hip OA patients with data complete at 3 months.

Source of support: "Supported by the following statutory health insurance funds in Germany: Techniker Krankenkasse,..."

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	Low risk	"Those who met the inclusion criteria, provided informed consent, and signed an agreement for randomization were randomized using a central telephone randomization procedure. For randomization we used blocks of 10, and the random list was generated with SAS software."
Allocation concealment?	Low risk	"Those who met the inclusion criteria, provided informed consent, and signed an agreement for randomization were randomized using a central telephone randomization procedure." "
Blinding? Versus sham	High risk	Acupuncture vs. waiting list. No sham control group.
Incomplete outcome data addressed? Short term	Low risk	"After 3 months, data were available on 93.2% of the patients (308(357) in the randomized acupuncture group, 289(355) in the control group...)"  At 6 months follow-up, there were 282(357) in acupuncture groups, 277(355) in waiting list group.  See <a href="#">Figure 1</a> .
Free of selective reporting?	Low risk	"The patients completed standardized questionnaires, including information on sociodemographic characteristics, at baseline and after 3 months and 6 months. The primary outcome measure was the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (11,12). The patients documented, at baseline, which was the most painful joint and assessed this during the whole study period. As a secondary outcome measure, we used the percent reduction in the WOMAC index."  "As further secondary outcome parameters we used the Short Form 36 (SF-36) component scales (13) to assess health-related quality of life. Side effects were recorded on patient and physician questionnaires after 3 months."  Findings reported in <a href="#">Table 3-4</a> and <a href="#">Figure 2</a> .
Free of other bias?	Unclear risk	Co-intervention was not reported.  Compliance acceptability was not reported.

**Characteristics of excluded studies** [ordered by study ID]

Study	Reason for exclusion
<a href="#">Ammer 1988</a>	results measured only after four weeks of treatment
<a href="#">Coan 1982</a>	cervical OA
<a href="#">Dickens 1989</a>	Results measured only after two weeks of treatment, and again two weeks later
<a href="#">Gaw 1975</a>	OA in a variety of joints, including spine and results for the patients with OA of the peripheral joints were not reported separately from the results of the patients with OA of the spine.
<a href="#">Junnila 1982</a>	not explicitly randomized
<a href="#">Kwon 2001</a>	compared two types of acupuncture (bee venom versus traditional needle)
<a href="#">Loy 1983</a>	cervical spondylosis
<a href="#">Lundeberg 1991</a>	OA of neck
<a href="#">McIndoe 1994</a>	results measured only after 4 weeks of treatment
<a href="#">Milligan 1981</a>	Could not find a copy of this abstract. All co-authors of the abstract were contacted and none had a copy. Also an attempt was made to locate the conference proceedings in which the abstract was published, but the attempt was unsuccessful.
<a href="#">Ng 2003</a>	Results measured only after two weeks of treatment, and again two weeks later
<a href="#">Petrie 1983</a>	cervical OA
<a href="#">Petrie 1986</a>	cervical OA
<a href="#">Petrou 1988</a>	results measured only after two and a half weeks of treatment
<a href="#">Thomas 1991</a>	cervical OA
<a href="#">Tillu 2001</a>	compared two types of acupuncture
<a href="#">Tillu 2002</a>	not explicitly randomized
<a href="#">Weiner 2007</a>	not Chinese style acupuncture -- needles not placed at meridian points
<a href="#">Yurtkuran 1999</a>	results measured only after two weeks of treatment
<a href="#">Zherebkin 1998</a>	not explicitly randomized; included both hip and knee joint patients

### Characteristics of studies awaiting assessment *[ordered by study ID]*

#### [Itoh 2008](#)

Methods	Small randomized trial with blinding between sham and true acupuncture and relatively low attrition (6 drop-outs out of 30)
Participants	Patients with OA (N = 30)
Interventions	Comparing trigger point acupuncture vs acupuncture at standard points vs sham acupuncture

#### [Acupuncture for peripheral joint osteoarthritis \(Review\)](#)

**Itoh 2008** (Continued)

Outcomes	pain intensity (VAS) and WOMAC index.
Notes	Five acupuncture treatment sessions. Total duration = 21 weeks

**Itoh 2008\***

Methods	Unblinded randomized trial comparing four different active treatments for knee OA
Participants	Patients with OA (N = 32)
Interventions	Comparing acupuncture vs TENS vs Acupuncture plus TENS vs topical poultice
Outcomes	pain intensity in a visual analogue scale (VAS) and knee function in WOMAC
Notes	Treatment for 5 weeks, once per week, a treatment session for 15 min. Follow-up evaluation at 10 weeks after the first treatment. Total duration = 10 weeks

**Jubb 2008**

Methods	Small randomized trial comparing acupuncture versus a non-penetrating sham for patients with knee OA
Participants	Patients with OA (N = 68)
Interventions	Comparing acupuncture (manual and electroacupuncture) with non-penetrating sham ('placebo' needle)
Outcomes	pain scale (VAS), the EuroQol score and plasma beta-endorphin.
Notes	Treatment for 5 weeks, 2 sessions/week, a treatment session was for a total of 30 minutes, Follow-up evaluation at 4 weeks after end of treatment. Total duration = 9 weeks.

**Characteristics of ongoing studies** [ordered by study ID]

**Bower**

Trial name or title	A randomized, controlled, single-blind trial to test the efficacy of acupuncture on pain relief in osteoarthritis of the knee
Methods	
Participants	Patients with OA of the knee (N = not specified)
Interventions	RCT comparing active acupuncture versus minimal acupuncture
Outcomes	Only stated as 'treatment-outcome'
Starting date	Start date: May 1999 Expected date of completion: September 2004
Contact information	Susanne.Bower@ncl.ac.uk

**Bower** (Continued)

Notes	Location: Study based at Royal Victoria Infirmary, Newcastle upon Tyne, UK Sponsor: Arthritis Research Campaign
-------	---

**Farrar**

Trial name or title	Efficacy of acupuncture with physical therapy for knee osteoarthritis
Methods	
Participants	Patients with OA of the knee (N = 300)
Interventions	Double-blind RCT comparing acupuncture versus placebo acupuncture as adjunct to physical therapy
Outcomes	pain , function
Starting date	Start date: March 2002 Expected date of completion: December 2005
Contact information	jfarrar@cceb.med.upenn.edu
Notes	Location: Study based at the University of Pennsylvania. Sponsor: National Center for Complementary and Alternative Medicine (Grant number: 5R01AT000304-04)

**McNally**

Trial name or title	A study of acupuncture, physiotherapy or non-intervention in management of painful (OA) knee
Methods	
Participants	Patients with OA of the knee (n = 150)
Interventions	RCT comparing acupuncture versus physiotherapy versus non-intervention
Outcomes	Not specified
Starting date	Start date: October 2001 Expected date of completion: June 2004
Contact information	Dr Jeremy Mc Nally, Department of Rheumatology; Battle Hospital; Oxford Road; Reading RG30 1AG; United Kingdom Telephone: 0118-958-3666
Notes	Location: Study based at Battle Hospital, Reading, UK Sponsor: UK NHS

**Suarez-Almazor**

Trial name or title	Interaction between patient and healthcare provider: response to acupuncture in knee osteoarthritis
Methods	
Participants	Patients with OA of the knee (N = 760)

**Acupuncture for peripheral joint osteoarthritis (Review)**

**Suarez-Almazor** (Continued)

Interventions	Single-blind RCT comparing acupuncture versus sham acupuncture
Outcomes	placebo effects, pain
Starting date	Start date: September 2002 Expected date of completion: August 2006
Contact information	med@bcm.tmc.edu
Notes	Location: Study based at the Baylor College of Medicine in Houston, Texas Sponsor: US National Center for Complementary and Alternative Medicine (Grant number: 5R01AR049999-02)

**van Arendok**

Trial name or title	Does the inclusion of acupuncture improve pain and outcome measures following total knee arthroscopy?
Methods	
Participants	Patients having undergone total knee arthroscopy (N = 90)
Interventions	Pragmatic RCT of acupuncture
Outcomes	WOMAC
Starting date	Start date: December 2001 Expected date of completion: September 2002
Contact information	Ms Norma <a href="#">van Arendok</a> , Physiotherapist; St. Mary's Hospital; Newport; Isle of Wight; PO30 5TG, United Kingdom
Notes	Location: study based at St. Mary's Hospital, Newport, Isle of Wight, UK Sponsor: Not specified

**White**

Trial name or title	The process of acupuncture: a randomized controlled trial and qualitative study to evaluate the relative contributions of specific and non-specific effects
Methods	
Participants	Patients with OA of the knee or hip (N = 288)
Interventions	RCT comparing acupuncture with 1) placebo acupuncture (non-penetrating needles) and 2) placebo transcutaneous electrical nerve stimulation
Outcomes	pain measured on VAS; WOMAC
Starting date	Start date: 1/9/2003 Anticipated end date: 1/9/2007
Contact information	School of Health Professions University of Southampton Highfield Southampton United Kingdom

**Acupuncture for peripheral joint osteoarthritis (Review)**

White (Continued)

SO17 1BJ  
pjw1@soton.ac.uk

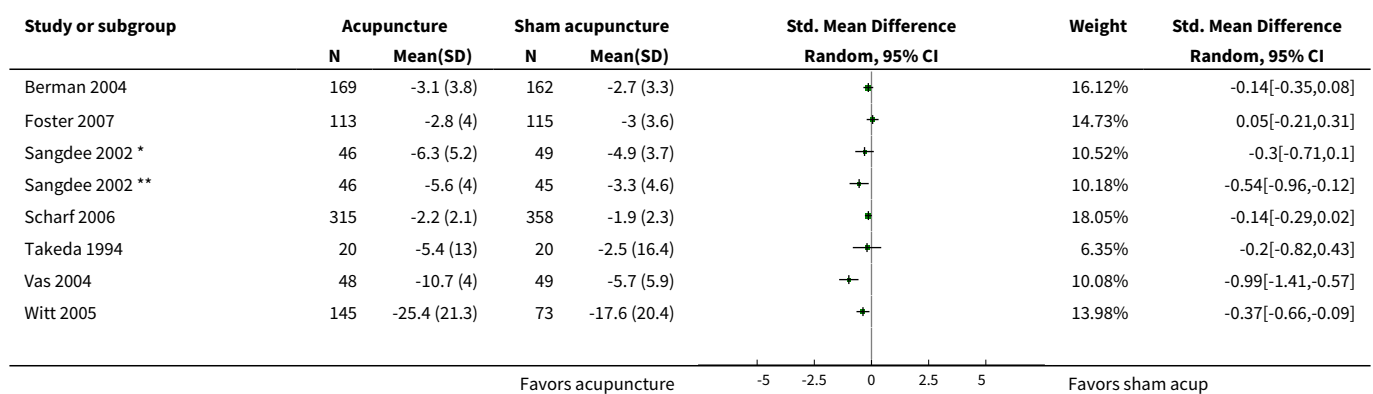
Notes Location: University of Southampton  
Sponsor: Department of Health-Funded Trials for Complementary Alternative Medicine (CAM)  
03/12

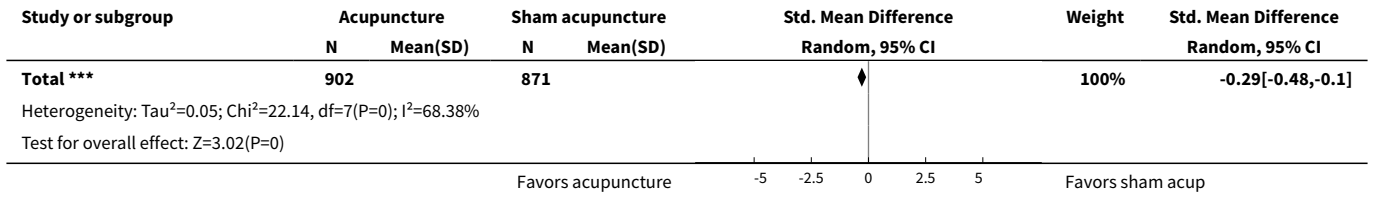
**DATA AND ANALYSES**

**Comparison 1. Acupuncture vs. sham acupuncture for knee OA change from baseline analysis**

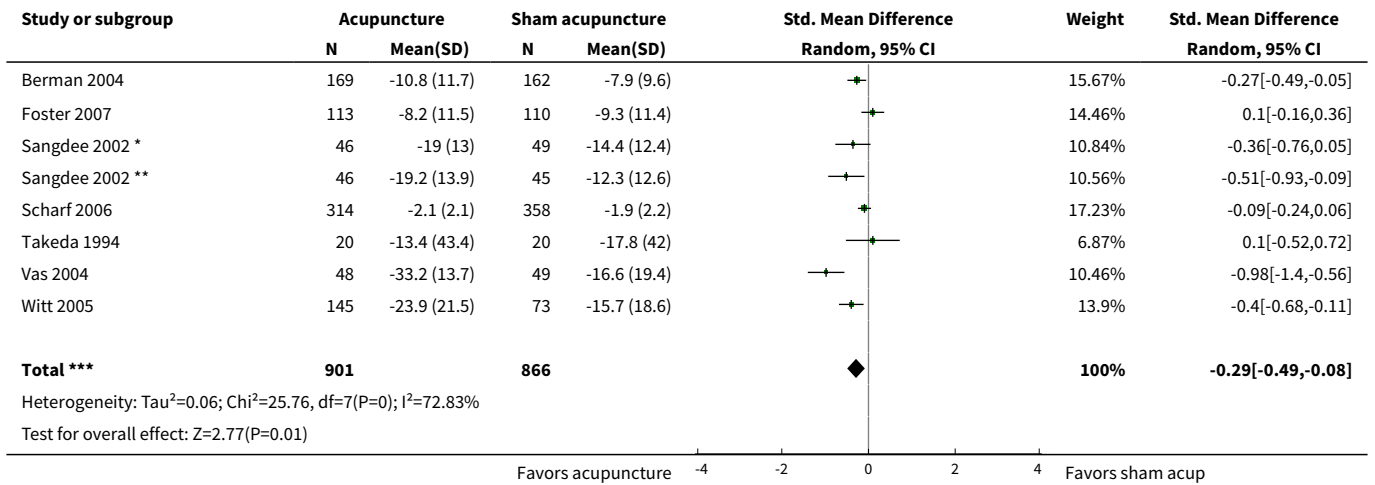
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)	8	1773	Std. Mean Difference (IV, Random, 95% CI)	-0.29 [-0.48, -0.10]
2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)	8	1767	Std. Mean Difference (IV, Random, 95% CI)	-0.29 [-0.49, -0.08]
3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization)	8	1767	Std. Mean Difference (IV, Random, 95% CI)	-0.29 [-0.50, -0.09]
4 Pain (26 weeks after baseline)	4	1399	Std. Mean Difference (IV, Random, 95% CI)	-0.10 [-0.21, 0.01]
5 Function (26 weeks after baseline)	4	1398	Std. Mean Difference (IV, Random, 95% CI)	-0.11 [-0.22, 0.00]
6 Total score (26 weeks after baseline)	4	1398	Std. Mean Difference (IV, Random, 95% CI)	-0.11 [-0.22, -0.00]

**Analysis 1.1. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**

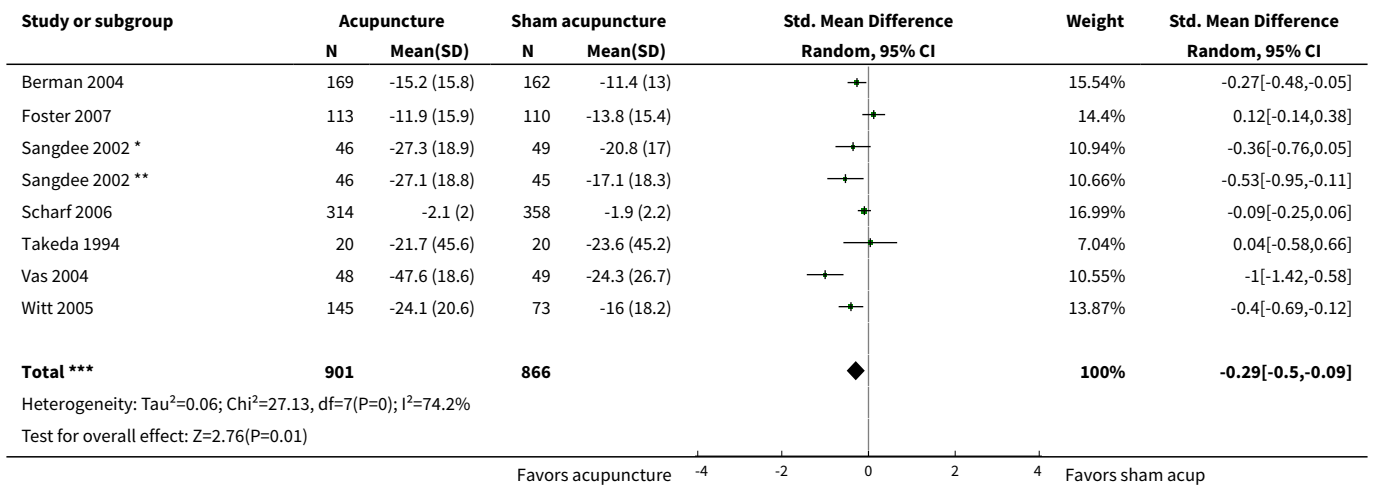




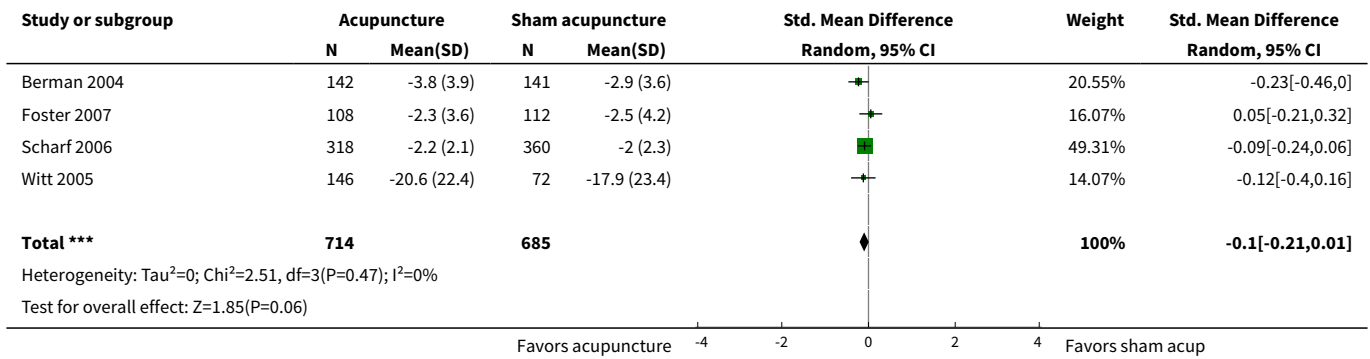
**Analysis 1.2. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**



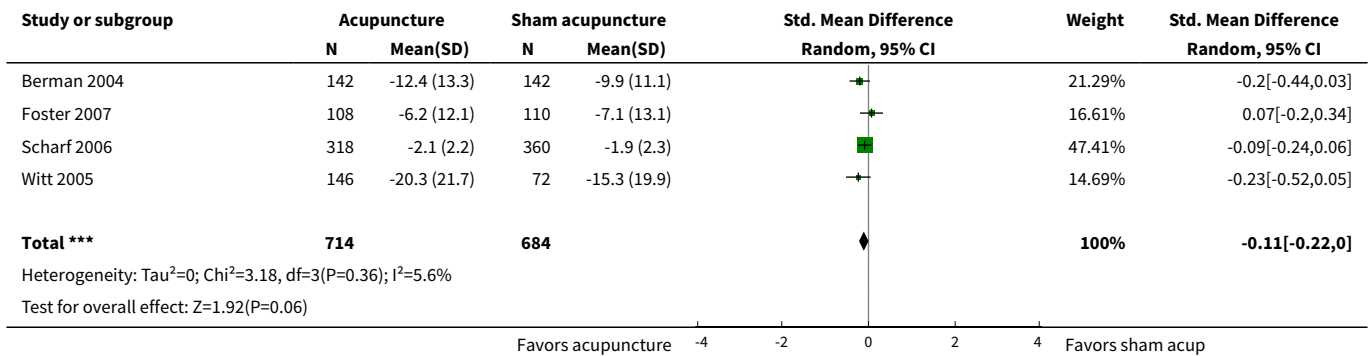
**Analysis 1.3. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization).**



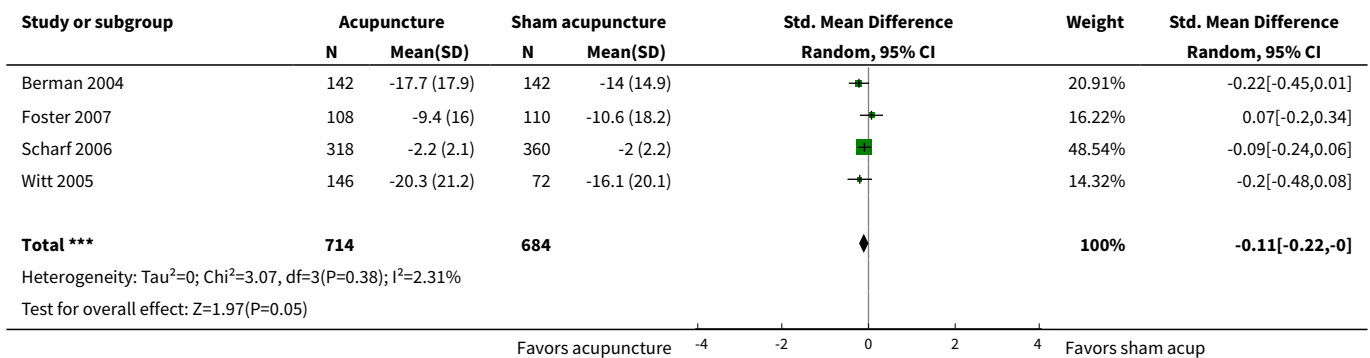
**Analysis 1.4. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 4 Pain (26 weeks after baseline).**



**Analysis 1.5. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 5 Function (26 weeks after baseline).**



**Analysis 1.6. Comparison 1 Acupuncture vs. sham acupuncture for knee OA change from baseline analysis, Outcome 6 Total score (26 weeks after baseline).**

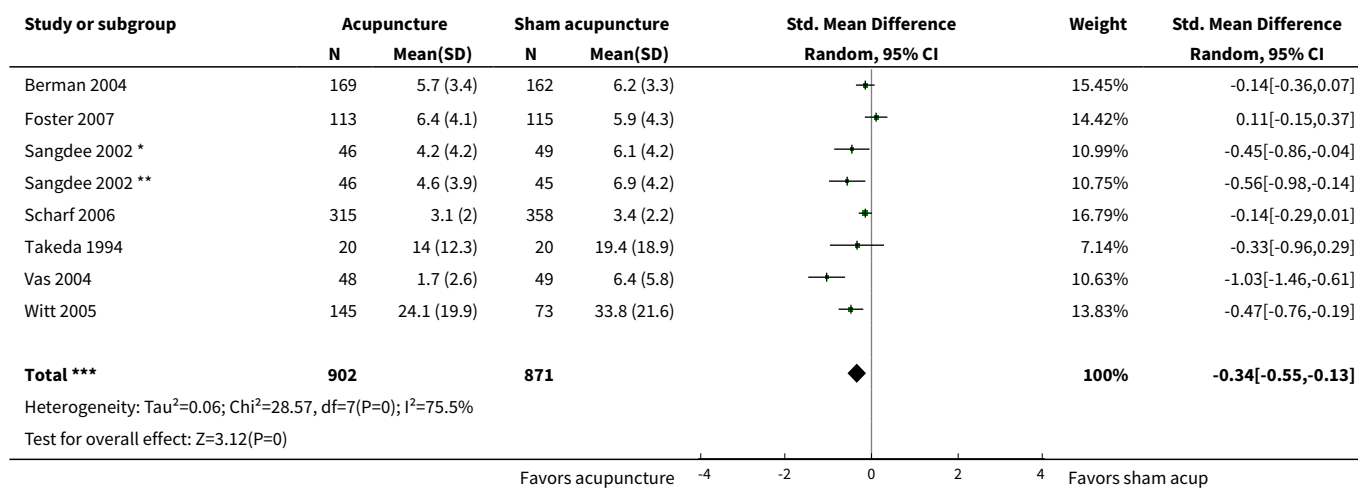




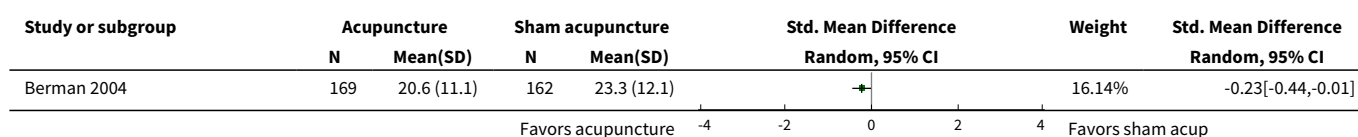
**Comparison 2. Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis**

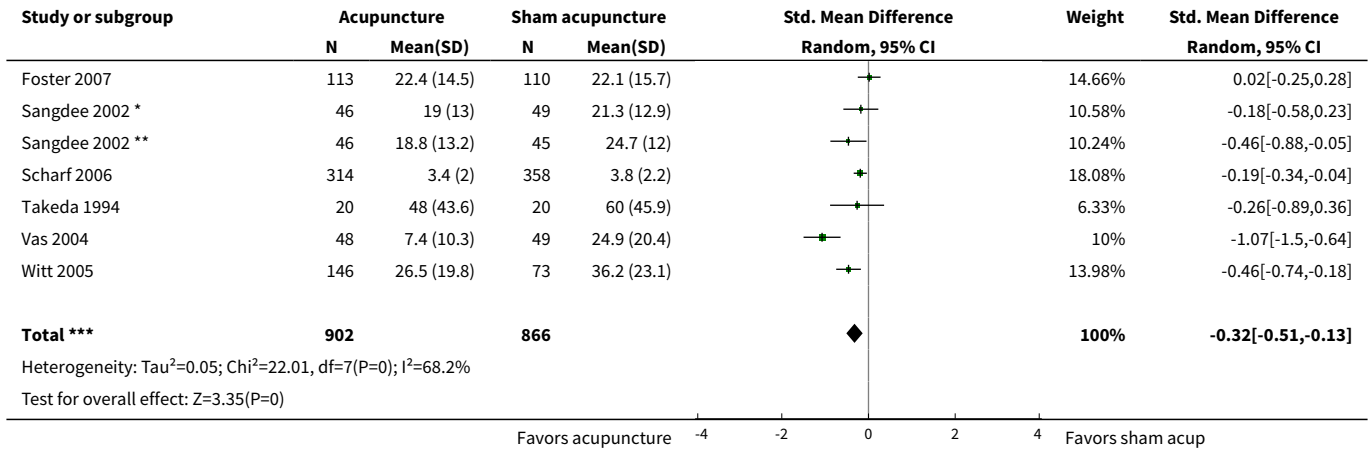
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)	8	1773	Std. Mean Difference (IV, Random, 95% CI)	-0.34 [-0.55, -0.13]
2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)	8	1768	Std. Mean Difference (IV, Random, 95% CI)	-0.32 [-0.51, -0.13]
3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization)	8	1767	Std. Mean Difference (IV, Random, 95% CI)	-0.34 [-0.55, -0.14]
4 Pain (26 weeks after baseline)	4	1400	Std. Mean Difference (IV, Random, 95% CI)	-0.14 [-0.31, 0.02]
5 Function (26 weeks after baseline)	4	1398	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.29, -0.02]
6 Total score (26 weeks after baseline)	4	1398	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.30, -0.01]

**Analysis 2.1. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**

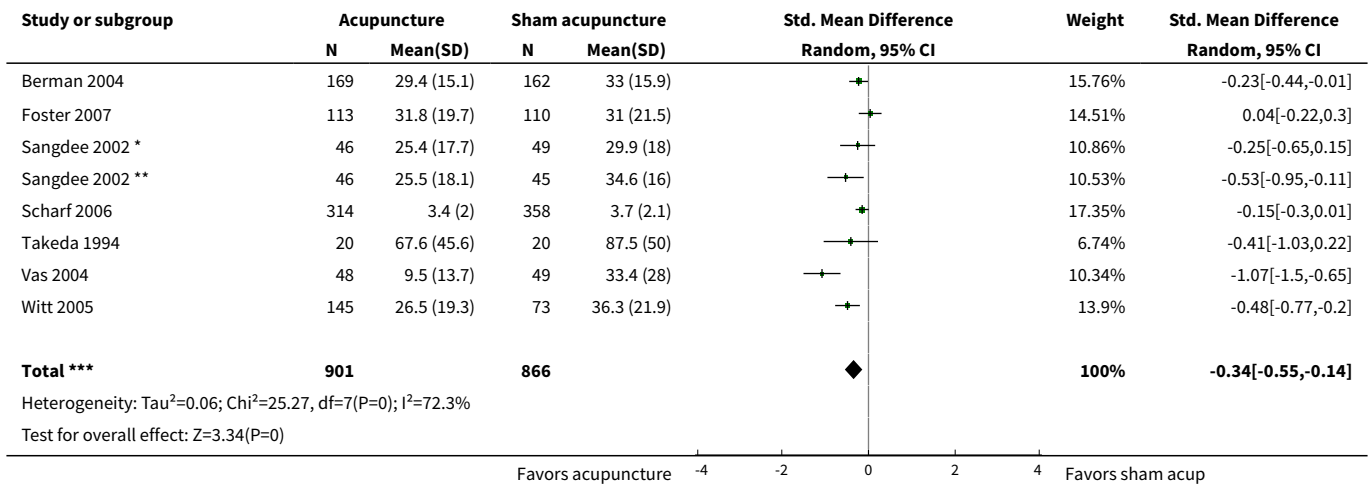


**Analysis 2.2. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**

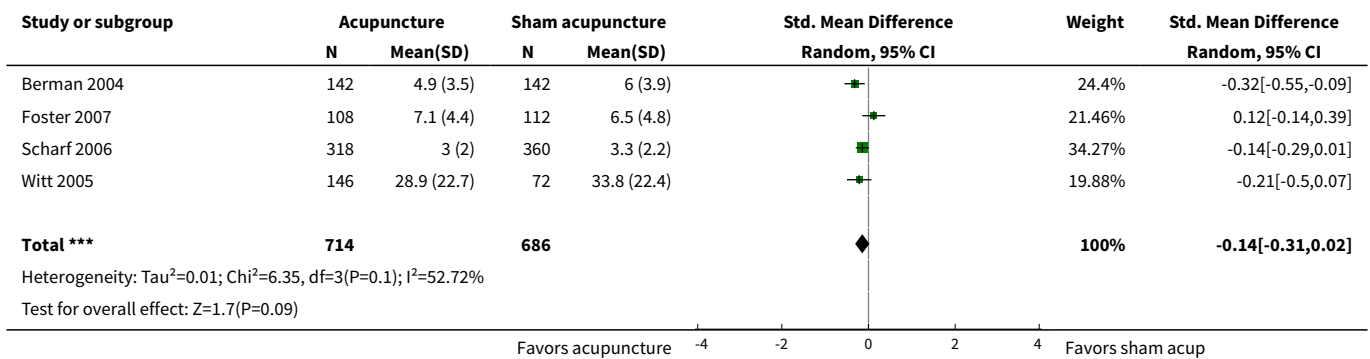




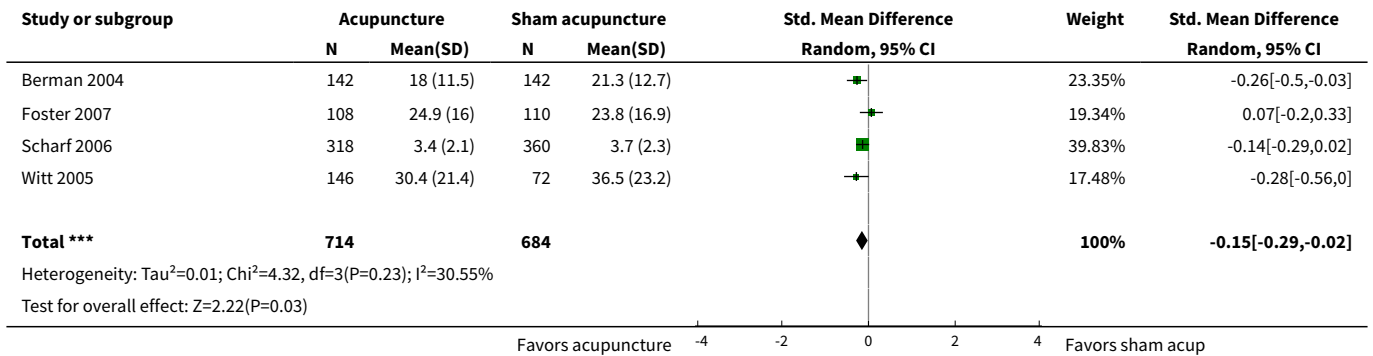
**Analysis 2.3. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization).**



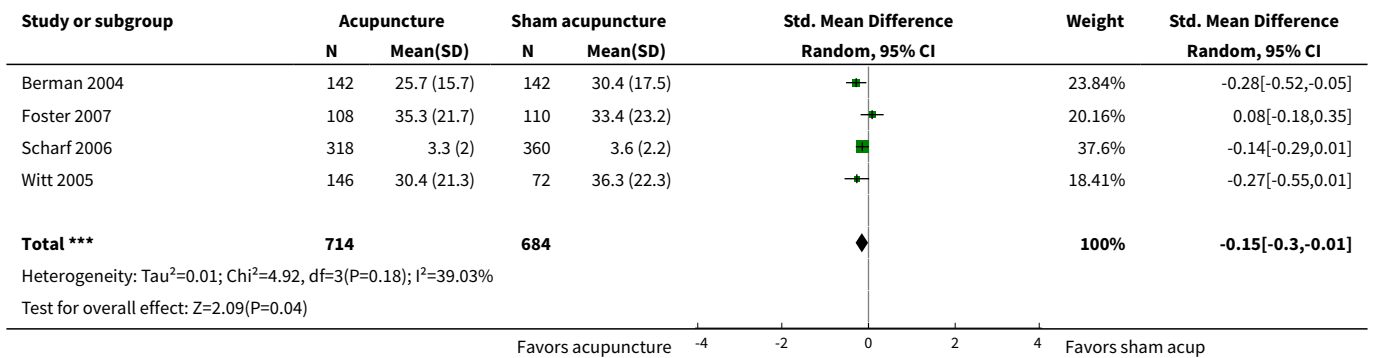
**Analysis 2.4. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 4 Pain (26 weeks after baseline).**



**Analysis 2.5. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 5 Function (26 weeks after baseline).**



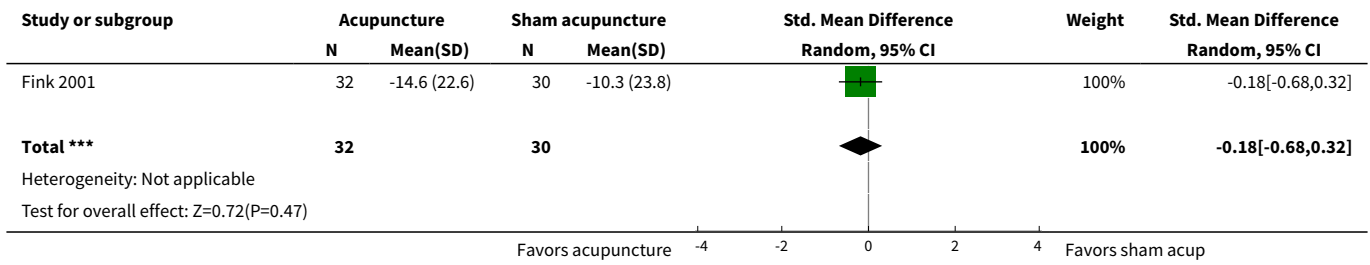
**Analysis 2.6. Comparison 2 Acupuncture vs. sham acupuncture for knee OA post-treatment scores analysis, Outcome 6 Total score (26 weeks after baseline).**



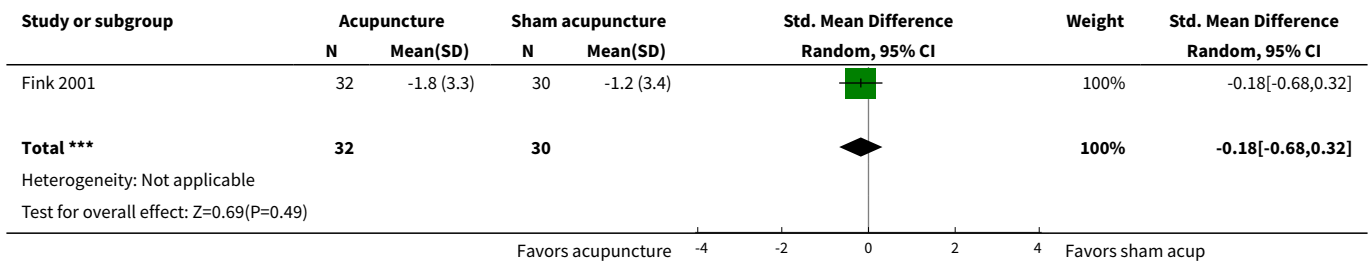
**Comparison 3. Acupuncture vs. sham acupuncture for hip OA change from baseline analysis**

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)	1	62	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.68, 0.32]
2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)	1	62	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.68, 0.32]

**Analysis 3.1. Comparison 3 Acupuncture vs. sham acupuncture for hip OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**



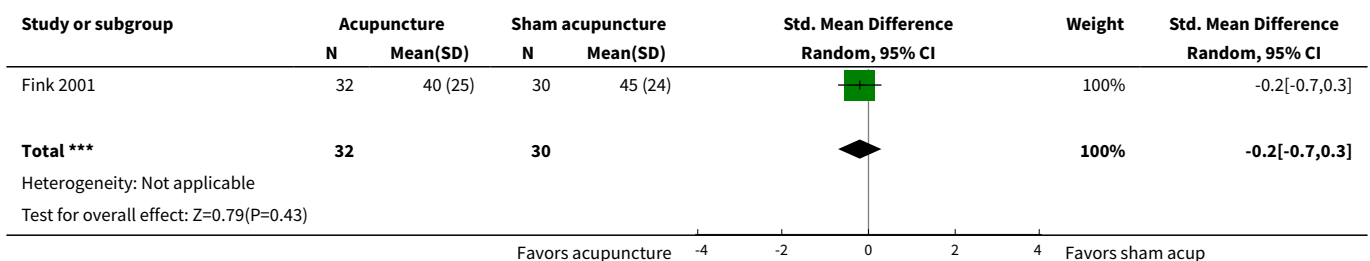
**Analysis 3.2. Comparison 3 Acupuncture vs. sham acupuncture for hip OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**



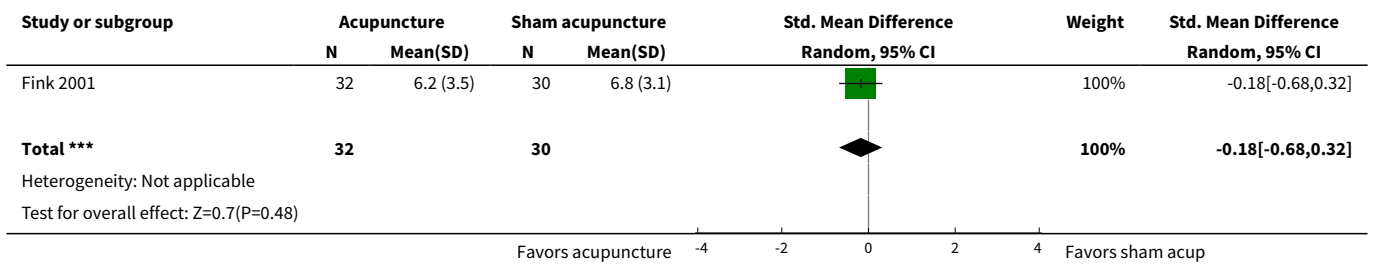
**Comparison 4. Acupuncture vs. sham acupuncture for hip OA post-treatment scores post-analysis**

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)	1	62	Std. Mean Difference (IV, Random, 95% CI)	-0.20 [-0.70, 0.30]
2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)	1	62	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.68, 0.32]

**Analysis 4.1. Comparison 4 Acupuncture vs. sham acupuncture for hip OA post-treatment scores post-analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**



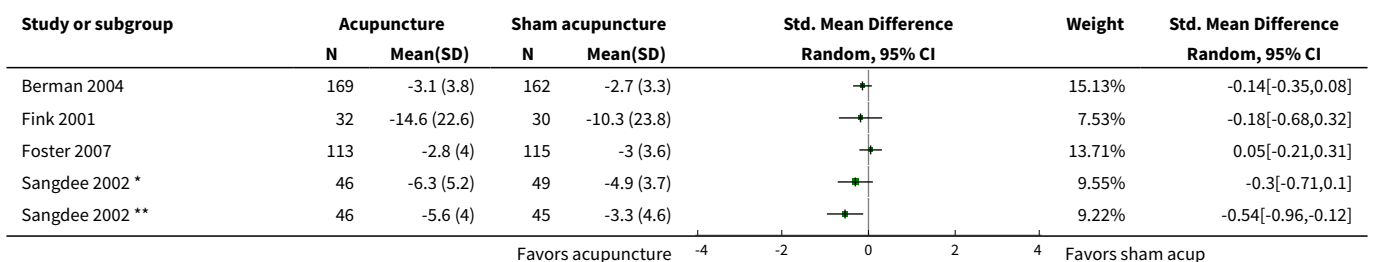
**Analysis 4.2. Comparison 4 Acupuncture vs. sham acupuncture for hip OA post-treatment scores post-analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**

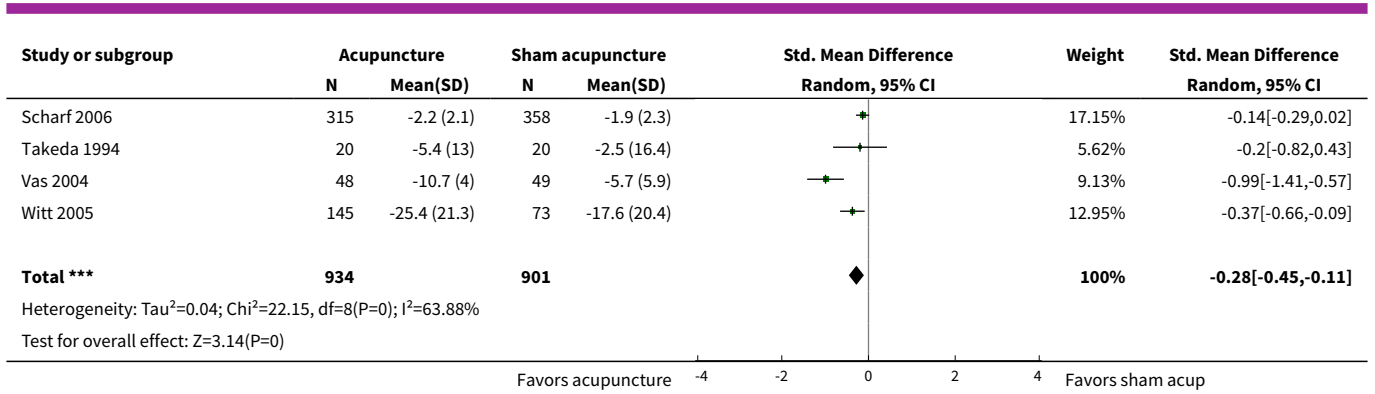


**Comparison 5. Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis**

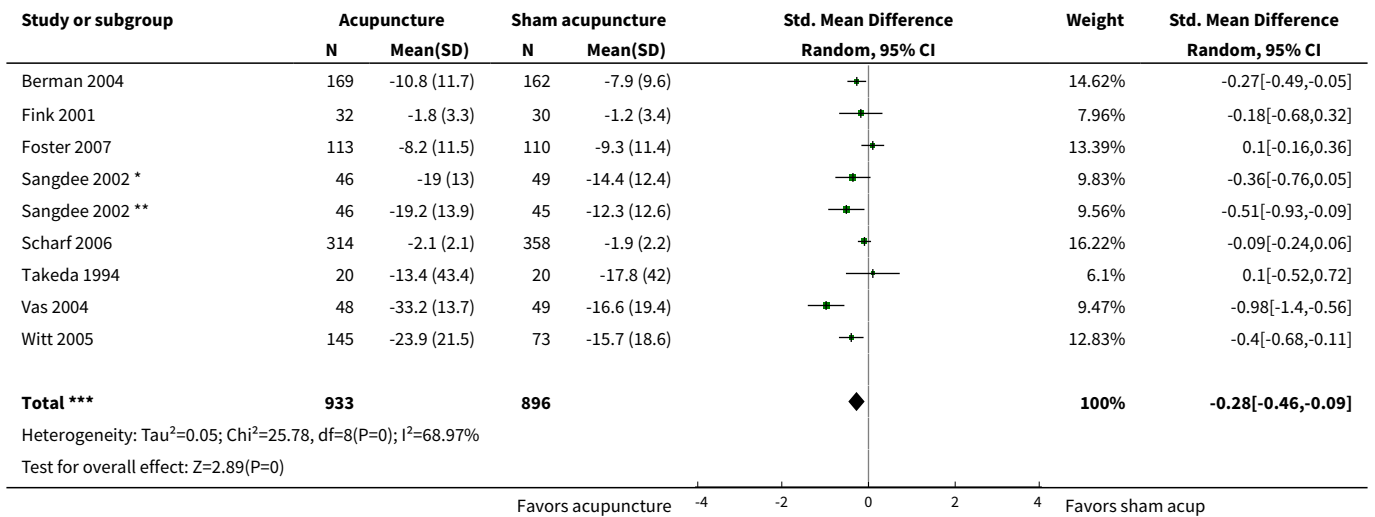
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)	9	1835	Std. Mean Difference (IV, Random, 95% CI)	-0.28 [-0.45, -0.11]
2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)	9	1829	Std. Mean Difference (IV, Random, 95% CI)	-0.28 [-0.46, -0.09]
3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization)	8	1767	Std. Mean Difference (IV, Random, 95% CI)	-0.29 [-0.50, -0.09]
4 Pain (26 weeks after baseline)	4	1399	Std. Mean Difference (IV, Random, 95% CI)	-0.10 [-0.21, 0.01]
5 Function (26 weeks after baseline)	4	1398	Std. Mean Difference (IV, Random, 95% CI)	-0.11 [-0.22, 0.00]
6 Total score (26 weeks after baseline)	4	1398	Std. Mean Difference (IV, Random, 95% CI)	-0.11 [-0.22, -0.00]

**Analysis 5.1. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**

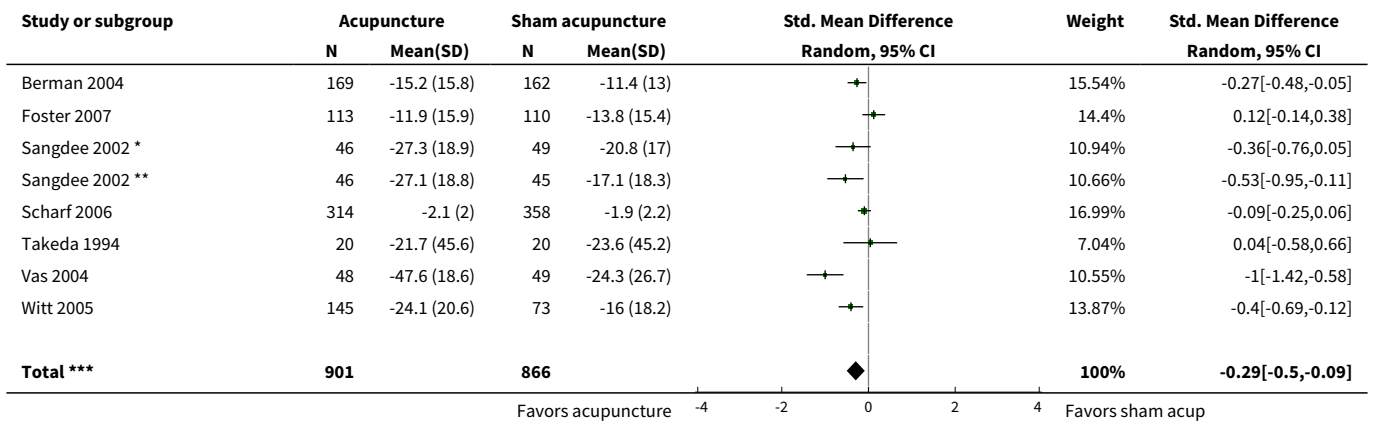


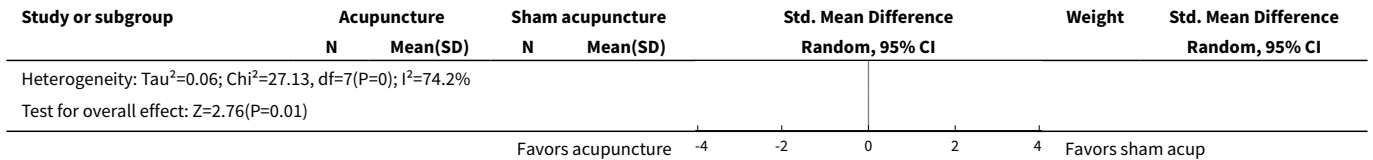


**Analysis 5.2. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**

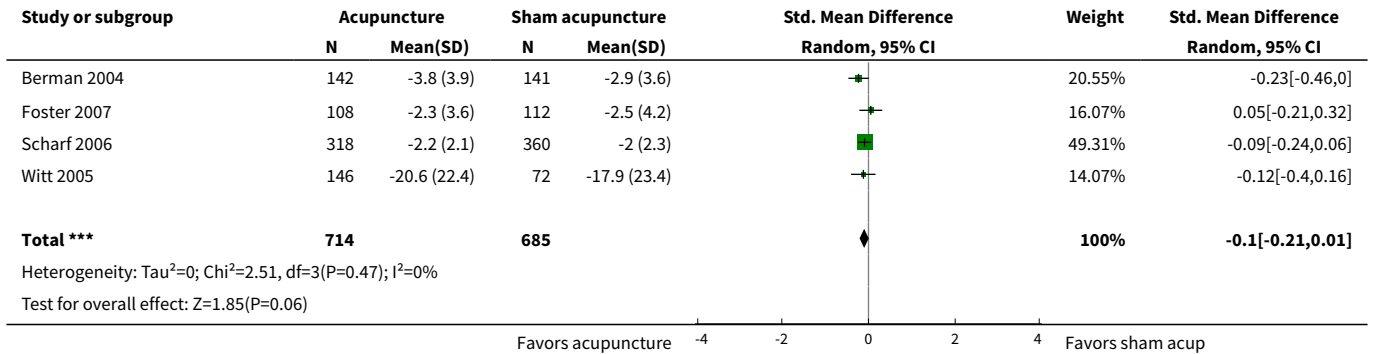


**Analysis 5.3. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization).**

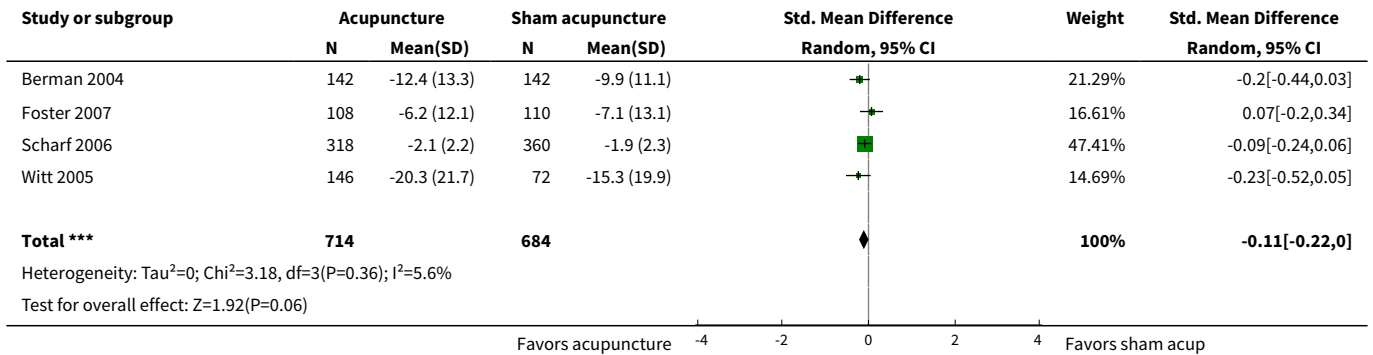




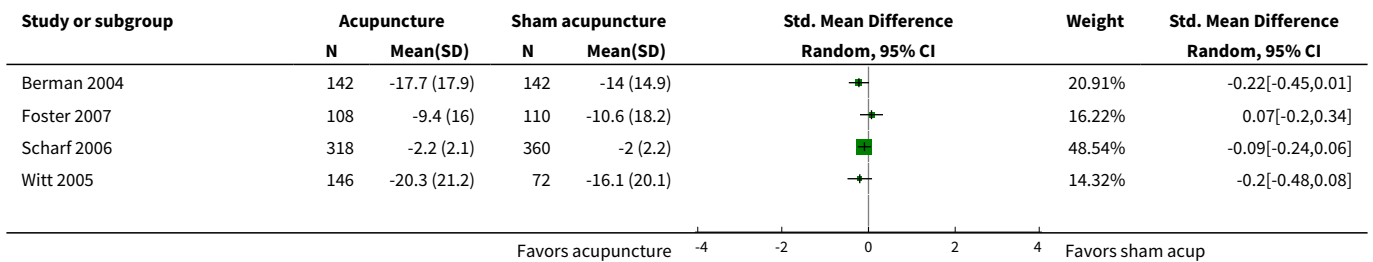
**Analysis 5.4. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 4 Pain (26 weeks after baseline).**

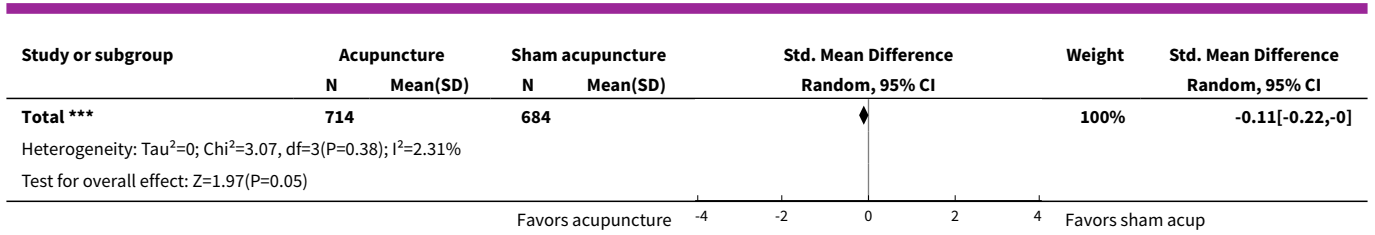


**Analysis 5.5. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 5 Function (26 weeks after baseline).**



**Analysis 5.6. Comparison 5 Acupuncture vs. sham acupuncture for peripheral joint OA change from baseline analysis, Outcome 6 Total score (26 weeks after baseline).**

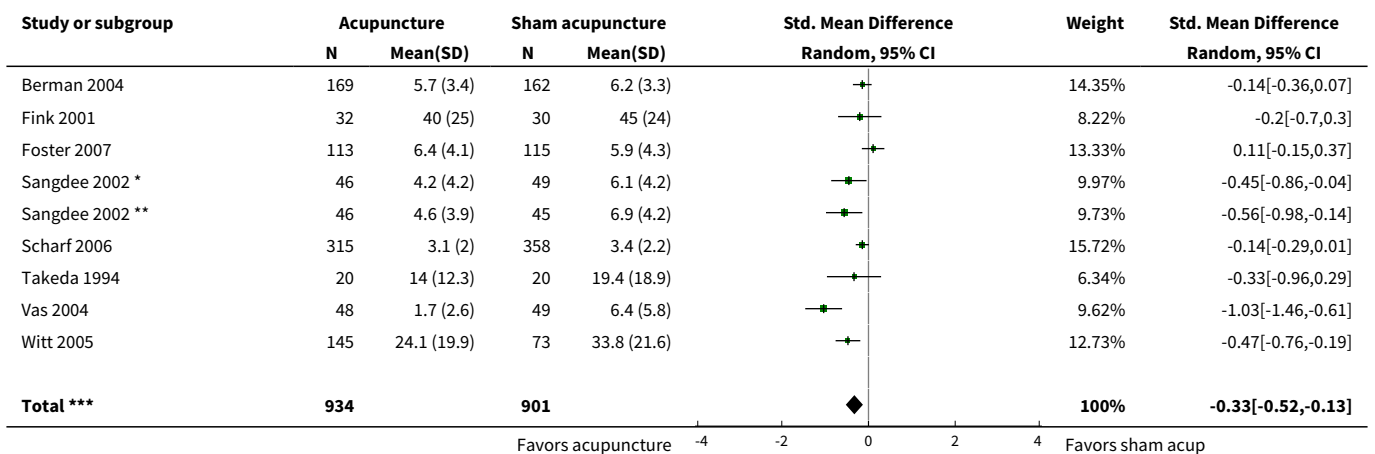




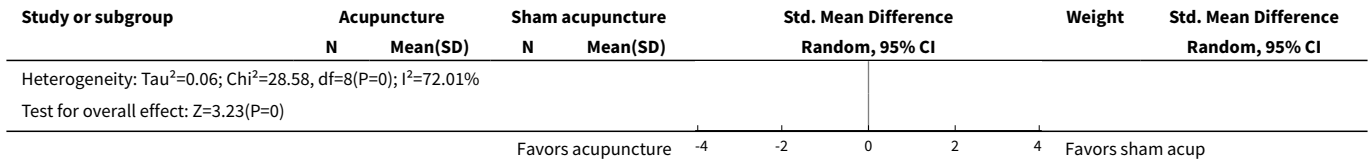
**Comparison 6. Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis**

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)	9	1835	Std. Mean Difference (IV, Random, 95% CI)	-0.33 [-0.52, -0.13]
2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)	9	1830	Std. Mean Difference (IV, Random, 95% CI)	-0.31 [-0.48, -0.13]
3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization)	8	1767	Std. Mean Difference (IV, Random, 95% CI)	-0.34 [-0.55, -0.14]
4 Pain (26 weeks after baseline)	4	1400	Std. Mean Difference (IV, Random, 95% CI)	-0.14 [-0.31, 0.02]
5 Function (26 weeks after baseline)	4	1398	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.29, -0.02]
6 Total score (26 weeks after baseline)	4	1398	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.30, -0.01]

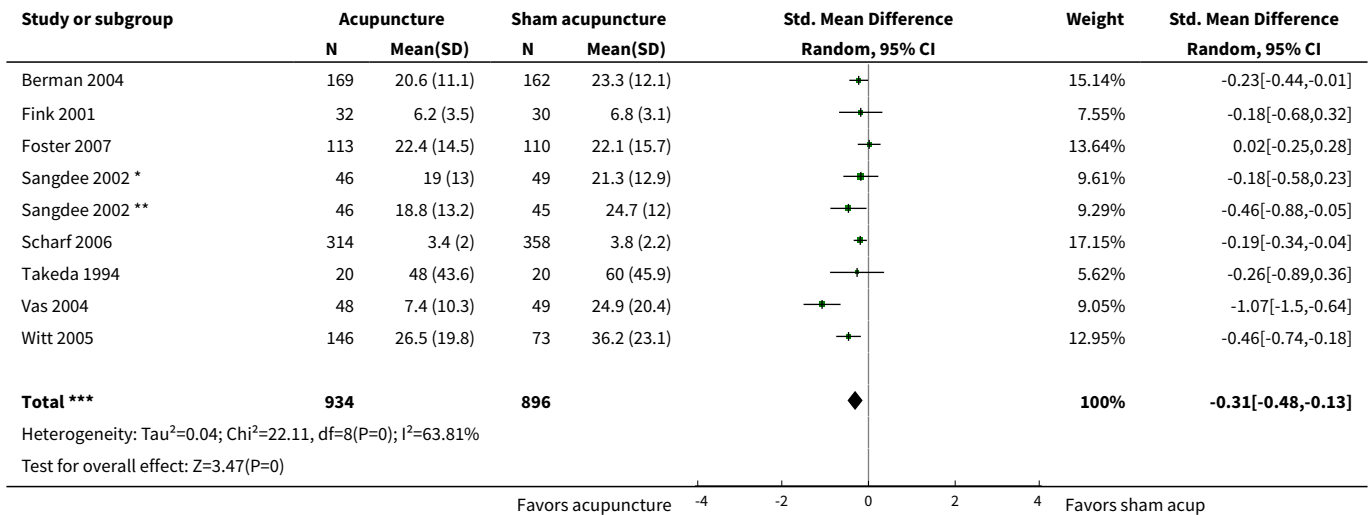
**Analysis 6.1. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**



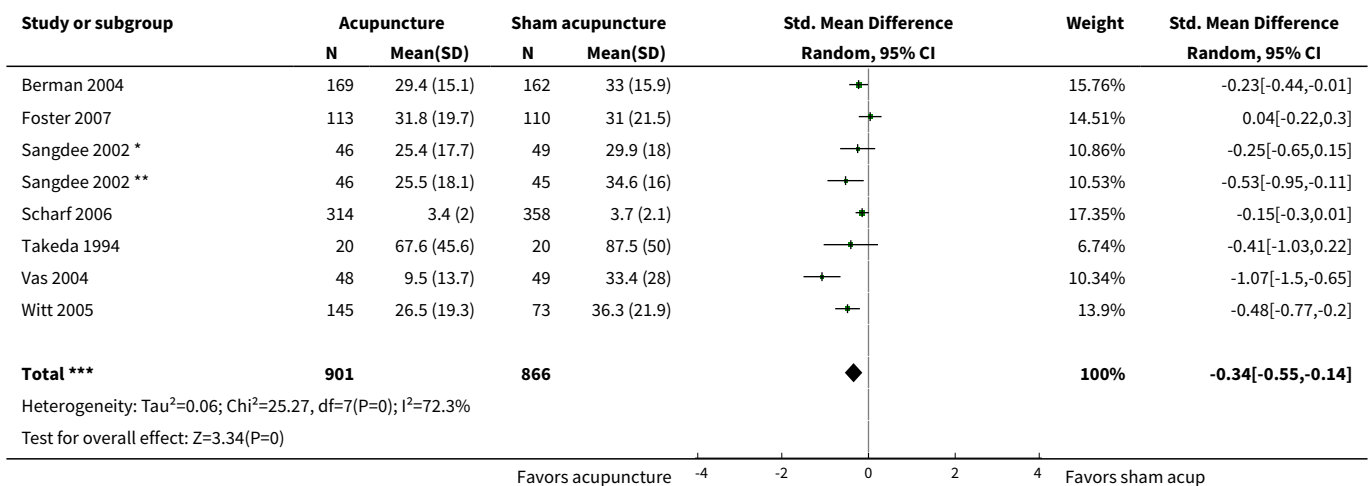




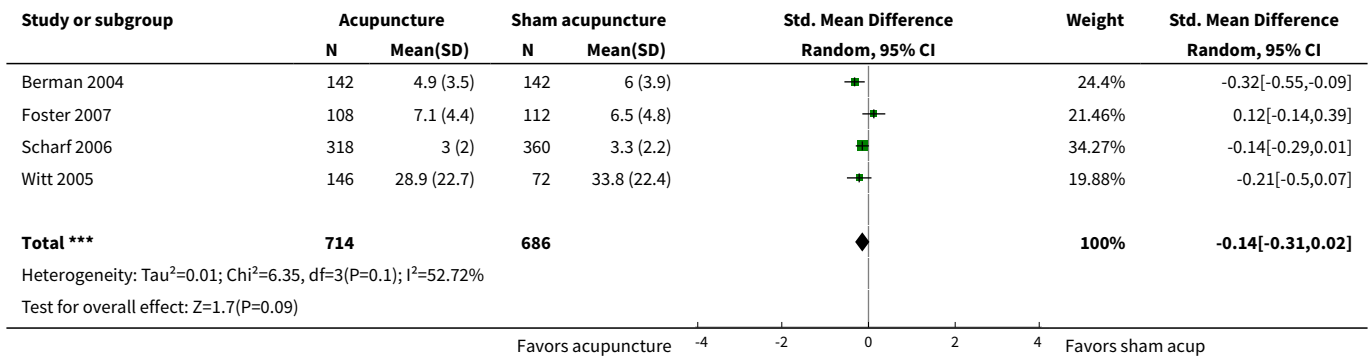
**Analysis 6.2. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**



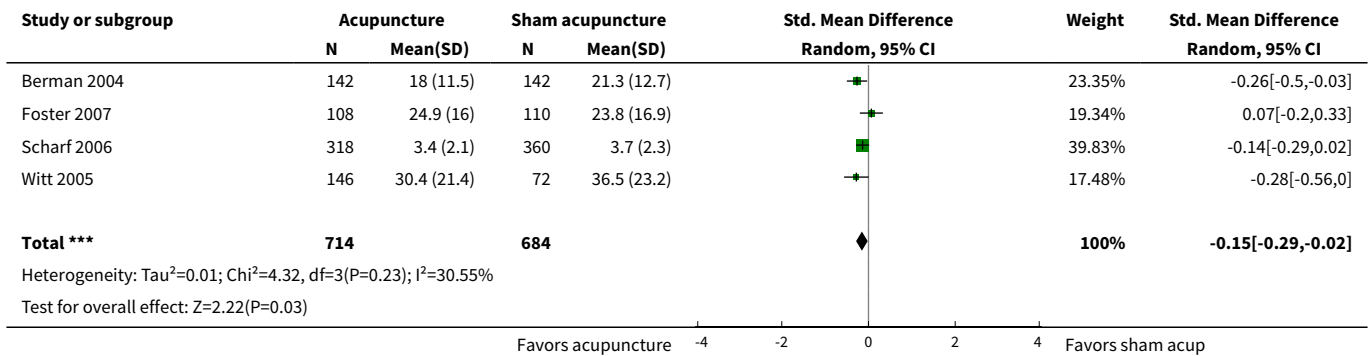
**Analysis 6.3. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization).**



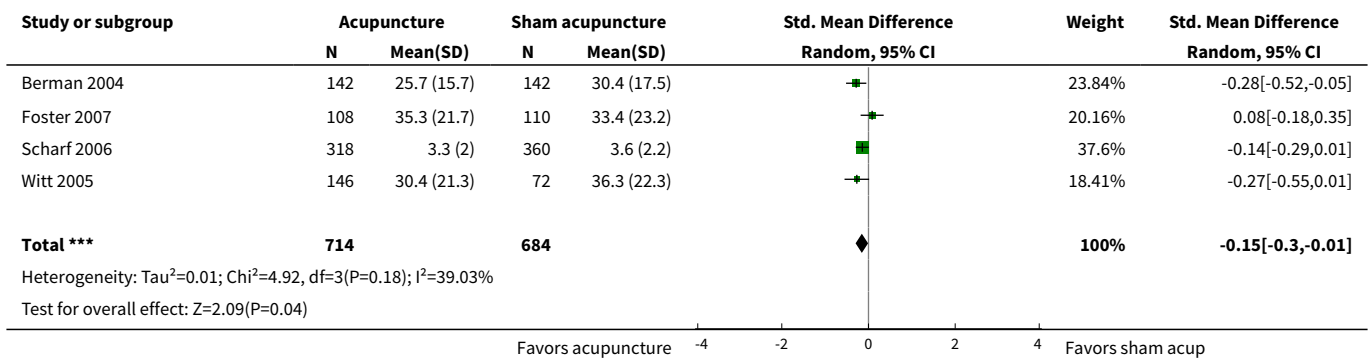
**Analysis 6.4. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 4 Pain (26 weeks after baseline).**



**Analysis 6.5. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 5 Function (26 weeks after baseline).**



**Analysis 6.6. Comparison 6 Acupuncture vs. sham acupuncture for peripheral joint OA post-treatment scores analysis, Outcome 6 Total score (26 weeks after baseline).**



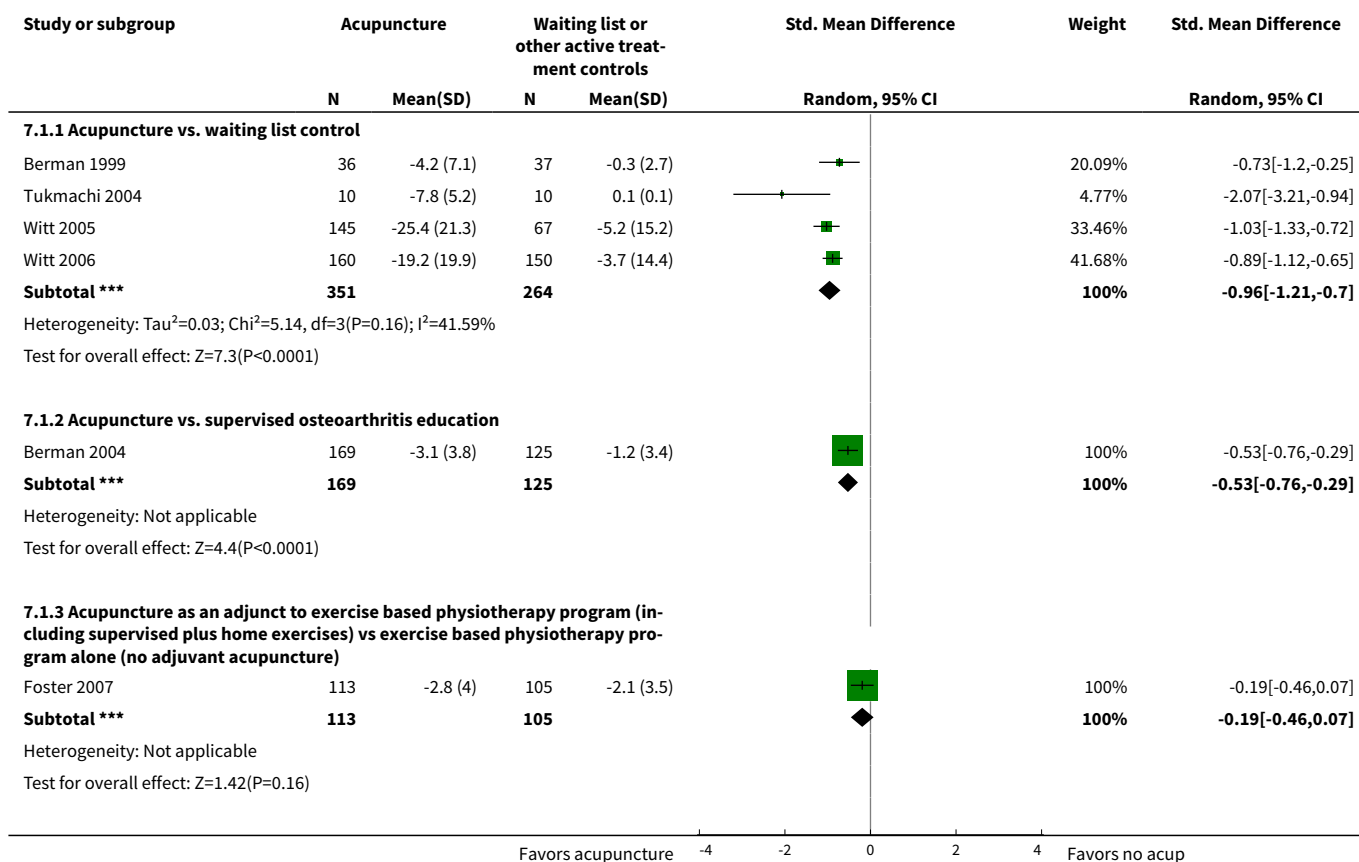
**Comparison 7. Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis**

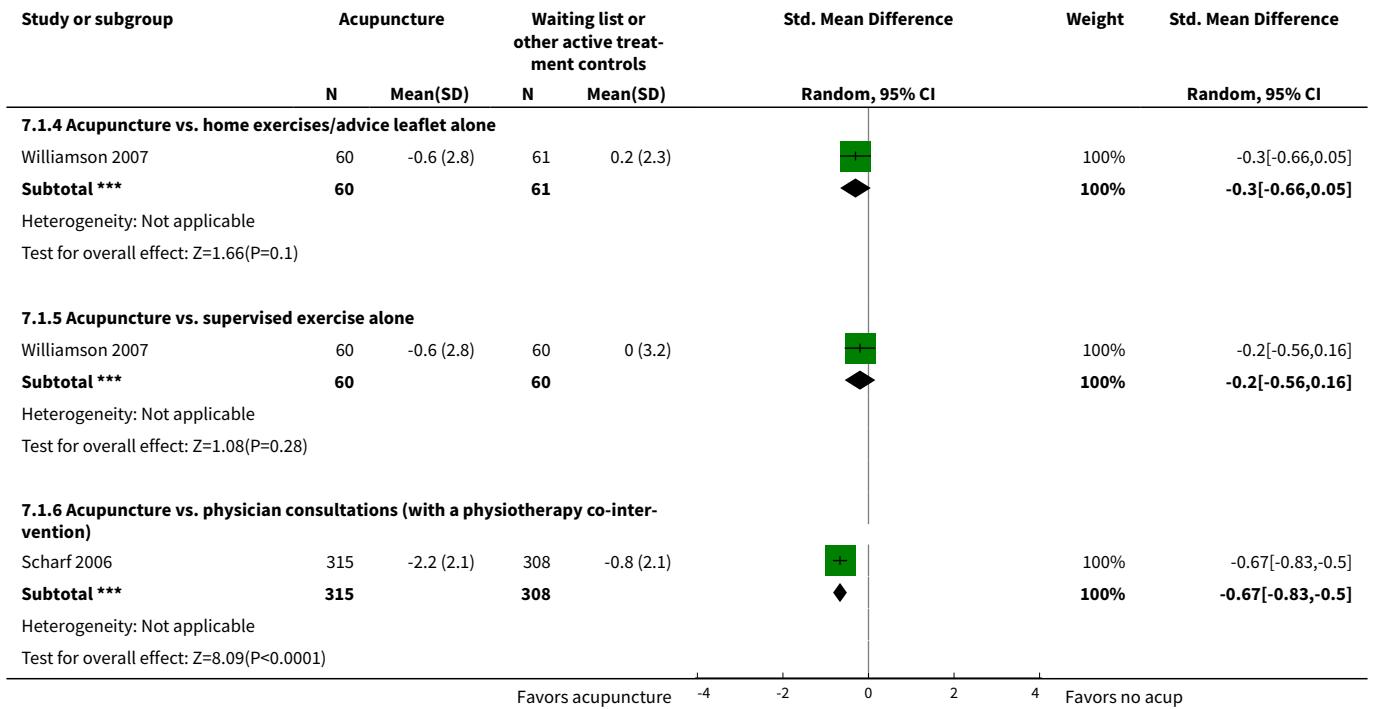
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
<b>1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	8		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Acupuncture vs. waiting list control	4	615	Std. Mean Difference (IV, Random, 95% CI)	-0.96 [-1.21, -0.70]
1.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.53 [-0.76, -0.29]
1.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.19 [-0.46, 0.07]
1.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.30 [-0.66, 0.05]
1.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.20 [-0.56, 0.16]
1.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	623	Std. Mean Difference (IV, Random, 95% CI)	-0.67 [-0.83, -0.50]
<b>2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	7		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 Acupuncture vs. waiting list control	3	587	Std. Mean Difference (IV, Random, 95% CI)	-0.93 [-1.16, -0.69]
2.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.48 [-0.72, -0.25]
2.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.17 [-0.44, 0.09]
2.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.28 [-0.64, 0.07]
2.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.13 [-0.49, 0.23]
2.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	622	Std. Mean Difference (IV, Random, 95% CI)	-0.60 [-0.76, -0.44]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
<b>3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	7		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1 Acupuncture vs. waiting list control	3	581	Std. Mean Difference (IV, Random, 95% CI)	-0.96 [-1.17, -0.74]
3.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.76, -0.29]
3.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.45, 0.08]
3.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.37 [-0.73, -0.01]
3.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.20 [-0.56, 0.16]
3.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	622	Std. Mean Difference (IV, Random, 95% CI)	-0.61 [-0.78, -0.45]
<b>4 Pain (26 weeks after baseline)</b>	3	1087	Std. Mean Difference (IV, Random, 95% CI)	-0.37 [-0.68, -0.06]
4.1 Acupuncture vs. supervised osteoarthritis education	1	250	Std. Mean Difference (IV, Random, 95% CI)	-0.56 [-0.81, -0.30]
4.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	213	Std. Mean Difference (IV, Random, 95% CI)	-0.01 [-0.28, 0.26]
4.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	624	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.67, -0.35]
<b>5 Function (26 weeks after baseline)</b>	3	1083	Std. Mean Difference (IV, Random, 95% CI)	-0.36 [-0.55, -0.18]
5.1 Acupuncture vs. supervised osteoarthritis education	1	250	Std. Mean Difference (IV, Random, 95% CI)	-0.42 [-0.67, -0.17]
5.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	209	Std. Mean Difference (IV, Random, 95% CI)	-0.14 [-0.41, 0.13]

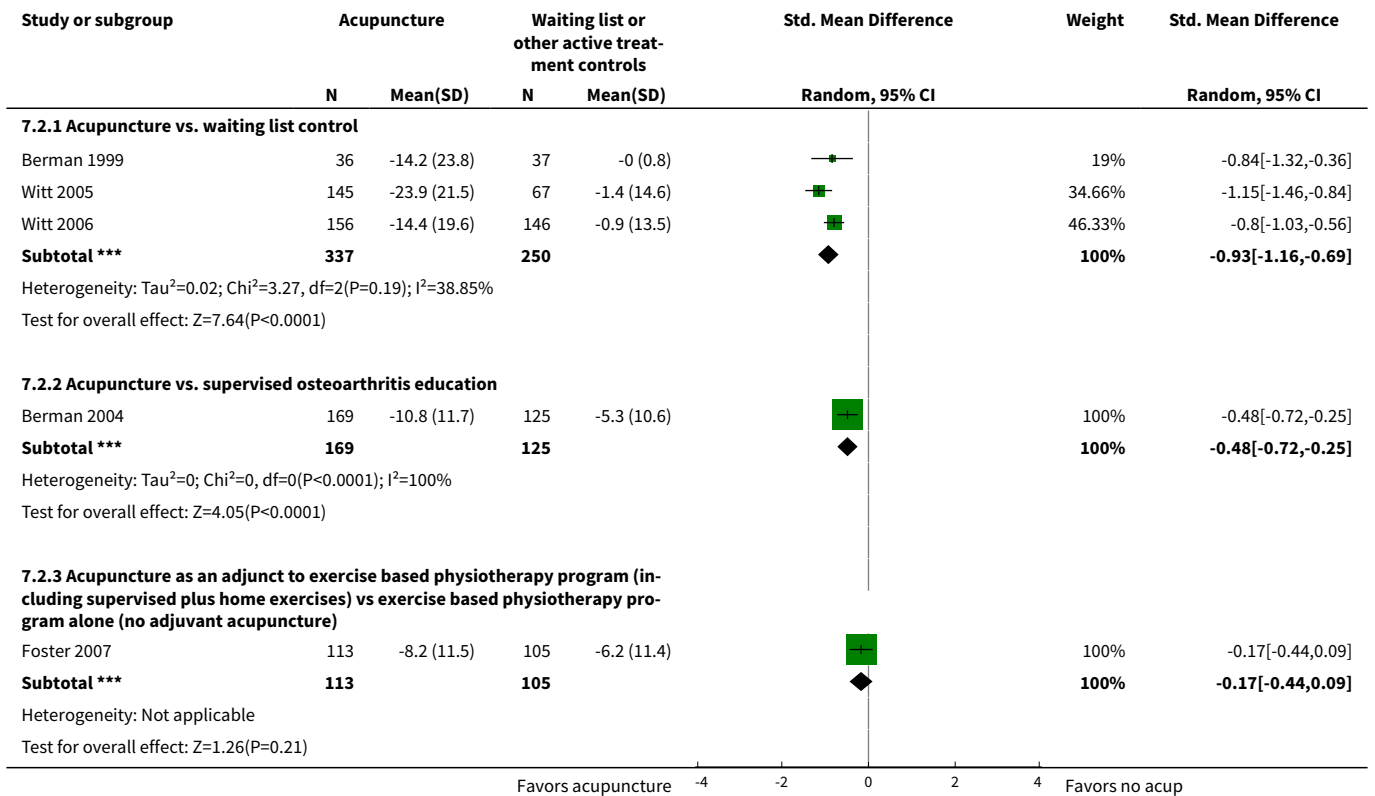
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
5.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	624	Std. Mean Difference (IV, Random, 95% CI)	-0.46 [-0.62, -0.31]
<b>6 Total score (26 weeks after baseline)</b>	<b>3</b>	<b>1083</b>	Std. Mean Difference (IV, Random, 95% CI)	-0.38 [-0.62, -0.15]
6.1 Acupuncture vs. supervised osteoarthritis education	1	250	Std. Mean Difference (IV, Random, 95% CI)	-0.46 [-0.71, -0.20]
6.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	209	Std. Mean Difference (IV, Random, 95% CI)	-0.12 [-0.39, 0.15]
6.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	624	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.68, -0.36]

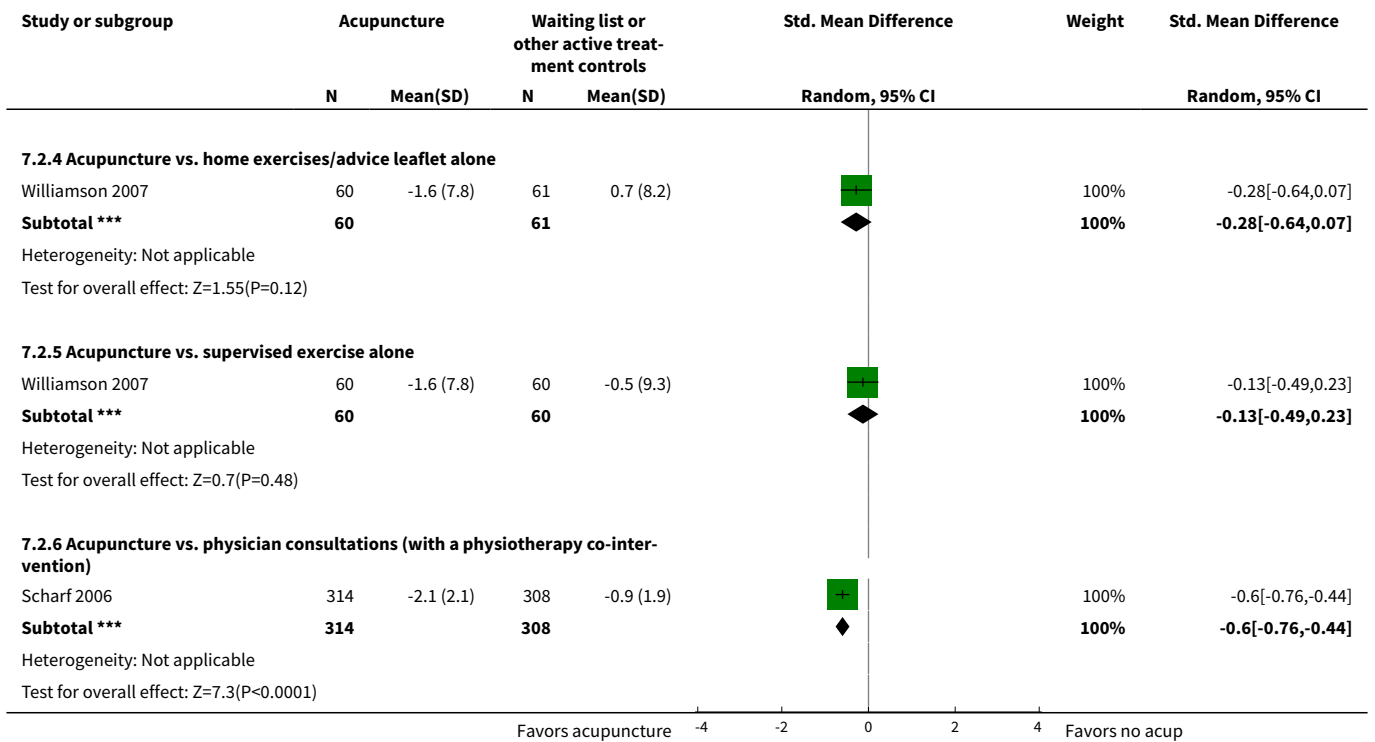
**Analysis 7.1. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**



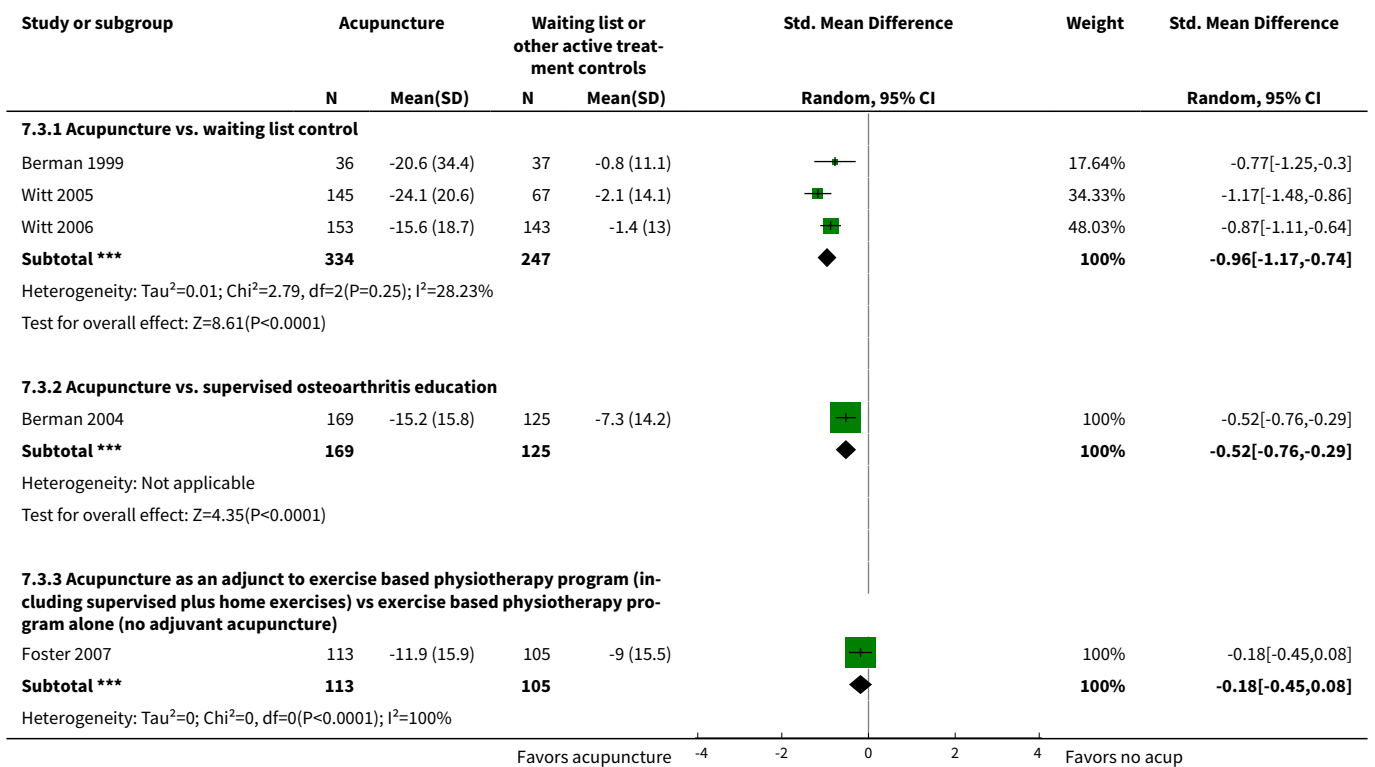


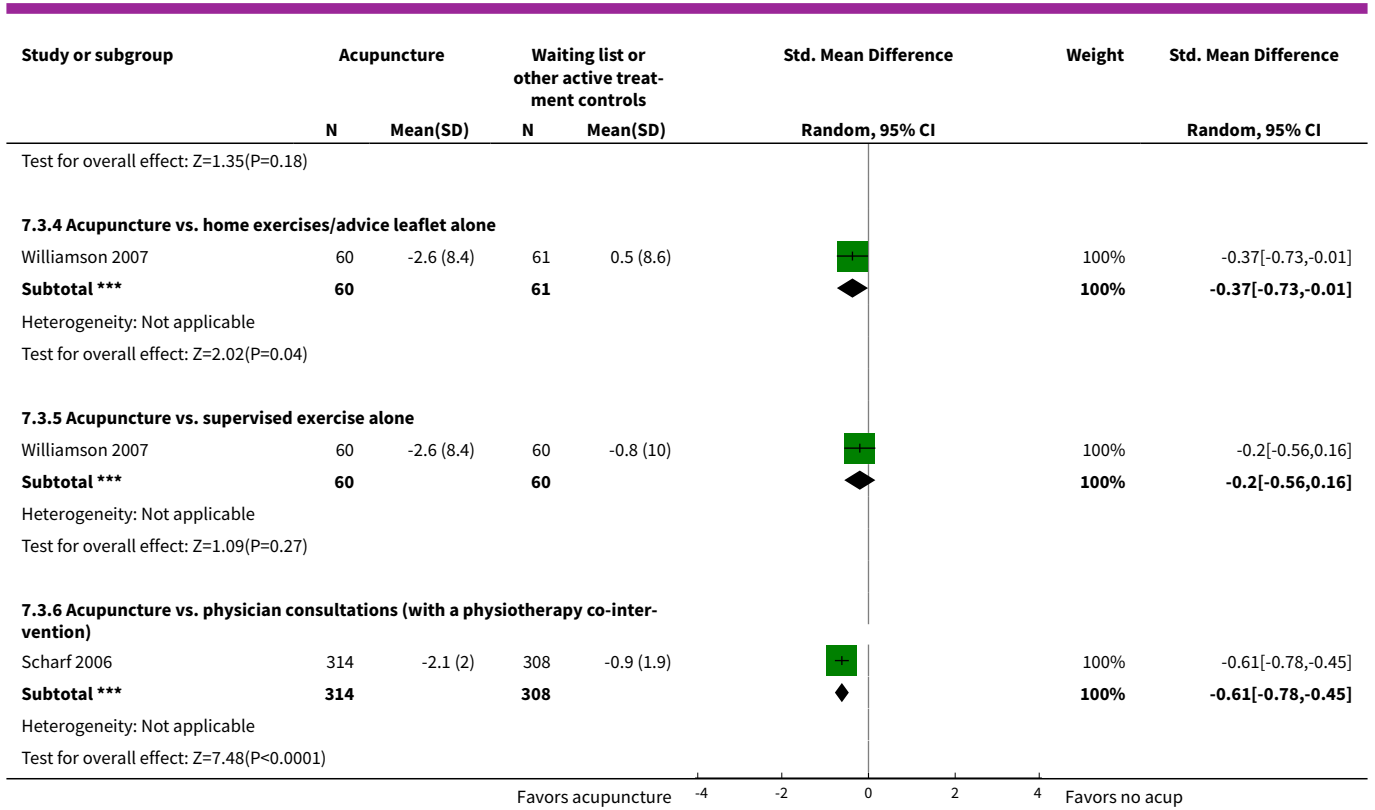
**Analysis 7.2. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**



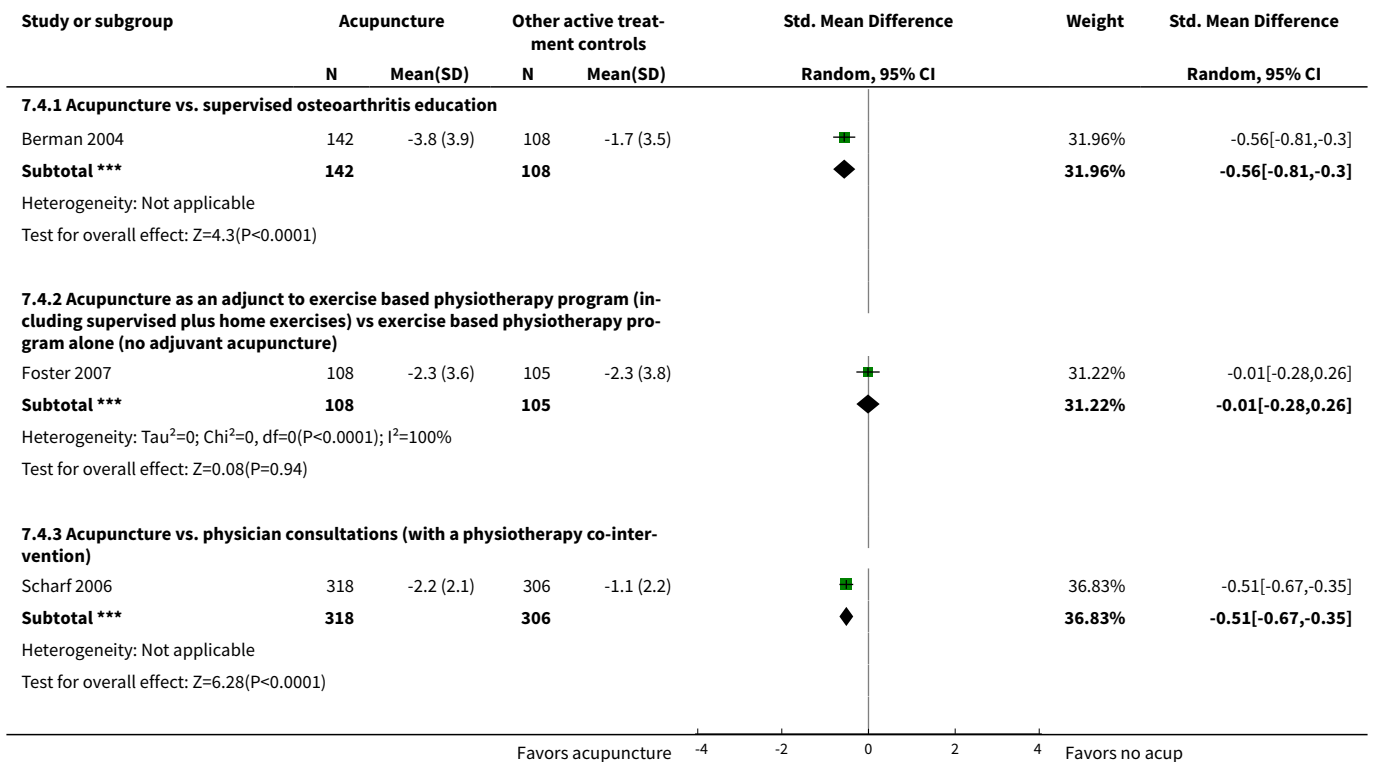


**Analysis 7.3. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization).**

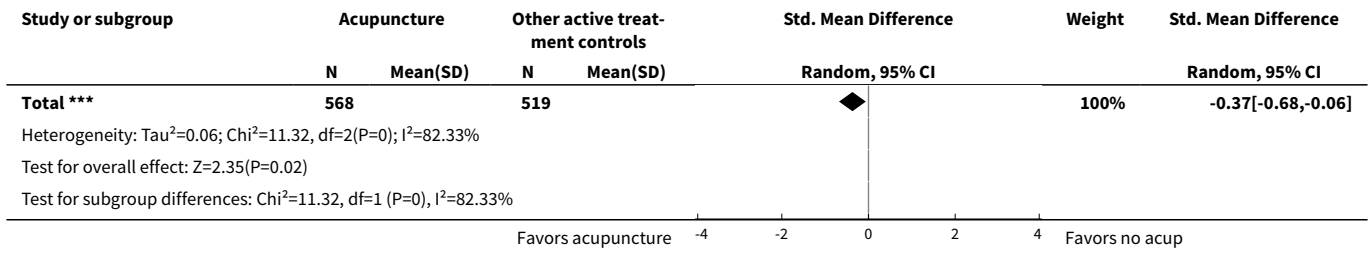




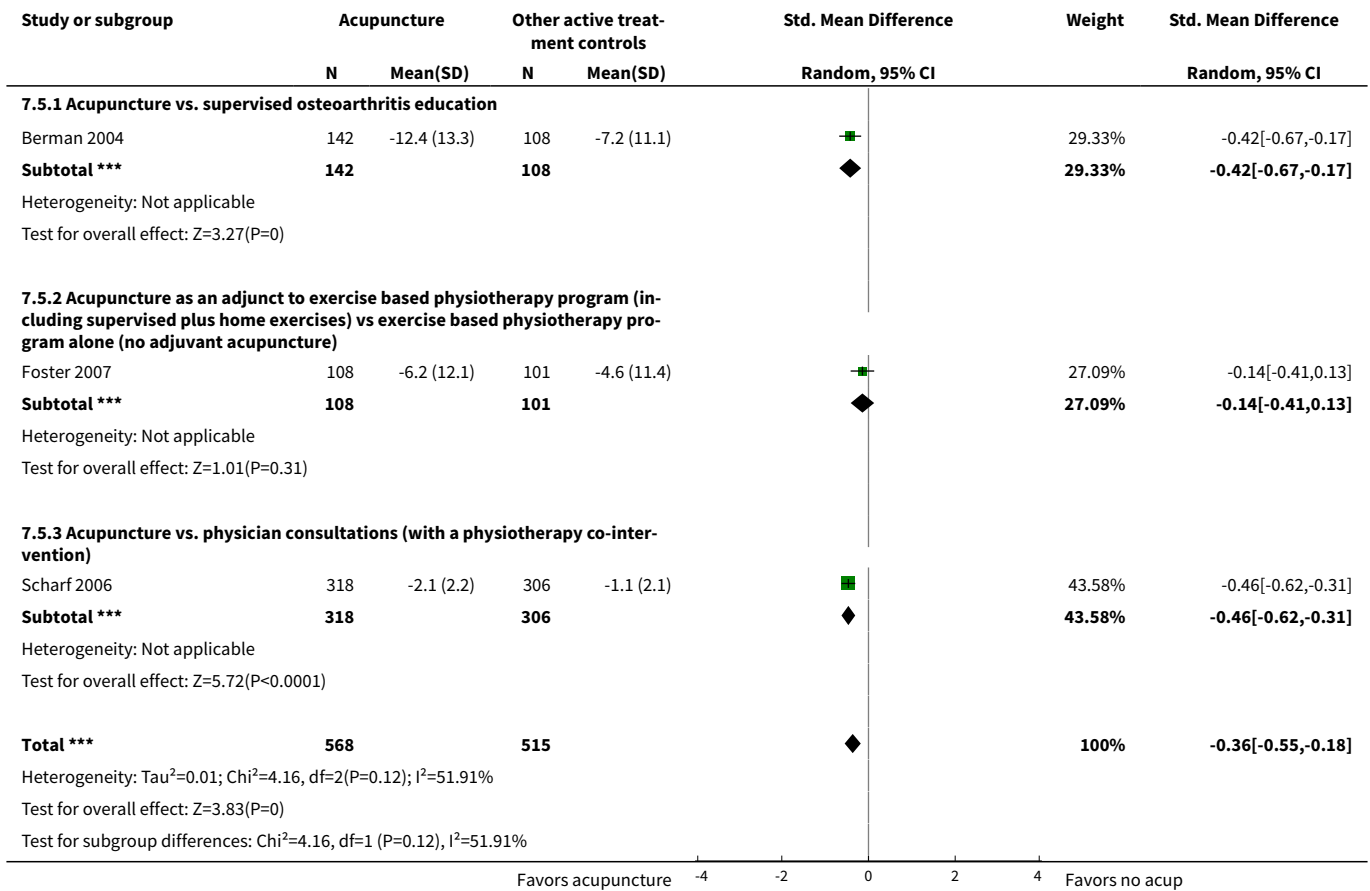
**Analysis 7.4. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 4 Pain (26 weeks after baseline).**



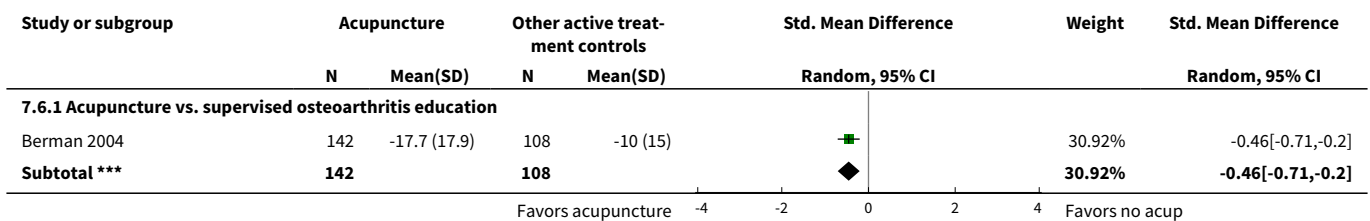


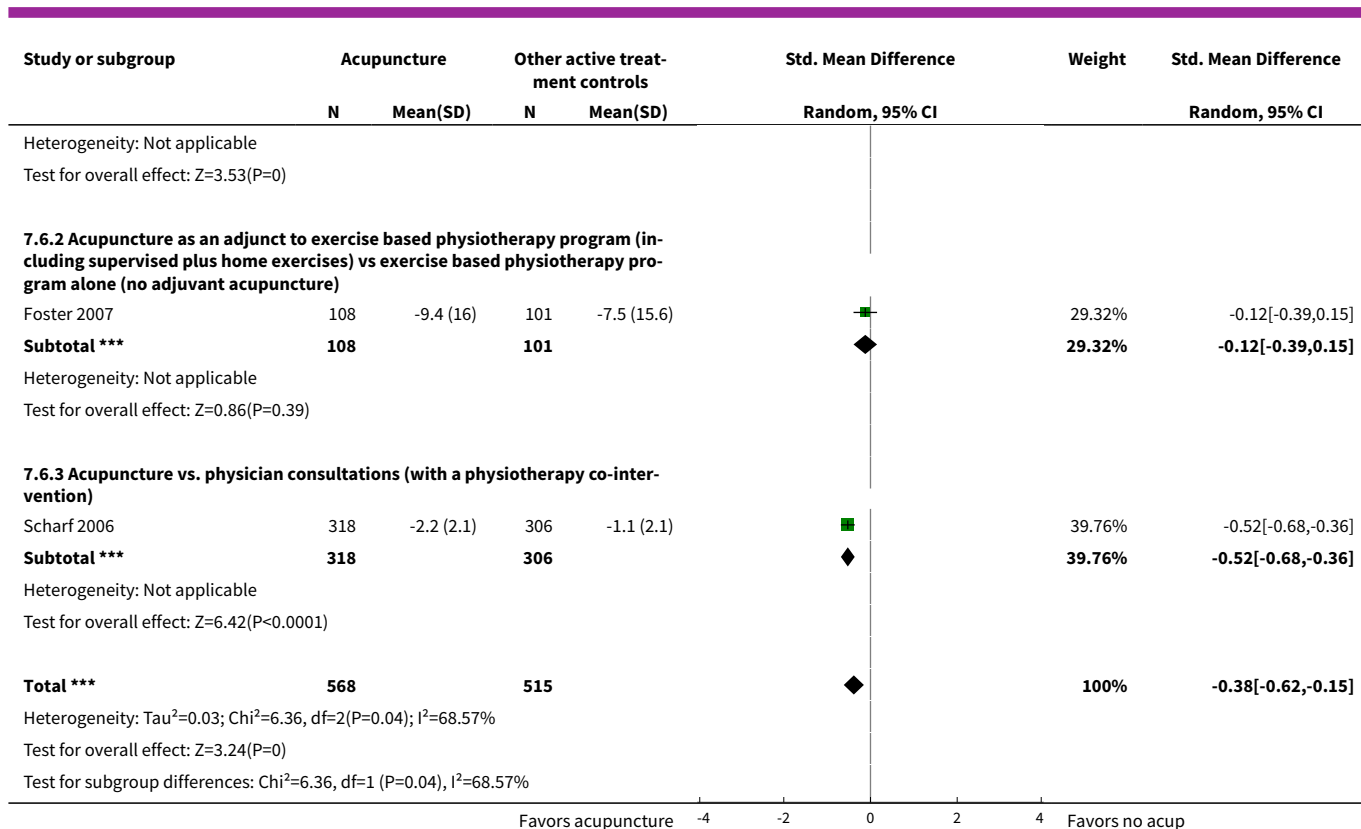


**Analysis 7.5. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 5 Function (26 weeks after baseline).**



**Analysis 7.6. Comparison 7 Acupuncture vs. waiting list or other active treatment controls for knee OA change from baseline analysis, Outcome 6 Total score (26 weeks after baseline).**





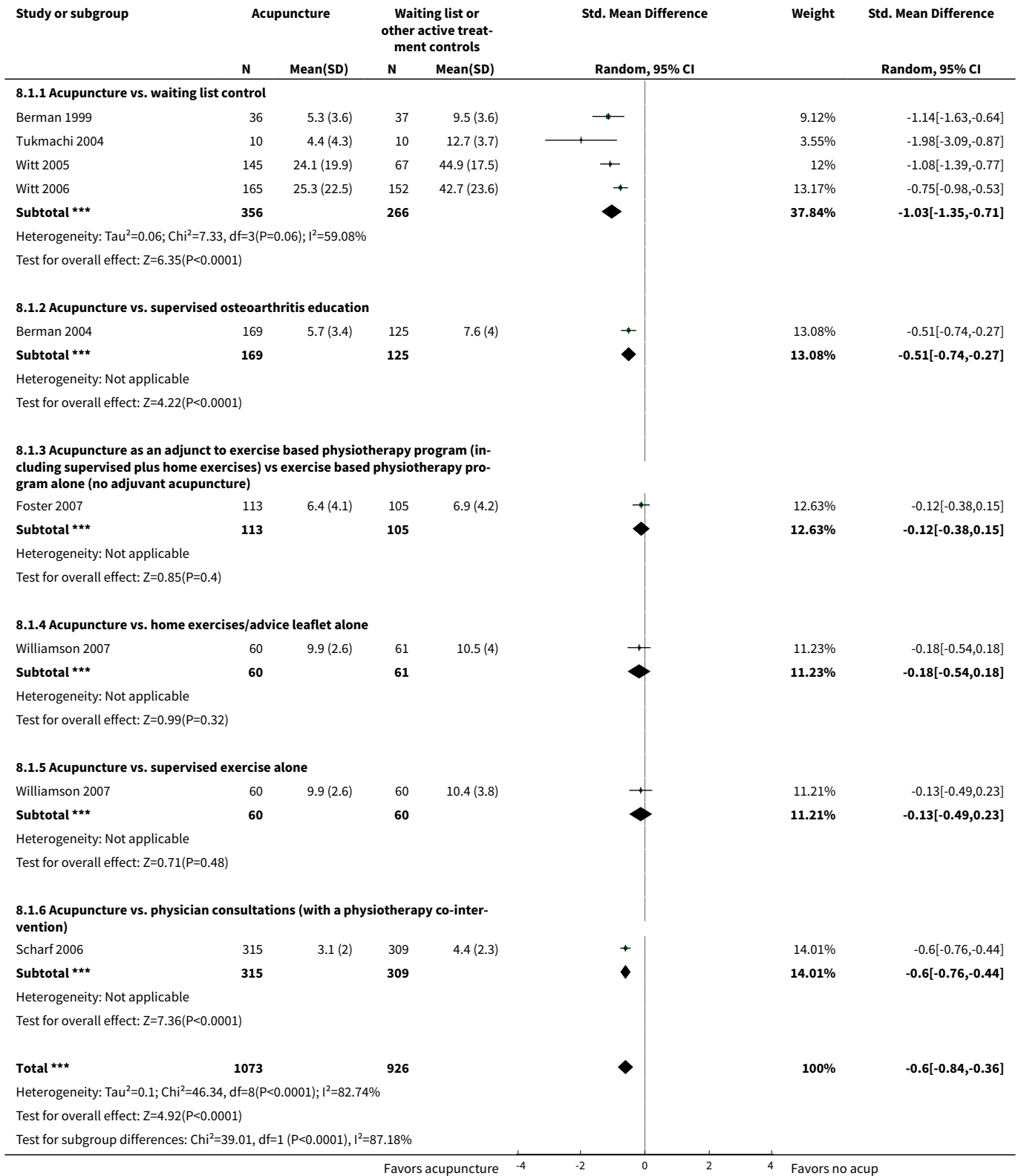
**Comparison 8. Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis**

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
<a href="#">1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)</a>	8	1999	Std. Mean Difference (IV, Random, 95% CI)	-0.60 [-0.84, -0.36]
1.1 Acupuncture vs. waiting list control	4	622	Std. Mean Difference (IV, Random, 95% CI)	-1.03 [-1.35, -0.71]
1.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.74, -0.27]
1.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.12 [-0.38, 0.15]
1.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.54, 0.18]
1.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.13 [-0.49, 0.23]

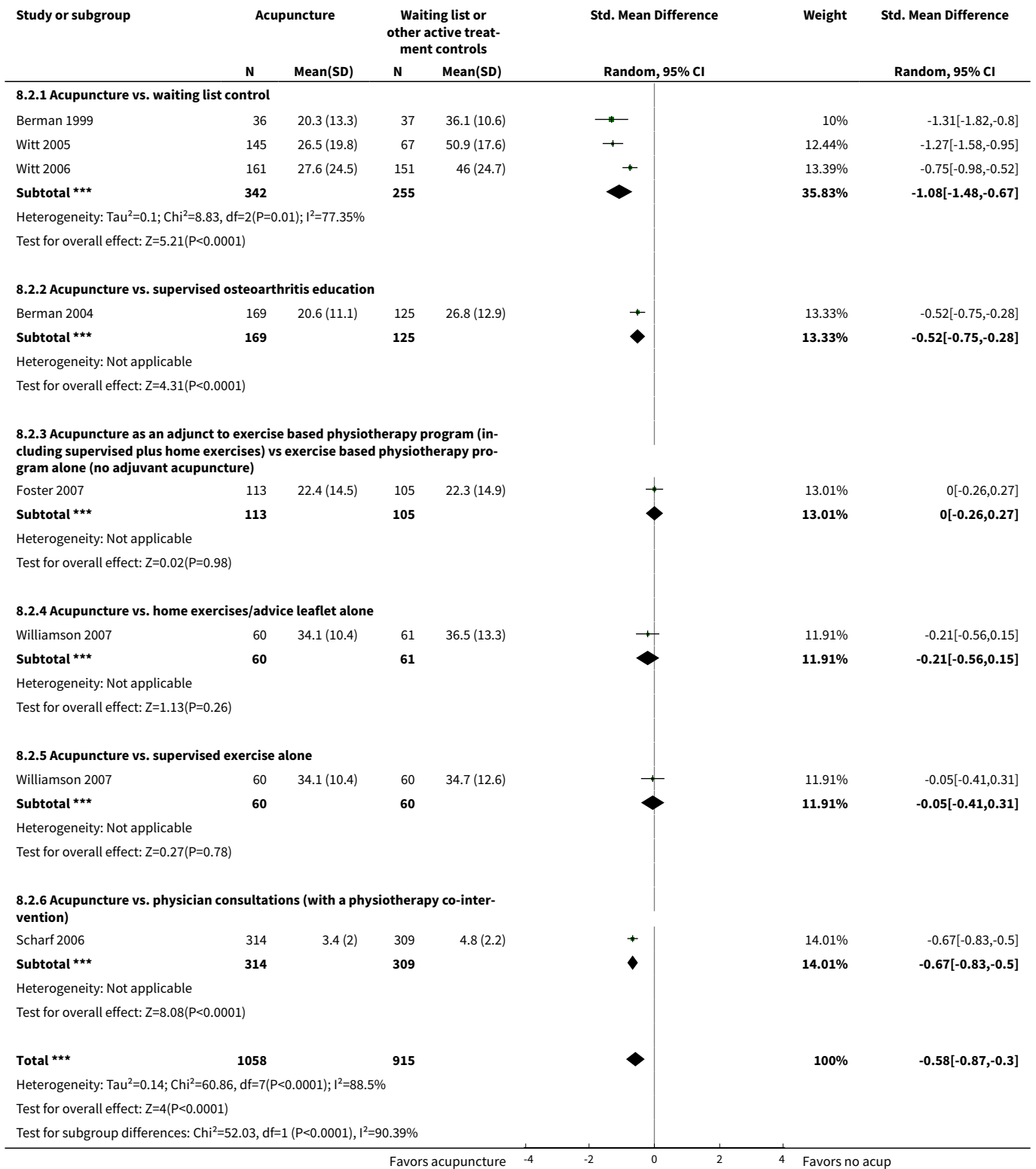
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	624	Std. Mean Difference (IV, Random, 95% CI)	-0.60 [-0.76, -0.44]
<b>2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	7	1973	Std. Mean Difference (IV, Random, 95% CI)	-0.58 [-0.87, -0.30]
2.1 Acupuncture vs. waiting list control	3	597	Std. Mean Difference (IV, Random, 95% CI)	-1.08 [-1.48, -0.67]
2.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.75, -0.28]
2.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	0.00 [-0.26, 0.27]
2.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.21 [-0.56, 0.15]
2.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.05 [-0.41, 0.31]
2.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	623	Std. Mean Difference (IV, Random, 95% CI)	-0.67 [-0.83, -0.50]
<b>3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	7	1969	Std. Mean Difference (IV, Random, 95% CI)	-0.58 [-0.85, -0.30]
3.1 Acupuncture vs. waiting list control	3	593	Std. Mean Difference (IV, Random, 95% CI)	-1.08 [-1.49, -0.68]
3.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.53 [-0.77, -0.30]
3.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.03 [-0.30, 0.24]
3.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.21 [-0.57, 0.15]
3.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.07 [-0.43, 0.28]
3.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	623	Std. Mean Difference (IV, Random, 95% CI)	-0.57 [-0.73, -0.41]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
<b>4 Pain (26 weeks after baseline)</b>	3	1088	Std. Mean Difference (IV, Random, 95% CI)	-0.34 [-0.70, 0.01]
4.1 Acupuncture vs. supervised osteoarthritis education	1	250	Std. Mean Difference (IV, Random, 95% CI)	-0.54 [-0.80, -0.29]
4.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	213	Std. Mean Difference (IV, Random, 95% CI)	0.06 [-0.20, 0.33]
4.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	625	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.68, -0.36]
<b>5 Function (26 weeks after baseline)</b>	3	1093	Std. Mean Difference (IV, Random, 95% CI)	-0.34 [-0.64, -0.04]
5.1 Acupuncture vs. supervised osteoarthritis education	1	250	Std. Mean Difference (IV, Random, 95% CI)	-0.50 [-0.75, -0.24]
5.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	0.00 [-0.26, 0.27]
5.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	625	Std. Mean Difference (IV, Random, 95% CI)	-0.50 [-0.66, -0.34]
<b>6 Total score (26 weeks after baseline)</b>	3	1084	Std. Mean Difference (IV, Random, 95% CI)	-0.34 [-0.67, -0.01]
6.1 Acupuncture vs. supervised osteoarthritis education	1	250	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.77, -0.26]
6.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	209	Std. Mean Difference (IV, Random, 95% CI)	0.04 [-0.23, 0.31]
6.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	625	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.67, -0.35]

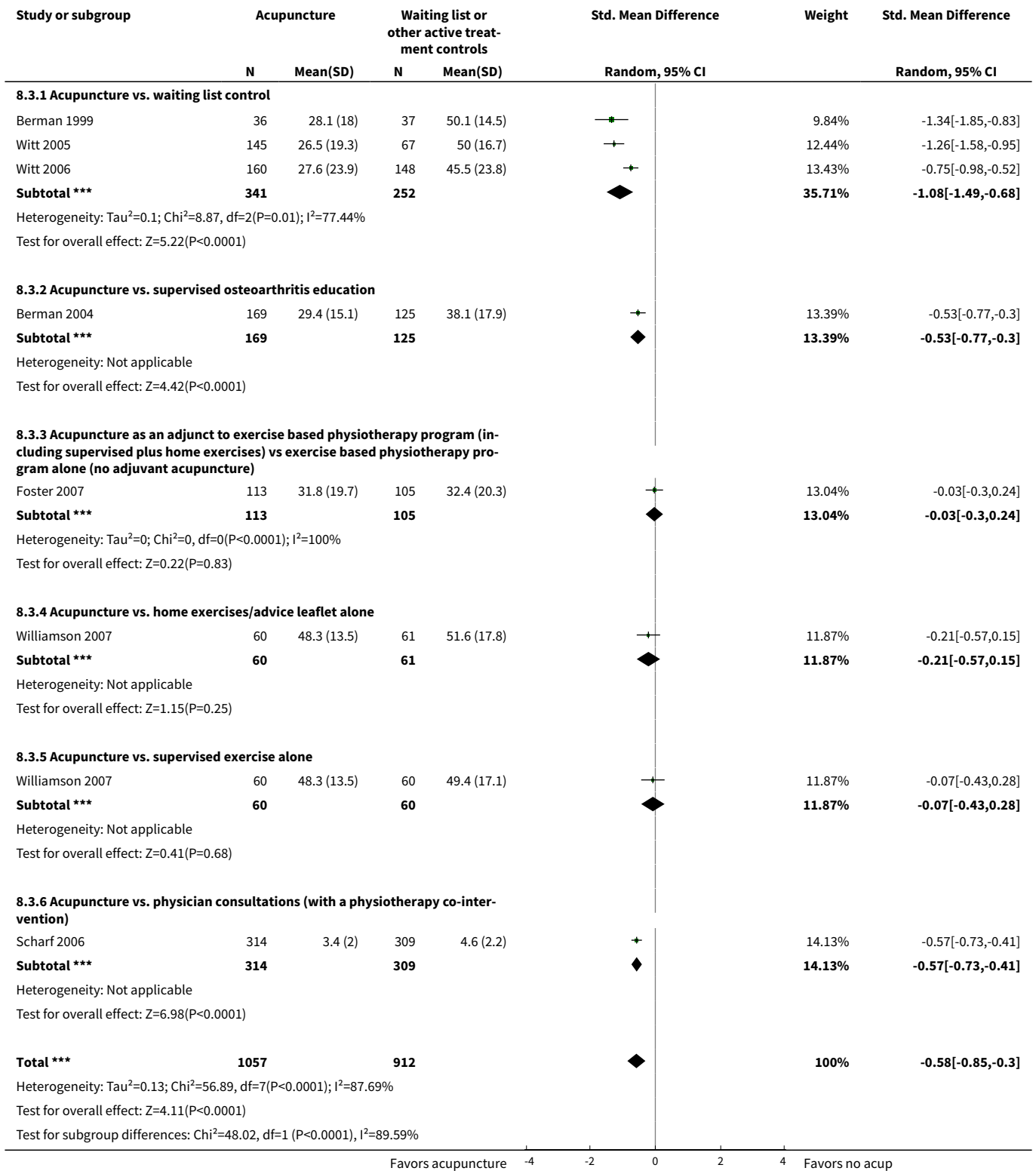
**Analysis 8.1. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**



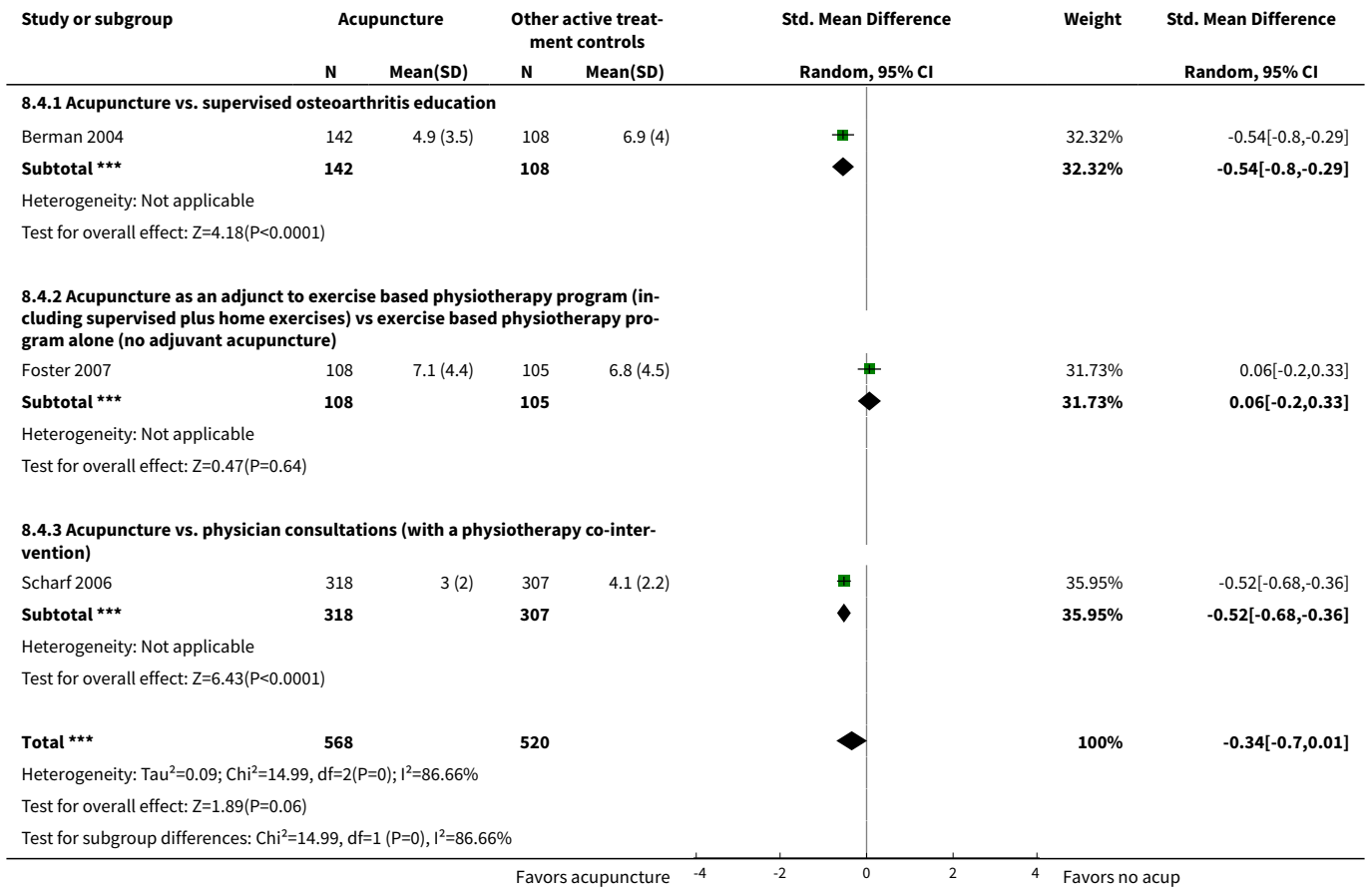
**Analysis 8.2. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**



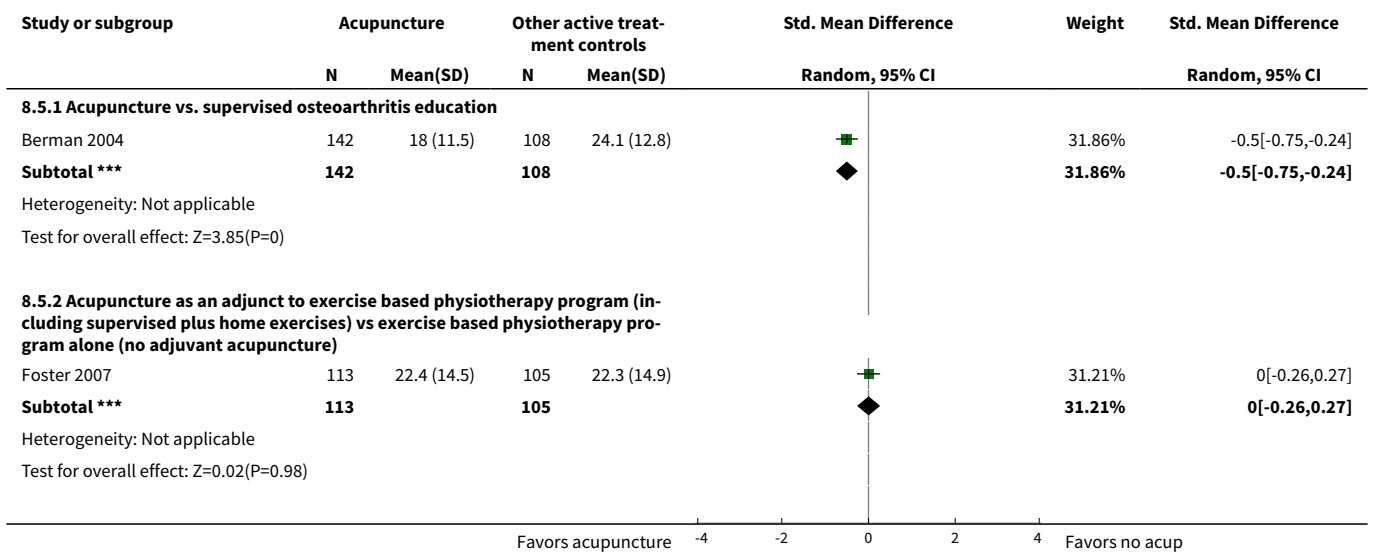
**Analysis 8.3. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization).**



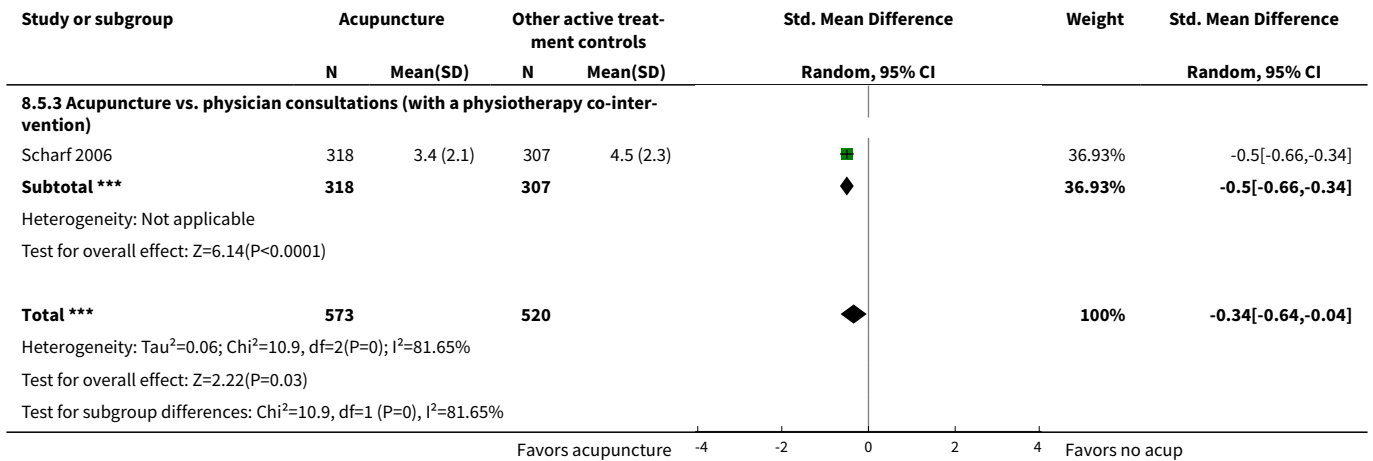
**Analysis 8.4. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 4 Pain (26 weeks after baseline).**



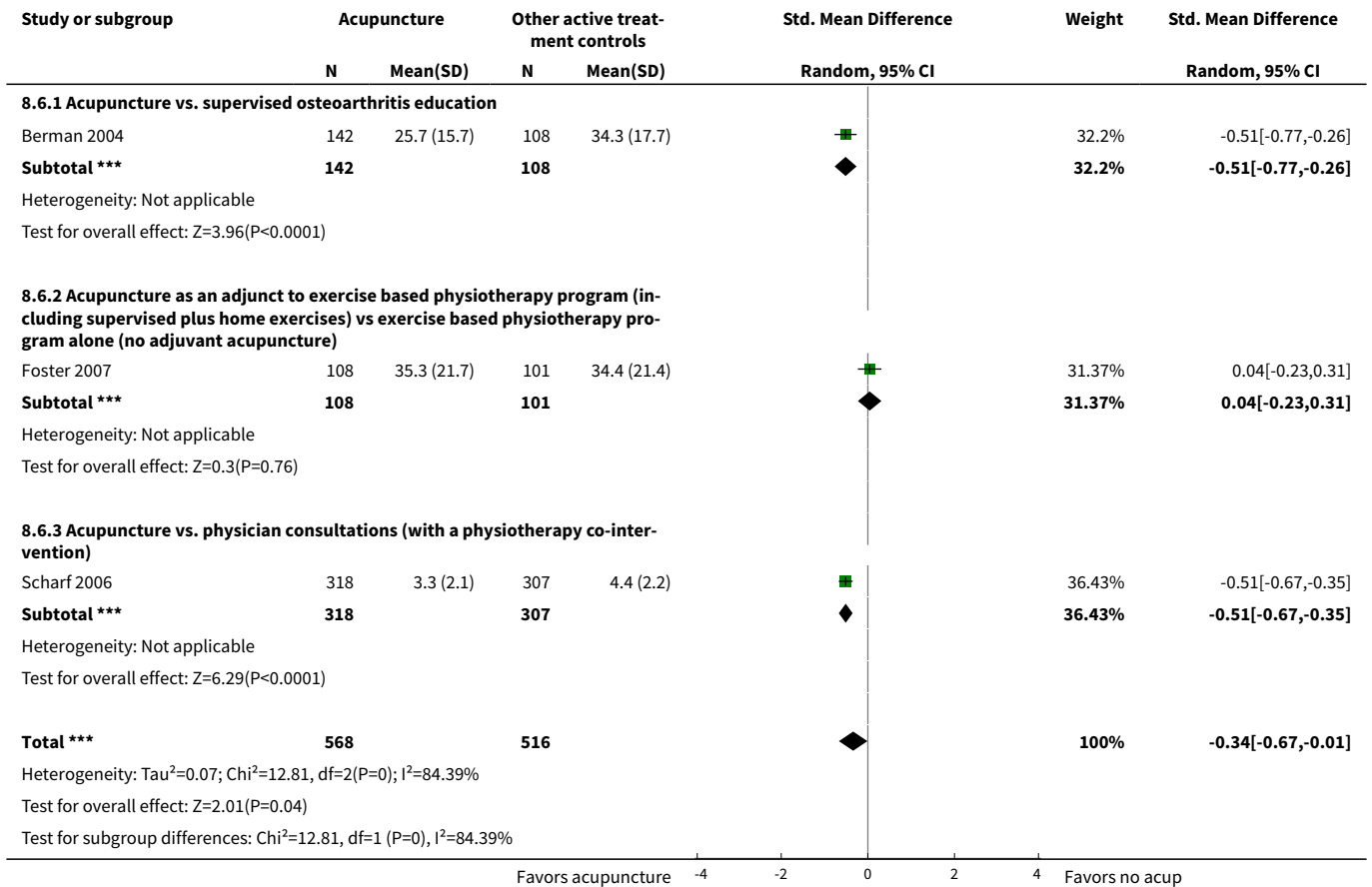
**Analysis 8.5. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 5 Function (26 weeks after baseline).**







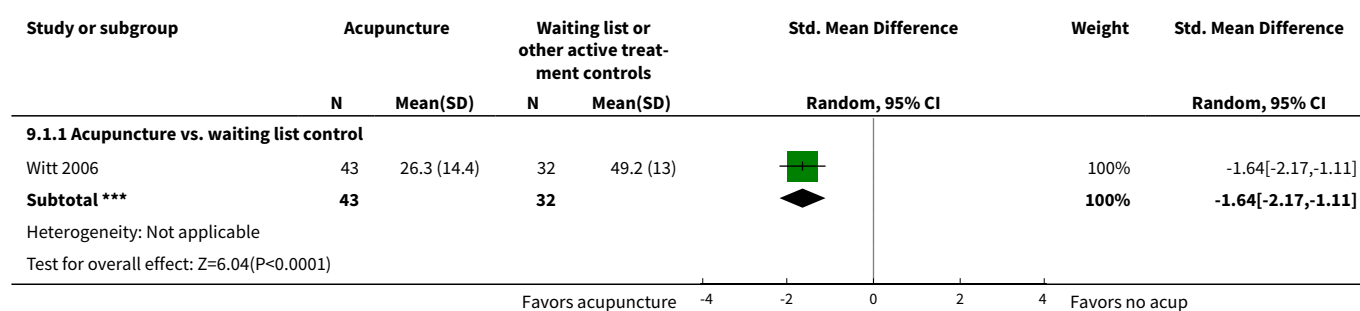
**Analysis 8.6. Comparison 8 Acupuncture vs. waiting list or other active treatment controls for knee OA post-treatment scores analysis, Outcome 6 Total score (26 weeks after baseline).**



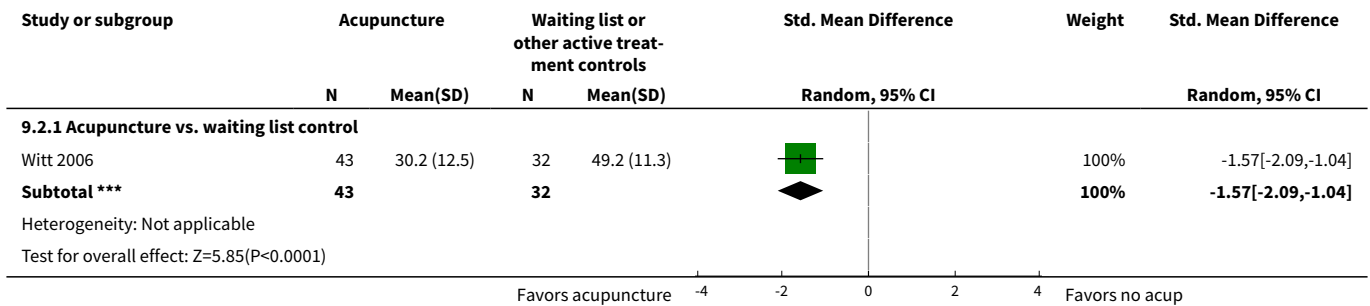
**Comparison 9. Acupuncture vs. waiting list or other active treatment controls for hip OA post-treatment scores analysis**

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)	1		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Acupuncture vs. waiting list control	1	75	Std. Mean Difference (IV, Random, 95% CI)	-1.64 [-2.17, -1.11]
2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)	1		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 Acupuncture vs. waiting list control	1	75	Std. Mean Difference (IV, Random, 95% CI)	-1.57 [-2.09, -1.04]
3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization)	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1 Acupuncture vs. waiting list control	1	75	Std. Mean Difference (IV, Random, 95% CI)	-1.62 [-2.15, -1.09]
3.2 Acupuncture vs. supervised exercise alone	1	28	Std. Mean Difference (IV, Random, 95% CI)	-0.54 [-1.30, 0.22]

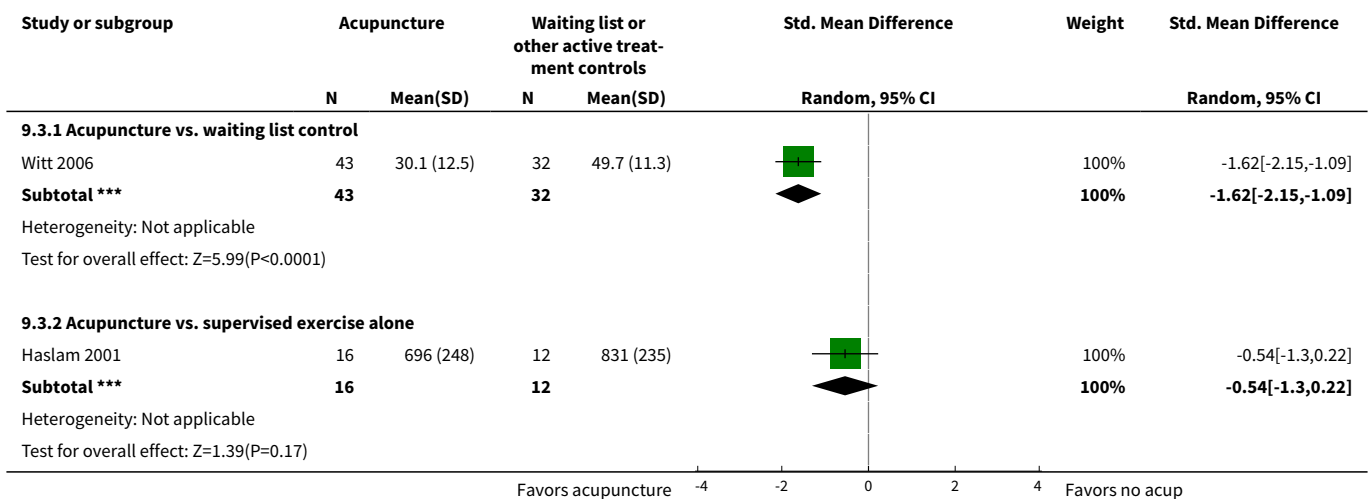
**Analysis 9.1. Comparison 9 Acupuncture vs. waiting list or other active treatment controls for hip OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**



**Analysis 9.2. Comparison 9 Acupuncture vs. waiting list or other active treatment controls for hip OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**



**Analysis 9.3. Comparison 9 Acupuncture vs. waiting list or other active treatment controls for hip OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization).**



**Comparison 10. Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis**

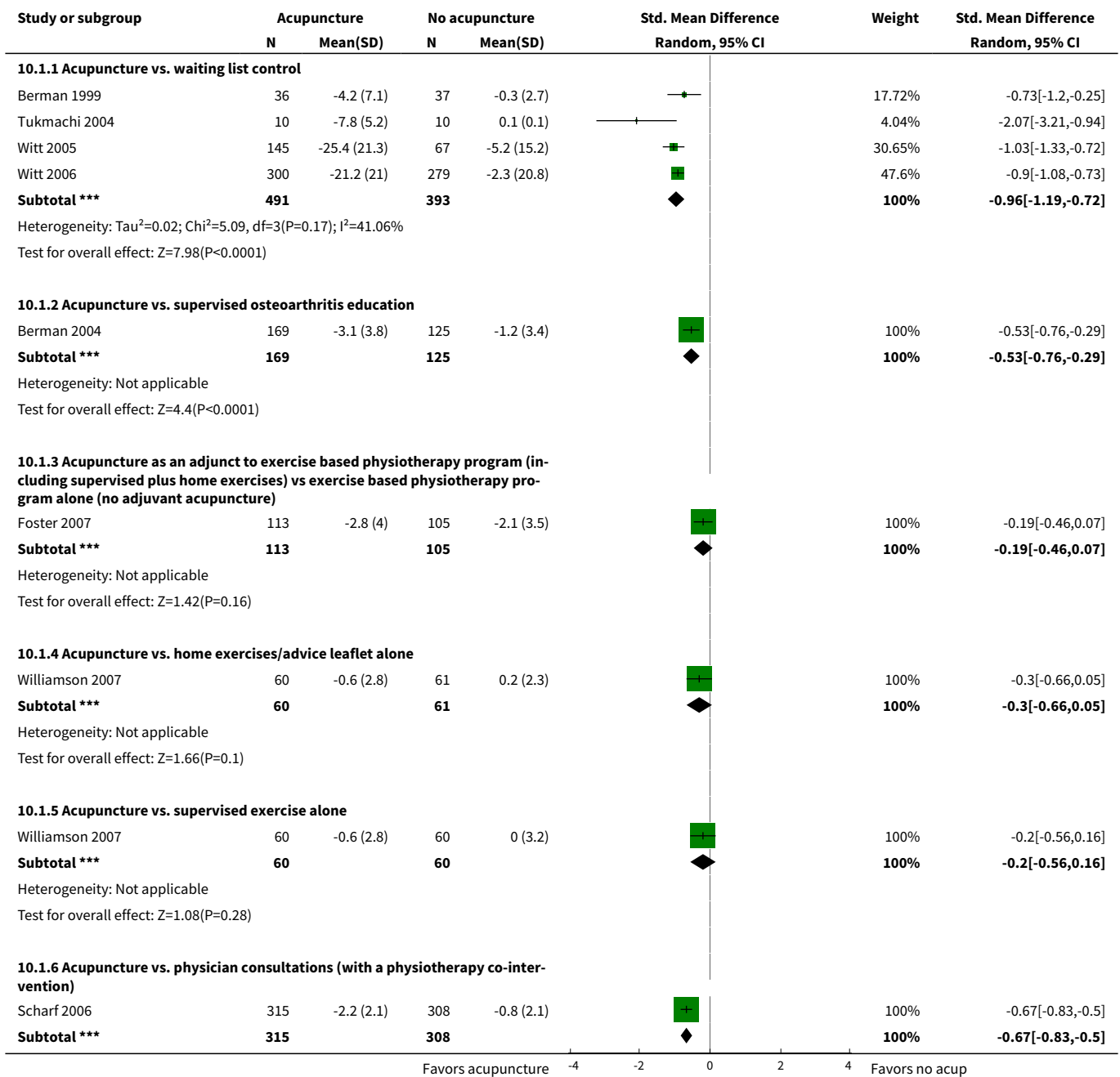
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)	8		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Acupuncture vs. waiting list control	4	884	Std. Mean Difference (IV, Random, 95% CI)	-0.96 [-1.19, -0.72]
1.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.53 [-0.76, -0.29]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.19 [-0.46, 0.07]
1.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.30 [-0.66, 0.05]
1.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.20 [-0.56, 0.16]
1.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	623	Std. Mean Difference (IV, Random, 95% CI)	-0.67 [-0.83, -0.50]
<b>2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	7		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 Acupuncture vs. waiting list control	3	864	Std. Mean Difference (IV, Random, 95% CI)	-0.89 [-1.18, -0.60]
2.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.48 [-0.72, -0.25]
2.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.17 [-0.44, 0.09]
2.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.28 [-0.64, 0.07]
2.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.13 [-0.49, 0.23]
2.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	622	Std. Mean Difference (IV, Random, 95% CI)	-0.60 [-0.76, -0.44]
<b>3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	7		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1 Acupuncture vs. waiting list control	3	864	Std. Mean Difference (IV, Random, 95% CI)	-0.92 [-1.16, -0.67]
3.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.76, -0.29]
3.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exer-	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.45, 0.08]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
cise based physiotherapy program alone (no adjuvant acupuncture)				
3.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.37 [-0.73, -0.01]
3.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.20 [-0.56, 0.16]
3.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	622	Std. Mean Difference (IV, Random, 95% CI)	-0.61 [-0.78, -0.45]
<b>4 Pain (26 weeks after baseline)</b>	<b>3</b>	<b>1087</b>	<b>Std. Mean Difference (IV, Random, 95% CI)</b>	<b>-0.37 [-0.68, -0.06]</b>
4.1 Acupuncture vs. supervised osteoarthritis education	1	250	Std. Mean Difference (IV, Random, 95% CI)	-0.56 [-0.81, -0.30]
4.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	213	Std. Mean Difference (IV, Random, 95% CI)	-0.01 [-0.28, 0.26]
4.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	624	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.67, -0.35]
<b>5 Function (26 weeks after baseline)</b>	<b>3</b>	<b>1083</b>	<b>Std. Mean Difference (IV, Random, 95% CI)</b>	<b>-0.36 [-0.55, -0.18]</b>
5.1 Acupuncture vs. supervised osteoarthritis education	1	250	Std. Mean Difference (IV, Random, 95% CI)	-0.42 [-0.67, -0.17]
5.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	209	Std. Mean Difference (IV, Random, 95% CI)	-0.14 [-0.41, 0.13]
5.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	624	Std. Mean Difference (IV, Random, 95% CI)	-0.46 [-0.62, -0.31]
<b>6 Total score (26 weeks after baseline)</b>	<b>3</b>	<b>1083</b>	<b>Std. Mean Difference (IV, Random, 95% CI)</b>	<b>-0.38 [-0.62, -0.15]</b>
6.1 Acupuncture vs. supervised osteoarthritis education	1	250	Std. Mean Difference (IV, Random, 95% CI)	-0.46 [-0.71, -0.20]
6.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)	1	209	Std. Mean Difference (IV, Random, 95% CI)	-0.12 [-0.39, 0.15]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
6.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	624	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.68, -0.36]

**Analysis 10.1. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**



Study or subgroup	Acupuncture		No acupuncture		Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
	N	Mean(SD)	N	Mean(SD)			
Heterogeneity: Not applicable Test for overall effect: Z=8.09(P<0.0001)							
Favors acupuncture    -4    -2    0    2    4    Favors no acup							

**Analysis 10.2. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**

Study or subgroup	Acupuncture		Waiting list or other active treatment controls		Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
	N	Mean(SD)	N	Mean(SD)			
<b>10.2.1 Acupuncture vs. waiting list control</b>							
Berman 1999	36	-14.2 (23.8)	37	-0 (0.8)		21.55%	-0.84[-1.32,-0.36]
Witt 2005	145	-23.9 (21.5)	67	-1.4 (14.6)		33.1%	-1.15[-1.46,-0.84]
Witt 2006	300	-16.5 (21.9)	279	-0.6 (21.8)		45.35%	-0.73[-0.9,-0.56]
<b>Subtotal ***</b>	<b>481</b>		<b>383</b>			<b>100%</b>	<b>-0.89[-1.18,-0.6]</b>
Heterogeneity: Tau <sup>2</sup> =0.04; Chi <sup>2</sup> =5.49, df=2(P=0.06); I <sup>2</sup> =63.6% Test for overall effect: Z=6.06(P<0.0001)							
<b>10.2.2 Acupuncture vs. supervised osteoarthritis education</b>							
Berman 2004	169	-10.8 (11.7)	125	-5.3 (10.6)		100%	-0.48[-0.72,-0.25]
<b>Subtotal ***</b>	<b>169</b>		<b>125</b>			<b>100%</b>	<b>-0.48[-0.72,-0.25]</b>
Heterogeneity: Tau <sup>2</sup> =0; Chi <sup>2</sup> =0, df=0(P<0.0001); I <sup>2</sup> =100% Test for overall effect: Z=4.05(P<0.0001)							
<b>10.2.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)</b>							
Foster 2007	113	-8.2 (11.5)	105	-6.2 (11.4)		100%	-0.17[-0.44,0.09]
<b>Subtotal ***</b>	<b>113</b>		<b>105</b>			<b>100%</b>	<b>-0.17[-0.44,0.09]</b>
Heterogeneity: Not applicable Test for overall effect: Z=1.26(P=0.21)							
<b>10.2.4 Acupuncture vs. home exercises/advice leaflet alone</b>							
Williamson 2007	60	-1.6 (7.8)	61	0.7 (8.2)		100%	-0.28[-0.64,0.07]
<b>Subtotal ***</b>	<b>60</b>		<b>61</b>			<b>100%</b>	<b>-0.28[-0.64,0.07]</b>
Heterogeneity: Not applicable Test for overall effect: Z=1.55(P=0.12)							
<b>10.2.5 Acupuncture vs. supervised exercise alone</b>							
Williamson 2007	60	-1.6 (7.8)	60	-0.5 (9.3)		100%	-0.13[-0.49,0.23]
<b>Subtotal ***</b>	<b>60</b>		<b>60</b>			<b>100%</b>	<b>-0.13[-0.49,0.23]</b>
Heterogeneity: Not applicable Test for overall effect: Z=0.7(P=0.48)							
<b>10.2.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)</b>							
Scharf 2006	314	-2.1 (2.1)	308	-0.9 (1.9)		100%	-0.6[-0.76,-0.44]
<b>Subtotal ***</b>	<b>314</b>		<b>308</b>			<b>100%</b>	<b>-0.6[-0.76,-0.44]</b>
Heterogeneity: Not applicable							
Favors acupuncture    -4    -2    0    2    4    Favors no acup							

Study or subgroup	Acupuncture		Waiting list or other active treatment controls		Std. Mean Difference	Weight	Std. Mean Difference
	N	Mean(SD)	N	Mean(SD)			

Test for overall effect: Z=7.3(P<0.0001)

**Analysis 10.3. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization).**

Study or subgroup	Acupuncture		Waiting list or other active treatment controls		Std. Mean Difference	Weight	Std. Mean Difference
	N	Mean(SD)	N	Mean(SD)			
<b>10.3.1 Acupuncture vs. waiting list control</b>							
Berman 1999	36	-20.6 (34.4)	37	-0.8 (11.1)		18.91%	-0.77[-1.25,-0.3]
Witt 2005	145	-24.1 (20.6)	67	-2.1 (14.1)		31.83%	-1.17[-1.48,-0.86]
Witt 2006	300	-17.7 (21)	279	-0.7 (21)		49.26%	-0.81[-0.98,-0.64]
<b>Subtotal ***</b>	<b>481</b>		<b>383</b>			<b>100%</b>	<b>-0.92[-1.16,-0.67]</b>
Heterogeneity: Tau <sup>2</sup> =0.02; Chi <sup>2</sup> =4.14, df=2(P=0.13); I <sup>2</sup> =51.71%							
Test for overall effect: Z=7.28(P<0.0001)							
<b>10.3.2 Acupuncture vs. supervised osteoarthritis education</b>							
Berman 2004	169	-15.2 (15.8)	125	-7.3 (14.2)		100%	-0.52[-0.76,-0.29]
<b>Subtotal ***</b>	<b>169</b>		<b>125</b>			<b>100%</b>	<b>-0.52[-0.76,-0.29]</b>
Heterogeneity: Not applicable							
Test for overall effect: Z=4.35(P<0.0001)							
<b>10.3.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)</b>							
Foster 2007	113	-11.9 (15.9)	105	-9 (15.5)		100%	-0.18[-0.45,0.08]
<b>Subtotal ***</b>	<b>113</b>		<b>105</b>			<b>100%</b>	<b>-0.18[-0.45,0.08]</b>
Heterogeneity: Tau <sup>2</sup> =0; Chi <sup>2</sup> =0, df=0(P<0.0001); I <sup>2</sup> =100%							
Test for overall effect: Z=1.35(P=0.18)							
<b>10.3.4 Acupuncture vs. home exercises/advice leaflet alone</b>							
Williamson 2007	60	-2.6 (8.4)	61	0.5 (8.6)		100%	-0.37[-0.73,-0.01]
<b>Subtotal ***</b>	<b>60</b>		<b>61</b>			<b>100%</b>	<b>-0.37[-0.73,-0.01]</b>
Heterogeneity: Not applicable							
Test for overall effect: Z=2.02(P=0.04)							
<b>10.3.5 Acupuncture vs. supervised exercise alone</b>							
Williamson 2007	60	-2.6 (8.4)	60	-0.8 (10)		100%	-0.2[-0.56,0.16]
<b>Subtotal ***</b>	<b>60</b>		<b>60</b>			<b>100%</b>	<b>-0.2[-0.56,0.16]</b>
Heterogeneity: Not applicable							
Test for overall effect: Z=1.09(P=0.27)							
<b>10.3.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)</b>							
Scharf 2006	314	-2.1 (2)	308	-0.9 (1.9)		100%	-0.61[-0.78,-0.45]
<b>Subtotal ***</b>	<b>314</b>		<b>308</b>			<b>100%</b>	<b>-0.61[-0.78,-0.45]</b>

Favors acupuncture      -4      -2      0      2      4      Favors no acup



Study or subgroup	Acupuncture		Waiting list or other active treatment controls		Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
	N	Mean(SD)	N	Mean(SD)			

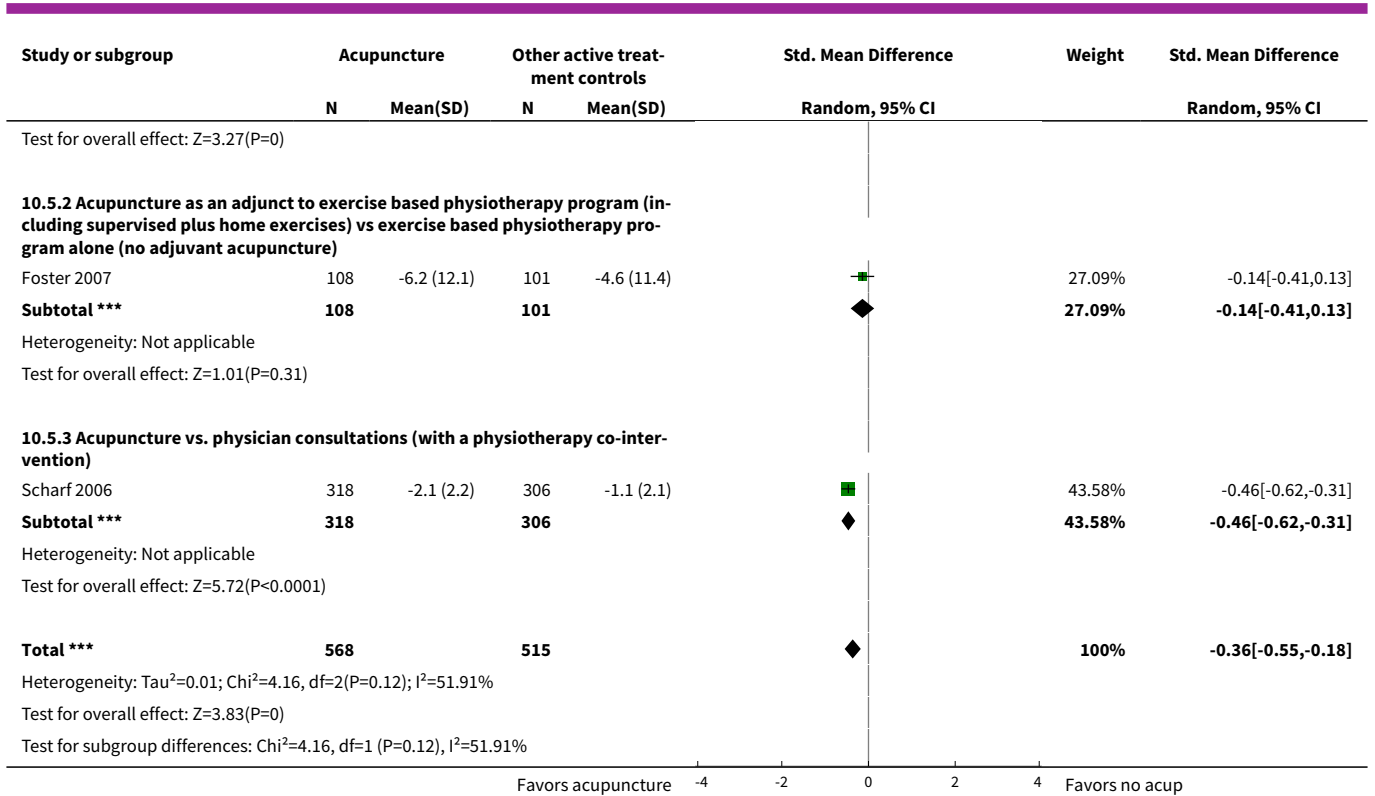
Heterogeneity: Not applicable  
Test for overall effect: Z=7.48(P<0.0001)

**Analysis 10.4. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 4 Pain (26 weeks after baseline).**

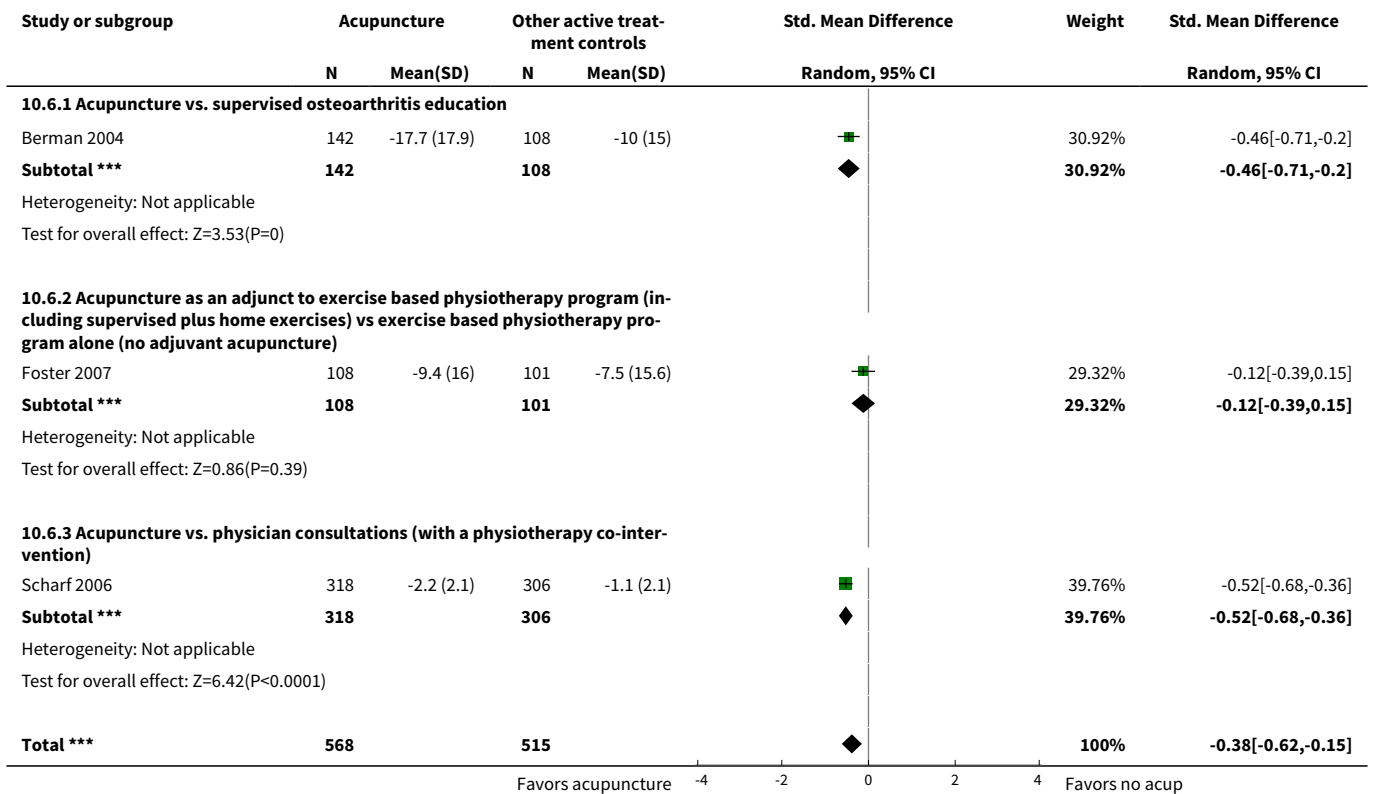
Study or subgroup	Acupuncture		Other active treatment controls		Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
	N	Mean(SD)	N	Mean(SD)			
<b>10.4.1 Acupuncture vs. supervised osteoarthritis education</b>							
Berman 2004	142	-3.8 (3.9)	108	-1.7 (3.5)	-0.56	31.96%	-0.56[-0.81,-0.3]
<b>Subtotal ***</b>	<b>142</b>		<b>108</b>		<b>-0.56</b>	<b>31.96%</b>	<b>-0.56[-0.81,-0.3]</b>
Heterogeneity: Not applicable Test for overall effect: Z=4.3(P<0.0001)							
<b>10.4.2 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjuvant acupuncture)</b>							
Foster 2007	108	-2.3 (3.6)	105	-2.3 (3.8)	-0.01	31.22%	-0.01[-0.28,0.26]
<b>Subtotal ***</b>	<b>108</b>		<b>105</b>		<b>-0.01</b>	<b>31.22%</b>	<b>-0.01[-0.28,0.26]</b>
Heterogeneity: Tau <sup>2</sup> =0; Chi <sup>2</sup> =0, df=0(P<0.0001); I <sup>2</sup> =100% Test for overall effect: Z=0.08(P=0.94)							
<b>10.4.3 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)</b>							
Scharf 2006	318	-2.2 (2.1)	306	-1.1 (2.2)	-0.51	36.83%	-0.51[-0.67,-0.35]
<b>Subtotal ***</b>	<b>318</b>		<b>306</b>		<b>-0.51</b>	<b>36.83%</b>	<b>-0.51[-0.67,-0.35]</b>
Heterogeneity: Not applicable Test for overall effect: Z=6.28(P<0.0001)							
<b>Total ***</b>	<b>568</b>		<b>519</b>		<b>-0.37</b>	<b>100%</b>	<b>-0.37[-0.68,-0.06]</b>
Heterogeneity: Tau <sup>2</sup> =0.06; Chi <sup>2</sup> =11.32, df=2(P=0); I <sup>2</sup> =82.33% Test for overall effect: Z=2.35(P=0.02) Test for subgroup differences: Chi <sup>2</sup> =11.32, df=1 (P=0), I <sup>2</sup> =82.33%							

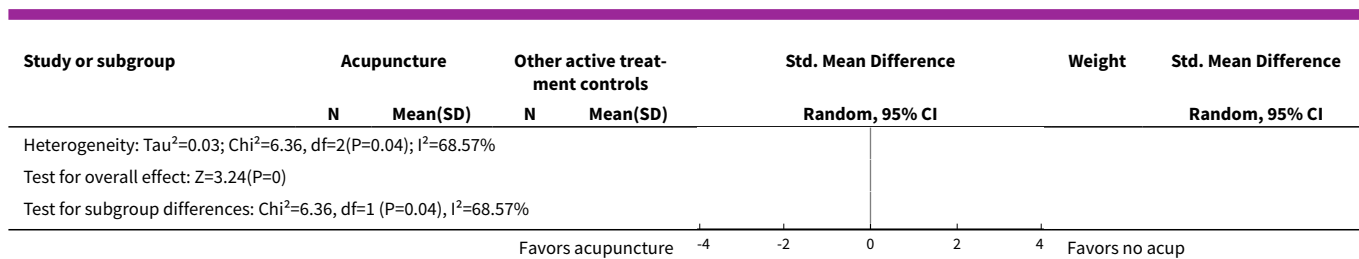
**Analysis 10.5. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 5 Function (26 weeks after baseline).**

Study or subgroup	Acupuncture		Other active treatment controls		Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
	N	Mean(SD)	N	Mean(SD)			
<b>10.5.1 Acupuncture vs. supervised osteoarthritis education</b>							
Berman 2004	142	-12.4 (13.3)	108	-7.2 (11.1)	-0.42	29.33%	-0.42[-0.67,-0.17]
<b>Subtotal ***</b>	<b>142</b>		<b>108</b>		<b>-0.42</b>	<b>29.33%</b>	<b>-0.42[-0.67,-0.17]</b>
Heterogeneity: Not applicable							



**Analysis 10.6. Comparison 10 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA change from baseline analysis, Outcome 6 Total score (26 weeks after baseline).**



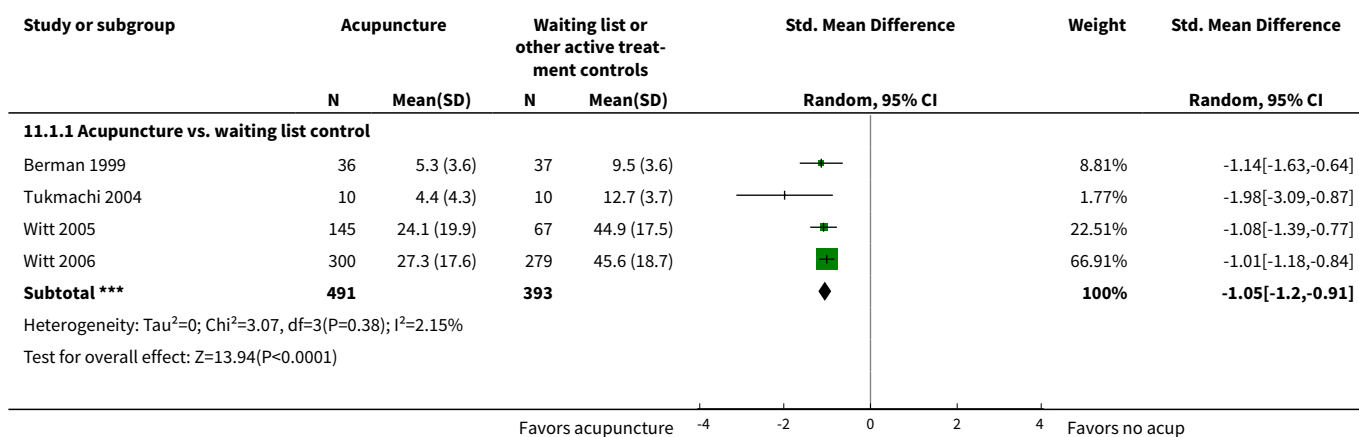


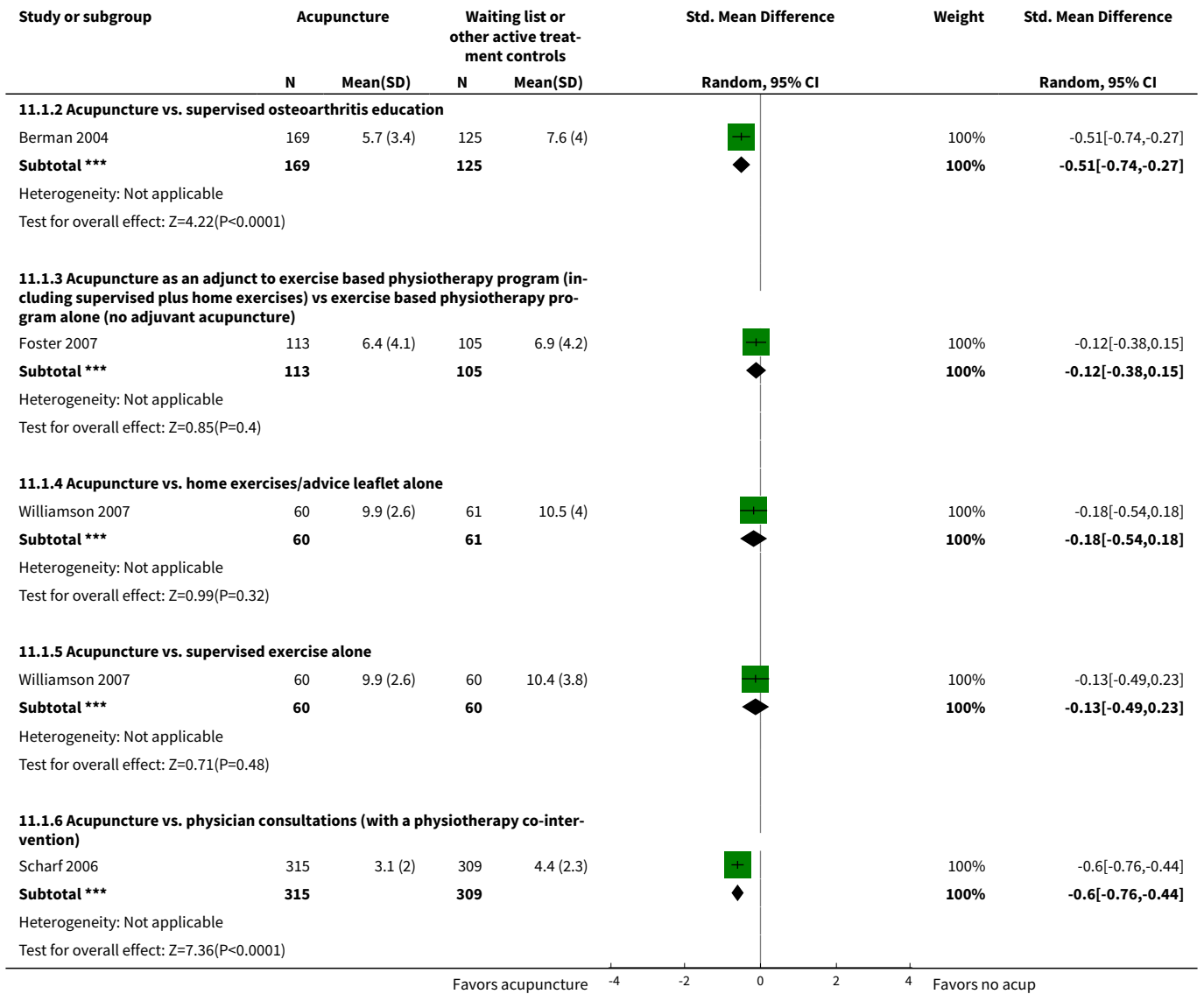
**Comparison 11. Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA post-treatment scores analysis**

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
<b>1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	8		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Acupuncture vs. waiting list control	4	884	Std. Mean Difference (IV, Random, 95% CI)	-1.05 [-1.20, -0.91]
1.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.74, -0.27]
1.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjunct acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.12 [-0.38, 0.15]
1.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.54, 0.18]
1.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.13 [-0.49, 0.23]
1.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	624	Std. Mean Difference (IV, Random, 95% CI)	-0.60 [-0.76, -0.44]
<b>2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	7		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 Acupuncture vs. waiting list control	3	864	Std. Mean Difference (IV, Random, 95% CI)	-1.12 [-1.38, -0.86]
2.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.75, -0.28]
2.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjunct acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	0.00 [-0.26, 0.27]

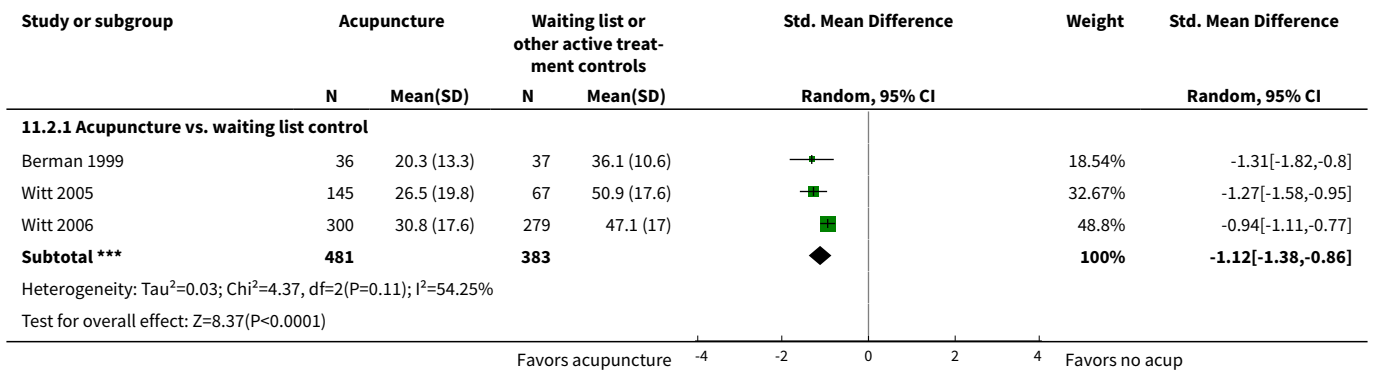
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.21 [-0.56, 0.15]
2.5 Acupuncture vs. supervised exercise alone	1	120	Std. Mean Difference (IV, Random, 95% CI)	-0.05 [-0.41, 0.31]
2.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	623	Std. Mean Difference (IV, Random, 95% CI)	-0.67 [-0.83, -0.50]
<b>3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization)</b>	<b>8</b>		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1 Acupuncture vs. waiting list control	3	864	Std. Mean Difference (IV, Random, 95% CI)	-1.13 [-1.37, -0.89]
3.2 Acupuncture vs. supervised osteoarthritis education	1	294	Std. Mean Difference (IV, Random, 95% CI)	-0.53 [-0.77, -0.30]
3.3 Acupuncture as an adjunct to exercise based physiotherapy program (including supervised plus home exercises) vs exercise based physiotherapy program alone (no adjunct acupuncture)	1	218	Std. Mean Difference (IV, Random, 95% CI)	-0.03 [-0.30, 0.24]
3.4 Acupuncture vs. home exercises/advice leaflet alone	1	121	Std. Mean Difference (IV, Random, 95% CI)	-0.21 [-0.57, 0.15]
3.5 Acupuncture vs. supervised exercise alone	2	148	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.56, 0.20]
3.6 Acupuncture vs. physician consultations (with a physiotherapy co-intervention)	1	623	Std. Mean Difference (IV, Random, 95% CI)	-0.57 [-0.73, -0.41]

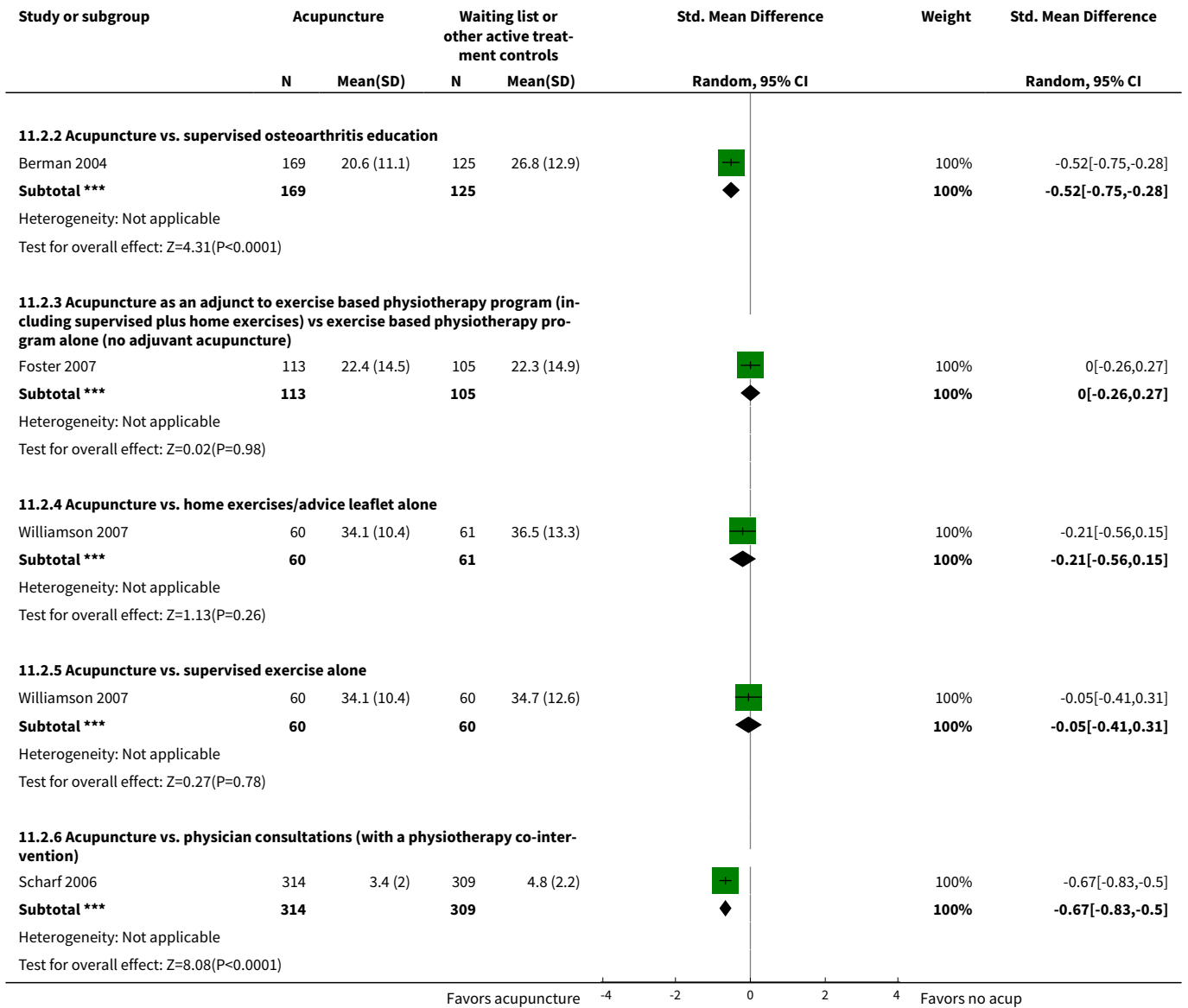
**Analysis 11.1. Comparison 11 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA post-treatment scores analysis, Outcome 1 Pain (Time point equal to or less than three months and closest to eight weeks post-randomization).**



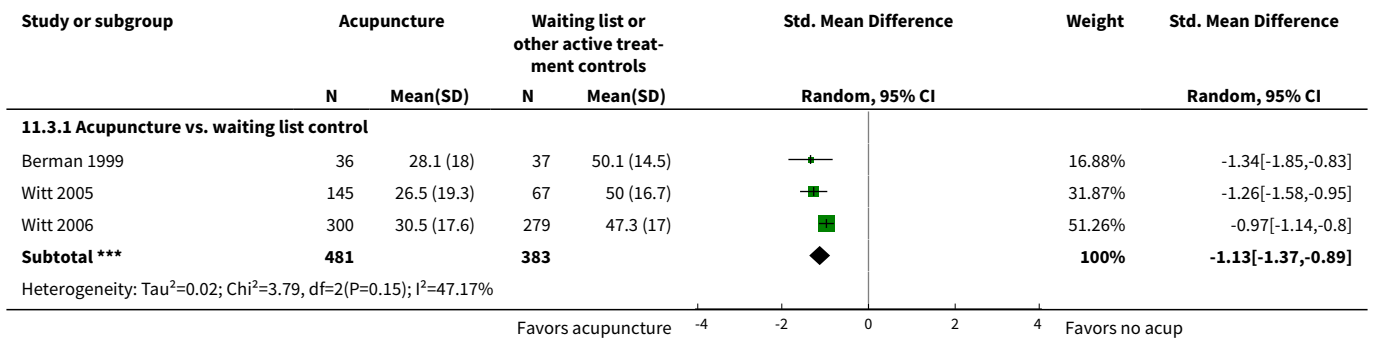


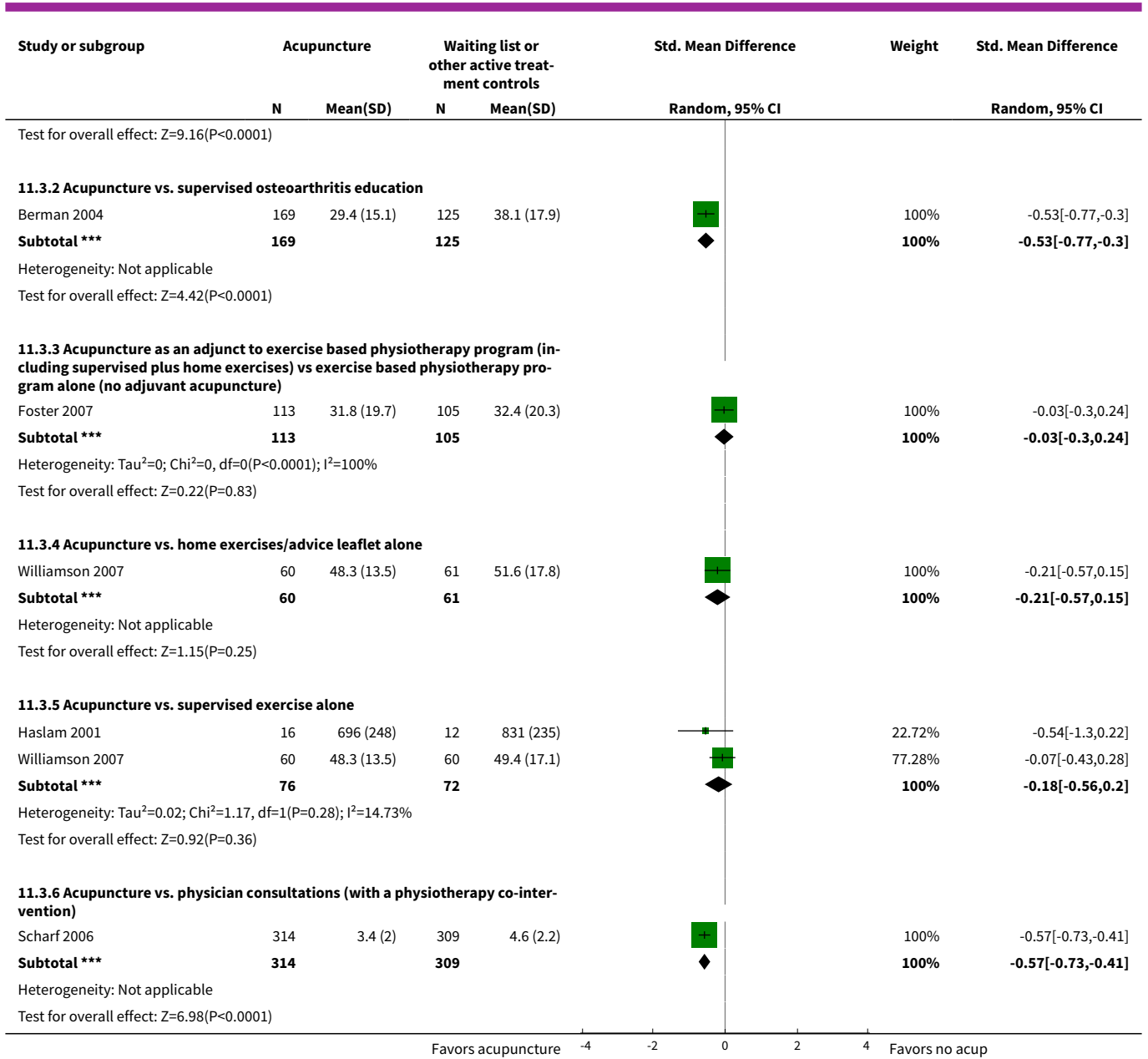
**Analysis 11.2. Comparison 11 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA post-treatment scores analysis, Outcome 2 Function (Time point equal to or less than three months and closest to eight weeks post-randomization).**





**Analysis 11.3. Comparison 11 Acupuncture vs. waiting list or other active treatment controls for peripheral joint OA post-treatment scores analysis, Outcome 3 Total score (Time point equal to or less than three months and closest to eight weeks post-randomization).**





**ADDITIONAL TABLES**

**Table 1. Search strategies**

MEDLINE (OVID) and Central (OVID)*	EMBASE (OVID)
1. Acupuncture/ or acupuncture.mp.	1. exp ACUPUNCTURE/ or acupuncture.mp.
2. acupuncture therapy.mp. or exp Acupuncture Therapy/	2. exp Acupuncture Analgesia/
3. auriculotherapy.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	3. auriculotherapy.mp.
4. electroacupuncture/ or electroacupuncture.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	4. electroacupuncture.mp. or ELEC-TROACUPUNCTURE/
5. moxibustion/ or moxibustion.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	5. moxibustion.mp.
	6. traditional medicine.mp. or exp Traditional Medicine/

**Table 1. Search strategies** (Continued)

6. medicine, oriental traditional/ or medicine, chinese traditional/ 7. Oriental Traditional Medicine.mp. [mp=title, original title, abstract, name of substance word, subject heading word] 8. Chinese traditional medicine.mp. [mp=title, original title, abstract, name of substance word, subject heading word] 9. or/1-8 10. arthritis.mp. [mp=title, original title, abstract, name of substance word, subject heading word] 11. arthritis/ or exp osteoarthritis/ 12. osteoarthritis.mp. [mp=title, original title, abstract, name of substance word, subject heading word] 13. joint diseases/ or arthralgia/ 14. (joint pain or chronic joint symptoms or gonarthrosis or osteoarthritis or ostoarthrosis or degenerative arthritis or joint diseases or arthralgia).mp. [mp=title, original title, abstract, name of substance word, subject heading word] 15. or/10-14 16. clinical trial.pt. 17. randomized.ab. or randomized controlled trial.pt. or controlled clinical trial.pt. 18. placebo.ab. 19. exp clinical trials/ 20. randomly.ab. 21. trial.ti. 22. or/16-21 23. animals/ 24. humans/ 25. 23 not (23 and 24) 26. 22 not 25 27. 9 and 15 and 26	7. 1 or 2 or 3 or 4 or 5 or 6 8. arthritis.mp. [mp=title, original title, abstract, name of substance word, subject heading word] 9. arthritis/ or exp osteoarthritis/ 10. osteoarthritis.mp. [mp=title, original title, abstract, name of substance word, subject heading word] 11. (joint diseases or arthralgia or joint pain or chronic joint symptoms or gonarthrosis or osteoarthritis or ostoarthrosis or degenerative arthritis).mp. [mp=title, original title, abstract, name of substance word, subject heading word] 12. or/8-11 13. exp clinical trial/ 14. exp randomized controlled trial/ 15. randomized.ab. 16. placebo.ab. 17. ct.fs. 18. randomly.ab. 19. trial.ti. 20. or/13-19 21. exp animal/ 22. human/ 23. 21 not (21 and 22) 24. 20 not 23 25. 7 and 12 and 24
---	--

\*Note: Cochrane CENTRAL, Issue 1, 2008 was searched through EBM Reviews (OVID). The strategy used is identical to that used to search MEDLINE (OVID).

**Table 2. Assessments of adequacy of acupuncture and control protocols\***

Study	Choice of acupoints	Number of sessions	Needling technique	Experience	Control Group	Guess of study
<a href="#">Berman 1999</a>	Adequate	Adequate	Adequate	Adequate†	No sham/placebo acupuncture was used. Patients were not blinded.	<a href="#">Berman 1999</a>
<a href="#">Berman 2004</a>	Adequate	Adequate	Adequate	Adequate	Sham/placebo‡	<a href="#">Berman 2004</a>
<a href="#">Christensen 1992</a>	Adequate	Inadequate (6 treatments in 3 weeks is too few and course is too short.)	Adequate	Don't know (Not stated)	No appropriate sham/placebo intervention.	Uncertain
<a href="#">Fink 2001</a>	Adequate	Adequate	Adequate	Don't know §	The insertion sham acupuncture with same number of points and same depths and was 5cm away from the "real points". This technique may cause non-specific endorphin release which has some effect on	Don't know



**Table 2. Assessments of adequacy of acupuncture and control protocols\*** (Continued)

					<p>pain. And the sham points stimulated are also probably located at the Twelve muscle region and twelve Cutaneous region where Qi and Blood are distributed.</p>	
Foster 2007	Adequate	Inadequate (Too few treatments and too short a course)	Adequate	Don't know (No description on the training and experience of the "therapist")	Non-insertion sham control was used and seems appropriate. The sham procedure may have had a physiological effect, including an acupressure massage effect, because the sham devices were placed and left at the true acupuncture points. This is suggested by the fact that the de qi sensation was reported at least once for 55% of patients receiving the non-penetrating acupuncture.	Don't know
Haslam 2001	Adequate	Inadequate	Adequate	Don't know	No sham control.	Don't know
Molsberger 1994	Adequate	Adequate	Adequate	Don't know (No qualification of the acupuncturist provided.)	Needling insertion sham control used. It may cause some non-specific needling effect such as DINC effect..	Uncertain
Scharf 2006	Adequate	Adequate	Adequate	Don't know (Not described)	The minimal sham insertion control without stimulation at 10 points at defined distances from TCA points. It may cause non-specific needling response such as diffuse noxious inhibitory control effect; with 315 practices, it is not clear how the principal investigator could ensure that these minimal sham treatments were properly implemented	One acupuncturist (MH) is unaware of this study or its results. The other acupuncturist (LL) guess this may be the German Scharf study, but not sure.
Sangdee 2002	Adequate	Adequate	Adequate	Don't know ¶	Non-insertion placebo control used. Dummy patch electrodes on surface of acupuncture point.	Uncertain
Stener-Victorin 2004	Adequate	Adequate	Adequate	Don't know **	No sham/placebo acupuncture control.	Don't know
Takeda 1994	Adequate	Inadequate ††	Adequate	Don't know ††	Sham acupuncture (minimal-insertion sham) was used. The sham points selected were too close to the real point (one inch away) and may produce therapeutic effect. The investigators told the patients in both groups that they were receiving two different types of acupuncture treatments.	Uncertain

**Table 2. Assessments of adequacy of acupuncture and control protocols\*** (Continued)

					Therefore, patients in both group thought they had received real treatment.	
Tukmachi 2004	Adequate	Adequate	Adequate	Don't know (No mentioning of acupuncturist's training or certifications.)	No sham/placebo acupuncture used. Patients were not blinded.	Uncertain
Vas 2004	Adequate	Adequate	Adequate	Don't know §§	Non-insertion placebo control used which seems appropriate	Uncertain
Williamson 2007	Adequate	Inadequate	Adequate	Don't know	No sham control.	Don't know
Witt 2005	Adequate	Adequate	Adequate	Don't know ¶¶	Minimal sham insertion control at non-acupuncture points was used. As the authors pointed out, this type of intervention may have a physiological effect. Credibility of the blinding to treatment was assessed.	Uncertain
Witt 2006	Don't know ***	Don't know †††	Don't know ‡‡‡	Don't know ¶¶	No sham/placebo acupuncture was used. Patients were not blinded.	Uncertain

\* Acupuncturists assessed acupuncture as adequate in terms of the choice of acupuncture points and needling technique for all trials except for the Witt 2006 trial, a pragmatic trial for which the point selection and needling technique were entirely at the discretion of the treating physician.

† The paper did not mention the qualification of the acupuncturist, although the reviewer knows that the acupuncturist was qualified.

‡ Insertion-sham and non-insertion sham combination method was adequately used. Credibility of the blinding to treatment was assessed.

§ No mentioning of the training of physician who carried out the procedure. However, the paper states the physician has sound knowledge of acupuncture.

¶ The treatments were performed "by the author". No information regarding the training background and practice experience of this author was provided.

¶¶ Physician acupuncturist who received acupuncture training in the People's Republic of China. It is too vague to tell the training background. It is not clear whether he or she had full extensive acupuncture training or just a short training course designed for physicians.

\*\* The authors reported that "experienced physiotherapists" performed treatments but the acupuncture training and experience of these physiotherapists was not reported.

†† The frequency of 3 times per week is adequate, but the length of 3 weeks might be too short for the given condition of OA.

‡‡ A physical therapist trained in acupuncture, but no length of the training and no length of the number of years of acupuncture practice was reported.

§§ "A doctor specialising in acupuncture accredited by the Beijing University of Medical Sciences in China." The criteria for accreditation by that university is not clear.

||| The treatments were performed "by a senior physiotherapist experienced in musculoskeletal acupuncture." It is not known how the physiotherapist trained in acupuncture.

¶¶¶ Physician acupuncturists trained at least 140 hours. The training requirement is low.

\*\*\* Acupuncture point selection was at the discretion of the physicians. No point selection was described.

††† 15 treatments within 3 months, but not known how frequent the treatments were given.

‡‡‡ Not described. 140 hour training many not be sufficient to have good needling techniques.

**Table 3. Cochrane Back Review Group Quality Items**

Quality items	Scoring of item	Notes about scoring
A. Was the method of randomization ade-	A random (unpredictable) assignment sequence. Examples of adequate methods are computer generated random number table and	

**Table 3. Cochrane Back Review Group Quality Items** (Continued)

<p>quate? Yes/No/Don't know</p>	<p>use of sealed opaque envelopes. Methods of allocation using date of birth, date of admission, hospital numbers, or alternation should not be regarded as appropriate.</p>	
<p>B. Was the treatment allocation concealed? Yes/No/Don't know</p>	<p>Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.</p>	
<p>C. Were the groups similar at baseline regarding the most important prognostic indicators? Yes/No/Don't know</p>	<p>In order to receive a "yes," groups have to be similar at baseline regarding demographic factors, duration and severity of complaints, percentage of patients with neurologic symptoms, and value of main outcome measure(s).</p>	
<p>D. Was the patient blinded to the intervention? Yes/No/Don't know</p>	<p>The reviewer determines if enough information about the blinding is given in order to score a "yes."</p>	<p>We scored sham controlled trials as .5 (rather than 1) for patient blinding because even though these trials used a sham control, we cannot be certain that the sham was sufficiently credible to allow patients to be blinded to the true treatment being evaluated. The only exception was trials that evaluated the credibility of the sham, and found the sham to be indistinguishable from the acupuncture.</p>
<p>E. Was the care provider blinded to the intervention? Yes/No/Don't know</p>	<p>The reviewer determines if enough information about the blinding is given in order to score a "yes."</p>	
<p>F. Was the outcome assessor blinded to the intervention? Yes/No/Don't know</p>	<p>The reviewer determines if enough information about the blinding is given in order to score a "yes."</p>	<p>For determining whether the outcomes assessor was blinded, we always considered the patient to be the outcomes assessor for patient rated outcomes (ie, WOMAC), even if the outcomes data was collected by someone else. We scored sham controlled trials as .5 (rather than 1) for outcomes assessor blinding because even though these trials used a sham control, we cannot be certain that the sham was sufficiently credible to allow patients to be blinded to the true treatment being evaluated. The only exception was trials that evaluated the credibility of the sham, and found the sham to be indistinguishable from the acupuncture.</p>

**Table 3. Cochrane Back Review Group Quality Items** (Continued)

G. Were co-interventions avoided or similar? Yes/No/Don't know	Co-interventions should either be avoided in the trial design or similar between the index and control groups.
H. Was the compliance acceptable in all groups? Yes/No/Don't know	The reviewer determines if the compliance to the interventions is acceptable, based on the reported intensity, duration, number and frequency of sessions for both the index intervention and control intervention(s).
I. Was the drop-out rate described and acceptable? Yes/No/Don't know	The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of withdrawals and drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up and does not lead to substantial bias a "yes" is scored. (N.B. these percentages are arbitrary, not supported by literature).
J. Was the timing of the outcome assessment in all groups similar? Yes/No/Don't know	Timing of outcome assessment should be identical for all intervention groups and for all important outcome assessments.
K. Did the analysis include an intention-to-treat analysis? Yes/No/Don't know	All randomized patients are reported/analyzed in the group they were allocated to by randomization for the most important moments of effect measurement (minus missing values) irrespective of noncompliance and co-interventions.

Note: There is a maximum total of 11 points on the Cochrane Back Review Group scale, with each item scored as 1 for 'Yes', and 0 for 'No' or 'Don't know' (except where indicated in 'Notes about scoring' column). Lower quality = 0-6; higher quality = 7-11.

**Table 4. Cochrane Back Review Group Quality Scores**

<b>BStudy, Year</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>	<b>e</b>	<b>f</b>	<b>g</b>	<b>h</b>	<b>i</b>	<b>j</b>	<b>k</b>	<b>Total</b>
Berman 1999	1	[1]	1	0	0	0	1	1	1	1	1	8
Berman 2004	1	1	1	.5/0	0	.5/0	1	1/0	1/0	1	1	9/6*
Christensen 1992	[1]	[1]	1	0	0	1/0	?	?	1	1	0	6/5†
Molsberger 1994	?	0	1	.5	0	.5	1	?	0	1	0	4
Fink 2001	1	?	1	1	0	1	?	0	0	1	0	5
Foster 2007	1	1	1	1/0	0	1/0	1/0	1	1	1	0	9/6‡
Haslam 2001	1	?	1	0	0	0	?	?	0	1	0	3
Sangdee 2002	?	0	1	.5	0	.5	1	1	1	1	0	6
Scharf 2006	1	1	1	1/0	0	1/0	0	1	1	1	1	9/7§
Stener-Victorin 2004	[1]	[1]	0	0	0	0	1	?	0	0	0	3
Takeda 1994	[1]	[1]	0	.5	0	.5	?	?	1	1	0	5
Tukmachi 2004	1	1	?	0	0	0	?	?	1	1	0	4
Vas 2004	1	1	1	.5	0	.5	1	[1]	0	1	1	8
Williamson 2007	1	1	1	0	0	0	?	?	0	1	1	5
Witt 2005	1	1	1	1/0	0	1/0	1	1	1	1	1	10/8
Witt 2006	1	1	1	0	0	0	?	?	1	1	1	6

\* Number preceding / is score for sham control group and number following / is score for education control group.

† Number preceding / is score for objective outcome assessed by blinded observer and number following / is score for patient assessed outcome.

‡ Number preceding / is sham control group score and number following / is advice and exercise control group score.

§ Number preceding / is score for sham control group and number following / is score for standard care control group.

|| Number preceding / is score for sham control group and number following / is score for waiting list control group.

**Table 5. Adverse events in the acupuncture and control groups and minor side effects of acupuncture \***

Study, Year	Assessment Methods for Adverse Events	Minor Adverse Events per Group, n/n†		Assessment Methods for Side Effects of Acupuncture	Side Effects of Acupuncture‡
		Acupuncture	Control		
Berman 1999	NR	NR	NR	NR	No patients reported side effects from the 16 acupuncture sessions
Berman 2004	At each measurement time, investigators assessed patients' self-reports of adverse events that were potentially related to acupuncture; assessment instrument and definition of serious adverse event were not reported	14/190	Sham, 5/191; usual care, 7/189	Participants were asked to report subjective symptoms that could be attributed to acupuncture (such as dizziness, nausea, and numbness) during the study	Changes in subjective symptoms after baseline assessment did not statistically significantly differ among the 3 groups, and the incidence of these symptoms was low throughout the RCT (numbers not reported)
Christensen 1999	NR	NR	NR	NR	Approximately 16% of patients reported minor side effects of acupuncture (e.g., nausea, dizziness, and bruising); no patients withdrew from study because of side effects
Fink 2001	At each session, independent observer documented any adverse events	0§	0§	At each session, independent observer documented side effects	No side effects occurred
Foster 2007	At each measurement time, physiotherapists assessed participants' self-reports of adverse events that were potentially related to acupuncture; physiotherapists recorded any adverse reactions	5/117 (five events, reported as: "(pain, sleepiness, fainting, nausea, and swelling around the treated knee).")	Sham, 0/119; usual care, 0/116	Participants were asked to report side effects of treatment during the study, and physiotherapists recorded them	No side effects occurred
Haslam 2001	At each measurement time, investigators assessed patients' self-reports of adverse events that were potentially related to acupuncture	0	0	Participants were asked to report subjective symptoms that could be attributed to acupuncture during the study	Patients were compliant with treatment, and there were no reported side-effects
Molsberger 1994	NR	NR	NR	NR	NR
Sangdee 2002	Nondirective questioning for adverse	0	0	Physical examination of the knee for contusions	Approximately 45% of patients in the electroacupuncture groups had local bruises

**Table 5. Adverse events in the acupuncture and control groups and minor side effects of acupuncture \*** (Continued)

	events performed weekly for 4 wk				around the knee; however, the bruises usually disappeared within 5-7 d
Scharf 2006	At each visit, the investigator documented all serious and non-serious events since the last visit; events were coded according to the Medical Dictionary for Regulatory Activities	20/330	Sham, 9/367; usual care, 16/342	Same assessment method as that used for adverse events (i.e., adverse events and minor side effects were not differentiated); all events coded for acupuncture and control groups	Of all events coded, only bruising was reported more often in the acupuncture and sham groups than in the usual care group; approximately 7% of patients in the acupuncture group and 10% of patients in the sham group reported bruising
Stener-Victorin 2004	NR	NR	NR	NR	NR
Takeda 1994	NR	NR	NR	NR	NR
Tukmachi 2004	NR	NR	NR	NR	NR
Vas 2004	NR	NR	NR	NR	3 patients reported bruising at 1 acupuncture point (SP6); it is assumed, but was not explicitly stated, that these 3 patients were among the 48 patients randomly assigned to the true acupuncture group
Williamson 2007	At each measurement time, investigators assessed patients' self-reports of adverse events that were potentially related to acupuncture	0	0	Participants were asked to report subjective symptoms that could be attributed to acupuncture during the study	No side effects to treatment occurred in any of the groups, beyond occasional minor bruising and bleeding in the acupuncture group, and this was never to the extent to cause any concern or discomfort for the patient
Witt 2005	Documented by physicians at each session; assessment instrument and definition of serious adverse event were not reported; assessments were made at each measurement point, but publication reports only adverse events measured at 26 wk	3/150	Sham, 2/76; waiting list, 4/74	Documented by physicians at each session and reported by patients at the end of wk 8; assessment instrument was not reported	Approximately 14% of patients in the acupuncture group and 18% of patients in the sham group reported minor side effects, primarily bruising, bleeding, or pain at needle insertion site
Witt 2006	NR	NR	NR	Side effects were recorded on patient and physician questionnaires after 3 mo	Approximately 5% of patients reported side effects of acupuncture, which were primarily local bleeding, bruising, or pain at needle insertion site;

**Table 5. Adverse events in the acupuncture and control groups and minor side effects of acupuncture** \* (Continued)  
 no effects were life-threatening

\*Definitions of adverse events and side effects of acupuncture and methods of assessments varied among trials. NR = not reported; RCT = randomized, controlled trial; SP6 = spleen meridian point 6.

†In 3 (Berman 2004; Witt 2005; Foster 2007) of the 4 RCT publications that reported number of adverse events per group, it was not clear whether the numbers reported in the publications (as included in the table above) were the numbers of adverse events in each group or the numbers of patients in each group who had adverse events. For the fourth RCT (Scharf 2006), numbers reported in this table are the numbers of patients in each group who had adverse events. Of the serious adverse events observed in the acupuncture groups (Berman 2004; Witt 2005; Scharf 2006), none was interpreted as treatment-related. Of the five adverse events reported in the Foster 2007 trial, only one appeared to be serious (i.e. "fainting"), according to the description in the paper (copied into table above).

‡The severity or duration of the minor side effects associated with the needling was not systematically reported in any RCT.

§Percentages of patients who experienced less than serious adverse events (e.g., hypertension, rash, or gastrointestinal symptoms) did not differ between groups (numbers not reported by authors).

||The number of patients with serious adverse events. The investigators did not define what was considered a serious adverse event. They also reported "notable" serious adverse events (also undefined), which occurred for 3 patients in the acupuncture group, 1 in the sham group, and 3 in the usual care group.



**Table 6. Results of the subgroup meta-analyses for pain outcome\***

Variable	Total Trials, n	Patients who were randomly assigned, n	Effect size(95% CI)	I <sup>2</sup> ,%	P value for interaction†	Not met‡
<b>All trials</b>	9	1835	-0.28 (-0.45,-0.11)	63.9	-	-
<b>Methodological variables</b>						
<b>Generation of random sequence</b>					0.141	SS
Adequate	7	1649	-0.25 (-0.45,-0.05)	69		
Unclear or no	2	186	-0.42 (-0.71,-0.13)	0		
<b>Allocation concealment</b>					0.215	SSFi
Adequate	6	1587	-0.26 (-0.48,-0.04)	74.2		
Unclear	3	248	-0.36 (-0.61,-0.11)	0		
<b>Blinding success</b>					0.042	BSSV
Yes	5	1221	-0.15 (-0.28,-0.01)	14.5		
Uncertain	4	614	-0.47 (-0.84,-0.10)	77.5		
<b>Intention-to-treat analysis</b>					0.532	FiFoSST
Yes	4	1319	-0.35 (-0.63,-0.07)	80.7		
Unclear or no	5	516	-0.20 (-0.43,0.03)	35.2		
<b>Clinical variables</b>					0.047	FoT
<b>Sufficient number of sessions delivered over an adequate treatment duration</b>						
Yes	7	1567	-0.34 (-0.54, -0.15)	66.1		
No	2	268	0.01 (-0.23, 0.25)	0		

**Table 6. Results of the subgroup meta-analyses for pain outcome\*** (Continued)

					0.042	BSSV met
<b>Electrical stimulation was used with the acupuncture</b>						
Yes	4	614	-0.50 (-0.81,-0.20)	66		
No	5	1215	-0.11 (-0.29,0.07)	42.3		
<b>Physiological activity of sham, as judged by acupuncturist</b>					0.042	BSSV
Likely	5	1221	-0.15 (-0.28,-0.01)	14.5		
Not likely	4	614	-0.47 (-0.84,-0.10)	77.5		
<b>Formula versus flexible formula for point selection</b>					0.057	BFISSTV used formula
Formula	6	716	-0.39 (-0.66,-0.12)	72.8		
Flexible formula	3	1119	-0.14 (-0.34,0.05)	77		

\* Subgroup analyses for sham controlled peripheral joint OA trials based on changes of pain scores from baseline at the short-term (i.e. closest to 8 wks) follow up. The structure and organization of this table was largely modeled after a similar table in a recent review (Reichenbach 2007).

†A statistical test for interaction could not be performed for the following prespecified subgroups because there were one or fewer studies in one of the strata of each of these subgroups: comparability of baseline, acceptability of drop-out rate, CBRG score  $\geq 6$ , follow-up length  $\geq 3$  months, choice of acupoints, needling technique, acupuncturists' experience.

‡"Not met" column lists the first (or first two) letters of the author of the studies that do not meet the criterion.

**Table 7. Results of the subgroup meta-analyses for function outcome\***

Variable	Total Trials, n	Patients who were randomly assigned, n	Effect size(95% CI)	I <sup>2</sup> ,%	P value for interaction†	Not met‡
All trials	9	1829	-0.28 (-0.46,-0.09)	69	-	
<b>CBRG Scale elements</b>						
<b>Generation of random sequence</b>					0.115	SS

**Table 7. Results of the subgroup meta-analyses for function outcome\*** (Continued)

	Adequate	7	1643	-0.24 (-0.46,-0.02)	73.9		
	Unclear or no	2	186	-0.43 (-0.72,-0.14)	0		
	<b>Allocation concealment</b>					0.187	SSFi
	Adequate	6	1581	-0.25 (-0.49,-0.01)	78.3		
	Unclear	3	248	-0.37 (-0.62,-0.12)	0		
	<b>Blinding success</b>					0.002	BSSV
	Yes	5	1215	-0.11 (-0.29,0.07)	42.3		
	Uncertain	4	614	-0.50 (-0.81,-0.20)	66		
	<b>Intention-to-treat analysis</b>					0.258	FiFoSST
	Yes	4	1318	-0.38 (-0.67,-0.09)	81.9		
	Unclear or no	5	511	-0.17 (-0.43,0.10)	49.3		
	<b>Other elements</b>					0.007	FoT
	<b>Sufficient number of sessions delivered over an adequate treatment duration</b>						
	Yes	7	1566	-0.37 (-0.57, -0.16)	67.4		
	No	2	263	0.10 (-0.14,0.34)	0		
	<b>Electrical stimulation was used with the acupuncture</b>					0.002	BSSV met
	Yes	4	614	-0.47 (-0.84,-0.10)	77.5		
	No	5	1215	-0.15 (-0.28,-0.01)	14.5		
	<b>Physiological activity of sham, as judged by acupuncturist</b>				14.5	0.002	BSSV



**Table 7. Results of the subgroup meta-analyses for function outcome\*** (Continued)

Likely	5	1215	-0.11 (-0.29,0.07)	42.3		
Not likely	4	614	-0.50 (-0.81,-0.20)	66		
<b>Formula versus flexible formula for point selection</b>					0.007	BFISSTV used formula
Formula	6	716	-0.39 (-0.64,-0.13)	58.4		
Flexible formula	3	1113	-0.12 (-0.369,0.12)	68.9		

\* Subgroup analyses for sham controlled peripheral joint OA trials based on changes of pain scores from baseline at the short-term (closest to 8 wks) follow up. The structure and organization of this table was largely modeled after a similar table in a recent review (Reichenbach 2007).

†A statistical test for interaction could not be performed for the following prespecified subgroups because there were one or fewer studies in one of the strata of each of these subgroups: comparability of baseline, acceptability of drop-out rate, CBRG score  $\geq 6$ , follow-up length  $\geq 3$  months, choice of acupoints, needling technique, acupuncturists' experience.

‡"Not met" column lists the first (or first two) letters of the author of the studies that do not meet the criterion.

## WHAT'S NEW

Date	Event	Description
11 November 2009	Amended	CMSSG ID C104-R

## CONTRIBUTIONS OF AUTHORS

Eric Manheimer determined inclusion eligibility of trials, developed the data extraction form, and extracted data/assessed methodological quality of all included trials (with the exception of the German language trial). For the version of this review that was previously published in *Annals of Internal Medicine*, Klaus Linde acted as a second assessor, and determined inclusion eligibility of trials, extracted data from the trials, and assessed the methodological quality of the trials. For this major update for *The Cochrane Library*, Ke Cheng acted as a second assessor, and determined inclusion eligibility of newer trials, as well as trials that potentially met the expanded eligibility criteria (i.e. any peripheral joints), and she also extracted data/assessed methodological quality of the three new hip OA trials, and the two most recently published knee OA trials. Eric Manheimer organized the data in RevMan, conducted and oversaw the analyses, and drafted the manuscript. Ke Cheng assisted with data entry and data analysis; specifically, Ke Cheng entered all the subgroup analyses data in Tables 5 and 6, and conducted these subgroup analyses, under the supervision of Eric Manheimer. Susan Wieland participated in risk of bias assessments and completed the risk of bias tables for included studies. Susan Wieland also completed the summary of findings tables, under the supervision of Eric Manheimer. Lixing Lao (with Marcos Hsu) rated the adequacy of the acupuncture treatments. Lex Bouter, Brian Berman, Ke Cheng, Klaus Linde, Junghee Yoo, Lixing Lao, Susan Wieland and Danielle van der Windt provided comments on the drafts of the manuscript, and endorsed its final version.

## DECLARATIONS OF INTEREST

This review includes trials in which some of the reviewers were involved: [Berman 1999](#) and [Berman 2004](#): Brian Berman and Lixing Lao; [Witt 2005](#): Klaus Linde. These trials were reviewed by at least two other members of the review team. Eric Manheimer was one of the two reviewers for all included trials, including the two Berman trials. Mr. Manheimer works at the same research center at which the two Berman trials were conducted, but he was not involved in the conduct of either of these two trials. Lixing Lao uses acupuncture in his clinical work. Klaus Linde has received travel reimbursement and in two cases fees for speaking on research at meetings of acupuncture societies (British, German and Spanish Medical Acupuncture Society, Society of Acupuncture Research). Brian Berman, Lixing Lao, and Eric Manheimer received honoraria for preparing and delivering presentations on acupuncture at the 2007 meeting of the Society for Acupuncture Research.

## SOURCES OF SUPPORT

### Internal sources

- No sources of support supplied

### External sources

- National Institutes of Health Grant R24 AT001293, USA.
- National Institutes of Health Grant P01AT002605, USA.

## DIFFERENCES BETWEEN PROTOCOL AND REVIEW

For the earlier journal version of this review ([Manheimer 2007](#)), we prepared a protocol which was internally reviewed amongst the co-authors but which was not published in *The Cochrane Library*. That protocol is available upon request from the contact author of this review. Because the methods that we planned to use in preparing the original journal version (for which the unpublished protocol served as the review protocol) are similar to the methods used for this update, we have described below differences between that unpublished protocol and this update. All of these differences relate to the quantitative data synthesis, as follows:

- In the protocol, “we defined the short-term outcome using two alternate follow-up time points: first, as the measurement point closest to eight weeks, and less than or equal to three months, following randomization, and second, as the measurement point closest to the end of treatment. We planned to conduct the meta-analysis at both points in order to determine whether the choice of short-term time points affected the meta-analysis results.” For the review, we have removed the ‘end of treatment’ analysis because co-authors raised the concern that ‘end of treatment’ may vary in terms of time since randomization across studies, and thus may not be a suitable outcome time point. Also the ‘end of treatment’ analysis was mostly duplicative because the ‘end of treatment’ and ‘closest to 8 weeks’ analyses used almost identical outcome data. There were differences in data used for only two trials, as described below. For a draft version of this review, we prepared forest plots which showed that the combined and individual trial estimates calculated using the

alternate definition of the short term effects measurement time point (i.e., end of acupuncture treatment) were very similar to those calculated using the time point defined for the primary analyses (i.e., less than or equal to 3 months and closest to eight weeks after baseline). Because the end of treatment analyses forest plots did not affect any of our results or conclusions, and only contributed to the complexity of the review by the addition of several new figures, and also because there were concerns about the validity of this analysis because end of treatment may vary in terms of time since randomization across studies, these plots were removed from the final review for the sake of simplicity.

--For the [Takeda 1994](#) trial, the 8 week post-randomization analysis data included the data from the time point four weeks after the end of the three weeks period of treatment while the end of treatment analysis data included the data from the time point immediately after the three weeks of treatment. (It turns out that for the [Takeda 1994](#) trial, the end of treatment data is very similar to the 4 week post-end of treatment data, so it did not matter whether we used end of treatment data or the four week post end of treatment data for this trial.)

--For the [Berman 2004](#) trial, we used the 14 week time point when we defined short-term effects as end of treatment whereas we used the 8 week time point when we defined short-term effects as the closest to 8 weeks. (For the [Berman 2004](#) trial, at 8 weeks, acuapunctures relative effects on function were slightly larger and pain slightly smaller, as compared with the results at 14 weeks.)

- In the protocol, we had specified “To calculate standardized mean differences, we used a comparison of post-treatment mean values and SDs for each group as well as a comparison of means and SDs of changes from baseline for each group. For the included trials, which were often incompletely reported, post-treatment scores analyses required making fewer assumptions than the changes from baseline analyses. Therefore, we will present only the results from the post-treatment scores analyses; however, we will also report any outcomes at any time points for which the statistical significance of the pooled result changes depending on the choice of analysis.” However, after the protocol was drafted, some authors noted a preference for calculating standardized mean differences as a comparison of means and SDs of changes from baseline as opposed to means and SDs of post-treatment values. Based on these preferences, and also because we were eventually able to obtain SDs of change from most trials, it was decided to use changes from baseline as the primary analyses, although the final version of the internal protocol did not reflect this decision. The post-treatment values showed slightly larger pooled benefits than the changes from baseline values, as noted in the Results section, but there were no outcomes at any time points for which the statistical significance of the pooled result changed depending on the choice of analysis.
- In the protocol, we specified that “Sensitivity analyses were performed between overall trial quality (ie, higher versus lower) and trial outcome (ie, positive versus neutral).” However, because these analyses were largely uninformative and also because some researchers consider subgroup analyses on overall quality scores to be problematic ([Juni 1999](#)), for this review, we also conducted subgroup analyses on each of 15 dichotomized methodological and clinical variables to investigate whether differences in effects of acupuncture between any two subgroups for any variable were statistically significant.
- In the protocol, we specified that we would place “studies into one of three categories according to which of the following comparisons were evaluated:

- 1) Acupuncture versus a sham intervention;
- 2) Acupuncture versus no acupuncture; or
- 3) Acupuncture versus another active treatment.

In the review, we found that several studies compared acupuncture with a waiting list control. Therefore, for the review, we created a new comparison, not prespecified in the protocol, of acupuncture versus a waiting list control. In addition, in the protocol, we did not specify whether we would meta-analyze together trials that used different usual care or active intervention controls. Indeed, for the earlier version of this review, trials classified as usual care were meta-analyzed together, even though the types of usual care were different. This was because for that version there were only two included trials that used a usual care control (i.e. supervised osteoarthritis education in [Berman 2004](#) trial and physician consultations and adjuvant physiotherapy in [Scharf 2006](#) trial), and each had a similar effect size relative to acupuncture. However, with the new trials included in this update, there are several different types of usual care controls (which we now label as other active treatments) and they have different effect sizes relative to acupuncture, so it does not seem as reasonable to lump them all together in the same analysis. Setting up the structure of the review in this way (i.e. with each active treatment control analyzed separately) will also accommodate future review updates for which accumulating data may allow us to draw conclusions about the relative effects of acupuncture compared to different active treatment comparators. Lastly, there was only one included trial ([Foster 2007](#)) that evaluated acupuncture strictly as an adjuvant, in this case, as an adjuvant to advice and exercise. Rather than label this trials comparison group as no acupuncture for the forest plots, we instead labeled it as 'Supervised exercise plus home exercises/advice leaflet alone (no adjuvant acupuncture)'. Finally, in the protocol we did not specify that the 'acupuncture versus a sham intervention' would be considered the primary analysis. However, based on the request of the Cochrane Musculoskeletal Group statistical reviewer, in the review, we now refer to the acupuncture versus sham comparison as the primary analysis.

- As an additional method of quality assessment not included in the protocol, we included in this review the Cochrane Musculoskeletal Group (MSG) grading system for ranking the strength and quality of the evidence. The MSG grading system was added because this grading system is required for all reviews submitted to the MSG.
- As a final point, there was also a protocol for this title published in *The Cochrane Library* in 1998, which had been peer-reviewed at that time by the Cochrane Musculoskeletal Review Group. However, that protocol had not been modified since the date of its original publication, and was not consulted in the preparation of the earlier review ([Manheimer 2007](#)), nor for this update. There is no similarity

between any of the methods described in that protocol and the methods used in preparing the earlier review or this update. The citation information for this 1998 published protocol is copied below:

"\*Ezzo J, Hadhazy V, Berman B, Birch S, Kaplan G, Hochberg M. Acupuncture for osteoarthritis. (Protocol) *Cochrane Database of Systematic Reviews* 1998, Issue 4. Art. No.: CD001977. DOI: 10.1002/14651858.CD001977. This version first published online: 26 October 1998 in Issue 4, 1998. Date of most recent substantive amendment: 14 August 1998."

## INDEX TERMS

### Medical Subject Headings (MeSH)

Acupuncture Therapy [\*methods]; Arthralgia [therapy]; Osteoarthritis, Hip [\*therapy]; Osteoarthritis, Knee [\*therapy]; Randomized Controlled Trials as Topic; Recovery of Function

### MeSH check words

Humans