Cochrane Library

Cochrane Database of Systematic Reviews

Acupuncture for the prevention of episodic migraine

Cochrane Systematic Review - **Intervention** │ Version published: 28 June 2016 **see what's new**

Am score ‹ 334   **View article information**

✉ Klaus Linde │ Gianni Allais │ Benno Brinkhaus │ Yutong Fei │ Michael Mehring │ Emily A. Vertosick │ Andrew Vickers │ Adrian R White
View authors' declarations of interest

Abstract *available in* English │ Français │ 한국어

Background

Acupuncture is often used for migraine prevention but its effectiveness is still controversial. We present an update of our Cochrane review from 2009.

Objectives

To investigate whether acupuncture is a) more effective than no prophylactic treatment/routine care only; b) more effective than sham (placebo) acupuncture; and c) as effective as prophylactic treatment with drugs in reducing headache frequency in adults with episodic migraine.

Search methods

We searched the Cochrane Central Register of Controlled Trials (CENTRAL: 2016, issue 1); MEDLINE (via Ovid, 2008 to January 2016); Ovid EMBASE (2008 to January 2016); and Ovid AMED (1985 to January 2016). We checked PubMed for recent publications to April 2016. We searched the World Health

Organization (WHO) Clinical Trials Registry Platform to February 2016 for ongoing and unpublished trials.

Selection criteria

We included randomized trials at least eight weeks in duration that compared an acupuncture intervention with a no-acupuncture control (no prophylactic treatment or routine care only), a sham-acupuncture intervention, or prophylactic drug in participants with episodic migraine.

Data collection and analysis

Two reviewers checked eligibility; extracted information on participants, interventions, methods and results, and assessed risk of bias and quality of the acupuncture intervention. The primary outcome was migraine frequency (preferably migraine days, attacks or headache days if migraine days not measured/reported) after treatment and at follow-up. The secondary outcome was response (at least 50% frequency reduction). Safety outcomes were number of participants dropping out due to adverse effects and number of participants reporting at least one adverse effect. We calculated pooled effect size estimates using a fixed-effect model. We assessed the evidence using GRADE and created 'Summary of findings' tables.

Main results

Twenty-two trials including 4985 participants in total (median 71, range 30 to 1715) met our updated selection criteria. We excluded five previously included trials from this update because they included people who had had migraine for less than 12 months, and included five new trials. Five trials had a no-acupuncture control group (either treatment of attacks only or non-regulated routine care), 15 a sham-acupuncture control group, and five a comparator group receiving prophylactic drug treatment. In comparisons with no-acupuncture control groups and groups receiving prophylactic drug treatment, there was risk of performance and detection bias as blinding was not possible. Overall the quality of the evidence was moderate.

Comparison with no acupuncture

Acupuncture was associated with a moderate reduction of headache frequency over no acupuncture after treatment (four trials, 2199 participants; standardised mean difference (SMD) -0.56; 95% CI -0.65 to -0.48); findings were statistically heterogeneous (I^2 = 57%; moderate quality evidence). After treatment headache frequency at least halved in 41% of participants receiving acupuncture and 17% receiving no acupuncture (pooled risk ratio (RR) 2.40; 95% CI 2.08 to 2.76; 4 studies, 2519 participants) with a corresponding number needed to treat for an additional beneficial outcome (NNTB) of 4 (95% CI 3 to 6); there was no indication of statistical heterogeneity (I^2 = 7%; moderate quality evidence). The only trial

with post-treatment follow-up found a small but significant benefit 12 months after randomisation (RR 2.16; 95% CI 1.35 to 3.45; NNT 7; 95% 4 to 25; 377 participants, low quality evidence).

Comparison with sham acupuncture

Both after treatment (12 trials, 1646 participants) and at follow-up (10 trials, 1534 participants), acupuncture was associated with a small but statistically significant frequency reduction over sham (moderate quality evidence). The SMD was -0.18 (95% CI -0.28 to -0.08; I^2 = 47%) after treatment and -0.19 (95% CI -0.30 to -0.09; I^2 = 59%) at follow-up. After treatment headache frequency at least halved in 50% of participants receiving true acupuncture and 41% receiving sham acupuncture (pooled RR 1.23, 95% CI 1.11 to 1.36; I^2 = 48%; 14 trials, 1825 participants) and at follow-up in 53% and 42%, respectively (pooled RR 1.25, 95% CI 1.13 to 1.39; I^2 = 61%; 11 trials, 1683 participants; moderate quality evidence). The corresponding NNTBs are 11 (95% CI 7.00 to 20.00) and 10 (95% CI 6.00 to 18.00), respectively. The number of participants dropping out due to adverse effects (odds ratio (OR) 2.84; 95% CI 0.43 to 18.71; 7 trials, 931 participants; low quality evidence) and the number of participants reporting adverse effects (OR 1.15; 95% CI 0.85 to 1.56; 4 trials, 1414 participants; moderate quality evidence) did not differ significantly between acupuncture and sham groups.

Comparison with prophylactic drug treatment

Acupuncture reduced migraine frequency significantly more than drug prophylaxis after treatment (SMD -0.25; 95% CI -0.39 to -0.10; 3 trials, 739 participants), but the significance was not maintained at follow-up (SMD -0.13; 95% CI -0.28 to 0.01; 3 trials, 744 participants; moderate quality evidence). After three months headache frequency at least halved in 57% of participants receiving acupuncture and 46%

receiving prophylactic drugs (pooled RR 1.24; 95% CI 1.08 to 1.44) and after six months in 59% and 54%, respectively (pooled RR 1.11; 95% CI 0.97 to 1.26; moderate quality evidence). Findings were consistent among trials with I^2 being 0% in all analyses. Trial participants receiving acupuncture were less likely to drop out due to adverse effects (OR 0.27; 95% CI 0.08 to 0.86; 4 trials, 451 participants) and to report adverse effects (OR 0.25; 95% CI 0.10 to 0.62; 5 trials 931 participants) than participants receiving prophylactic drugs (moderate quality evidence).

Authors' conclusions

The available evidence suggests that adding acupuncture to symptomatic treatment of attacks reduces the frequency of headaches. Contrary to the previous findings, the updated evidence also suggests that there is an effect over sham, but this effect is small. The available trials also suggest that acupuncture may be at least similarly effective as treatment with prophylactic drugs. Acupuncture can be considered a

treatment option for patients willing to undergo this treatment. As for other migraine treatments, long-term studies, more than one year in duration, are lacking.

Plain language summary

Acupuncture for preventing migraine attacks

Bottom line

The available evidence suggests that a course of acupuncture consisting of at least six treatment sessions can be a valuable option for people with migraine.

Background

Individuals with migraine have repeated attacks of severe headache, usually just on one side and often with vomiting. Acupuncture is a therapy in which thin needles are inserted into the skin at particular points. It originated in China, and is now used in many countries to treat people with migraine. We evaluated whether acupuncture reduces the number of episodes of migraine. We looked at the number of people in whom the number of migraine days per month was reduced by half or more than half.

Key results

For this update, we reviewed 22 trials with 4985 people, published up to January 2016. We omitted five trials from the original review because they included people who had had migraine for less than 12 months. We included five new trials in this update.

In four trials, acupuncture added to usual care or treatment of migraine on onset only (usually with pain-killers) resulted in 41 in 100 people having the frequency of headaches at least halved, compared to 17 of 100 people given usual care only.

In 15 trials, acupuncture was compared with 'fake' acupuncture, where needles are inserted at incorrect points or do not penetrate the skin. The frequency of headaches halved in 50 of 100 people receiving true acupuncture, compared with 41 of 100 people receiving 'fake' acupuncture. The results were dominated by three good quality large trials (with about 1200 people) showing that the effect of true acupuncture was still present after six months. There were no differences in the number of side effects of real and 'fake' acupuncture, or the numbers dropping out because of side effects.

In five trials, acupuncture was compared to a drug proven to reduce the frequency of migraine attacks, but only three trials provided useful information. At three months, headache frequency halved in 57 of 100 people receiving acupuncture, compared with 46 of 100 people taking the drug. After six months, headache frequency halved in 59 of 100 people receiving acupuncture, compared with 54 of 100 people taking the drug. People receiving acupuncture reported side effects less often than people receiving drugs, and were less likely to drop out of the trial.

Our findings about the number of days with migraine per month can be summarized as follows. If people have six days with migraine per month on average before starting treatment, this would be reduced to five days in people receiving only usual care, to four days in those receiving fake acupuncture or a prophylactic drug, and to three and a half days in those receiving true acupuncture.

Quality of the evidence

Overall the quality of the evidence was moderate.

Authors' conclusions

Implications for practice

Acupuncture seems to be effective for migraine prophylaxis. The effects over sham acupuncture found in this review were small, but there were clinically relevant effects over no acupuncture/no prophylactic treatment, and acupuncture compared well with prophylactic drugs regarding effectiveness and side effects. As the findings of our main analysis on headache frequency use standardized mean differences as an effect measure they are somewhat difficult to interpret clinically. In terms of number of migraine days, our findings approximately indicate the following: assuming a frequency of six migraine days per month at baseline, this would be reduced to five days in the no-treatment control group, to four in the sham group and the prophylactic drug group, and to three and a half in the acupuncture group. Acupuncture can be considered as a treatment option for people with migraine needing prophylactic treatment because of frequent or inadequately controlled migraine attacks, particularly people refusing prophylactic drug treatment or experiencing adverse effects from such treatment.

Implications for research

As migraine is a chronic condition, it would be important for clinicians to know how long improvements associated with acupuncture treatment last, whether continued intermittent treatment sustains the effect, and whether a further treatment cycle again leads to improvement. These latter questions might be best investigated in cohort studies. In principle, it seems important to know which types of acupuncture work best, what is the optimal frequency and duration of sessions, and so on. Some studies have not shown important differences in the effects of different acupuncture techniques (Jena 2008; Weidenhammer 2006), but this review found an influence of number of treatment sessions, in line with other evidence on dose (number of needles, number of sessions) of treatment (MacPherson 2013); these issues could also be investigated in observational studies. For decision-makers it would be important to know who is sufficiently qualified to deliver acupuncture. Randomized trials comparing outcomes after treatment by different types of practitioner are desirable, although very large sample sizes would be needed. Such studies would also be interesting from a more scientific perspective because it is unclear to what extent the effects of acupuncture are mainly mediated by context variables and generalized (i.e. not specific to traditional points) needling effects, and what contribution correct point location makes. Although further sham-controlled trials are desirable, we think that such studies should not have the highest priority unless they also address other important questions. Further comparisons with prophylactic drug treatment and other non-pharmacological interventions are needed. To facilitate future meta-analyses, it would be helpful if some standards for reporting outcome data were established.

Summary of findings

Open in table viewer

Summary of findings for the main comparison. Acupuncture compared to no treatment/usual care

Acupuncture compared to no treatment/usual care

Patient or population: people with episodic migraine
Setting: primary care or outpatient care
Intervention: acupuncture
Comparison: no treatment/usual care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect	№ of participants	Quality of the

	Risk with no treatment/usual care	Risk with Acupuncture	(95% CI)	(studies)	evidence (GRADE)	
Headache frequency (after treatment) assessed with days per month follow-up: median 3 months	Headache frequency was 0.56 SDs (-0.65 to -0.48) lower than in the groups receiving no/usual treatment		-	2199 (4 RCTs)	⊕⊕⊕⊖ MODERATE [1][2]	As a rule of thumb 0.2 SD represents a small difference, 0.5 a moderate, and 0.8 a large difference. Size of difference open to change with more trials
Headache frequency (follow-up) assessed with days per month follow-up: 12 months	Headache frequency was 0.36 SDs (-0.59 to -0.12) lower than in the groups receiving no/usual treatment		-	284 (1 RCT)	⊕⊕⊖⊖ LOW [2][3]	Only single large trial available. As a rule of thumb 0.2 SD represents a small difference, 0.5 a moderate, and 0.8 a large difference. Size of difference open to change with more trials.
Response (after treatment) assessed with proportion of participants with at least 50% headache frequency reduction follow-up: median 3 months	Study population		RR 2.40 (2.08 to 2.76)	2519 (4 RCTs)	⊕⊕⊕⊖ MODERATE [2]	No blinding, variable care in control groups, variable size of effects, but moderate to large effects in all three larger trials
	171 per 1000	410 per 1000 (355 to 472)				
Response (follow-up) assessed with proportion of participants with at least 50% headache frequency reduction follow-up: 12 months	Study population		RR 2.16 (1.35 to 3.45)	377 (1 RCT)	⊕⊕⊖⊖ LOW [2][3]	Only single large trial available
	98 per 1000	212 per 1000 (133 to 339)				

*The risk in the intervention group** (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; OR: Odds ratio

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] Interventions in control groups and study findings variable (I^2 = 57%; Chi^2 = 6.96, P value = 0.07), but effects moderate to large in all three larger trials

[2] Downgraded once: no blinding

[3] Downgraded once: only one study

Summary of findings 2. Acupuncture compared to sham interventions

Open in table viewer

Acupuncture compared to sham acupuncture

Patient or population: people with episodic migraine
Setting: primary care or outpatient care
Intervention: acupuncture
Comparison: sham acupuncture

Outcomes	Anticipated absolute effects[*] (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with sham acupuncture	Risk with Acupuncture				
Headache frequency (after treatment) assessed with days per month follow-up: median 12	Headache frequency was 0.18 SDs (-0.28 to -0.08) lower than in the groups receiving sham treatment		-	1646 (12 RCTs)	⊕⊕⊕⊖ MODERATE [1]	As a rule of thumb 0.2 SD represents a small difference, 0.5 a

Outcome			Relative effect (95% CI)	№ of participants (studies)	Certainty	Comments
weeks						moderate, and 0.8 a large difference
Headache frequency (follow-up) assessed with days per month follow-up: median 6 months	Assuming a mean number of 3.5 (SD 3.0) migraine days in the sham group, participants in the acupuncture group would have 0.6 days (95% CI 0.3 to 1.1 days) less (SMD = -0.19; 95% CI -0.30 to -0.09; 896 patients receiving acupuncture, 638 sham)		-	1534 (10 RCTs)	⊕⊕⊕⊖ MODERATE [2]	As a rule of thumb 0.2 SD represents a small difference, 0.5 a moderate, and 0.8 a large difference
Response (after treatment) assessed with proportion of participants with at least 50% headache frequency reduction follow-up: median 12 weeks	Study population		RR 1.23 (1.11 to 1.36)	1825 (14 RCTs)	⊕⊕⊕⊖ MODERATE [3]	Variable results between studies; modest effect size leaves magnitude of effect open to change with further large trials
	408 per 1000	502 per 1000 (453 to 555)				
Response (follow-up) assessed with proportion of participants with at least 50% headache frequency reduction follow-up: median 6 months	Study population		RR 1.25 (1.13 to 1.39)	1683 (11 RCTs)	⊕⊕⊕⊖ MODERATE [4]	Variable results between studies; modest effect size leaves magnitude of effect open to change with
	423 per 1000	529 per 1000 (479 to 589)				

					further large trials
Number of participants dropping out due to adverse effects	Study population		RR 2.84 (0.43 to 18.71)	931 (7 RCTs)	⊕⊕⊖⊖ LOW [5]	Relevant uncertainty due to low event rates
	Only 3/621 participants receiving acupuncture and 0/310 receiving sham dropped out due to adverse effects					
Number of participants reporting adverse effects	Study population		RR 1.15 (0.85 to 1.56)	1414 (4 RCTs)	⊕⊕⊕⊕ HIGH	Only 4 large trials report this outcome adequately; variable methods to document adverse effects, yet results of trials are consistent
	173 per 1000	199 per 1000 (147 to 270)				

***The risk in the intervention group** (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; **OR:** Odds ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] Downgraded once: pronounced heterogeneity of study results (I^2 = 47%; Chi^2 = 20.69; P value = 0.04)

[2] Downgraded once: pronounced heterogeneity of study results (I^2 = 59%; Chi^2 = 27.10; P value = 0.0003)

[3] Downgraded once: pronounced heterogeneity of study results (I^2 = 48%; Chi^2 = 25.09; P value = 0.02)

[4] Downgraded once: pronounced heterogeneity of study results (I^2 = 61%; Chi^2 = 25.50; P value = 0.004)

[5] Downgraded twice: only very few events

Summary of findings 3. Acupuncture compared to prophylactic drugs

Acupuncture compared to prophylactic drugs

Patient or population: people with episodic migraine
Setting: primary care or outpatient care
Intervention: acupuncture
Comparison: prophylactic drug treatment

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with prophylactic drug treatment	**Risk with acupuncture**				
Headache frequency assessed with days per month follow-up median 3 months	Headache frequency was 0.25 SDs (-039 to -0.10) lower than in the groups receiving prophylactic drug treatment		-	739 (3 RCTs)	⊕⊕⊕⊖ MODERATE [1]	As a rule of thumb 0.2 SD represents a small difference, 0.5 a moderate, and 0.8 a large difference. Size of difference open to change with more trials
Headache frequency assessed with days per month follow-up: median 6 months	Headache frequency was 0.13 SDs (-0.28 to 0.01) lower than in the groups receiving prophylactic drug treatment		-	744 (3 RCTs)	⊕⊕⊕⊖ MODERATE [1]	As a rule of thumb 0.2 SD represents a small difference, 0.5 a moderate, and 0.8 a large difference. Size of difference open to change with more trials
Response assessed with proportion of participants with at least 50% headache frequency reduction follow-up: median 3	Study population		RR 1.24 (1.08 to 1.44)	743 (3 RCTs)	⊕⊕⊕⊖ MODERATE [1]	Due to the limited number of trials and risk of bias size of differences open to change with more trials
	461 per 1000	572 per 1000 (498 to 664)				

	Study population		Relative effect	№ of participants	Certainty	Comments
Response assessed with proportion of participants with at least 50% headache frequency reduction follow-up: median 6 months	Study population		RR 1.11 (0.97 to 1.26)	744 (3 RCTs)	⊕⊕⊕⊖ MODERATE [1]	Due to the limited number of trials and risk of bias size of differences open to change with more trials
	536 per 1000	595 per 1000 (520 to 675)				
Number of participants dropping out due to adverse effects	Study population		OR 0.27 (0.08 to 0.86)	451 (4 RCTs)	⊕⊕⊕⊖ MODERATE [2]	Consistent results between studies, but uncertainty about size of difference due to low frequency of events in acupuncture group
	71 per 1000	20 per 1000 (6 to 62)				
Number of participants reporting adverse effects	Study population		OR 0.25 (0.10 to 0.62)	931 (5 RCTs)	⊕⊕⊕⊖ MODERATE [3]	Consistently fewer adverse effects in acupuncture groups, but strong variability of size of differences (probably due to different assessment methods)
	341 per 1000	114 per 1000 (49 to 243)				

*__The risk in the intervention group__ (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; __OR:__ Odds ratio;

__GRADE Working Group grades of evidence__

__High quality:__ We are very confident that the true effect lies close to that of the estimate of the effect

__Moderate quality:__ We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

---

[1] Downgraded once: in two of three studies a relevant proportion of participants randomized to drug treatment dropped out early (analysis included only participants receiving at least a minimal amount of treatment); no blinding of participants

[2] Downgraded once: few events in acupuncture group; wide confidence interval

[3] Downgraded once: size of differences highly variable ($I^2 = 78\%$; $Chi^2 = 17.95$, P value = 0.001), but consistently more adverse effects in drug groups

---

Background

This review is an update of a previously published review in The Cochrane Database of Systematic Reviews [Issue 1, 2009] on acupuncture for migraine (Linde 2009).

Description of the condition

Migraine is a disorder with recurrent headaches manifesting in attacks lasting from four to 72 hours. Typical characteristics of the headache are unilateral location, pulsating quality, moderate or severe intensity, aggravation by routine physical activity and association with nausea, photophobia or phonophobia, or any combination of all three (IHS 2013). Epidemiological studies have consistently shown that migraine is a common disorder with a one-year prevalence of around 10% to 12% and a lifetime prevalence of between 15% and 20% (Oleson 2007). In Europe, the economic cost of migraine is estimated at EUR 27 billion per year (Andlin-Sobocki 2005). Migraine is subclassified into the more frequent episodic migraine (fewer than 15 days with migrainous headaches per month) and the less frequent chronic migraine (more than 15 days per month). Most people with migraine can be adequately managed by treatingof acute headaches alone, but a relevant minority need prophylactic interventions, as their attacks are either very frequent or are insufficiently controlled by acute therapy. Several drugs, such as propranolol, metoprolol, flunarizine, valproic acid and topiramate, have been shown to reduce attack frequency in some people (Dodick 2007; Linde M 2013a; Linde M 2013b), however, all these drugs are associated with adverse effects. Dropout rates in most clinical trials are high, suggesting that the drugs are not well accepted by patients. There is some evidence that behavioural interventions such as relaxation or biofeedback are beneficial (Holroyd 1990; Nestoriuc 2007), but additional effective, low-risk

relaxation or biofeedback are beneficial (Holroyd 1990; Nestoriuc 2007), but additional effective, low-risk treatments are clearly desirable.

Description of the intervention

Acupuncture in the context of this review is defined as the needling of specific points of the body. It is one of the most widely used complementary therapies in many countries (Bodeker 2005). For example, according to a population-based survey in 2002 in the United States of America (USA), 4.1% of respondents reported lifetime use of acupuncture, and 1.1% reported recent use (Burke 2006). A similar survey in Germany performed in the same year found that 8.7% of adults between 18 and 69 years of age had received acupuncture treatment in the previous 12 months (Härtel 2004). Acupuncture was originally developed as part of Chinese medicine wherein the purpose of treatment was to bring the patient back to the state of equilibrium postulated to exist prior to illness (Endres 2007). Some acupuncture practitioners have dispensed with these concepts and understand acupuncture in terms of conventional neurophysiology. Acupuncture is often used to treat headache, especially migraine. For example, 9.9% of the acupuncture users in the US survey mentioned above stated that they had been treated for migraine or other headaches (Burke 2006).

How the intervention might work

Many studies have shown that acupuncture has short-term effects on a variety of physiological variables relevant to analgesia (Bäcker 2004; Endres 2007). However, it is unclear to what extent these observations from experimental settings are relevant to the long-term effects reported by practitioners. It is assumed that a variable combination of local effects; spinal and supraspinal mechanisms; and cortical,

psychological or 'placebo' mechanisms contribute to the clinical effects in routine care (Carlsson 2002). While there is little doubt that acupuncture interventions cause neurophysiological changes in the organism, the traditional concepts of acupuncture involving specifically located points on a system of 'channels' called meridians are controversial (Kaptchuk 2002). As for many non-pharmacological interventions, it is difficult to create sham interventions for acupuncture which are both indistinguishable and physiologically inert. This is due both to technical reasons and the unclear mechanism of action. Consequently, trials using sham acupuncture controls must be interpreted carefully, as sham treatments might not be inactive placebos, while trials comparing acupuncture with no prophylactic treatment, prophylactic drugs or other interventions must also be interpreted carefully, as they have a higher risk of bias due to lack of blinding.

Why it is important to do this review

Despite acupuncture's widespread use its effectiveness is still discussed controversially (Da Silva 2015, McGeeney 2015). Since the publication of the previous version of our Cochrane review (Linde 2009) a number of new trials have been published. Therefore, an update of the review was necessary. To sharpen the focus of our review we narrowed our selection criteria. In particular, we now focus on episodic migraine.

Objectives

To investigate whether acupuncture is a) more effective than no prophylactic treatment/routine care only; b) more effective than 'sham' (placebo) acupuncture; and c) as effective as prophylactic treatment with drugs in reducing headache frequency in patients with episodic migraine.

Methods

Criteria for considering studies for this review

Types of studies

We included controlled trials investigating the prophylactic effect of acupuncture in which allocation to treatment was explicitly randomized, and in which participants were followed up for at least eight weeks after randomisation. We excluded trials in which a clearly inappropriate method of randomisation was used, for example, open alternation.

Types of participants

We included trials in which study participants had been diagnosed with episodic migraine (the word episodic did not have to be mentioned in the report explicitly; see exclusion criteria below to exclude trials focusing on chronic migraine). Studies focusing on migraine but including participants with additional tension-type headache were included. We included studies including participants with headaches of various types (for example, some participants with migraine, some with tension-type headache) only if findings for participants with migraine were available separately, or if more than 90% of

participants suffered from migraine.

The duration of the condition had to be longer than one year in the great majority (more than 80%) of participants. This criterion was considered met if:

- duration for longer than year was an inclusion criterion; or

- the mean duration minus one standard deviation was more than one year; or

- the mean duration (standard deviation not reported) was more than 10 years; or

- other information was presented that made it highly likely that the criterion was met (e.g. study authors presented proportions with duration ranges).

We excluded trials in patients with chronic migraine, chronic daily headache or in which at baseline more than half of participants had more than 15 days with migrainous headache per month. We also excluded trials in which there was no information of the duration of headache complaints.

Changes to previous version

In this update of the review we have excluded trials focusing on chronic migraine, as the definition of chronic migraine is still debated and the separation from other diagnoses, for example headache due to medication overuse, is difficult (in the previous version of this review (Linde 2009) we were not aware of any trials on chronic migraine and they were not explicitly excluded). In the current update we have also excluded trials in which a relevant proportion of participants had been suffering from migraine for less than one year or in which duration was unclear.

Types of interventions

Experimental interventions

- Any treatment involving needle insertion (with or without manual or electrical stimulation) at acupuncture points, pain points or trigger points, described as acupuncture. The planned treatment course must have had at least six treatment sessions, and been given at least once per week. Trials with individualised strategies were included if the median or mean number of treatments was at least six sessions, and there was no reason to believe that treatments were given less frequently than once per week in the majority of participants.

- We excluded studies that:

  - exclusively investigated acupuncture at specific 'micro-systems' (e.g. scalp or ear acupuncture), although we included trials using micro-system points in addition to body acupuncture;

  - investigated other methods of stimulating acupuncture points without needle insertion, for

example, acupressure, laser stimulation or transcutaneous electrical stimulation;

- injected fluids at acupuncture or trigger points.

Control interventions

- No treatment other than treatment of acute migraine attacks or routine care (which typically includes treatment of acute attacks, but might also include other treatments; however, trials normally require that no new experimental or standardized treatment be initiated during the trial period).

- Sham interventions (interventions mimicking 'true' acupuncture/true treatment, but deviating in at least one aspect considered important by acupuncture theory, such as skin penetration or correct point location).

- Prophylactic pharmacological treatment (for example, β-blocking agents, calcium channel antagonists, anti-epileptic drugs) given for at least eight weeks.

- We excluded trials comparing acupuncture to food supplements, herbal drugs or combinations of herbal drugs, and trials that only compared different forms of acupuncture.

Changes to previous version

In the previous version of the review (Linde 2009) we included trials using any prophylactic treatment other than acupuncture as comparison. With a slowly increasing number of trials using a wide range of different treatments (mainly various herbal medicines) we decided to concentrate on conventional

prophylactic pharmacological treatment to keep the review focused. We have defined a minimum number and frequency of acupuncture treatment sessions to ensure that treatments meet basic quality criteria.

Types of outcome measures

We included studies if they measured at least one of the following outcome measures for at least eight weeks after randomisation:

- headache frequency (attacks, days, hours, headache-free days) per defined time period;

- response (≥ 50% frequency reduction documented in a headache diary);

- disability or quality of life with a validated measure.

We excluded trials that:

- focussed on the treatment and measurement of acute attacks;

- reported only measures such as "total effectiveness rate" (e.g. proportion of participants healed, much improved, improved, unchanged);

- reported only physiological or laboratory parameters;

- had outcome measurement periods of less than eight weeks (from randomisation to final observation).

Changes to previous version

We have defined outcome measures more precisely to ensure that measurement methods meet current standards of migraine research.

Primary outcomes

The primary efficacy outcome of our systematic review was headache frequency at completion of treatment and at follow-up. The primary safety/acceptability outcomes were the number of participants dropping out due to adverse effects and the number of participants reporting at least one adverse event or effect (see Measures of treatment effect for details).

Secondary outcomes

The secondary efficacy outcome of our systematic review was the proportion of 'responders' at completion of treatment and at follow-up.

Search methods for identification of studies

Electronic searches

For this update we searched the following databases without language restrictions (date of the last search 20 January 2016):

- Cochrane Central Register of Controlled Trials (CENTRAL; *The Cochrane Library* 2016, Issue 1), searched from 2008 to 2016;

- MEDLINE (via Ovid) 2008 to week 1 of January 2016;

- EMBASE (via Ovid) 2008 to 19 January 2016;

- AMED (via OVID) 1985 to January 2016.

The search strategies are reported in Appendix 1. In addition, the first author checked PubMed monthly for new publications, screening all hits for 'acupuncture AND (headache OR migraine)' (last search 12 April 2016). For previous versions of this review (Melchart 2001; Linde 2009) we had searched the Cochrane Complementary Medicine Field Trials Register (whose results are now included in CENTRAL without relevant delay) and the Cochrane Pain, Palliative & Supportive Care Trials Register (no longer updated).

Searching other resources

We searched the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP; apps.who.int/trialsearch/) and ClinicalTrials.gov (ClinicalTrials.gov) for completed or ongoing trials using the search string 'acupuncture AND (headache OR migraine)'. The last search was on February 10, 2016. We also searched the reference lists of all eligible studies for additional studies.

Data collection and analysis

Selection of studies

Two review authors screened all abstracts identified by the updated search and excluded those that were clearly irrelevant (for example, studies focusing on other conditions, reviews, etc.). We obtained full texts of all remaining references and, again, screened them to exclude clearly irrelevant papers. At least two review authors formally checked all remaining articles and all trials included in the previous version of our review (Linde 2009) for eligibility according to the above-mentioned selection criteria. We resolved any disagreements by discussion.

Data extraction and management

At least two review authors independently extracted information on participants, methods, interventions, outcomes and results using a specially designed form before entry into Review Manager (RevMan) (RevMan 2014). In particular, we extracted exact diagnoses; headache classifications used; number and type of centres; age; sex; duration of disease; number of participants randomized, treated and analysed; number of, and reasons for dropouts; duration of baseline, treatment and follow-up periods; details of acupuncture treatments (such as selection of points; number, frequency and duration of sessions; achievement of de-chi (an irradiating feeling considered to indicate effective needling); number, training and experience of acupuncturists); and details of control interventions (sham technique, type and dosage of drugs). For details regarding methodological issues and study results, see below.

Where necessary, we sought additional information from the first or corresponding authors of the

For six trials (Diener 2006; Jena 2008; Li 2012; Linde K 2005; Streng 2006; Vickers 2004) included in the individual patient database of the Acupuncture Trialists' Collaboration (ATC), an international collaborative network for high quality randomized trials of acupuncture for chronic pain (see Vickers 2010; Vickers 2012), we obtained uniformly re-analysed summary data for numeric variables and the number of responders for calculation of effect sizes. We used these data to ensure that we obtained the most precise estimate of treatment effect. For each trial, we created an analysis of covariance (ANCOVA) model for each numeric outcome at each time point and adjusted for the baseline value of that outcome, treatment group (acupuncture or control), and any variables that were used to stratify randomisation in the original trial. Using this model, we calculated the adjusted mean outcome values for each group (acupuncture and control), and we used the standard error for the effect of treatment from the ANCOVA model to calculate the standard deviation for the difference in adjusted means. Therefore, effect sizes calculated in our analyses might to some degree deviate from those in the original publications of the six trials. Use of raw data also allowed us to calculate response rates, such as for a 50% reduction in pain, even if this was not reported in the original trial publication.

Assessment of risk of bias in included studies

For the assessment of study quality, the risk of bias approach for Cochrane reviews was used (Higgins 2011). We used the following six separate criteria:

- adequate sequence generation;

- allocation concealment;

- blinding;

- incomplete outcome data addressed (up to three months after randomisation);

- incomplete follow-up outcome data addressed (four to 12 months after randomisation);

- free of selective reporting.

We did not include the item 'other potential threats to validity' in a formal manner, but noted if relevant flaws were detected.

In a first step, we copied information relevant to making a judgment on a criterion from the original publication into an assessment table. We entered any additional information from the study authors into the table, if it was available, , along with an indication that this was unpublished information. At least two reviewers independently made a judgment on whether the risk of bias for each criterion was considered low, high or unclear. We resolved any disagreements by discussion.

For the first five criteria (above), we followed the recommendations of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). For 'selective reporting', we decided to use a more liberal definition. Headache trials typically measure a multiplicity of headache outcomes at several time points using diaries, and there is a plethora of slightly different outcome measurement methods. While a single primary endpoint is sometimes predefined, the overall pattern of a variety of outcomes is necessary to get a clinically interpretable picture. If we had applied the strict guidelines in the *Cochrane Handbook for Systematic Reviews of Interventions*, almost all trials would have been rated 'unclear' for 'selective reporting'. We considered trials as having a low risk of bias for selective reporting if they reported the results of the most relevant headache outcomes assessed (typically a frequency measure, intensity, analgesic use and response) for the most relevant time points (end of treatment and, if done, follow-up), and if the outcomes and time points reported made it unlikely that study investigators had picked them out because they were particularly favourable or unfavourable.

If trials had both blinded sham control groups and unblinded comparison groups receiving no prophylactic treatment or drug treatment, in the risk of bias tables, the 'Judgement' column always relates to the comparison with sham interventions. In the 'Description' column, we included the assessment for the other comparison group(s). As the risk of bias table does not include a 'not applicable' option, we rated the item 'incomplete follow-up outcome data addressed (four to 12 months after randomisation)?' as 'unclear' for trials that did not follow participants longer than three months.

We also assessed the adequacy of concealment of allocation according to the criteria of the ATC (Vickers 2010) which are stricter than those in the *Cochrane Handbook for Systematic Reviews of Interventions*. In particular, in the case of envelope randomisation, investigators must have established and described detailed procedures to ensure that allocation could neither be predicted nor changed post hoc. For example, there should have been procedures to prevent investigators resealing and reusing an envelope after it had been opened (e.g. envelopes were held by an independent party). As the level of information needed for this assessment was often not available in publications, we contacted study authors for clarification. If information was not available, we did not consider adequacy of concealment to be "unambiguously adequate".

Assessment of the adequacy of the acupuncture intervention

We also attempted to provide a crude estimate of the quality of acupuncture. At least two reviewers who are trained in acupuncture and have several years of practical experience (GA, BB, YF, AW) answered two questions. First, they were asked how they would treat the participants included in the study. Answer options were 'exactly or almost exactly the same way', 'similarly', 'differently', 'completely differently' or

could not assess' due to insufficient information (on acupuncture or on the participants). Second, they were asked to rate their degree of confidence that acupuncture was applied in an appropriate manner on a 100-mm visual scale (with 0% = complete absence of evidence that the acupuncture was appropriate, and 100% = total certainty that the acupuncture was appropriate). A member of the review team (AW) proposed the latter method , which was used in a systematic review of clinical trials of acupuncture for back pain (Ernst 1998). In the Characteristics of included studies table, the acupuncturists' assessments are summarized under 'Methods' (for example, "similarly/70%" indicates a trial where the acupuncturist-reviewer would treat 'similarly' and is '70%' confident that acupuncture was applied appropriately).

Measures of treatment effect

Main analysis

Our primary efficacy outcome was headache frequency at completion of treatment and at follow-up (closest to six months after randomisation). As studies may report either attacks, migraine days or headache days as a measure of headache frequency, we used a system where various frequency measures could be used. As available, we used (in descending order of preference) absolute values from four-week periods or other periods for (again, in descending order of preference) migraine days, migraine attacks or headache days. Due to the variability of outcomes, standardized mean differences (SMD) were calculated as effect size measures. Negative values indicate better outcomes in the acupuncture group.

Our secondary efficacy outcome was the proportion of 'responders' at completion of treatment and at follow-up (closest to six months after randomisation). Response was defined as a reduction in migraine days of at least 50% compared to baseline. If the number of responders regarding migraine days was not available we used at least 50% reduction in number of migraine attacks (second preference), or at least 50% reduction in number of headache days (third preference). We calculated risk ratios (RR) of having a response and 95% confidence intervals (CI) as effect size measures. Risk ratios greater than 1 indicate that there were more responders in the acupuncture group compared to the comparator group. When reporting results on response in this review (in the abstract, the plain language summary, the results section and the 'Summary of findings' tables) these are based on the observed proportion (sum of participants with response divided by the sum of participants randomized) in the control group and the expected proportion based on the pooled risk ratio from meta-analysis.

As primary safety/acceptability outcomes we used the number of participants dropping out due to adverse effects and the number of participants reporting at least one adverse event or effect. Further safety/acceptability outcomes were the number of participants not reaching the primary endpoint (we originally had planned to extract the number of participants dropping out but this proved difficult due to multiple measurement time points and reporting issues) and the number of participants with serious

adverse events. As the number of events was typically low we calculated odds ratios (OR) instead of risk ratios. Odds ratios greater than 1 indicate more events (e.g. dropouts) in the acupuncture group.

Time window analysis

In the previous version of this review (Linde 2009) we analysed findings according to the four time windows described below. This had the advantage that measurement times used were similar across trials. However, it had two disadvantages. Firstly, duration of treatment periods was quite variable, so while in some trials treatment was already completed (e.g. at 8 weeks) it was still ongoing (e.g. until week 16) in others; secondly, four time windows for each outcome made the 'Summary of findings' tables very complex. Therefore, in this update we have reported the time window analyses as additional analyses only.

We used the following time windows:

- up to eight weeks/two months after randomisation;

- three to four months after randomisation;

- five to six months after randomisation; and

- more than six months after randomisation.

If more than one data point was available for a given time window, we used: for the first time window, preferably data closest to eight weeks; for the second window, data closest to the four weeks after completion of treatment (for example, if treatment lasted eight weeks, data for weeks nine to 12); for the third window, data closest to six months; and for the fourth window, data closest to 12 months.

The following outcomes were used in the time windows analysis.

- Frequency of migraine attacks (means and standard deviations) per four-week period. Mean differences were calculated as effect size measures

- Response (risk ratio of having a response).

- Number of migraine days (means and standard deviations) per four-week period (mean differences).

- Number of headache days (means and standard deviations) per four-week period (mean differences).

- Headache intensity (any measures available, extraction of means and standard deviations, calculation of SMDs).

- Frequency of analgesic use (any continuous or rank measures available, extraction of means and standard deviations, calculation of SMDs).

- Headache scores (SMDs)

All these outcomes rely on participant reports, mainly collected in headache diaries.

Unit of analysis issues

The unit of analysis was the individual participant.

Dealing with missing data

If publications reported study findings with insufficient detail or in an inconsistent manner we attempted to obtain further information from the study authors.

Regarding missing participant data due to dropout or loss to follow-up in the included studies we used the following strategies.

- Efficacy outcomes:

  - for comparisons of acupuncture with no acupuncture and sham we used for continuous measures, if available, the data from intention-to-treat analyses with missing values replaced; otherwise, we used the presented data on available cases;

  - for response we used the number of responders divided by the number of participants randomized to the respective group (counting missing information as non-response). In studies comparing acupuncture with drug treatment, we used as first preference analyses of participants having at least started treatment as first preference, available cases as second preference and intention-to-treat analyses as third preference.

- Safety outcomes:

  - for all comparisons we used the number of participants randomized as denominator for the outcomes number of participants dropping out due to adverse effects, not reaching the primary endpoint and experiencing serious adverse events;

  - for the outcome number of participants reporting adverse effects we used the number of participants having received at least one treatment as denominator.

Assessment of heterogeneity

We assessed heterogeneity with the Chi² test (Deeks 2011) and the I² statistic (Higgins 2003).

Assessment of reporting biases

In forest plots studies are ordered according to their weight in meta-analysis. The weight depends on the standard errors of the point estimate (precision) which is dependent on sample size and variability/frequency of events. This gives readers a crude impression whether more and less precise trials yield similar findings.

Data synthesis

For the purposes of summarizing results, we categorized the included trials according to control groups:

- comparisons with no acupuncture (acute treatment only or routine care);

- comparisons with sham acupuncture interventions;

- comparisons with prophylactic drug treatment.

If a trial included more than one acupuncture group, we pooled results of the groups so that participants in the control group were counted more than once.

We calculated pooled fixed-effect estimates, their 95% confidence intervals, the Chi^2 test for heterogeneity and the I^2 statistic. If the P value of the Chi^2 test for heterogeneity was less than 0.2 or I^2 greater than 40%, or both, we also reported random-effects estimates.

Change to previous version

Based on the recommendation of the statistician in our team (AV) we have used fixed-effect models for calculating pooled estimates in this updated review. This is primarily because the fixed-effect analysis constitutes a valid test of the null hypothesis. Moreover, due to very large discrepancies in sample size, a

random-effects model would have resulted in participants in small studies being given greater weight that participants in large studies. . Nonetheless, if the P value of the Chi^2 test for heterogeneity was less than 0.2 or I^2 was greater than 40%, or both, we have also reported random-effect estimates.

We used the GRADE approach to assess the quality of evidence related to each of the key outcomes as appropriate (GRADEpro GDT 2015; Schünemann 2011). GRADE Working Group grades of evidence are:

- High quality: we are very confident that the true effect lies close to that of the estimate of the effect.

- Moderate quality: we are moderately confident in the effect estimate, the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

- Low quality: our confidence in the effect estimate is limited, the true effect may be substantially different from the estimate of the effect.

- Very low quality: we have very little confidence in the effect estimate, the true effect is likely to be

substantially different from the estimate of effect.

Subgroup analysis and investigation of heterogeneity

To investigate potential sources of heterogeneity and the robustness of our findings we performed subgroup analyses for the primary outcome, headache frequency, and for the secondary outcome, response, both after treatment and at follow-up for the comparison vs. sham (the number of trials being too small for the other two comparisons) for four variables. These variables were selected after reviewing the trials qualitatively but before running analyses: unambiguously adequate randomisation vs. other; larger (sample size above median of the trials included in the analysis) vs. smaller trials; number of treatment sessions up to 12 vs. 16 and more; and sham penetrating the skin vs. non-penetrating sham.

Results

Description of studies

Results of the search

Update searches identified 528 hits (518 by database searches, six by checking references and alerts, and four from checking entries in trials registries not identified otherwise). Thirty-six full-text publications and four entries in clinical trial registries that were deemed potentially eligible were formally checked against the eligibility criteria (see Figure 1). As we had modified selection criteria for this update, we reassessed the 22 trials included in the previous version for eligibility. Five new trials met the revised selection criteria (Facco 2013; Li 2012; Wallasch 2012; Wang 2015; Zhao 2014), while five trials included in the previous version or our review had to be excluded (Baust 1978; Doerr-Proske 1985; Dowson 1985; Henry 1985; Wylie 1997; see Characteristics of excluded studies).

Figure 1   Open in figure viewer | **Download as PowerPoint**
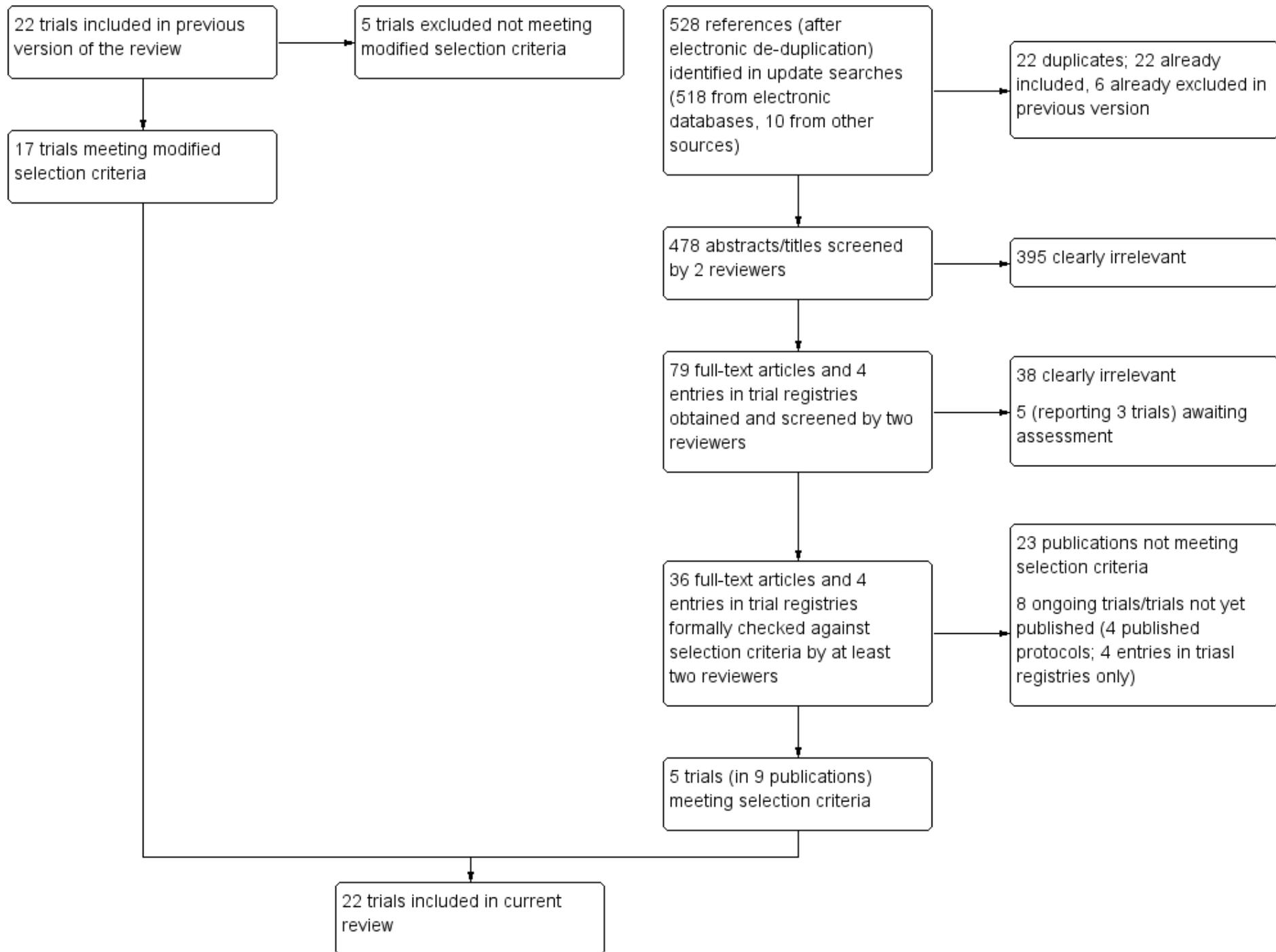
Flow diagram

Included studies

General characteristics

Twenty-two trials including 4985 participants in total (median 71, range 30 to 1715) met our selection criteria; 18 studies were two-armed, two were three-armed and a further two were four-armed (see Characteristics of included studies). All trials used parallel-group designs; there were no cross-over studies. Fifteen trials included a sham acupuncture control group (Alecrim 2005; Alecrim 2006; Alecrim 2008; Ceccherelli 1992; Diener 2006; Facco 2008; Li 2012; Linde K 2005; Linde M 2004; Vincent 1989; Wallasch 2012; Wang 2015; Weinschütz 1993; Weinschütz 1994; Zhao 2014), five a no-acupuncture control

group (Facco 2008; Jena 2008; Linde K 2005; Linde M 2000; Vickers 2004), and five a comparator group receiving prophylactic drug treatment (Allais 2002; Diener 2006; Facco 2013; Hesse 1994; Streng 2006). Sixteen trials were performed in a single centre and six were multicentre trials. Seven trials were performed in Germany, four in Italy, three in Brazil, two each in China, Sweden and the UK, and one each in Denmark and Australia. Seven trials were published between 1989 and 2002 and 15 between 2004 and 2015. We tried to contact corresponding authors of all trials at least once (either for previous versions of this review or for the current update). For one trial we could not obtain a valid contact address (Hesse 1994) and three study authors or co-authors did not provide additional information before completion of this update (Wallasch 2012; Weinschütz 1993; Weinschütz 1994). For the remaining 18 trials we obtained some additional information. Detailed additional data for effect size calculation was obtained from study authors or from the individual patient database of the ATC for 11 trials (Alecrim 2005; Alecrim 2006; Alecrim 2008; Allais 2002; Diener 2006; Jena 2008; Li 2012; Linde K 2005; Streng 2006; Vickers 2004; Vincent 1989).

Study participants

Fifteen trials included participants diagnosed as having migraine with or without aura, six exclusively participants without aura, and one recruited only women with menstrually-related migraine (Linde M 2004). In two large, pragmatic multicentre trials (Jena 2008; Vickers 2004) baseline headache frequency and the reported diagnoses make it likely that, in spite of the use of the criteria of the International Headache Society, there was some diagnostic misclassification (i.e. some participants were likely to suffer from tension-type headache and not migraine). This applied to a minor extent also to three other multicenter trials (Diener 2006; Linde K 2005; Streng 2006).

Acupuncture interventions

The acupuncture interventions tested in the included trials varied to a great extent. Five trials (Allais 2002; Ceccherelli 1992; Li 2012; Wallasch 2012; Zhao 2014) standardized acupuncture treatments (all participants were treated at the same points); seven (Alecrim 2006; Diener 2006; Facco 2013; Linde K 2005; Linde M 2000; Linde M 2004; Wang 2015) semi-standardized treatments (either all participants were treated at some basic points and additional individualized points, or there were different predefined needling schemes depending on symptom patterns); and 10 trials individualized the selection of acupuncture points (Alecrim 2005; Alecrim 2008; Facco 2008; Hesse 1994; Jena 2008; Streng 2006; Vickers 2004; Vincent 1989; Weinschütz 1993; Weinschütz 1994). The number of treatment sessions was between six and 12 in 13 trials, and 16 or more in nine trials. Most trials reporting the duration of sessions, left needles in place for between 20 and 30 minutes; one trial (Hesse 1994) investigated brief needling for a few seconds. Electro-stimulation of needles was used in one trial (Li 2012). Agreement among

acupuncturists on whether they would do acupuncture similarly to that used in the study assessed and whether they had confidence in the quality of the acupuncture was low (intra-class correlation coefficients -0.08 and 0.24). For two studies (Hesse 1994; Linde M 2004) both acupuncturists rating the study had 50% or less confidence that the acupuncture had adequate quality. For a further six studies (Ceccherelli 1992; Li 2012; Linde M 2000; Wallasch 2012; Weinschütz 1993; Weinschütz 1994) at least one acupuncturist gave a rating of 50% or lower. We could not assess four trials using individualized treatments not described in detail (Alecrim 2005; Alecrim 2008; Facco 2008; Jena 2008).

Comparator interventions

Five trials included a group which either received treatment of acute attacks only (Facco 2008; Linde K 2005; Linde M 2000) or 'routine care' that was not specified by protocol (Jena 2008; Vickers 2004), while the experimental group received acupuncture in addition. In the 15 trials with a sham control, techniques varied considerably. Four trials superficially needled recognized acupuncture points considered inadequate for the treatment of migraine (Alecrim 2005; Alecrim 2006; Alecrim 2008; Zhao 2014); seven trials used needling (mostly superficial) of non-acupuncture points at variable distance from true points (Diener 2006; Li 2012; Linde K 2005; Vincent 1989; Wallasch 2012; Weinschütz 1993; Weinschütz 1994). Two trials (Facco 2008; Linde M 2004) used 'placebo' needles (telescopic needles with blunt tips not penetrating the skin). In Linde M 2004 these were placed at the same predefined points as in the true treatment group. Facco 2008 had two sham groups: in one group the placebo needles were placed at correct, individualized points after the same process of Chinese diagnosis as in the true treatment group. In the second group placebo needles were placed at standardized points without the 'Chinese ritual' (to investigate whether the different interaction and process affected outcomes). One study (Ceccherelli 1992) used a complex procedure without real needling. One study used a mix of superficial needling at non-acupuncture points and a non-penetrating technique (with a blunted cocktail stick) for points on the head (Wang 2015). Five trials compared acupuncture to prophylactic drug treatment, using metoprolol (Hesse 1994; Streng 2006), flunarizine (Allais 2002), valproic acid (Facco 2013) or individualized treatment according to guidelines (Diener 2006). In four of these trials participants were unblinded, while one blinded trial used a double-dummy approach (true acupuncture + metoprolol placebo vs. metoprolol + sham acupuncture; Hesse 1994).

Excluded studies

Results were not yet available for eight studies registered in trial registries likely to meet selection criteria. For four of these, detailed protocols have been published (Chen 2013; Lan 2013; Vas 2008; Zhang 2013); for the other four only the registry entries were available (Li 2007; Liang 2013; Wang J 2015; Xing 2015).

For at least four trials recruitment has been completed (Lan 2013; Li 2007; Vas 2008; Wang J 2015) (see Characteristics of ongoing studies).

Twenty studies (described in 23 publications) did not meet selection criteria (Agro 2005; Boutouyrie 2010; Ceccherelli 2012; Deng 2006; Ferro 2012; Foroughipour 2014; Han 2011; Jia 2009; Matra 2012; Qin 2006; Vijayalakshmi 2014; Wang 2011; Wu 2011; Yang 2009; Yang 2011; Zhang 2006; Zhang 2009; Zheng 2013; Zhong 2009; Zhou 2007). A number of Chinese trials were excluded due to inadequate duration of prophylactic drug treatment (several Chinese trials gave flunarizine or other drugs for four weeks only), overall observation of less than eight weeks, inclusion of participants with recent onset of migraine, and lack of relevant outcome measures. Furthermore, five trials included in the previous version of our review were excluded (Baust 1978; Doerr-Proske 1985; Dowson 1985; Henry 1985; Wylie 1997). Reasons for exclusions are reported in the Characteristics of excluded studies.

Studies awaiting classification

We classified three trials (five publications) identified by our most recent update search as awaiting assessment (see Characteristics of studies awaiting classification). One (Giannini 2015) is an abstract of an interim analysis of a trial comparing acupuncture and individualized prophylactic drug treatment. The abstract does not provide sufficient information but based on background information available to one of us (KL) it seems likely that the trial will meet our eligibility criteria when a full publication with final data becomes available. A second trial originating from China (Li 2016) was published in February and April 2016 after all analyses for this review had been completed. The two publications focus on functional magnetic resonance imaging (fMRI) outcomes but also report headache frequency data for participants completing all fMRI measurements. It seems likely that these trials will meet inclusion criteria. A third trial of uncertain eligibility (participants with "menstrual headache") is available only in Chinese (Sun 2015). Full text translation has to be available before final assessment of eligibility.

Risk of bias in included studies

We discuss the methodological quality of trials (risk of bias) for the three comparisons separately, as problems differ according to control groups. The risk of bias assessments of single trials are displayed in Figure 2; a summary across trials is presented in Figure 3. It should be noted that three trials rated unclear for the item 'incomplete follow-up outcome data' actually did not include a follow-up (Ceccherelli 1992; Jena 2008; Zhao 2014).

Figure 2   **Open in figure viewer** | **Download as PowerPoint**

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding (performance bias and detection bias)	Incomplete outcome data (attrition bias)	Incomplete follow-up outcome data (attrition bias)	Selective reporting (reporting bias)
Alecrim 2005	?	?	?	?	?	?
Alecrim 2006	+	+	+	+	?	+
Alecrim 2008	+	+	+	+	+	+
Allais 2002	+	+	−	+	+	+
Ceccherelli 1992	+	+	+	?	?	+
Diener 2006	+	+	+	+	+	+
Facco 2008	+	+	+	−	−	+
Facco 2013	+	?	−	?	?	+
Hesse 1994	?	?	?	+	?	+
Jena 2008	+	+	−	+	?	+
Li 2012	+	+	+	+	+	+
Linde K 2005	+	+	+	+	+	+
Linde M 2000	+	−	−	−	−	+
Linde M 2004	+	?	+	+	+	+
Streng 2006	+	+	−	−	−	+
Vickers 2004	+	+	−	+	+	+
Vincent 1989	?	?	+	+	?	+
Wallasch 2012	?	?	+	−	−	?
Wang 2015	+	?	+	+	+	+

Risk of bias summary: review authors' judgements about each risk of bias item for each included study. Note: for trials including both a comparison with sham and a no-acupuncture control/prophylactic drugs (Diener 2006, Facco 2008, Linde K 2005) blinding was assessed for the comparisons with sham. For the comparisons with no acupuncture/prophylactic drugs the risk of bias is high (no blinding).

Figure 3   Open in figure viewer | Download as PowerPoint
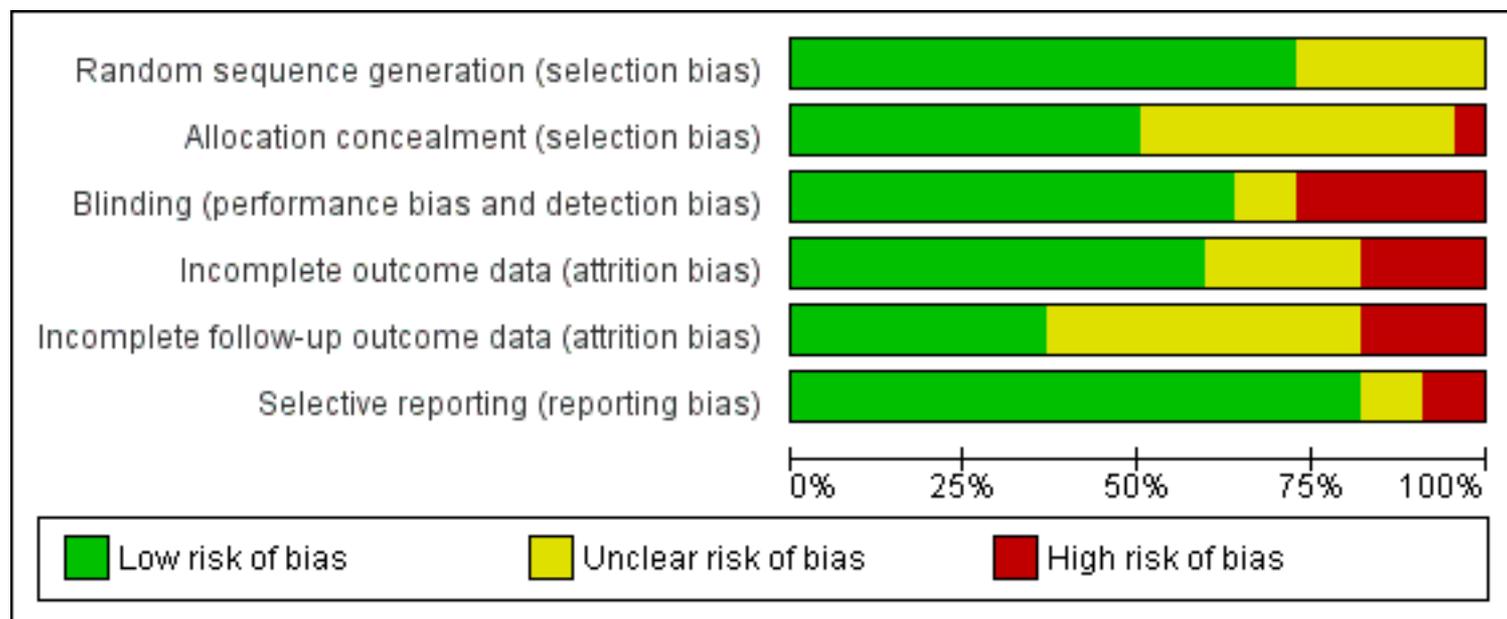
Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies

Comparisons with no acupuncture (acute treatment only or routine care)

Four trials (Facco 2008; Jena 2008; Linde K 2005; Vickers 2004) used adequate methods for allocation sequence generation and concealment of allocation when judged according to the Cochrane 'Risk of bias' tool (Higgins 2011). According to the definition of the ATC, three trials (Jena 2008; Linde K 2005; Vickers 2004) were "unambiguously adequately concealed". For the two other trials sequence generation was adequate but concealment was inadequate (Linde M 2000) or not fully adequate (Facco 2008). Given the

comparison between acupuncture and no acupuncture, the participants (who were also assessing all relevant outcomes) were unblinded in all six trials. In consequence, bias could not be ruled out. The use of headache diaries to monitor symptoms closely over a long period of time (Linde K 2005; Linde M 2000; Vickers 2004) might be less prone to bias than the use of questionnaires with retrospective assessment of symptoms for the preceding weeks (Facco 2008; Jena 2008). Attrition in the first three months was high in Linde M 2000 and minor to moderate in the remaining trials. The analyses of Jena 2008, Linde K 2005 and Vickers 2004 took account of attrition, suggesting a low risk of bias. This also applied to the long-term follow-up in Vickers 2004. Facco 2008 presented only a per-protocol analysis. Although presentation of results was not always optimal, we considered the risk of selective reporting to be low as the most important outcome measures were always presented and consistent. Overall, due to the lack of blinding in all studies there was some risk of performance and detection bias for this comparison.

Comparisons with sham interventions

We could not formally assess the quality of Alecrim 2005, for which only an abstract and additional unpublished information provided by the authors were available. Unpublished information provided by the authors and published information from the two other trials (Alecrim 2006; Alecrim 2008) conducted by the same group suggested that the risk of bias in this trial was low. Among the 13 trials formally assessed, the risk of bias regarding sequence generation was low for 10 (Alecrim 2006; Alecrim 2008; Ceccherelli 1992; Diener 2006; Facco 2008; Li 2012; Linde K 2005; Linde M 2004; Wang 2015; Zhao 2014) and unclear in five. Publications for five trials reported adequate methods of allocation concealment (Alecrim 2006; Alecrim 2008; Diener 2006; Li 2012; Linde K 2005); for a further two trials, such information was provided by the authors (Ceccherelli 1992; Facco 2008). All the trials attempted to blind participants.

Several trials that used sham interventions which were not strictly indistinguishable from 'true' acupuncture (Ceccherelli 1992; Diener 2006; Facco 2008; Linde K 2005) did not mention explicitly the use of a sham or placebo control in the informed consent procedure. This is ethically problematic, but enhances the credibility of the sham interventions. Taking into account also the results of the trials, we considered the risk of bias to be low in all trials. Reporting of dropouts was insufficient in several older trials. We considered the risk of bias to be low regarding short-term outcomes (up to three months) in nine trials (Alecrim 2006; Alecrim 2008; Diener 2006; Li 2012; Linde K 2005; Linde M 2004; Vincent 1989; Wang 2015; Zhao 2014), and low regarding long-term outcomes in six (Alecrim 2008; Diener 2006; Li 2012; Linde K 2005; Linde M 2004; Wang 2015). For two trials (Weinschütz 1993; Weinschütz 1994) outcomes were reported so inadequately that selective reporting could not be ruled out. Overall, the risk of bias was variable, but, particularly in the three largest trials, good quality. However, as acupuncturists could not be blinded in any trial performance, bias could not be ruled out completely.

Comparisons with prophylactic drug treatment

Three trials (Allais 2002; Diener 2006; Streng 2006) used adequate methods for sequence generation and concealment, one trial reported an adequate method for sequence generation but insufficient detail regarding concealment (Facco 2013), and one trial (Hesse 1994) did not describe the methods. Four trials (Allais 2002; Diener 2006; Facco 2013; Streng 2006) compared acupuncture and drug treatment in an open manner, which implies that bias on this level is possible. The use of a double-dummy technique allowed participant blinding in Hesse 1994, but this approach might be associated with other problems (see Discussion). While there is little risk of bias due to low attrition rates in Allais 2002 and Hesse 1994, and unclear risk in Facco 2013, a relevant problem occurred in the two German trials (Diener 2006; Streng 2006). The recruitment situation for these trials made it likely that participants had a preference for acupuncture. This resulted in a high proportion of participants allocated to drug treatment withdrawing informed consent immediately after randomisation (34% in Diener 2006 and 13% in Streng 2006), as well as high treatment discontinuation (18% in Diener 2006) or dropout rates due to adverse effects (16% in Streng 2006). These trials did not include participants refusing informed consent immediately after randomisation in analyses, and one (Streng 2006) also excluded early dropouts. Such analyses should normally tend to favour drug treatment. Both trials presented additional analyses restricted to participants complying with the protocol. All five trials presented the most important outcomes measured, so we considered the risk of bias of selective reporting to be low. Overall, as four of the trials were not blinded and two trials had a problem with relevant attrition in the drug group there is a considerable risk of bias (see also Discussion).

Effects of interventions

See: **Summary of findings for the main comparison** Acupuncture compared to no treatment/usual care; **Summary of findings 2** Acupuncture compared to sham interventions; **Summary of findings 3** Acupuncture compared to prophylactic drugs

Comparisons with no acupuncture (acute treatment only or routine care)

The five trials comparing acupuncture with a control group receiving either treatment of acute migraine attacks only or routine care are clinically very heterogeneous. Facco 2008 performed a four-armed trial (n = 160) in which participants in the control group all received acute treatment of attacks with rizatriptan. Jena 2008 is a very large, highly pragmatic study which included a total of 15,056 headache sufferers recruited by more than 4000 physicians in Germany. A total of 11,874 people not giving consent to randomisation received up to 15 acupuncture treatments within three months and were followed for an

additional three months. This was also the case for 1613 participants randomized to immediate acupuncture, while the remaining 1569 participants remained on routine care (not further defined) for three months and then received acupuncture. The published analysis of this trial is on all randomized participants, but we received unpublished results of subgroup analyses on the 1715 participants with migraine from the study authors for the previous version of our review and we re-analysed the data from the ATC for this update. Linde M 2000 was a small pilot trial (n = 39) performed in a specialized migraine clinic in Sweden in which control participants continued with their individualized treatment of acute attacks but did not receive additional acupuncture. A similar approach was used for the waiting-list control group in the three-armed (also sham control group) Linde K 2005 (n = 302) trial. Finally, in the Vickers 2004 trial (n = 401), acupuncture in addition to routine care in the British National Health Service was compared to a strategy, 'avoid acupuncture'. In addition to the strong clinical heterogeneity, the methods and timing of outcome measurement in these trials also differed considerably.

Therefore, pooled effect size estimates have to be interpreted with caution. Nevertheless, the findings show that acupuncture treatment is associated with a moderately large short-term benefit compared to no acupuncture control groups (Figure 4; Figure 5).

Figure 4   Open in figure viewer | **Download as PowerPoint**
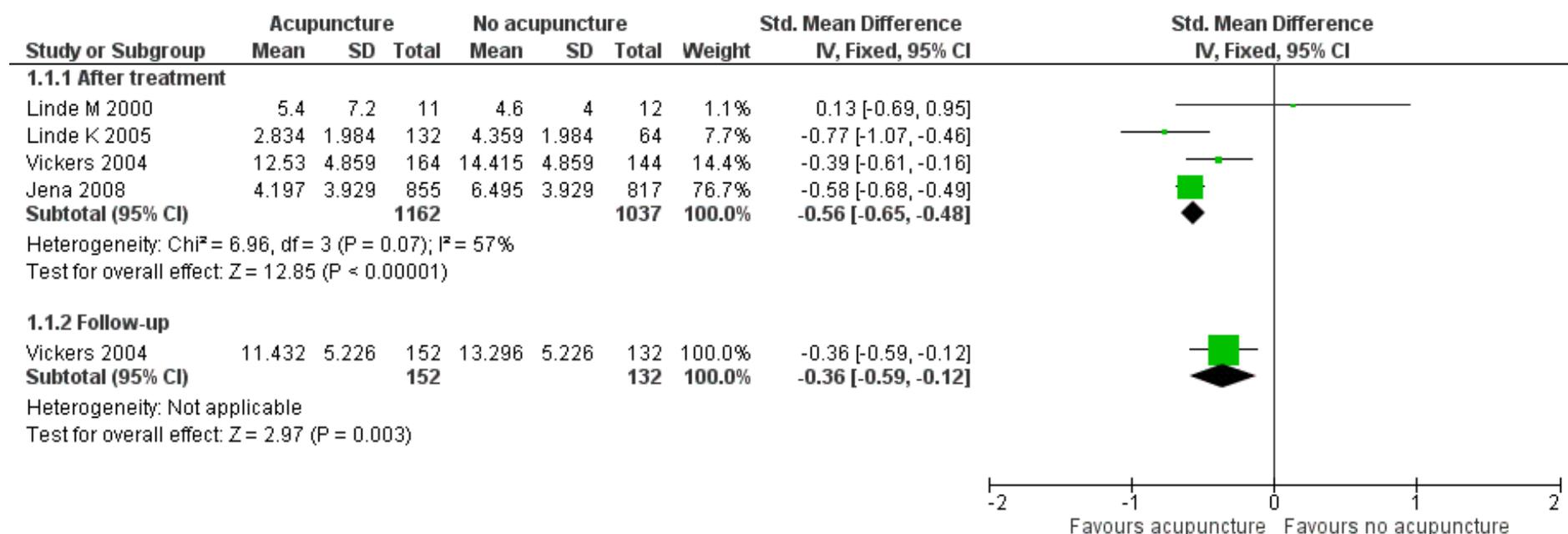
Forest plot of comparison: 1 Acupuncture vs. no acupuncture, outcome: 1.1 Headache frequency

Figure 5

Open in figure viewer | Download as PowerPoint

Study or Subgroup	Acupuncture Events	Total	No acupuncture Events	Total	Weight	Risk Ratio IV, Fixed, 95% CI
1.2.1 After treatment						
Linde M 2000	3	20	2	19	0.7%	1.43 [0.27, 7.61]
Linde K 2005	75	145	11	76	6.1%	3.57 [2.02, 6.31]
Vickers 2004	35	194	18	183	7.0%	1.83 [1.08, 3.12]
Jena 2008	436	966	173	916	86.2%	2.39 [2.05, 2.78]
Subtotal (95% CI)		1325		1194	100.0%	2.40 [2.08, 2.76]
Total events	549		204			

Heterogeneity: Chi² = 3.24, df = 3 (P = 0.36); I² = 7%
Test for overall effect: Z = 12.20 (P < 0.00001)

Study or Subgroup	Acupuncture Events	Total	No acupuncture Events	Total	Weight	Risk Ratio IV, Fixed, 95% CI
1.2.2 Follow-up						
Vickers 2004	48	194	21	183	100.0%	2.16 [1.35, 3.45]
Subtotal (95% CI)		194		183	100.0%	2.16 [1.35, 3.45]
Total events	48		21			

Heterogeneity: Not applicable
Test for overall effect: Z = 3.19 (P = 0.001)

Risk Ratio IV, Fixed, 95% CI

0.1   0.2   0.5   1   2   5   10

Favours no acupuncture   Favours acupuncture

Forest plot of comparison: 1 Acupuncture vs. no acupuncture, outcome: 1.2 Response (at least 50% frequency reduction)

Among the four trials providing sufficient data the pooled fixed-effects standardized mean difference (SMD) was -0.56 (95% CI -0.65 to -0.48; 2199 participants); findings were statistically heterogeneous (P value = 0.07; I^2 = 57%; random-effects estimate -0.53; 95% CI -0.72 to -0.34).

After treatment, headache frequency at least halved in 41% of participants receiving acupuncture and 17% receiving no acupuncture. The fixed effects risk ratio (RR) was 2.40 (95% CI 2.08 to 2.76; 4 trials, 2519 participants); there was no indication of statistical heterogeneity (P value = 0.36; I^2 = 7%). We consider these findings after treatment as moderate quality evidence because as the large trials consistently show clinically relevant differences, in spite of the risk of bias due to lack of blinding, we found some indication of heterogeneity (headache frequency) and clinical differences between trials. The corresponding number needed to treat for an additional beneficial outcome (NNTB) was 4 (95% CI 3 to 6). There was only one trial with a follow-up beyond three months (Vickers 2004; 12 month follow-up). The SMD (frequency) was -0.36 (95% CI -0.59 to -0.12; 284 participants with data) and the RR for response was 2.16

(95% CI 1.35 to 3.45; 377 participants). The NNTB based on this trial was 7 (95% CI 4.00 to 25.00; proportion of participants with response in the sham group 11%). Although the trial was large we consider its long-term findings to be low quality evidence as, given the variable effect sizes after treatment in the available trials, future trials performed in different settings might well yield different effect sizes. Findings in the time window analyses are consistent with those of the main analysis (Analysis 1.3; Analysis 1.4). The single specific frequency outcomes, migraine attacks, migraine days and headache days were not measured or reported in any trials but findings were consistent with those in our primary outcome, headache frequency (Analysis 1.5; Analysis 1.6; Analysis 1.7). This also applies to the outcomes headache intensity, analgesic use and headache scores (Analysis 1.8; Analysis 1.9; Analysis 1.10). We did not explore reasons for heterogeneity due to the small number of trials.

The number of participants not reaching the primary endpoint was slightly lower in acupuncture than in non-acupuncture groups (OR 0.69; 95% CI 0.46 to 1.05); there was some heterogeneity (P value = 0.13; I^2 = 47%). In the two trials reporting reasons for attrition there were no dropouts due to adverse effects. Information on other safety/acceptability outcomes was reported insufficiently (see Analysis 1.11).

Comparisons with sham interventions

Both after treatment (12 trials providing data from 1646 participants) and at follow-up (10 trials, 1534 participants) acupuncture was associated with a small but statistically significant frequency reduction over sham in the fixed-effect analyses (Figure 6). The SMD was -0.18 (95% -0.28 to -0.08; P value from the Chi² test for heterogeneity = 0.04, I^2 = 47%) after treatment and -0.19 (95% -0.30 to -0.09; P value from the Chi² test for heterogeneity = 0.01, I^2 = 59%) at follow-up. The results of the random-effects models were similar (SMD -0.24; 95% CI -0.41 to -0.07 for post-treatment, SMD-0.16; 95% CI -0.37 to 0.04 at follow-up).

Figure 6

Study or Subgroup	Acupuncture Mean	SD	Total	Sham Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
2.1.1 After treatment								
Ceccherelli 1992	11	11	15	35	28	15	1.7%	-1.10 [-1.87, -0.32]
Linde M 2004	2.5	1	15	1.9	0.95	13	1.8%	0.60 [-0.17, 1.36]
Wallasch 2012	4.13	2.55	13	4.82	1.98	14	1.8%	-0.29 [-1.05, 0.47]
Alecrim 2006	6.4	3.3	14	6.3	3.1	13	1.8%	0.03 [-0.72, 0.79]
Vincent 1989	-2.73	4.85	15	-0.73	3.22	15	1.9%	-0.47 [-1.20, 0.25]
Alecrim 2008	4.8	3	18	6.3	3.7	17	2.3%	-0.44 [-1.11, 0.24]
Wang 2015	5.2	5	26	10.1	7.1	24	3.1%	-0.79 [-1.37, -0.21]
Alecrim 2005	4.8	3.6	32	7.3	5	32	4.1%	-0.57 [-1.07, -0.07]
Zhao 2014	3.51	2.66	40	3.91	2.82	40	5.3%	-0.14 [-0.58, 0.29]
Linde K 2005	2.795	2.184	132	2.711	2.184	76	12.9%	0.04 [-0.24, 0.32]
Li 2012	2.787	2.89	332	3.46	2.89	110	21.9%	-0.23 [-0.45, -0.02]
Diener 2006	3.117	2.871	300	3.455	2.871	325	41.5%	-0.12 [-0.27, 0.04]
Subtotal (95% CI)			**952**			**694**	**100.0%**	**-0.18 [-0.28, -0.08]**

Heterogeneity: Chi² = 20.69, df = 11 (P = 0.04); I² = 47%
Test for overall effect: Z = 3.52 (P = 0.0004)

2.1.2 Follow up

Study										
Alecrim 2006	6.2	4.3	13	4.2	4.8	13	1.8%	0.43 [-0.35, 1.20]		
Wallasch 2012	3.76	3.03	13	4.34	2.63	14	1.9%	-0.20 [-0.96, 0.56]		
Linde M 2004	2.3	1.7	15	2.1	0.8	13	2.0%	0.14 [-0.60, 0.89]		
Vincent 1989	-4.87	4.6	15	-3.4	2.87	15	2.1%	-0.37 [-1.10, 0.35]		
Alecrim 2008	7.4	4.4	17	6.2	4.5	17	2.4%	0.26 [-0.41, 0.94]		
Wang 2015	5.1	3.7	26	9.5	6.2	24	3.3%	-0.86 [-1.44, -0.28]		
Alecrim 2005	5.2	3.6	32	6.5	4.8	32	4.5%	-0.30 [-0.80, 0.19]		
Linde K 2005	3.144	2.251	131	2.682	2.251	72	13.3%	0.20 [-0.08, 0.49]		
Li 2012	2.216	2.737	329	3.355	2.737	110	23.3%	-0.42 [-0.63, -0.20]		
Diener 2006	3.048	3.1	305	3.651	3.1	328	45.3%	-0.19 [-0.35, -0.04]		
Subtotal (95% CI)			**896**			**638**	**100.0%**	**-0.19 [-0.30, -0.09]**		

Heterogeneity: Chi² = 21.70, df = 9 (P = 0.010); I² = 59%
Test for overall effect: Z = 3.62 (P = 0.0003)
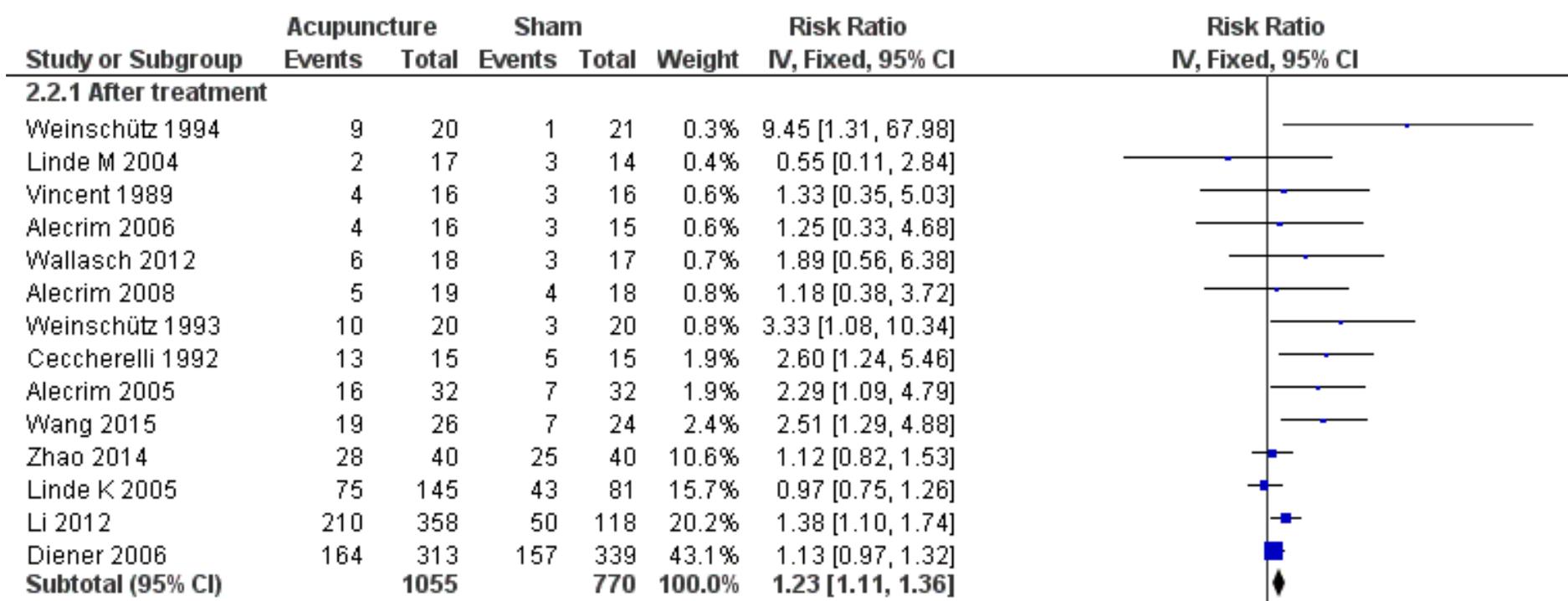
Favours acupuncture   Favours sham
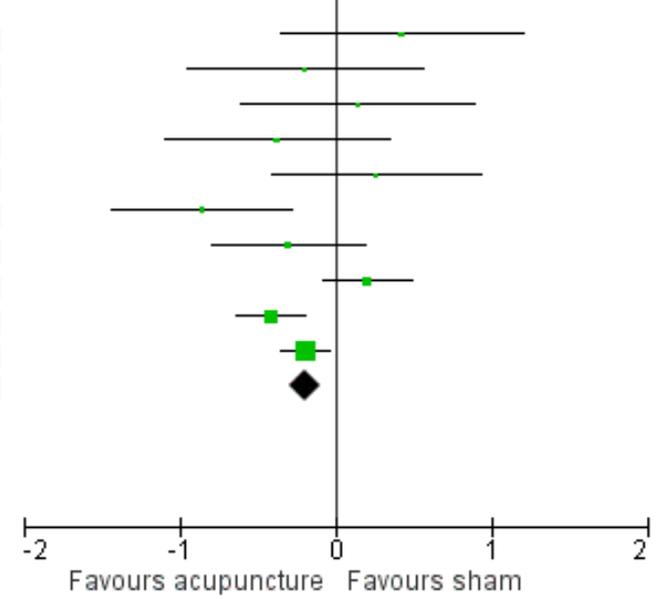
Forest plot of comparison: 2 Acupuncture vs. sham interventions, outcome: 2.1 Headache frequency

After treatment, headache frequency at least halved in 50% of participants receiving true acupuncture and 41% receiving sham acupuncture (pooled RR 1.23, 95% CI 1.11 to 1.36; P value = 0.02, I^2 = 48%; 14 trials, 1825 participants) and at follow-up in 53% and 42%, respectively. The pooled fixed effects RR was 1.23 (95% CI 1.11 to 1.36; P value from the Chi² test for heterogeneity = 0.02, I^2 = 48%; 14 trials, 1825 participants) after treatment and 1.25 (95% CI 1.13 to 1.39; P value from the Chi² test for heterogeneity = 0.004, I^2 = 61%; 11 trials, 1683 participants) at follow-up (Figure 7). The corresponding NNTB would be 11 (95% CI 7.00 to 20.00) after treatment and 10 (95% CI 6.00 to 18.00) at follow-up. Random-effects RRs were 1.39 (95% CI 1.14 to 1.69) and 1.33 (95% CI 1.05 to 1.70). The results were dominated by the three large, high-quality trials (Diener 2006; Li 2012; Linde K 2005; 75% and 82% weight, respectively, in the meta-analyses). We consider the findings for the outcomes headache frequency and response both after treatment and at follow-up as moderate quality evidence (indication of heterogeneity and small effect sizes leaving magnitude and statistical significance of effect open to some change with more trials). The time windows analyses yielded findings which were consistent with our main analyses (Analysis 2.3; Analysis 2.4). Specific frequency outcomes as well as intensity, analgesic use and headache scores were typically available for less than half of the trials (Analysis 2.5; Analysis 2.6; Analysis 2.7; Analysis 2.8; Analysis 2.9; Analysis 2.10).

Figure 7

Open in figure viewer   Download as PowerPoint

Study or Subgroup	Acupuncture Events	Total	Sham Events	Total	Weight	Risk Ratio IV, Fixed, 95% CI	Risk Ratio IV, Fixed, 95% CI
2.2.1 After treatment							
Weinschütz 1994	9	20	1	21	0.3%	9.45 [1.31, 67.98]	
Linde M 2004	2	17	3	14	0.4%	0.55 [0.11, 2.84]	
Vincent 1989	4	16	3	16	0.6%	1.33 [0.35, 5.03]	
Alecrim 2006	4	16	3	15	0.6%	1.25 [0.33, 4.68]	
Wallasch 2012	6	18	3	17	0.7%	1.89 [0.56, 6.38]	
Alecrim 2008	5	19	4	18	0.8%	1.18 [0.38, 3.72]	
Weinschütz 1993	10	20	3	20	0.8%	3.33 [1.08, 10.34]	
Ceccherelli 1992	13	15	5	15	1.9%	2.60 [1.24, 5.46]	
Alecrim 2005	16	32	7	32	1.9%	2.29 [1.09, 4.79]	
Wang 2015	19	26	7	24	2.4%	2.51 [1.29, 4.88]	
Zhao 2014	28	40	25	40	10.6%	1.12 [0.82, 1.53]	
Linde K 2005	75	145	43	81	15.7%	0.97 [0.75, 1.26]	
Li 2012	210	358	50	118	20.2%	1.38 [1.10, 1.74]	
Diener 2006	164	313	157	339	43.1%	1.13 [0.97, 1.32]	
Subtotal (95% CI)		**1055**		**770**	**100.0%**	**1.23 [1.11, 1.36]**	
Total events	565		314				

Heterogeneity: Chi² = 25.09, df = 13 (P = 0.02); I² = 48%
Test for overall effect: Z = 3.93 (P < 0.0001)

2.2.2 Follow-up

Alecrim 2008	1	19	3	18	0.2%	0.32 [0.04, 2.76]	
Linde M 2004	4	17	1	14	0.3%	3.29 [0.41, 26.21]	
Alecrim 2006	2	16	8	15	0.6%	0.23 [0.06, 0.93]	
Wallasch 2012	7	18	4	17	1.0%	1.65 [0.59, 4.65]	
Wang 2015	17	26	5	24	1.6%	3.14 [1.37, 7.19]	
Alecrim 2005	11	32	7	32	1.7%	1.57 [0.70, 3.54]	
Weinschütz 1993	13	20	8	20	2.9%	1.63 [0.87, 3.04]	
Weinschütz 1994	15	20	8	21	3.1%	1.97 [1.08, 3.59]	
Linde K 2005	67	145	42	81	15.0%	0.89 [0.68, 1.17]	
Li 2012	249	358	55	118	26.7%	1.49 [1.22, 1.83]	
Diener 2006	169	313	155	339	46.9%	1.18 [1.01, 1.38]	
Subtotal (95% CI)		**984**		**699**	**100.0%**	**1.25 [1.13, 1.39]**	
Total events	555		296				

Heterogeneity: Chi² = 25.50, df = 10 (P = 0.004); I² = 61%
Test for overall effect: Z = 4.18 (P < 0.0001)

0.01 — 0.1 — 1 — 10 — 100
Favours sham   Favours acupuncture
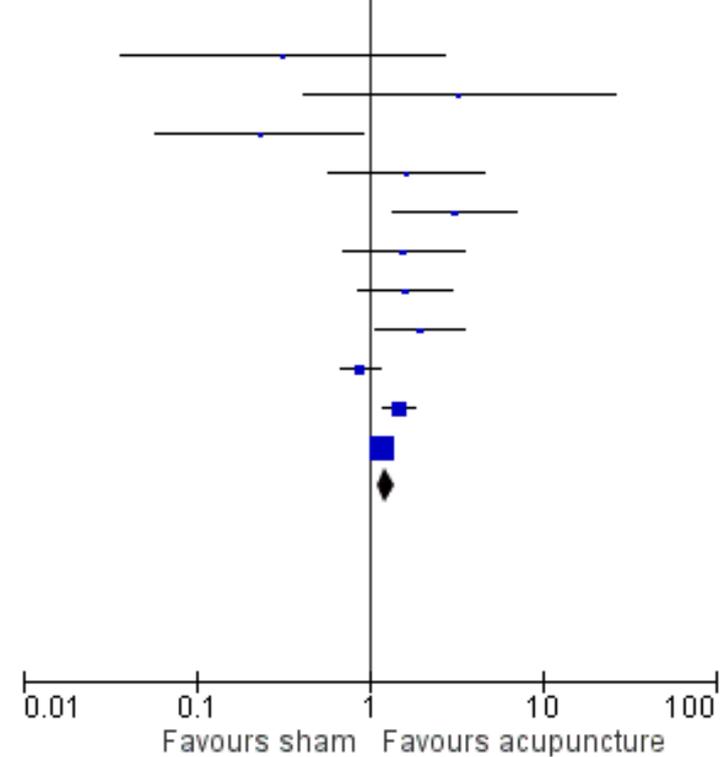
Forest plot of comparison: 2 Acupuncture vs. sham interventions, outcome: 2.2 Response (at least 50% frequency reduction)

We performed subgroup analyses to investigate four potential sources of heterogeneity for both frequency and response, both after treatment and follow-up (four analyses for each potential source of heterogeneity). While there were suggestions of subgroup differences, heterogeneity tended to remain considerable in most subgroups across analyses. Effects of acupuncture over sham were significantly smaller in the three unambiguously adequately concealed trials (which were also by far the three largest trials) than in the remaining trials in three (Analysis 2.11; Analysis 2.13Analysis 2.14) of four analyses (no significant difference in Analysis 2.12). In the analyses grouping trials into smaller and larger (number of participant up to or above the median number of participants in the trials included in the analysis) studies' differences tended to be somewhat smaller in larger trials but these findings were mainly driven by the three larger, unambiguously adequately concealed trials (Analysis 2.15; Analysis 2.16; Analysis 2.17; Analysis 2.22). Consistently, effects over sham tended to be larger in trials with 16 or more treatment sessions compared to trials with up to 12 sessions (Analysis 2.18; Analysis 2.19; Analysis 2.20; Analysis 2.21;). Effects also tended to be somewhat larger in trials using non-penetrating sham techniques, however, only three relatively small trials used such sham techniques (Analysis 2.23; Analysis 2.24; Analysis 2.25; Analysis 2.26). Re-including the five trials excluded for this update (but included in the previous version of this review) had only minimal impact on results.
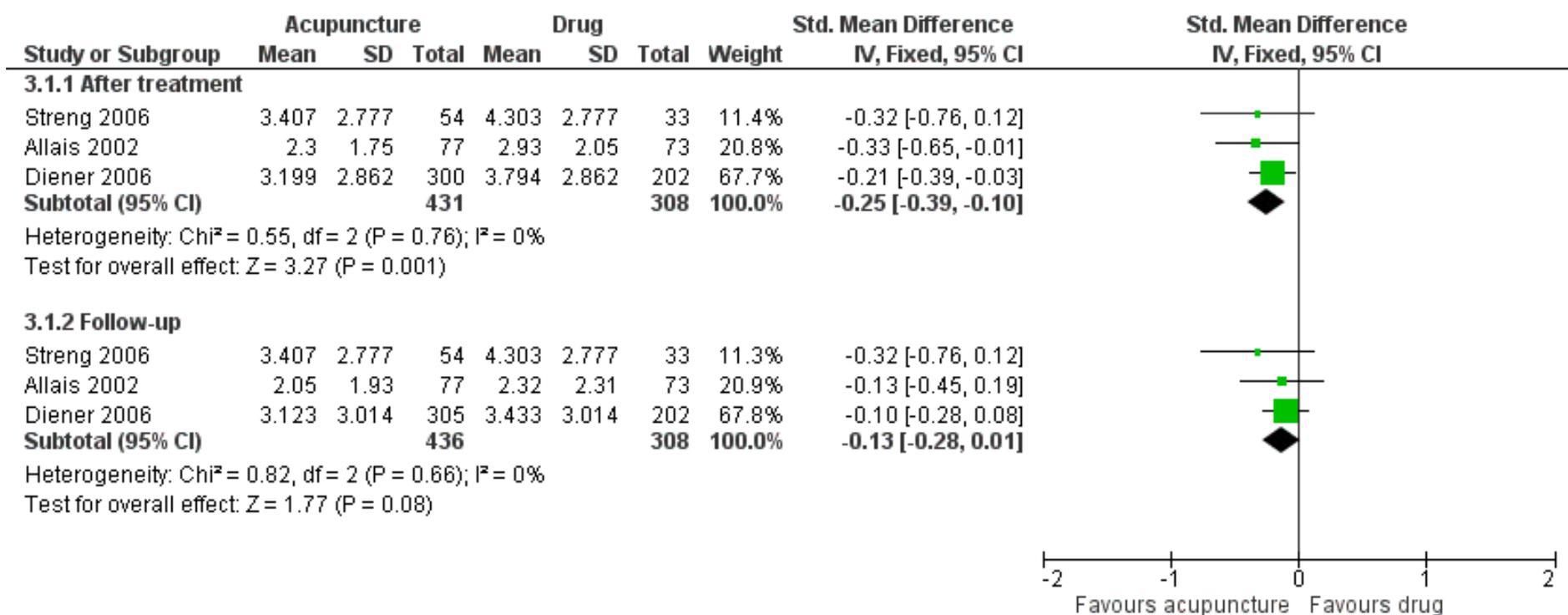
In the seven trials reporting this outcome only three of 621 participants receiving acupuncture and none

of 310 in control groups dropped out due to adverse effects (OR 2.84; 95% CI 0.43 to 18.71; 7 trials, 931 participants; I^2 = 0%). We consider this low quality evidence as there is great uncertainty regarding the effect estimate due to the very small number of such dropouts. Only the four largest trials reported the number of participants reporting adverse effects. Among 847 participants receiving acupuncture 138 (16%) reported adverse effects compared to 98 (17%) receiving sham (OR 1.15; 95% CI 0.85 to 1.56; I^2 = 0%; moderate quality evidence; Analysis 2.27). There were also no significant differences in the number of participants not reaching the primary endpoint of the trial (OR 1.14; 95% CI 0.78 to 1.67; 11 trials, 1770 participants) and experiencing serious adverse events (OR 1.29; 95% CI 0.43 to 3.83; 6 trials, 1071 participants; Analysis 2.27).

Comparisons with prophylactic drug treatment

The results of Hesse 1994 regarding treatment effectiveness were not reported in a manner that allowed effect size estimation. Overall, the findings of this trial, which used a double-dummy design (true acupuncture plus metoprolol placebo versus sham acupuncture plus metoprolol), showed similar improvements in both groups, slightly favouring the sham acupuncture plus metoprolol group. The acupuncture technique used in this trial (very brief needling of individual trigger points) was rather unusual and was considered with skepticism by our acupuncturists. The pragmatic trial, Facco 2013, which compared a traditional acupuncture strategy with valproic acid, did not use a headache diary, but only a questionnaire including the Migraine Disability (MIDAS) instrument (Stewart 2001). Participants were asked to report the number of headache days over periods of three months. The publication reported very large improvements (the median number of headache days went down from 18 days before the trial, to four days during the treatment phase and during the follow-up phase in the

acupuncture group, and from 17 to three and six days in the group receiving valproic acid), and very little variation (narrow interquartile ranges). The first author provided means and standard deviations, but standard deviations were very small. As we were uncertain about the reliability of the data we decided to not include it in meta-analysis.

The remaining three trials could be entered into meta-analyses. Acupuncture reduced migraine frequency significantly more than drug prophylaxis after treatment (SMD -0.25; 95% CI -0.39 to -0.10; 739 participants) while differences were smaller at follow-up and were not significantly different (SMD -0.13; 95% CI -0.28 to 0.01; P value = 0.08; 744 participants; see Figure 8).

Figure 8   Open in figure viewer | Download as PowerPoint

Study or Subgroup	Acupuncture Mean	SD	Total	Drug Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI	Std. Mean Difference IV, Fixed, 95% CI
3.1.1 After treatment									
Streng 2006	3.407	2.777	54	4.303	2.777	33	11.4%	-0.32 [-0.76, 0.12]	
Allais 2002	2.3	1.75	77	2.93	2.05	73	20.8%	-0.33 [-0.65, -0.01]	
Diener 2006	3.199	2.862	300	3.794	2.862	202	67.7%	-0.21 [-0.39, -0.03]	
Subtotal (95% CI)			431			308	100.0%	-0.25 [-0.39, -0.10]	

Heterogeneity: Chi² = 0.55, df = 2 (P = 0.76); I² = 0%
Test for overall effect: Z = 3.27 (P = 0.001)

Study or Subgroup	Acupuncture Mean	SD	Total	Drug Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI	Std. Mean Difference IV, Fixed, 95% CI
3.1.2 Follow-up									
Streng 2006	3.407	2.777	54	4.303	2.777	33	11.3%	-0.32 [-0.76, 0.12]	
Allais 2002	2.05	1.93	77	2.32	2.31	73	20.9%	-0.13 [-0.45, 0.19]	
Diener 2006	3.123	3.014	305	3.433	3.014	202	67.8%	-0.10 [-0.28, 0.08]	
Subtotal (95% CI)			436			308	100.0%	-0.13 [-0.28, 0.01]	

Heterogeneity: Chi² = 0.82, df = 2 (P = 0.66); I² = 0%
Test for overall effect: Z = 1.77 (P = 0.08)

Favours acupuncture Favours drug

Forest plot of comparison: 3 Acupuncture vs. prophylactic drug treatment, outcome: 3.1 Headache frequency

Findings were similar for response (Figure 9). After a median follow-up of three months headache frequency at least halved in 57% of participants receiving acupuncture and 46% receiving prophylactic drugs, and after six months in 59% and 54%, respectively. The RR was 1.24 (95% CI 1.08 to 1.44; 743 participants) after treatment and 1.11 (95% CI 0.97 to 1.26; 744 participants) at follow-up. Findings were consistent among trials with P values from the Chi² test for heterogeneity being above 0.5 and I² being 0% in all analyses. While there was risk of bias due to lack of blinding in all three trials and relevant attrition

in two trials, and future trials might not confirm the small effects of acupuncture over prophylactic drug treatment, we consider the highly consistent trial findings for both outcomes and both time points as moderate quality evidence that acupuncture is non-inferior to prophylactic drug treatment.The time window analyses are consistent with these findings as are the findings on additional outcomes (Analysis 3.5; Analysis 3.6; Analysis 3.7; Analysis 3.8; Analysis 3.9; Analysis 3.10).

Figure 9

Open in figure viewer | Download as PowerPoint

Study or Subgroup	Acupuncture Events	Total	Drug Events	Total	Weight	Risk Ratio IV, Fixed, 95% CI	Risk Ratio IV, Fixed, 95% CI
3.2.1 After treatment							
Streng 2006	26	56	12	35	7.2%	1.35 [0.79, 2.32]	
Allais 2002	54	77	38	73	29.8%	1.35 [1.03, 1.75]	

| Diener 2006 | 164 | 300 | 93 | 202 | 63.1% | 1.19 [0.99, 1.42] |
| Subtotal (95% CI) | | 433 | | 310 | 100.0% | 1.24 [1.08, 1.44] |

Total events 244 143
Heterogeneity: Chi² = 0.70, df = 2 (P = 0.71); I² = 0%
Test for overall effect: Z = 2.98 (P = 0.003)

3.2.2 Follow-up

Streng 2006	29	54	13	33	6.8%	1.36 [0.83, 2.23]
Allais 2002	56	77	46	73	33.1%	1.15 [0.92, 1.44]
Diener 2006	169	305	106	202	60.0%	1.06 [0.89, 1.25]
Subtotal (95% CI)		436		308	100.0%	1.11 [0.97, 1.26]

Total events 254 165
Heterogeneity: Chi² = 1.14, df = 2 (P = 0.57); I² = 0%
Test for overall effect: Z = 1.55 (P = 0.12)

Favours drug Favours acupuncture
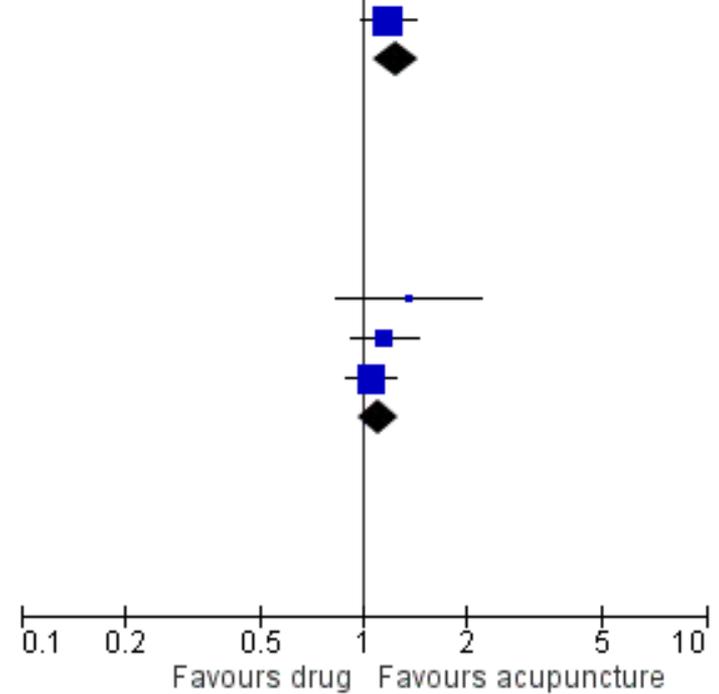
Forest plot of comparison: 3 Acupuncture vs. prophylactic drug treatment, outcome: 3.2 Response (at least 50% frequency reduction)

In the four studies reporting this outcome three (1%) of 227 participants receiving acupuncture dropped out due to adverse effects compared to 16 (7%) receiving prophylactic drugs (OR 0.27; 95% CI 0.08 to 0.86; I² = 0%; Analysis 3.11). All five trials provided the number of participants reporting adverse effects. Probably due to different methods for documenting this outcome the absolute frequency of adverse effects in both groups varied greatly between trials. A total of 90 (17%) of 520 participants receiving acupuncture reported adverse effects compared to 140 (34%) of 411 participants receiving prophylactic

drug treatment (OR 0.25; 95% CI 0.10 to 0.62; I² = 78%, P value from Chi² test for heterogeneity = 0.001). Despite some limitations (uncertainty regarding dropouts due to low event rates and heterogeneity regarding reporting of adverse effects) we consider this moderate quality evidence as study findings consistently favour acupuncture over prophylactic drug treatment. Furthermore, study findings also favoured acupuncture for the number of participants not reaching the primary endpoint (OR 0.28; 95% CI 0.10 to 0.78; 4 trials, 995 participants; I² = 80%; P value = 0.002). Serious adverse events were reported in seven (2%) of 313 participants receiving acupuncture compared to four (1%) of 307 participants receiving prophylactic drugs (OR 1.33; 95% CI 0.38 to 4.73; 4 trials; I² = 0%).

Discussion

Summary of main results

Several trials using quite variable methods and interventions consistently showed that the addition of acupuncture to treatment of acute migraine attacks or to routine care was beneficial for at least three months. Compared to no treatment or routine care only (which includes treatment of acute migraine attacks and possibly other interventions) the size of the effect seemed to be moderate according to usual standards for classifying effect size measures such as standardized mean differences. The only trial that investigated long-term effects showed a sustained small to moderate response to acupuncture in addition to routine care provided by a GP (Vickers 2004). Compared to sham acupuncture, true acupuncture interventions were associated with small but statistically significant effects both after treatment and at follow-up, but findings were statistically heterogeneous. In the largest, adequately concealed trials differences were even smaller (but still statistically significant). The pooled analyses of the available trials comparing acupuncture interventions with evidence-based prophylactic drug treatment found a superiority of acupuncture at completion of treatment, though at follow-up differences were no longer statistically significant. Compared to drug prophylaxis fewer participants dropped out due adverse effects or reported adverse effects.

Possible explanations of the findings

The interpretation of the findings of our review remains challenging. While, contrary to the results of the previous version of our review, differences between true acupuncture and sham interventions became statistically significant (after the inclusion of four new sham-controlled trials), it seems still surprising

that the size of the effect over sham is similar to that over prophylactic treatment with drugs that have been shown to be superior to placebo (Schürks 2008). Three factors could explain these findings (probably in combination). Firstly, sham acupuncture might have direct physiological effects on mechanisms relevant to migraine symptoms, secondly, acupuncture might be a particularly potent placebo, and thirdly, due to the lack of blinding, comparisons with routine care and prophylactic drug treatment might be biased.

We consider each of these possible explanations in turn.

Physiological effects of sham acupuncture

Many sham acupuncture procedures involve needling locations that are not traditional points with the same frequency and duration as in the true acupuncture group. In some studies needles are inserted into classical acupuncture points not indicated in migraine. Most physiological mechanisms proposed for

acupuncture do not necessarily imply point specificity (Bäcker 2004). Even the non-penetrating 'placebo' needles might activate unmyelinated (C 'tactile') afferent nerves which can influence pain perception (Lund 2006). Several researchers have argued that some effects of acupuncture might not be point-specific (Han 1997; Lundeberg 2007), and that these might be particularly relevant for treating conditions other than localized nociceptive pain (Thomas 1996; Borud 2010). In individual patient data meta-analysis, acupuncture was significantly superior to all categories of control group. For trials that used penetrating needles for sham control, acupuncture had smaller effect sizes than for trials with non-penetrating sham or sham control without needles (MacPherson 2014).

Sham acupuncture as a strong placebo

According to the available evidence, the most important mechanisms for placebo effects are expectations, conditioning, anxiety reduction and social support (Crow 1999; Benedetti 2008). These elements are likely to be influenced by the treatment setting, its context and its meaning. Acupuncture - with its repeated sessions, intense provider contact, slightly painful procedure, an often 'exotic' model of symptom explanation and associated relaxation during sessions - might maximize such effects.

While the average clinical effect of placebo interventions seems to be small (Hróbjartsson 2010), there is some evidence that sham acupuncture is associated with larger effects than, for example, a placebo pill or other non-pharmacological sham interventions. This evidence comes from one of the few randomized trials directly comparing different types of placebo (Kaptchuk 2006), from indirect comparison of trials including both a sham and a no-treatment control (Linde 2010a), and from a network meta-analysis of pharmacological and non-pharmacological treatments and their placebos in migraine prophylaxis (Meissner 2013). Furthermore, a systematic review of randomized trials of acupuncture including both a

sham and a no-treatment control found on average moderately large (SMD 0.45) differences (Linde 2010b). It seems highly plausible that both the physiological and strong placebo effects contribute to these considerable 'non-specific' effects of sham acupuncture. For example, a recent trial showed that the size of the effect associated with a sham acupuncture intervention can vary with the amount and characteristics of the patient-provider interaction (Kaptchuk 2008). Both the above explanations would also imply that it would be difficult to detect any small, point-specific effects in addition to potent placebo effects and non-specific needling effects.

Possible bias due to lack of blinding

While participants in the sham-controlled trials were blinded, this was (with the exception of the trial by Hesse 1994) not the case for the comparisons with treatment of acute migraine attacks only, routine care or other treatments. All clinically relevant outcome measures in clinical trials in migraine are patient-

reported (IHS 2000; IHS 2012). Preferably, outcomes are documented in diaries for at least four weeks before treatment and for longer time periods during and after treatment. It cannot be ruled out that participants allocated to acupuncture reported positively biased outcomes, while participants allocated to control reported negatively biased outcomes. However, response rates in participants allocated to drug treatment in the trials included in this review were comparable to those reported in drug trials (Van der Kuy 2002). Also, in groups receiving acute treatment only, response rates were within the range of placebo groups in drug trials (Van der Kuy 2002). In two trials comparing acupuncture and drug treatment (Diener 2006; Streng 2006), a relevant proportion of participants withdrew informed consent immediately after allocation to drug treatment. Additional participants dropped out during the study. This indicates that study participants had a preference for acupuncture. These problems could seriously bias the findings. However, participants not starting treatment were not included in the analyses, and per-protocol analyses confirmed the study findings. Still, these trials must be interpreted with caution.

A fourth possible explanation for the lack of larger effects of true acupuncture over sham comes from the perspective of acupuncture practitioners. The quality of acupuncture interventions in clinical trials is often disputed. Study protocols often limit the flexibility of treatment procedures, particularly in sham-controlled trials, and it is argued that better acupuncturists would have achieved better results. However, response rates in sham-controlled trials were on average similar to those in pragmatic trials with flexible treatments. Furthermore, while there is always the possibility that some expert acupuncturists are particularly successful, in several of the larger trials included in this review the training of treatment providers was at least comparable to that of the average acupuncturists in their country. Still, it cannot be ruled out that inadequate study interventions contribute to the lack of differences compared to sham interventions.

Overall completeness and applicability of evidence

Acupuncture is a therapy which is applied in a variable manner in different countries and settings. For example, in Germany, where the majority of the large trials included in this review were performed, acupuncture is mainly provided by general practitioners and other physicians. Their approach to acupuncture is based on the theories of traditional Chinese medicine, although the amount of training they receive in traditional Chinese medicine is limited (Weidenhammer 2007). In the UK, the providers are likely to be non-medical acupuncturists with a comparatively intense traditional training, physiotherapists or medical doctors with a more 'Western' approach (Dale 1997). The trials included in our review come from a variety of countries, and study designs range from very pragmatic (Jena 2008; Vickers 2004) to more experimental (Linde M 2004). Despite this distinct heterogeneity, within comparisons the findings seem broadly consistent.

Acupuncture is widely used in Asian countries, particularly China. We have not systematically searched Chinese databases for this version of the review, but plan to do so in the future. There is considerable skepticism toward clinical trials from China, as in the past results were almost exclusively positive (Vickers 1998). However, the quality and number of randomized trials published in Chinese have improved over recent years (Wang 2007), and it seems inadequate to neglect this evidence without examining it critically. Most of the identified, registered ongoing trials originate from China. Our update search (in non-Chinese databases) identified a number of trials from China but only two met inclusion criteria (Li 2012; Zhao 2014). When reading excluded trials we noticed several characteristics which suggest that at least some migraine studies from China are different and problematic from the point of view of Western headache research. Excluded trials often included participants with recent-onset of migraine, given acupuncture with the aim of 'curing' the condition. These trials seem hardly comparable to the many trials that included participants who had been suffering from migraine for a long time. Furthermore, most trials, including a group receiving prophylactic drugs, gave these only for four weeks, a period considered much too short by Western headache specialists (IHS 2012). Chinese trials also tend to use a higher number of treatment sessions and higher treatment frequency.

Large-scale observational studies (Jena 2008; Melchart 2006) and a systematic comparison of findings from a randomized and an observational study (Linde 2007a) suggest that the response rates observed in clinical trials are also seen in conditions similar to routine practice. However, as the overall evidence also suggests that factors other than the correct selection of acupuncture points and needling procedures play an important role in outcomes, treatment setting and participant selection could have a strong impact and might vary considerably. For example, a pooled analysis of four trials on chronic pain

(including Linde K 2005) found that even four months after completion of treatment, participants who had started acupuncture with a positive attitude and expectation had significantly better outcomes than participants with lower expectations (Linde 2007b).

People with migraine typically suffer from their headaches over many years. A general shortcoming of almost all randomized trials of any prophylactic treatments is their limited duration (rarely ever more than 12 months). Therefore, based on our review nothing can be said on sustainability of effects beyond 12 months.

Quality of the evidence

The methodological quality of the included trials was variable. Methods for sequence generation, allocation concealment, handling of dropouts and withdrawals and reporting of findings were adequate in most of the recent trials. Still, designing and performing clinical trials of acupuncture is a challenge,

particularly with respect to blinding and selection of control interventions. We have mentioned that bias cannot be ruled out in the unblinded studies, and that comparisons with prophylactic drug treatment have to be interpreted with caution due to high dropout rates in two of the trials. Blinding in comparisons with drug treatment could be achieved by double-dummy designs (drug plus sham acupuncture versus acupuncture plus drug placebo) as in the trials by Hesse 1994. However, if it is the case that sham acupuncture interventions are strong placebos and not physiologically inert, this approach would also be problematic.

We considered the overall quality of the evidence for most outcomes to be moderate. Reasons for not considering the quality of evidence to be high were lack of blinding of participants (for comparisons with no acupuncture controls and prophylactic drugs), unblinded treatment providers (all comparisons), indications of heterogeneity for some outcomes or major imprecision in the case of the outcome dropouts due to adverse effects. We did not further downgrade our rating because findings consistently showed clinically relevant effects in spite of variable effect sizes (efficacy outcomes for the comparison with no acupuncture controls) or very similar findings (efficacy outcomes for the comparison with prophylactic drugs).

Potential biases in the review process

We are confident that we have identified the existing large clinical trials relevant to our question, but we cannot rule out the possibility that there are additional small trials which are unpublished or published in sources not accessible to our search.

A relevant problem for systematic reviews on prophylactic treatments of migraine is the highly variable outcome measurement and the often inadequate reporting of results. Various measures of frequency, intensity, analgesic use and other outcomes are used, and as these measures have to be observed over longer time periods, the amount of data needed to obtain a good overview of the course of symptoms is considerable. Most trials in our review reported several outcome measures at different time points without evidence that these were selected in a biased way. Nevertheless, we were confronted with a complex mosaic of data. Several authors kindly provided unpublished data. Some sort of response and frequency measure was available for almost all trials, although the timing of the measurement and details of the measure often differed. As overall results are rather consistent, it seems unlikely that our results would have changed in a relevant manner if missing data had been available.

Four members of the review team were involved in at least one of the included trials. These trials were assessed by other members of the review team. All reviewers currently have affiliations to a CAM

(complementary and alternative medicine) research centre, or have had such an affiliation in the past.

Agreements and disagreements with other studies or reviews

We are not aware of any systematic reviews published after the previous version of our review (Linde 2009) focusing exclusively on randomized trials of acupuncture for the prophylaxis of episodic migraine. The analysis of the pooled individual database of the Acupuncture Trialists' Collaboration analysing high quality trials on chronic pain (Vickers 2012) included three trials also included here. The findings are consistent with those presented here. This applies also to a review of placebo- and sham-controlled trials of a variety of pharmacological and non-pharmacological prophylactic treatment focusing on the differential effectiveness of the placebo treatments (Meissner 2013).

Compared to the previous version of our review, findings are similar for the comparisons with no acupuncture and prophylactic drug treatment, while, at a first glance, our current results seem more positive for the comparison with sham. In our previous review there were no statistically significant differences between true and sham acupuncture, neither for frequency nor for response. The main reason is clearly that our current analyses have considerably more power. This is primarily due to inclusion of new trials (particularly, the large Li 2012 trial). Furthermore, the approach to group our analysis by four time windows in our previous version further decreased the number of trials per analysis. Finally, based on the advice of our statistician, in this updatewe have used fixed-effect models instead of random-effects models for the main analyses, which leads to narrower confidence intervals (yet, we also present random-effects estimates confirming our overall results). If one compares effect estimates

qualitatively, findings are very similar for the time points after treatment (current version) and two/four months after treatment (previous version). Instead, the addition of the new trials made our findings more positive for six-months' follow-up.

It should also be noted that the original publication of the Li 2012 trial, which compared each of the three tested acupuncture interventions against sham separately, did not report significant differences after treatment. We pooled the data of the three acupuncture groups (ensuring that the sham group was not counted more than once as a control) which explains that the observed difference is statistically significant for this trial in our analyses.