Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Acupuncture treatments for infantile colic: a systematic review and individual patient data meta-analysis of blinding test validated randomised controlled trials

Holgeir Skjeie[a], Trygve Skonnord[a], Mette Brekke[a], Atle Klovning[a], Arne Fetveit[a], Kajsa Landgren[b], Inger Kristensson Hallström[b] and Kjetil Gundro Brurberg[c,d]

[a]Department of General Practice, Institute of Health and Society University of Oslo, Oslo, Norway; [b]Faculty of Medicine Department of Health Sciences, Lund University, Lund, Sweden; [c]Norwegian Institute of Public Health Division for Health Services, Oslo, Norway; [d]Western Norway University for Applied Sciences Centre for Evidence Based Practice, Bergen, Norway

ABSTRACT

Objective: Needle acupuncture in small children has gained some acceptance in Western medicine. It is controversial, as infants and toddlers are unable to consent to treatment. We aimed to assess its efficacy for treating infantile colic.

Design: A systematic review and a blinding-test validation based on individual patient data from randomised controlled trials. Primary end-points were crying time at mid-treatment, at the end of treatment and at a 1-month follow-up. A 30-min mean difference (MD) in crying time between acupuncture and control was predefined as a clinically important difference. Pearson's chi-squared test and the James and Bang indices were used to test the success of blinding of the outcome assessors [parents].

Eligibility criteria and data sources: We included randomised controlled trials of acupuncture treatments of infantile colic. Systematic searches were conducted in Cochrane CENTRAL, MEDLINE, EMBASE, CINAHL and AMED, and in the Chinese language databases CNKI, VIP, Wang fang, SinoMed and Chinese Clinical Trial Registry.

Results: We included three randomised controlled trials with data from 307 participants. Only one of the included trials obtained a successful blinding of the outcome assessors in both the acupuncture and control groups. The MD in crying time between acupuncture intervention and no acupuncture control was −24.9 min [95% confidence interval, CI −46.2 to −3.6; three trials] at mid-treatment, −11.4 min [95% CI −31.8 to 9.0; three trials] at the end of treatment and −11.8 min [95% CI −62.9 to 39.2; one trial] at the 4-week follow-up. The corresponding standardised mean differences [SMDs] were −0.23 [95% CI −0.42 to −0.06], −0.10 [95% CI −0.29 to 0.08] and −0.09 [95% CI −0.48 to 0.30]. The heterogeneity was negligible in all analyses. The statistically significant result at mid-treatment was lost when excluding the apparently unblinded study in a sensitivity analysis: MD −13.8 min [95%CI −37.5 to 9.9] and SMD −0.13 [95%CI −0.35 to 0.09]. The registration of crying during treatment suggested more crying during acupuncture [odds ratio 7.7; 95% CI 2.7–20.6; one trial]. **GRADE**-Moderate quality evidence.

Conclusions: Percutaneous needle acupuncture treatments should not be recommended for infantile colic on a general basis.

Systematic review registration: PROSPERO 2015:CRD42015023253

KEY POINTS
- The role of acupuncture in the treatment of infantile colic is controversial. Available trials are small and present conflicting results.
- There were no clinically important differences between infants receiving acupuncture and no acupuncture control in this IPD meta-analysis of randomised controlled trials.
- The data indicate that acupuncture induces some treatment pain in many of the children.
- The study results indicate that percutaneous needle acupuncture should not be recommended for treatment of infantile colic on a general basis.

CONTACT Holgeir Skjeie ✉ holgeir.skjeie@medisin.uio.no 🖅 Department of General Practice, Institute of Health and Society, University of Oslo, P.O. Box 1130 Blindern, N-0318 Oslo, Norway
🅑 Supplemental data for this article can be accessed here.

Introduction

Infantile colic is a painful and poorly understood ailment in early infancy. It is a self-limiting condition normally ending at 3—4 months of age. The definition still commonly used is Wessel's symptom definition of 1954: 'Paroxysmal, uncontrollable crying and fussing in an otherwise healthy infant under 3 months of age, with more than 3 h of crying per day in more than 3 days for more than 3 weeks' [1]. A modified version, Rome III [2], has been in place since 2006 [3] and a further extension, ROME IV [2], since 2016 [4]. Persistent painful crying is a severe strain on both the child and parents [5]. There is no clear aetiology. According to the Rome IV criteria, infantile colic is in most cases regarded as a behavioural syndrome representing the high spectrum of normal developmental crying, rather than symptoms of abdominal pain [4]. Physiological factors such as altered gut motility, immature digestive functions, altered intestinal macrobiotics or food sensitivity might be involved [6—8]. Psychological factors like inadequate parent—infant interaction or family tension have also been proposed as important factors [6—8]. There is no consensus on treatment strategies for the condition [5,9]. Strategies include counselling on specific management techniques, reduced stimulation, herbal teas, sucrose, simethicone, hypoallergenic diet, chiropractic manipulation, probiotics and acupuncture [5,8,10].

Acupuncture is performed using thin steel needles penetrating the skin and into connective tissue and muscle fibres. The neurophysiologic basis for its pain-inhibiting effects has been studied in detail, and is well understood [11—13]. Needle effect sizes between real and sham acupuncture in various chronic pain conditions in adults are small and consistent, usually with standardised mean differences (SMDs) in the range of 0.15—0.23 [14]. When comparing real acupuncture versus no acupuncture, SMDs are typically between 0.42 and 0.57 [14]. Current evidence does not substantiate that the choice of needle points or the type of acupuncture affects efficacy, but more sessions and more needle applications are associated with more favourable outcomes in adults [15]. Acupuncture is considered safe when offered by trained practitioners, both when used in adults [13,16] and in children [17]. Acupuncture for children has gained some acceptance in Western medicine, even though the evidence supporting the use of acupuncture in small children is sparse [18,19]. In some Western-based textbooks, it is argued that the effects of acupuncture in small children are swift and often stronger than in adults [20—23], but these notions are usually based on

tradition, personal views and clinical experience. Case reports and qualitative studies about the efficacy of acupuncture in small children are often very optimistic and recommend the use of acupuncture on an anecdotal basis [24—26]. Interestingly, contemporary traditional Chinese medicine (TCM) practitioners in China rarely used needles on infants [27], so acupuncture treatment of small children seems to have developed as a part of a Western practice.

It has been suggested that acupuncture might play a role in the treatment of infantile colic [5,18,24,28,29]. Thus, it is proposed that acupuncture can counteract gut dysmotility in infants with colic, possibly by affecting the parasympathetic vagal reflexes and the centrally opioid-mediated pain inhibitory pathway [30,31]. However, controlled trials investigating the efficacy of acupuncture in cases of infantile colic show conflicting results [31—34]. There are ethical concerns about the use of an intervention that can cause pain [32,34,35] in children who cannot consent to treatment [28,36,37]. There are no previous systematic reviews or meta-analyses concerning acupuncture treatment for infantile colic. We aimed to assess the efficacy and adverse effects of acupuncture for infantile colic in a systematic review with a meta-analysis based on individual patient data from all eligible randomised controlled trials (RCTs).

Methods

The protocol is registered at the University of York Centre for Reviews and Dissemination — PROSPERO 2015: CRD42015023253 [38] (Supplementary Appendix 1). The unabridged protocol is included in Supplementary Appendix 2. The study has been reported using PRISMA [39,40] and PRISMA-IPD [41] recommendations.

Eligibility criteria

We included full RCTs of acupuncture treatments for infantile colic [42,43]. The participants were infants fulfilling Wessel's criteria or Wessel's modified criteria of infantile colic. There were no exclusion criteria. The intervention was percutaneous needle acupuncture treatment. There were no limitations on variation on doses, intensity, administration or personnel giving the intervention. For comparators, we used no treatment, placebo/sham, standard care or waiting list control. The primary outcome was baseline-corrected differences in crying time in minutes between intervention and control. Secondary outcomes were baseline-corrected differences between intervention and control in

not fulfilling the colic criterion [>180 min crying/day], parental evaluation of effects, and adverse effects. No language restrictions were employed.

Literature search

We decided that a search restricted to English databases might be insufficient for a systematic review about acupuncture [44,45]. We sought assistance from the Chinese Centre for Evidence-Based Medicine, Beijing University of Chinese Medicine to search Chinese language databases. All searches were up to date as of January 2017 (English) and February 2017 (Chinese).

English language database search

Electronic scoping searches were conducted in Best Practice, UpToDate, Cochrane [CDSR, DARE, and HTA] and Prospero from inception to the search date (Supplementary Appendix 3). An electronic search for on-going clinical trials was conducted in ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform Search portal. Electronic searches were performed in Cochrane CENTRAL, MEDLINE, EMBASE, CINAHL and AMED, using both MeSH terms/subject headings and text words in the title/abstract. Truncating and Boolean searching were used. RCT filtering was used at the end of the search. The search terms were: acupuncture; needle acupuncture; infant colic; infantile colic; child; and abdominal pain. The search results were merged in the reference manager software EndNote®, and duplicates and multiplicities were removed. Three authors, two content area experts (H. S. and T. S.) and one methodological expert (A. K.), independently assessed the potential relevance of all titles and abstracts collected through the searches. Relevant articles according to the predefined eligibility criteria were assessed as full text copies. The same three authors independently searched reference lists and one (A. K.) used citation tracking on the Web of Science database. We searched for unpublished and on-going studies by correspondence with field experts. If there was disagreement on which studies to include, this was resolved by discussion, or if agreement could not be reached, by consulting a fourth author (M. B.).

Chinese language database search

Electronic scoping searches were performed in five databases: CNKI, VIP, Wanfang, the Chinese BioMedical Literature Service System [SinoMed] and the Chinese Clinical Trial Registry [ChiCTR], from inception to

search date. An electronic search for on-going clinical trials was conducted to identify Chinese trials, thus bypassing English language bias. A full electronic search in these five databases using the same words was conducted. Infantile colic is not a TCM term. The following search strategy was used, and was revalidated by A. K., using Google translate on Chinese language options, finding that the search was optimal [evidence provided on request]: [针灸 [zhenjiu = acupuncture/moxibustion] OR 针刺 [zhen ci = acupuncture] OR 针 [zhen = needle] OR 刺 [ci = thorn] OR 电针 [dian zhen = electroacupuncture]] AND [[腹痛 [futong = stomach ache] OR 肠痉挛 [chang jingluan = bowel spasm] OR 夜啼 [ye ti = night crying] OR 疝气 [shanqi = colic]] AND [[婴儿(yinger = baby] OR 儿童 [ertong = child] OR 幼儿 [youer = child care] OR 小儿 [xiaoer = children]]. We searched CNKI with MeSH, Wanfang with keywords, VIP with title or keywords and SinoMed common field. One methodological expert (Y. S.) at the Centre for Evidence-Based Chinese Medicine conducted the search and assessed the potential relevance of all titles and abstracts. These were discussed by email and by formal meeting with one content area expert (H. S.) in Beijing in May 2016. Relevant articles according to the predefined eligibility criteria were translated and further discussed with H. S. An updated search was conducted on 17 February 2017 with no changes.

Study selection

To select eligible publications, two authors independently read all titles and abstracts in the records retrieved by the searches. We obtained publications in full text if the abstract was deemed eligible by at least one review author. At least two authors independently read the full text papers and selected studies according to the inclusion criteria. Any disagreement between review authors was resolved by discussion.

Data extraction and management

Our original protocol was constructed for a systematic review and meta-analysis based on aggregated data. We realised that varying strategies for adjustment for baseline imbalances would impair an analysis based on aggregated data. Hence, trialists of eligible trials were invited to take part in a collaborative group and asked to provide their raw data. Before receiving the data, we arranged a consensus meeting at Lund University, Sweden, in February 2017 where trialists representing the eligible trials agreed on the individual patient data (IPD) protocol and defined limits for

clinical important differences in crying time [38]. All datasets were stored securely and pseudo-anonymously; that is, identifiers that could be linked directly to the actual participants were deleted. Once the raw data had been received, K. G. B. checked the data for consistency and comparability with the results presented in the journal papers. Any queries arising from these checks were resolved in co-operation with the trialists. K. G. B., independent of the trialists, performed all analyses, based on the IPD from the included trials and carried out the meta-analysis.

Risk of bias

Two authors (T. S., A. K.) used the Cochrane Collaboration's tool for assessing risk of bias [46]. This tool encourages consideration of how the allocation sequence was generated, how allocation was concealed, the integrity of blinding at outcome level, the completeness of outcome data, and selective outcome reporting and other potential sources of bias. Regarding blinding, we distinguished between performance bias [blinding of participants and personnel] and detection bias [blinding of outcome assessors]. Selective outcome reporting was in general not a problem as we had access to all data from the included studies with IPD. Furthermore, we reduced the risk of bias arising from non-completeness of outcome data for studies with the IPD by using statistical methods that did not exclude participants based on missing data. Each item in the 'Risk of bias' assessment was assessed as low, high or unclear. The quality of the blinding procedure [detection bias-blinding of outcome assessors] was tested for each of the included studies by performing Pearson's chi-squared tests and by calculating James and Bang blinding indices [47]. We performed a sensitivity analysis in which studies assessed to be at high risk of bias [across all items] were excluded. The decision on which studies were considered to be at high risk of bias was taken retrospectively, and guided by the results of the blinding tests.

Data synthesis

Data from the included studies were analysed using a two-step approach [48]. At the first step, we analysed the IPD for each trial separately. For continuous outcomes, the study-level analyses were based on repeated measurements with a reference group coding of independent factors, thus considering the correlations between baseline and post-intervention measurements. Data from all measurement points were

included in a single model. The post-intervention measurements were modelled as depending on the baseline measurement, time, group [intervention or control] and the interaction between time and group. Repeated measurements [from the same person] were assumed to have an unstructured covariance structure. The analyses of data from each of the included trials were conducted using the NLME and CONTRAST libraries in R [49]. For each trial, the estimate of effect at any given measurement point was calculated as the difference between the estimated value of the dependent variable in the intervention and control groups. The corresponding 95% confidence intervals (CI) were also calculated. SMDs were calculated based on the repeated measures standardised to a mean of 0 and a standard deviation of 1. For dichotomous outcomes, we modelled odds ratios [ORs] by logistic regression using the function GLM in R. The results are presented as ORs with 95% CIs and were adjusted for baseline differences in crying time. At the second step, we combined the estimates of effect across studies in the meta-analysis. The estimates of effect from all included studies were pooled using the generic inverse variance technique in a random-effects model in RevMan version 5.3.5.

Measures of treatment effects and harm

We did not detect any papers trying to establish guidelines on clinically relevant changes or minimal important differences (MID) in trials on pain or crying in infants. In large IPD meta-analysis of acupuncture in chronic pain, Vickers et al. considered SMD of 0.2 as being too small to make a clinically meaningful difference, whereas an SMD of 0.5 was considered sufficient [14]. Dworkin et al. suggested that 10—20% reduction in an anchor based numerical rating scale as minimally important reduction of chronic pain, and 30% as moderately important [50]. Furlan et al. established 30% difference on a VAS/NRS of back pain as clinically significant [51]. In studies on children, Carl von Bayer suggested that 10—20% reduction or 10—20 mm on a VAS scale to be the smallest meaningful change [52,53], whereas Dhanani et al. estimated that a minimum of 8 mm on a 100 mm VAS improvement was needed to achieve a meaningful improvement among children with rheumatic disease [54]. Guided by these references, we considered the minimally important difference in baseline-corrected crying time between acupuncture and control to be about 30 min, a number that is roughly equivalent to an SMD of 0.3.

The primary end point was as follows: *Baseline-corrected differences in crying time in minutes between*

intervention and control as measured during treatment one week after treatment ended and one month after treatment ended was the preferred outcome A 30-min difference in reduction in crying time between intervention and control groups was considered as a clinically relevant effect. We changed '1 week after treatment' in the original protocol to 'during the first week after the end of the treatment period' to be able to measure all included trials.

The secondary outcomes were as follows. (A) *The infantile colic 3-h crying criterion*: Baseline-corrected differences between intervention and control in not fulfilling the colic criterion at the end of the treatment period. (B) *Parental evaluation of effects:* Parents' evaluation of improvement in the infant is an important contextual outcome. All studies measured the parental evaluation on the last treatment day using a five-point Likert scale. (C) *Adverse effects:* We registered any serious adverse effects. Minor adverse effects other than crying during treatment were reported descriptively. We specifically wanted to analyse crying during interventions.

Other *Blinding validation of outcome assessors [parents]:* Trials with subjective outcome-effect estimates have been shown to be exaggerated when there was a lack of blinding (ratio of ORs of 0.75 [0.61—0.93]) [43]. Blinding of the outcome assessor is argued to be important in trials with subjective outcomes such as pain [55]. Blinding of the practitioner is not an option in manual treatments of infants [56]. We performed a statistical assessment of blinding validation questions from outcome assessors as registered in the different studies, using both chi-squared tests with ORs and Bang's blinding index with coefficient. The chi-squared test may indicate adequate blinding if $p > 0.05$, but the sensitivity becomes poor if both groups are fully unblinded. Bang's blinding index is calculated for each intervention group separately, and reflects adequate blinding if it centres around 0 [47]. James blinding index was added *post hoc* as we realised it could add information. James' blinding index suggests adequate blinding if it centres around 0.5, but the sensitivity is impaired if the degree of blinding varies between the two groups. All the different blinding tests were taken into account before making any conclusions about the success of blinding.

Assessment of heterogeneity

Analysis of heterogeneity and inconsistencies was performed on all primary and secondary outcomes using chi-squared tests and I^2 analysis to describe the heterogeneity between trials in relation to the total variation.

Fixed and random-effects models

We assumed the random-effects model to be the analysis of choice, representing a valid test of the null hypothesis of no clinically relevant treatment effect of acupuncture for infantile colic. We could not assume one fixed effect irrespective of treatment intensity, duration and point selection for acupuncture in infants, and there were no previous meta-analyses to guide us. There are opinions among acupuncturists that an individualised treatment of infantile colic is the correct one, and that different acupuncture point selections could have different effects [26]. This is contrary to meta-analysis of chronic pain conditions in adults, where a fixed intervention effect of acupuncture in large IPD meta-analysis has proven to approximate the random-effects model, and no significant differences have been shown for different intervention characteristics [14,15].

Subgroup and sensitivity analysis

We did not undertake subgroup analysis. As reported in the full protocol (Supplementary Appendix 2), we performed sensitivity analyses based on the risk of bias in included studies. The result of the blinding validation tests were used to guide the risk of bias assessments in the blinding of outcome assessor domain.

GRADE: Two methods experts (A. K., K. G. B.) assessed the overall quality of evidence according to Grading of Recommendations Assessment, Development and Evaluation [GRADE] [57,58].

Results

Study selection

We identified 384 English language and 24 Chinese language studies after removal of duplicates (Figure 1), but only three studies fulfilled all eligibility criteria. Three English language controlled trials of acupuncture for infantile colic were excluded: one because it was not properly randomised [quasi-randomised] [31]; one reported on data concerning feeding and stooling changes from the same study as reporting on crying time changes [59]; and one was an open pilot study with seven patients and changes during the trial [60]. Individual patient data were sought and obtained for all eligible RCTs
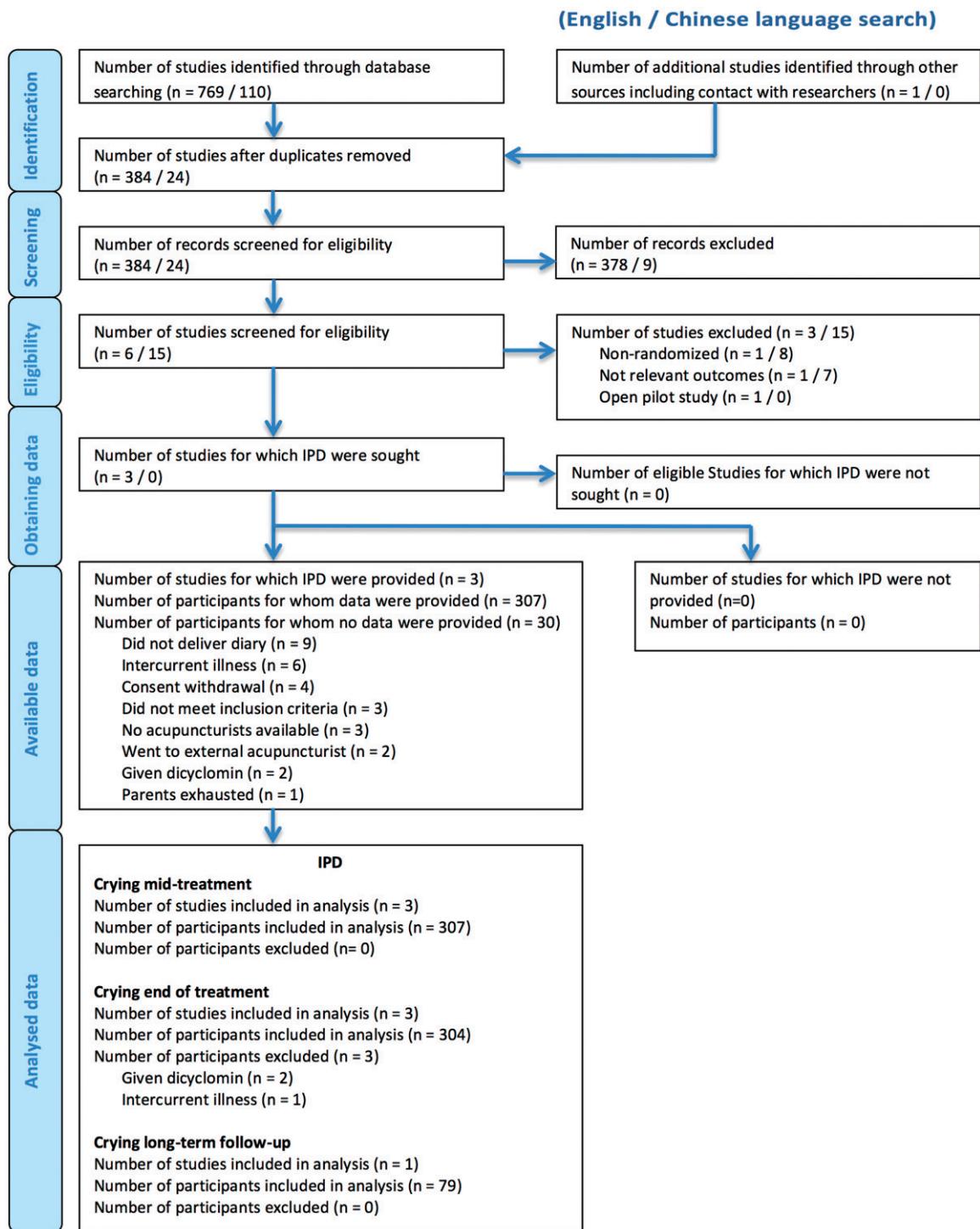
(English / Chinese language search)

Figure 1. PRISMA IPD Flow Diagram.

[32—34]. The characteristics of the included studies are described in detail in Table 1. One trial [34] consisted of two active treatment groups receiving standardised and semi-individualised acupuncture, respectively, and in accordance with the protocol [61], the two active arms were treated as separate comparisons by randomly splitting the control group.

All inconsistencies related to data checking and cleaning were easily resolved following correspondence with the primary authors. For two of the included studies, we report previously unpublished data, i.e. the results of blinding validation in Landgren et al. [32] and the result of parental evaluation of effects in Landgren et al. [32,34].

Table 1. Characteristics of included studies.

Paper Methods	Landgren et al. [32] Sweden RCT	Landgren et al. [34] Sweden Multicentre RCT	Skjeie et al. [33] Norway Multicentre RCT
Participants	Healthy infants 2—8 weeks old, born after week 36, with appropriate weight gain, fulfilling the modified Wessel's criteria of crying/fussing ≥3 h/day for ≥3 days a week. *Exclusion criterion*: medications other than dimethicone or *Lactobacillus reuteri*	Healthy infants, 2—8 weeks old, born after week 36, with appropriate weight gain, crying/fussing ≥3 h/day for ≥3 days at baseline week, after a diet without cow's milk protein either in formulas or from breast-feeding mother's diet ≥5 days. *Exclusion criteria*: Any medication or acupuncture treatment	Healthy infants born at full term and <3 months of age at inclusion. Fulfilling Wessel's criteria of paroxystic uncontrollable crying/fussing ≥3 h/day for ≥3 days a week in ≥3 weeks. No exclusion criteria
Intervention	*Intervention*: Standardised manual acupuncture, unilateral needling of LI4 at 2 mm depth for 2 s *Control*: Identical procedure, except for no needle insertions *Schedule*: two treatments per week for 3 weeks [6 in total]	*Intervention 1*: Standardised manual acupuncture unilaterally at LI4. One needle to a depth of 3 mm unilaterally for 2—5 s before withdrawal without stimulation *Intervention 2*: Semi-standardised individualised acupoint could include Sifeng, LI4 and ST36. Maximum of five insertions up to 30 s with minimal stimulation *Control*: Identical procedure, except for no needle insertions *Schedule*: two treatments per week for 2 weeks [4 in total]	*Intervention*: Standardised manual acupuncture, ST36 needled bilaterally to 12 mm depth without manipulation for 30 s. Point mark 3 mm, and waterproof circular adhesive dressing applied to hide insertion and erythema *Control*: Identical procedure, except for no needle insertions *Schedule*: 1 treatment each day for 3 consecutive days
Outcomes	*Primary*: Relative difference in the number of infants fulfilling colic criteria. *Secondary*: Difference in total crying time during the 3 intervention weeks, and adverse effects	*Primary*: Difference in total crying time at end of treatment *Secondary*: Relative difference in the number of infants fulfilling colic criteria at end of each intervention week, parents' assessment of the child and adverse effects	*Primary*: Difference in total crying time. *Secondary*: Relative difference in the number of infants fulfilling colic criteria, parents' assessment of the child and adverse effects

Figure 2. Risk of bias summary report.

Risk of bias and blinding validation

The risk of bias summary is presented in Figure 2. Detailed risks of bias assessments are available in Supplementary Appendix 4. Briefly, all included RCTs had adequate randomisation procedures reported and allocation concealment described (Supplementary Appendix 4). Acupuncturists were not blinded in any

of the studies. Parents acted as outcome assessors in all studies, and a thorough blinding validation (Table 2) indicated that the parents in Landgren et al.'s study from 2010 [32] seemed to be unblinded to treatment allocation. In contrast, Landgren et al. in 2017 [34] achieved blinding outcome assessment in the control group, but not in the acupuncture group, whereas Skjeie et al. [33] were able to mask all parents irrespective of the group to which the infant was allocated. The interpretation of these results is complicated because the timing of the blinding validation varied between the studies. Parents in Landgren et al. [32] were asked about allocation beliefs after the last treatment session. In Skjeie et al. [33], parents were questioned about allocation beliefs after the first treatment session, and after 4 weeks they were asked if they had noticed any needle marks. Landgren et al. [34] questioned parents about allocation beliefs after the second, the third and the fourth treatment session, as well as follow up. We decided to calculate blinding indexes obtained after the last treatment session to facilitate comparison with Landgren et al. [32], but we also calculated blinding indexes as they appeared after the second treatment session. During standardized acupuncture the quality of the blinding seemed to decrease throughout the intervention period, whereas individualised

Table 2. Blinding index with three different blinding tests.

		Blinding index			
Study	n	James	Bang [Acupuncture]	Bang [Control]	Chi-squared test p
Landgren et al. [32]	43	0.21[a]	0.77[a]	0.67[a]	0.87[b]
Landgren et al. [34] (1)	96	0.56[b]	0.44[a]	−0.08[b]	0.007[a]
Landgren et al. [34] (2)	96	0.51[b]	0.63[a]	−0.08[b]	<0.0001[a]
Skjeie et al. [33]	83	0.46[b]	0.00[b]	0.18[c]	0.55[b]

[a]Not blinded for outcome assessors [parents].
[b]Blinded for outcome assessors.
[c]Unsure blinding of outcome assessors.

Table 3. Summary of findings for primary outcome: differences in crying time.

	Average min. crying per day		SMD		
Timing	Control	Acupuncture	[95% CI]	No. of participants	Quality of evidence
Mid-treatment	193 min	25 min less [46 less to 4 less]	−0.23 [−0.42 to −0.05]	307 [3 studies]	⊕⊕⊕⊙ Moderate[a]
End of treatment	156 min	11 min less [32 less to 9 more]	−0.10 [−0.29 to 0.08]	304 [3 studies]	⊕⊕⊕⊙ Moderate[a]
Long-term follow-up	97 min	12 min less [63 less to 39 more]	−0.09 [−0.48 to 0.30]	79 [1 study]	⊕⊕⊙⊙ Low[a,b]

[a]Wide confidence intervals [CI] and imprecision.
[b]One study with few participants. SMD, standardised mean differences.

Table 4. Summary of findings for secondary outcomes.

	Absolute effect per 100 [95% CI]		Relative effect OR [95% CI]	No. of participants	Quality of evidence
Outcome	Control	Acupuncture			
No colic[a]	60	70 [57—80]	1.54 [0.88—2.70]	304 [3 studies]	⊕⊕⊕⊙ Moderate[b]
Much improvement	26	52 [35—67]	3.03 [1.56—5.89]	264 [3 studies]	⊕⊕⊕⊙ Moderate[b]
Some improvement	65	83 [73—90]	2.67 [1.43—4.97]	264 [3 studies]	⊕⊕⊕⊙ Moderate[b]
Worsening	30	26 [9—58]	0.83 [0.22—3.18]	46 [1 study]	⊕⊕⊙⊙ Low[b,c]
Crying during treatment	37	81 [61—92]	7.5 [2.7—20.6]	81 [1 study]	⊕⊕⊕⊙ Moderate[b]

[a]Defined as >180 min of crying per day.
[b]Wide confidence intervals [CI] and imprecision.
[c]Not possible to estimate in two studies because of too few events.

acupuncture was associated with fairly stable blinding indexes.

End-points

In the following, we report primary and secondary end-points in accordance with the protocol for this review (Supplementary Appendices 1 and 2). Tables 3 and 4 summarises findings for all investigated outcomes and shows that the quality of evidence was rated as moderate for most end-points. Because the heterogeneity was negligible in all analysis, the major reasons for downgrading were few participants and wide confidence intervals.

Reduction in crying time

We did not detect important differences in crying time between acupuncture and no acupuncture control at any of the pre-specified time periods (Figure 3). There

was a statistically significant difference in mean crying time [MD −24.88 min/day; 95% CI −46.20 to −3.57] and SMD [−0.23; 95% CI −0.42 to −0.05] at mid-treatment, but this was lost [MD −13.82; 95% CI −37.50 to 9.86] and [SMD −0.13; 95% CI −0.35 to 0.09] when the study assessed as unblinded was excluded in a sensitivity analysis (Figure 4).

Disappearance of colic

We did not detect statistically significant differences between acupuncture and no acupuncture control groups when comparing the odds of not fulfilling the colic criterion at the end of the treatment [OR 1.54; 95% CI 0.88—2.70] (Figure 5).
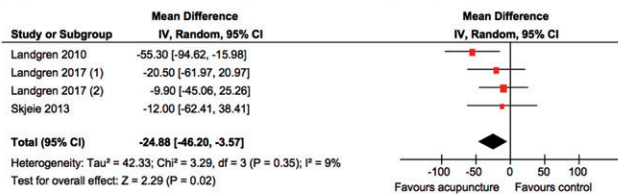
Parental evaluation

Parents of the infants in the acupuncture groups were more likely to report that the colic had improved at
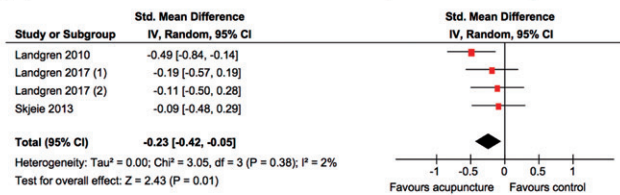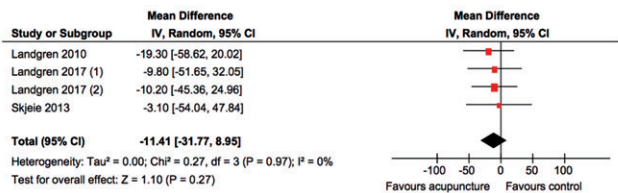
Primary end-point: crying time
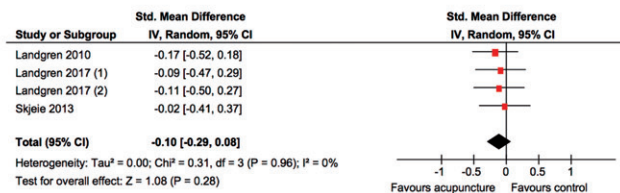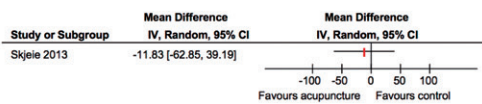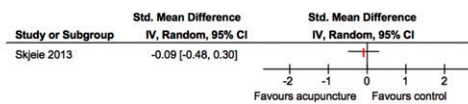
Figure 3. Primary end-point.
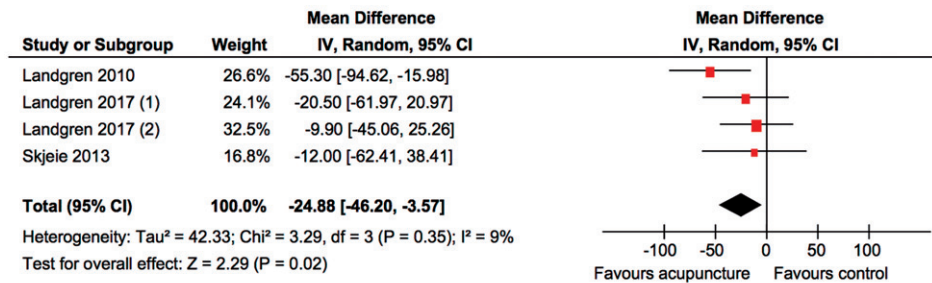
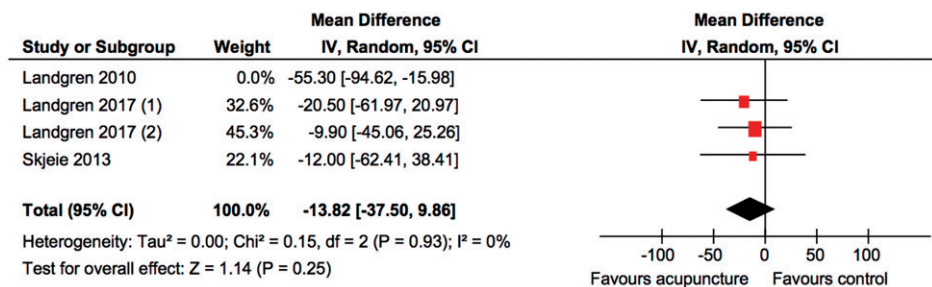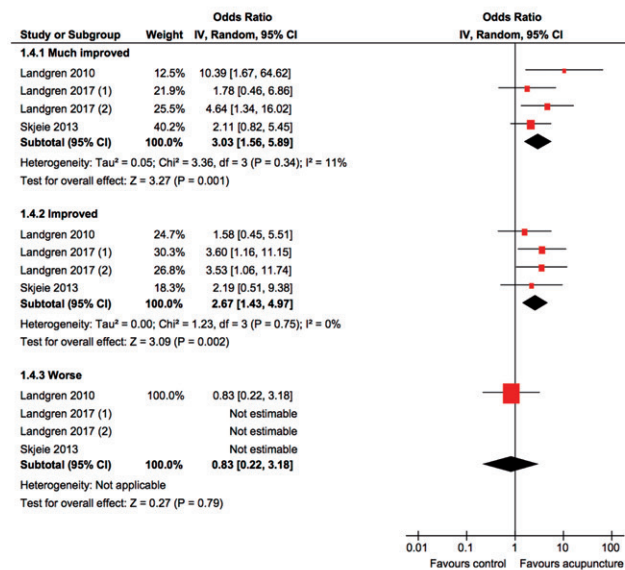**Figure 4.** Sensitivity analysis.

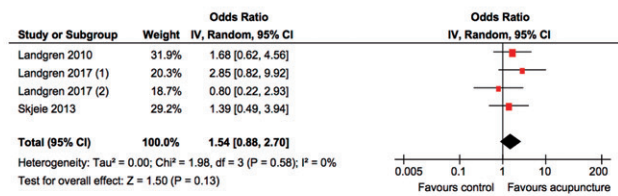the end of the treatment (Figure 5), with an OR of 3.03 (95% CI 1.56—5.89) for rating the condition as much improved and OR 2.67 (95% CI 1.43—4.97) for improved. The odds ratio for worsening was only available from the Landgren et al. [32] trial (OR 0.83; 95% CI 0.22—3.18).
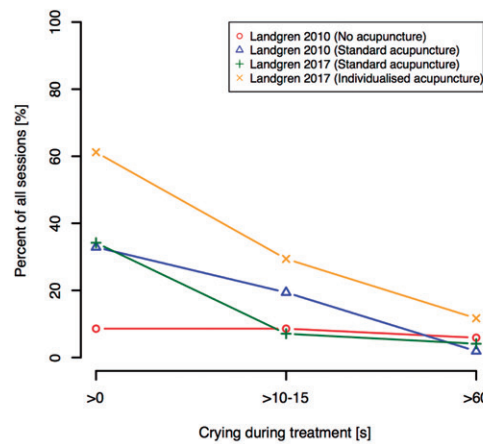
Figure 5. Secondary end-points.

Adverse effects

No major adverse effects were reported in the included trials. With regard to minor adverse effects other than crying during treatment, Landgren et al. [32] observed one minor bleeding in the acupuncture group. Skjeie et al. [33] observed two possible adverse events in the acupuncture group (hiccups, increased regurgitation), and eight events in the control group (small haematoma, restlessness, restlessness, excessive stools, frequent defecation, light sedation, abdominal pain and unease). Landgren et al. [34] registered one drop of blood on clothes and one mark on a hand, both in the acupuncture groups. Crying during treatment was assessed by Landgren et al. in 2010 [32] and 2017 [34]. In the first study, crying during treatment was assessed in both the acupuncture and the no acupuncture control group, and showed that infants in the acupuncture group were more likely to cry during treatment (OR 7.50; 95% CI 2.73—20.64), although the majority stopped crying within seconds [24] (Figure 5). In the second study, crying during treatment was assessed in the two acupuncture groups, but not in the control group, and showed some signs that crying occurred more frequently during semi-individualised acupuncture (up to five needles) than during standardised acupuncture (one needle) [25]. The OR was 2.53

(95% CI 0.72—8.86), but most infants also stopped crying within seconds in this study (Figure 5).

Discussion

This is the first IPD meta-analysis of acupuncture treatments of any condition in small children, and we included three trials where 307 infants with colic were randomised to receive either acupuncture or no acupuncture control. All included trials tested minimal acupuncture: 1—5 insertions for 2—30 s, and the results of the included trials were in general consistent. Considering our primary end-point, that is, total crying time, we detected a small change in favour of acupuncture at mid-treatment, but the significance was lost when the apparently unblinded study was left out in a sensitivity analysis. We did not find important differences in crying time between acupuncture and no acupuncture control measured after the treatment ended. At the long-term follow-up, we did not see a statistically significant difference between acupuncture and no acupuncture control. No major adverse effects occurred, but acupuncture induced some crying during treatments. For the other secondary outcomes, we were not able to detect a statistically significant difference in the number of infants who did not fulfil the colic criterion [>180 min crying/day] at the end of the

treatment period. Interestingly, however, parents of infants in acupuncture groups more frequently evaluated the colic symptoms as improved than parents in the control groups.

The positive results from parental evaluation contradict the lack of important differences in total crying times, and it is tempting to speculate why. One possible explanation relates to the type of crying. Not all crying is colicky crying, and it is possible that acupuncture changes the quality of crying or the degree to which the infant can be soothed. Such a change could be sensed by parents without being detected by crying time assessments. Another possible explanation is that acupuncture works by other means than crying time reduction, but the observation that acupuncture was not associated with changes in frequency of feeding, stooling and sleeping, as compared with controls does not support this hypothesis [59]. However, more subjective outcomes like 'normalised stooling' were reported more frequently among parents in the acupuncture group. The tendency towards more positive results on more subjective outcomes might of course be related to inadequate blinding, but ad hoc sensitivity analysis based on blinding validation data did not suggest such a relationship.

Despite our efforts to achieve thorough blinding validation of all included studies, we cannot be sure that existing blinding tests are sensitive enough to detect all relevant differences [47] Depending on when parents are asked blinding questions, the validity of the blinding tests can also be impaired by differences in efficacy between groups. Data from the trial that asked blinding questions at multiple time point [34] and small, non-significant improvements at end of treatment, does not support that the timing of the blinding questioning is essential for the results reported here, nor that unblinding because of efficacy are probable. The effect sizes [SMD 0.02—0.09—0.11—0.17], are so small that they would not normally be possible to detect. The differences of Parent evaluation of effect between the apparently unblinded study [32] with odds ratio 10.4, and the partially blinded study [34] with odds ratio 1.8, both using the same treatment method, would normally suggest influence of unblinding.

A major strength with this IPD-based analysis is that we could include all properly randomised controlled trials about acupuncture for infantile colic, but there are few trials and the number of participants in each trial was small. The results should be evaluated in this context [58,62]. All trials tested acupuncture treatment versus no acupuncture, thereby eliminating heterogeneity problems caused by different control group regimens [14]. Between-study heterogeneity was

negligible, adding robustness to the results. Our blinding validations also revealed some differences between the included trials. One trial did not seem to be blinded at all [32], one seemed to be blinded [33], and the third had managed to blind participants in the control group whereas participants in the acupuncture group remained unblinded [34]. If the results had shown an effect of treatment, this would have lowered our confidence in the results. Our study did not show clinically relevant results in spite of inadequate blinding, which normally favours active intervention [43].

We are aware of some other studies and reviews that need to be discussed in the light of the results presented in our review. First, one small controlled trial on acupuncture for infantile colic was excluded because it was not an adequately randomised trial [31,63]. Even though Reinthal et al. [31] presented results in favour of acupuncture, it should be recognised that that study was associated with substantial baseline differences between the groups that were not adjusted for. Although we intended to include that study [31] in our sensitivity analysis, differences in outcome reporting prevented us from doing so. Second, we identified an existing systematic review about acupuncture in infants [19], in which the authors conclude: 'In summary current evidence suggests that acupuncture is safe, effective and a cheap method to treat infantile colic'. This positive conclusion contradicts the results presented here in our review. Many factors can explain this difference. One crucial difference is that new evidence has accumulated [33,34], but we also show that it is important to adjust for baseline differences between the treatment groups. When using individual patient data to correct for baseline differences, the heterogeneity between the trials included in our meta-analysis seemed to disappear. Additionally, the differences in blinding regimens may explain some of the variation in results seen across trials.

A large IPD meta-analysis based on 17,922 adults receiving acupuncture for chronic pain found a needle-specific treatment effect SMD of 0.15—0.23 [14,15] between real and sham acupuncture, which is remarkably similar to the SMDs reported in our review [0.09—0.23]. Effect sizes of this small magnitude are not considered clinically relevant for pain conditions in either adults [50,51,64,65] or children [52—54,66]. Three different needling techniques were used in the trials included in our review, but with similar results. These observations correspond well to the large-scale IPD review of acupuncture in adults where it was reported that differences in techniques or acupuncture points did not have an impact on the results [14,15]. The same large-scale IPD study [15] highlighted an association

between the number of acupuncture needles applied and the effect of treatment. The minimal acupuncture method used in all trials of infantile colic could reduce the difference between acupuncture and no acupuncture treatment controls in the primary outcome, that is, differences in crying time. However, ethical concerns of potential needle pain, validated in this IPD, would prove an obstacle to more intense needle treatment.

When planning new RCTs about the efficacy of percutaneous needle acupuncture for infants, we consider it important to facilitate adequate blinding of outcome assessors, and to validate the quality of any blinding procedures. The risk of causing pain should be evaluated and taken into consideration when assessing the benefits and harms of acupuncture in small children who cannot consent to treatment [28,37,67]. Our findings also emphasise the need for more quantitative and also qualitative [68,69] research to explore parents' experiences and the possible positive impact of acupuncture on outcomes other than reduction in crying time. Defining thresholds for clinically important differences is important for evaluating implications of results, and should be put into practice in all acupuncture trials of pain conditions in children.

Conclusions

Our blinding test validated IPD meta-analysis of minimal acupuncture treatments of infantile colic did not show clinically relevant effects in pain reduction as estimated by differences in crying time between needle acupuncture intervention and no acupuncture control. Analyses indicated that acupuncture treatment induced crying in many of the children. Caution should therefore be exercised in recommending potentially painful treatments with uncertain efficacy in infants. The studies are few, the analysis is made on small samples of individuals, and conclusions should be considered in this context. With this limitation in mind, our findings do not support the idea that percutaneous needle acupuncture should be recommended for treatment of infantile colic on a general basis.

Acknowledgements

Ethics approval

Not required. This is a systematic review and IPD meta-analysis based on published trials. Local ethics committees have approved all included trials.

Disclosure statement

Funding

Notes on contributors

Holgeir Skjeie is a GP specialist and PhD student at the Department of General Practice, Institute of Health and Society, University of Oslo, Norway.

Trygve Skonnord is a GP specialist and PhD student at the Department of General Practice, Institute of Health and Society, University of Oslo, Norway.

Mette Brekke is a professor at the Department of General Practice, Institute of Health and Society, University of Oslo, Norway.

Atle klovning is an associate professor at the Department of General Practice, Institute of Health and Society, University of Oslo, Norway.

Arne Fetveit is an associate professor at the Department of General Practice, Institute of Health and Society, University of Oslo, Norway.

Kajsa Landgren is a lecturer and PhD at the Faculty of Medicine, Department of Health Sciences, Lund University, Lund, Sweden.

Inger Kristensson Hallstrom is a professor at the Faculty of Medicine, Department of Health Sciences, Lund University, Lund, Sweden.

Kjetil Gundro Brurberg is a senior researcher at the Norwegian institute of Public Health, Division for Health Services, Oslo, Norway, and an associate professor at the Wesern Norway University of Applied Sciences, Centre of Evidence Based Practice, Bergen, Norway.

References

[1] Wessel MA, Cobb JC, Jackson EB, et al. Paroxysmal fussing in infancy, sometimes called colic. Pediatrics. 1954;14:421–435.

[2] The Rome Foundation. 2016. Available from: http://theromefoundation.org/and http://www.webcitation.org/6mG5pslnl.

[3] Hyman PE, Milla PJ, Benninga MA, et al. Childhood functional gastrointestinal disorders: neonate/toddler. Gastroenterology. 2006;130:1519–1526.

[4] Benninga MA, Faure C, Hyman PE, et al. Childhood functional gastrointestinal disorders: neonate/toddler. Gastroenterology. 2016;150:1443–1455.e2.

[5] Drug and Therapeutics Bulletin. Management of infantile colic. BMJ. 2013;347:f4102.

[6] Gupta SK. Update on infantile colic and management options. Curr Opin Investig Drugs. 2007;8:921–926.

[7] Savino F. Focus on infantile colic. Acta Paediatr. 2007;96:1259–1264.

[8] Savino F, Ceratto S, De Marco A, et al. Looking for new treatments of infantile colic. Ital J Pediatr. 2014;40:53.

[9] Lucassen P. Colic in infants. BMJ Clin Evid. 2010;2010:pii0309.

[10] Savino FTV, Sorrenti M, Lingua C, et al. Dietary modifications for infantile colic. Cochrane Database Syst Rev. 2014; 3:CD011029.

[11] Kaptchuk TJ. Acupuncture: theory, efficacy, and practice. Ann Intern Med. 2002;136:374–383.

[12] NIH Consensus Conference. Acupuncture. JAMA. 1998;280:1518–1524.

[13] Chen L, Michalsen A. Management of chronic pain using complementary and integrative medicine. BMJ. 2017;357:j1284

[14] Vickers AJ, Cronin AM, Maschino AC, et al. Acupuncture for chronic pain: individual patient data meta-analysis. Arch Intern Med. 2012;172:1444–1453.

[15] MacPherson H, Maschino AC, Lewith G, et al. Characteristics of acupuncture treatment associated with outcome: an individual patient meta-analysis of 17,922 patients with chronic pain in randomised controlled trials. PLoS One. 2013;8:e77438.

[16] White A. A cumulative review of the range and incidence of significant adverse events associated with acupuncture. Acupunct Med. 2004;22:122–133.

[17] Adams D, Cheng F, Jou H, et al. The safety of pediatric acupuncture: a systematic review. Pediatrics. 2011;128:e1575–e1587.

[18] Gold JI, Nicolaou CD, Belmont KA, et al. Pediatric acupuncture: a review of clinical research. Evid Based Complement Alternat Med. 2009;6:429–439.

[19] Raith W, Urlesberger B, Schmolzer GM. Efficacy and safety of acupuncture in preterm and term infants. Evid Based Complement Alternat Med. 2013;2013: 739414.

[20] Scott J, Barlow T. Acupuncture in the treatment of Children. Seattle (WA): Eastland Press; 1999. 628 p.

[21] Loo M. Pediatric acupuncture. London: Elsevier, Churchill Livingstone; 2002. p. 96.

[22] Birch S. Shonishin: Japanese pediatric Acupuncture. Stuttgart: Georg Thieme; 2011. p. 3.

[23] Rossi E. Pediatrics in Chinese medicine. Milan: Donica Publishing Ltd.; 2011. p. 145–150.

[24] Gentry KR, McGinn KL, Kundu A, et al. Acupuncture therapy for infants: a preliminary report on reasons for consultation, feasibility, and tolerability. Paediatr Anaesth. 2012;22:690–695.

[25] Reinthal M, Lund I, Ullman D, et al. Gastrointestinal symptoms of infantile colic and their change after light needling of acupuncture: a case series study of 913 infants. Chin Med. 2011;6:28.

[26] Landgren K. Acupuncture in practice: investigating acupuncturists' approach to treating infantile colic. Evid-Based Complement Altern Med. 2013;2013: 456712.

[27] Skjeie H, Brekke M. 'Big needles, small bodies'-the absence of acupuncture treatment for infants in contemporary Shanghai: a qualitative study. BMJ Open. 2015;5:e009486.

[28] Gottsegen DN. Ethical integrative pediatric care: a new perspective. Eur J Integr Med. 2012;4:e1.

[29] Brittner M, Le Pertel N, Gold MA. Acupuncture in pediatrics. Curr Probl Pediatr Adolesc Health Care. 2016;46:179–183.

[30] Takahashi T. Acupuncture for functional gastrointestinal disorders. J Gastroenterol. 2006;41:408–417.

[31] Reinthal M, Andersson S, Gustafsson M, et al. Effects of minimal acupuncture in children with infantile colic – a prospective, quasi-randomised single blind controlled trial. Acupunct Med. 2008;26:171–182.

[32] Landgren K, Kvorning N, Hallström I. Acupuncture reduces crying in infants with infantile colic: a randomised, controlled, blind clinical study. Acupunct Med. 2010;28:174–179.

[33] Skjeie H, Skonnord T, Fetveit A, et al. Acupuncture for infantile colic: a blinding-validated, randomized controlled multicentre trial in general practice. Scand J Primary Health Care. 2013;31:190–196.

[34] Landgren K, Hallström I. Effect of minimal acupuncture for infantile colic: a multicentre, three-armed, single-blind, randomised controlled trial [ACU-COL]. Acupunct Med. 2017;35:171–179.

[35] Zhao X, Chen M, Du S, et al. Evaluation of stress and pain in young children with cerebral palsy during early developmental intervention programs: a descriptive study. Am J Phys Med Rehabil. 2015;94:169–175. quiz 176–179:

[36] Gilmour J, Harrison C, Cohen MH, et al. Pediatric use of complementary and alternative medicine: legal, ethical, and clinical issues in decision-making. Pediatrics. 2011;128:S149–S154.

[37] Markestad T. [Physicians alternative]. Tidsskr Nor Laegeforen. 2012;132:2409–2410.

[38] Skjeie H, Skonnord T, Klovning A, et al. Acupuncture treatments for infantile colic: a protocol for a systematic review and Individual Patient Data [IPD] metaanalysis of randomized controlled trials [RCT]. [Internet]; 2017. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD420150-23253.

[39] Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339: b2700.

[40] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6:e1000097.

[41] Stewart LA, Clarke M, Rovers M, et al. Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD statement. JAMA. 2015;313:1657–1665.

[42] Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA.1995;273:408–412.

[43] Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336:601–605.

[44] Cohen JF, Korevaar DA, Wang J, et al. Should we search Chinese biomedical databases when performing systematic reviews? Syst Rev. 2015;4:23.

[45] Xia J, Wright J, Adams CE. Five large Chinese biomedical bibliographic databases: accessibility and coverage. Health Info Libr J. 2008;25:55–61.

[46] Higgins J, Green S, eds. Cochrane handbook for systematic reviews of interventions. Chichester, West Sussex, England: Wiley-Blackwell; 2008. p. 187–241.

[47] Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. Control Clin Trials. 2004;25:143–156.

[48] Riley RD, Lambert PC, Staessen JA, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Stat Med. 2008;27:1870–1893.

[49] R Core Team. A language and environment for statistical computing. Vienna, Austria:R Foundation for Statistical Computing; 2015. Available from: https://www.R-project.org/.

[50] Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008;9:105–121.

[51] Furlan AD, Pennick V, Bombardier C, et al. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine [Phila, PA 1976]. 2009;34:1929–1941.

[52] von Baeyer CL. Children's self-reports of pain intensity: scale selection, limitations and interpretation. Pain Res Manag. 2006;11:157–162.

[53] von Baeyer CL. Children's self-report of pain intensity: what we know, where we are headed. Pain Res Manag. 2009;14:39–45.

[54] Dhanani S, Quenneville J, Perron M, et al. Minimal difference in pain associated with change in quality of life in children with rheumatic disease. Arthr Rheum. 2002;47:501–505.

[55] Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. BMJ. 2000;321:504.

[56] Dobson D, Lucassen PL, Miller JJ, et al. Manipulative therapies for infantile colic. Cochrane Database Syst Rev. 2012;12:CD004796.

[57] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–926.

[58] Guyatt GH, Oxman AD, Kunz R, et al. What is "quality of evidence and why is it important to clinicians?" BMJ. 2008;336:995–998.

[59] Landgren K, Kvorning N, Hallström I. Feeding, stooling and sleeping patterns in infants with colic – a randomized controlled trial of minimal acupuncture. BMC Complement Altern Med. 2011;11:93.

[60] Skjeie H, Skonnord T, Fetveit A, et al. A pilot study of ST36 acupuncture for infantile colic. Acupunct Med. 2011;29:103–107.

[61] Landgren K, Tiberg I, Hallström I. Standardized minimal acupuncture, individualized acupuncture, and no acupuncture for infantile colic: study protocol for a multicenter randomized controlled trial – ACU-COL. BMC Complement Altern Med. 2015;15:325.

[62] Higgins J, Green S, eds. Cochrane handbook for systematic reviews of interventions. Chichester, West Sussex, England: Wiley-Blackwell; 2008. p. 361–367.

[63] Higgins J, Green S, eds. Cochrane handbook for systematic reviews of interventions. Chichester, West Sussex, England: Wiley-Blackwell; 2008. p. 392–396.

[64] Madsen MV, Gotzsche PC, Hrobjartsson A. Acupuncture treatment for pain: systematic review of randomised clinical trials with acupuncture, placebo acupuncture, and no acupuncture groups. BMJ. 2009;338:a3115.

[65] Hinman RS, McCrory P, Pirotta M, , et al. Acupuncture for chronic knee pain: a randomized clinical trial. JAMA. 2014;312:1313–1322.

[66] Powell CV, Kelly AM, Williams A. Determining the minimum clinically significant difference in visual analog pain score for children. Ann Emerg Med. 2001;37:28–31.

[67] Beuchamp TL, Childress JF. Principles of biomedical ethics. Oxford: Oxford University Press; 2009. p. 90–91.

[68] Landgren K, Hallström I. Parents' experience of living with a baby with infantile colic – a phenomenological hermeneutic study. Scand J Caring Sci. 2011;25:317–324.

[69] Landgren K, Lundqvist A, Hallstrom I. Remembering the chaos – but life went on and the wound healed: a four year follow up with parents having had a baby with infantile colic. Open Nurs J. 2012;6:53–61.