Acupuncture treatments for infantile colic

PhD thesis

Holgeir Skjeie

Department of General practice, Institute of Health and Society,

Faculty of Medicine, University of Oslo

2017
“In science, it’s not a sin to change your mind when the evidence demands it. For some people, the tribe is more important than the truth. For the best scientists, the truth is more important than the tribe.”

Joel Achenbach, National Geographic, March 2015
4.3. METHODOLOGICAL CONSIDERATIONS .................................................................................. 70

4.3.1. Relevant methodological considerations in controlled clinical trials of acupuncture and pain ...................................................................................................................... 70
4.3.2. Papers I and II. The pilot study and the main RCT .......................................................... 72
4.3.3. Paper III. The systematic review and IPD meta-analysis .................................................... 74
4.3.4. Paper IV: “Big needles, small bodies”. The Shanghai qualitative study ......................... 76

4.4. LIMITATIONS AND STRENGTHS IN PAPERS I–IV ............................................................. 78

4.4.1. Overall considerations ........................................................................................................ 78
4.4.2. Papers I and II. The pilot study and the main RCT .......................................................... 79
4.4.3. Paper III. The systematic review and the IPD meta-analysis ............................................. 81
4.4.4. Paper IV. “Big needles, small bodies”: the Shanghai qualitative study ......................... 83

4.5. ETHICAL CONSIDERATIONS ............................................................................................... 85

4.5.1. Ethical considerations of percutaneous needle acupuncture treatment in small children .. 85
4.5.2. The pilot study and the main RCT ................................................................................... 86
4.5.3. The systematic review and IPD meta-analysis ................................................................. 87
4.5.4. “Big needles, small bodies”: the Shanghai qualitative study ......................................... 87

4.6. DISCUSSION OF RESULTS .................................................................................................. 87

5. CLINICAL IMPLICATIONS ....................................................................................................... 93

6. FUTURE RESEARCH .............................................................................................................. 94

7. CONCLUDING REMARKS ..................................................................................................... 95

LIST OF REFERENCES .............................................................................................................. 96

APPENDIX 1. REK ETHICS APPROVAL PILOT AND RCT (NORWEGIAN) ................................. 118

APPENDIX 2. NSD APPROVAL PILOT AND RCT (NORWEGIAN) .............................................. 120

APPENDIX 3. ETHICAL COUNCIL OF THE NORWEGIAN MEDICAL ASSOCIATION STATEMENT – CASE 42013 (NORWEGIAN) .............................................................................. 123

APPENDIX 4. PARENT CONSENT FORM RCT (NORWEGIAN) .................................................... 126

APPENDIX 5. CRYING TIME DIARY FORM RCT (NORWEGIAN) .................................................. 131

APPENDIX 6. PARENT INTERVIEW FORM RCT (NORWEGIAN) ................................................... 135

APPENDIX 7. REK-CONSIDERATION ON THE QUALITATIVE STUDY (NORWEGIAN) ............... 139

APPENDIX 8. INFORMATION AND CONSENT FORM – PARTICIPATING OBSERVATION – THE QUALITATIVE STUDY ........................................................................................................ 142

PAPER I–IV ................................................................................................................................ 144
Acknowledgements

I shall always be grateful for the warmth, interest, open-mindedness and inclusiveness all the people at the Department of General Practice, Institute of Health and Society, University of Oslo, headed by professor Jørund Straand, have given me all these years while I have drifted back and forth from the south coast to the capital in a prolonged quest for answers to the question of whether it is sound practice to insert thin steel needles into the bodies of small babies who cries of colic in order to make them cry less.

A heartfelt thank you to my main supervisor, professor Mette Brekke, who guided me with utmost care and patience for many years through methodological unchartered waters, to always arrive safe on shore. And to my co-supervisor, associate professor Arne Fetveit, the steady hand and the finishing line.

Lots of regards to my good friend and best research buddy Trygve Skonnord. Without him I would never have been able to master the considerable administrative and logistical challenge of conducting RCTs in clinical general practice. We have been the Acupuncture Boys at the department, running RCTs and having lots of fun. My man.

To Han Chouping and Heng Li in Shanghai. The best. You took me in and you never failed to deliver. Why, I do not know. But maybe trust is always the beginning.

To Kajsa Landgren, Grand Duchess of acupuncture of infantile colic in Scandinavia, who had the grace and the scientific preparedness to give full access to all data she had, to get to the answers.

To Kjetil Gundro Brurberg, who in spite of it all, delivered the IPD.

To Atle Klovning, best EBM in town, who insisted on the Chinese literature search, and delivered it.
To Oscar Heyerdahl and Nils Lystad, my mentors, who taught me acupuncture, taught me the philosophy of TCM, taught me the thin line of facts and fiction in a strange field. After almost thirty years my conclusions differ from theirs, but that is the way it goes.

Always in debt to Ann Kristin Grønvik, my head of office at Bystranda Legesenter, my research assistant, and the one who fixes everything. Every time.

To my peers, my fellow GP/acupuncturists at the former Reference group of Medical Acupuncture in NFA who did all the work of carrying out the RCT. No pay. Just hard work.

To my wife Lena Fatland. She questioned my priorities, but she stood by all the years. She encouraged me, listened to my doubts, heard my stories, and she is still here.

To my children, Hanna, Elias and Johannes, who went from children to adults in this time. They missed a lot of story-telling, boat rides and fishing trips on the way, because their father had to do his thing.

And at the end a thanks to my old brother in arms Bozidar Gardasevic, who offered me this Montenegrin saying: “A man must do three things. He must get married. He must have son. And he must write book.”

I have wife. I have sons. Now I have book.
Preface

My career as an acupuncturist is as long as my professional life as a GP. Over the years, I have practiced acupuncture at all levels, based on a variety of areas and indications. My clinical work has been based on textbooks in Traditional Chinese medicine (TCM), numerous clinical courses, and my own growing experience. I have taught acupuncture at the Norwegian Society of Medical Acupuncture from 2000 to 2013. Since 2009, I have done scientific research on the efficacy of acupuncture in infantile colic.

From when I started out in 1991, until today, the evidence base of acupuncture has changed fundamentally. The predominant opinion for many years, that acupuncture is a cure for all ills, has been opposed through a growing bulk of scientific evidence. Acupuncture today seems to be primarily a remedy for pain conditions in adults, with a small needle specific effect, and an overall medium effect size.

Acupuncture performed on children in the Western world is a story of ambiguity. It seems as if doctrines have been established based on a few well-spoken and out-spoken scholars of acupuncture who have translated TCM to encompass the treating of children's ills and diseases with needles. The lore is that acupuncture in children works fast, and works better, than in adults. The lack of evidence has for a long time been an ethical problem in treatment of children with acupuncture. Children, and especially small children, cannot consent to treatment. According to the four core principles of biomedical ethics: i) autonomy, ii) non-malevolence, iii) benevolence, and iv) justice, utmost care should be taken before undertaking potential painful medical treatment on vulnerable subjects. The evidence should be strong, and stronger than in consenting adults.

The basis of Western use of acupuncture in children has been based on what has been told to be TCM tradition and practice, as it originated and developed in China. The indications for
acupuncture in western textbooks are widespread, and encompass everything from sore throat to genetic disorders. The basis for the recommendations seems to be a mixture of various interpretations of the TCM tradition of diagnostic differentiation and treatment, and the author's own practicing experience.

However, lack of evidence of these widespread recommendations give rise to critical questions:

What if the small specific efficacy of the needle in clinical trials of acupuncture in adults are the same in infants and small children?
What if the efficacy of acupuncture in small children does not outweigh the needle pain?
What if needle acupuncture of infantile colic is a Western construct?

It is in this light that I wanted to test the practice of acupuncture treatment for infantile colic, which has been widely recommended because of its efficacy, and which I personally have practiced on my own infant patients for many years. My simple and primary research question was: Does it work?
Funding

The PhD project was financed through a grant from the Norwegian Research Fund for General Practice.
Abstract

Introduction

Infantile colic is a functional symptom diagnosis and a probable painful condition of the gastro-intestinal tract in the first months of infancy. Percutaneous needle acupuncture in small children has gained some acceptance in Western medicine. Needle acupuncture, originating from China, now uses sterile thin steel needles inserted through the skin and into the soft tissues to obtain the treatment effect. Acupuncture is used in Scandinavia and in other Western countries as a treatment for infantile colic. The treatment is controversial, as infants and toddlers are unable to consent to treatment. We aimed to evaluate the method and the efficacy of acupuncture treatments for infantile colic.

Method

We did a pilot study to assess the feasibility of our proposed design of a trial to assess the efficacy of percutaneous needle acupuncture.

We carried out a prospective, blinding-validated, randomised controlled multi-centre trial in general practice. Research assistants and parents were blinded. The intervention was standardised during three days, with no acupuncture treatment control. The patients fulfilled Wessel’s criteria and were born at full term. There were no exclusion criteria.

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

We did a qualitative study in Shanghai to explore contemporary practices and clinical
recommendations regarding the use of acupuncture for infants by contemporary TCM practitioners. This consisted of four field visits. Data was collected using participant observation, one focus group interview, one in-depth interview, textbook page analysis and informant validation.

**Results**

We improved the standardisation of the intervention procedure and changed the blinding procedure as a result of the pilot study. Blinding validation questions were considered necessary. These changes were implemented in the final trial protocol.

In the main RCT, blinding validation questions showed a random distribution in chi square tests with $p = 0.41$ and $0.60$, indicating likely true blinding. We found no statistically significant difference in crying time reduction between acupuncture and control group at any of the measured intervals, nor in the main linear mixed model analysis of differences in changes over time ($p= 0.26$). There was a tendency in favour of the acupuncture group, with a non-statistical significant total baseline-corrected mean of 13 minutes (95% CI - 24 to + 51) difference in crying time between the groups. This was not considered clinically relevant, independent of not reaching statistical significance, according to the protocol.

In the systematic review and IPD meta-analysis we conducted both an English-language and a Chinese-language literature search up to February 2017. There were no Chinese trials. We included three randomised controlled trials with data from 307 participants, all Scandinavian, and all using different percutaneous needle acupuncture methods. Only one of the included trials obtained a likely successful blinding of the outcome assessors in both the acupuncture and control groups.

The mean difference in crying time between acupuncture intervention and no acupuncture control was $-24.9$ minutes (95% confidence interval, CI $-46.2$ to $-3.6$; three trials) at mid-
treatment, −11.4 minutes (95% CI −31.8 to 9.0; three trials) at the end of treatment, and −11.8 minutes (95% CI −62.9 to 39.2; one trial) at the 4-week follow-up. The corresponding standardised mean differences 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). None of the results were clinically relevant according to protocol. The heterogeneity was negligible in all analyses.

The statistically significant result at mid-treatment was lost when excluding the apparently un-blinded study in a sensitivity analysis: Mean difference −13.8 minutes (95% CI −37.5 to 9.9) and standardised mean differences −0.13 (95% CI −0.35 to 0.099. The registration of crying during treatment suggested that more infants cried during acupuncture treatment versus control, indicating needle pain (odds ratio 7.7; 95% CI 2.7 to 20.6; one trial).

In the GRADE system, we rated the findings as moderate quality evidence.

In the qualitative study from Shanghai, during three days of participant observations at the Longhua Hospital outpatient acupuncture clinic in Shanghai, with 400 consultations on average per day, including referrals from the pediatric department, we saw two children. No infants. During three days at the pediatric department and one day at the Tui na (Chinese medical massage) department we saw no referrals of children for acupuncture. Formal interviews and informal conversations with acupuncturists and other TCM professionals at Longhua and at the Shanghai University of Traditional Chinese Medicine (SHUTCM) revealed that acupuncture was neither routinely practiced nor recommended for infants and small children. Percutaneous needle acupuncture was considered potentially painful for this young patient population.
Conclusions

Neither our RCT nor the IPD-meta-analysis did show any clinically relevant efficacy of percutaneous needle acupuncture in the treatment of infantile colic. The intervention induced treatment pain in many of the infants. Needle acupuncture on infants appears to be a Western practice with little basis in contemporary Chinese TCM practice.

Percutaneous needle acupuncture should not be recommended for infantile colic.
Sammendrag

Introduksjon


Metoder

Vi gjennomførte en pilotstudie for å vurdere gjennomførbarheten av en protokoll for en RCT om effekten av nåle-akupunktur i behandlingen av spedbarnskolikk.


Vi gjorde en systematisk oversikt og en blindingstest-validering basert på individuelle pasientdata fra alle randomiserte kontrollerte studier av effekt av akupunkturbehandling ved spedbarnskolikk. Primært endepunkt var gråtetid midtveis i behandlingen, ved avslutning av behandling og fire uker etter endt behandling. 30 minutters gjennomsnittsforskjell (MD) i gråtetid mellom akupunktur og kontroll var forhåndsbestemt som en klinisk relevant effekt (MID). Pearsons kji-kvadrat test og James and Bang-indeksene ble brukt for å teste
blindingen av utfallsmålerne (foreldrene).

Vi gjorde også en kvalitativ studie i Shanghai for å undersøke nåværende praksis og kliniske anbefalinger I forhold til bruk av akupunktur på spedbarn hos dagens kinesiske utøvere av Tradisjonell Kinesisk Medisin (TKM). Vi foretok fire feltbesøk. Data ble samlet gjennom deltakende observasjon, ett fokusgruppeintervju, ett dybdeintervju, kartlegging av læreboksider og informantvalidering.

**Resultater**

Vi forbedret standardiseringen av intervansjonsprosedyren og forandret blindingsprosedyren som et resultat av pilotstudien. Blindingsvalideringsspørsmål ble ansett nødvendig.

Endringene ble implementert i den endelige studieprotokollen.

I hovedstudien viste blindingsvalidering-spørsmålene en tilfeldig fordeling i kji –kvadrat testene med p= 0.41 og 0.60, som indikerte mulig sann blinding. Vi fant ingen statistisk signifikant forskjell i gråtetid mellom akupunktur og kontrollgruppe ved noen av de målte tidspunkt og heller ikke i den lineære blandede modellen som målte forandringer over tid (p=0.26). Der var en tendens i favør av akupunkturgruppen, med et ikke- statistisk signifikant total utgangs-korrigert gjennomsnitt på 13 minutter (95% KI – 24 to + 51) forskjell i gråtetid mellom gruppen. Dette var i henhold til protokollen heller ikke klinisk relevant, uavhengig om at det var statistisk signifikant eller ikke.

Gjennomsnittsforskjellen i gråtetid mellom akupunkturintervensjonen og ikke-akupunktur kontroll var -24.9 minutter (95% konfidensintervall, KI –46.2 to –3.6; tre studier) midtveis i behandling, –11.4 minutter (95% KI –31.8 to 9.0; tre studier) ved behandlings-slutt, og –11.8 minutter (95% KI –62.9 to 39.2; en studie) fire uker etter endt behandling. Det korrespondende standardiserte gjennomsnitt var –0.23 (95% KI –0.42 to –0.06), –0.10 (95% KI –0.29 to 0.08), and –0.09 (95% KI –0.48 to 0.30). Ingen av resultatene var klinisk relevante i henhold til protokoll.

Heterogeniteten var uten betydning i alle analyser.

Det statistisk signifikante resultatet midtveis in behandling forsvant når vi ekskluderte den tilsynelatende ublindede studien i en sensititetsanalyse: Gjennomsnitts-forskjell –13.8 minutter (95%KI –37.5 to 9.9) og standardisert gjennomsnitts-forskjell –0.13 (95%KI –0.35 to 0.09). Registeringen av gråt under behandling viste at flere spedbarn gråt under akupunkturbehandlingen enn i kontrollgruppen, noe som indikerte nålesmerte (odds ratio 7.7; 95% KI 2.7 to 20.6; en studie).

I GRADE systemet vurderte vi funnene som bevis av moderat kvalitet.

I den kvalitative studien fra Shanghai, i løpet av tre dager med deltagende observasjon på Longhua-sykehusets akupunktur-poliklinikk som mottar 400 pasienter i gjennomsnitt pr dag, inkludert henvisninger fra barneavdelingen, så vi to barn, ingen spedbarn. I løpet av tre dager på barneavdelingen og en dag på Tui na (kinesisk medisinsk massasje) avdelingen så vi ingen henvisninger av barn til akupunktur. Formelle intervju og uformelle samtaler med akupunktører og andre TKM klinikere ved Longhua og ved Shanghai Universitetet for Tradisjonell kinesisk Medisin (SHUTCM) viste at akupunktur hverken var rutinemessig praktisert eller anbefalt overfor spedbarn og småbarn. Hudpenetrerende nåleakupunktur ble vurdert som potensielt smertefullt for denne unge pasientgruppen.
Konklusjoner

Hverken vår RCT eller IPD-meta-analyse viste noen klinisk relevant effekt av hudpenetrerende nåleakupunktur i behandlingen av spedbarskolikk. Intervensjonen gav behandlings-smerte hos mange av spedbarna.

Nåleakupunktur på spedbarn ser ut til å være en vestlig praksis med liten basis i nåværende kinesisk TKM praksis. Hudpenetrerende nåle-akupunktur bør ikke anbefales for behandling av spedbarskolikk.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CONSORT</td>
<td>CONsolidated Standards Of Reporting Trials</td>
</tr>
<tr>
<td>COREQ</td>
<td>COnsolidated criteria for REporting Qualitative research</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>GP</td>
<td>General Practitioner</td>
</tr>
<tr>
<td>GRADE</td>
<td>Grading of Recommendations Assessment, Development and Evaluation</td>
</tr>
<tr>
<td>IPD</td>
<td>Individual Patient Data</td>
</tr>
<tr>
<td>LI4</td>
<td>Acupuncture point Large Intestine 4 – in the thenar soft tissue on the hand.</td>
</tr>
<tr>
<td>MD</td>
<td>Mean Difference</td>
</tr>
<tr>
<td>MCID</td>
<td>Minimal Clinically Important Difference</td>
</tr>
<tr>
<td>MIC</td>
<td>Minimal Important Change</td>
</tr>
<tr>
<td>MID</td>
<td>Minimal Important Difference</td>
</tr>
<tr>
<td>NFA</td>
<td>Norwegian College of General Practice</td>
</tr>
<tr>
<td>NMDA receptor</td>
<td>N-Methyl-D-Aspartate receptor</td>
</tr>
<tr>
<td>NNT</td>
<td>Numbers Needed to Treat</td>
</tr>
<tr>
<td>NSD</td>
<td>Norwegian Centre for research Data (former Norwegian Social Science Data Services)</td>
</tr>
<tr>
<td>OR</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>PRISMA</td>
<td>Preferred Reporting Items for Systematic reviews and Meta-Analyses</td>
</tr>
<tr>
<td>PRO</td>
<td>Patient Reported Outcomes</td>
</tr>
<tr>
<td>PROSPERO</td>
<td>International PROSPEctive Register Of systematic reviews</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised Controlled Trial</td>
</tr>
<tr>
<td>REK(REC)</td>
<td>Regional committees for medical and health research ethics</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>R</td>
<td>The R Project for Statistical Computing</td>
</tr>
<tr>
<td>SHUTCM</td>
<td>Shanghai University of Traditional Chinese Medicine</td>
</tr>
<tr>
<td>SMD</td>
<td>Standardised Mean Difference,</td>
</tr>
<tr>
<td>ST36</td>
<td>Acupuncture point Stomach 36 – in the soft tissue lateral of proximal tibia bone on the leg.</td>
</tr>
<tr>
<td>STRICTA</td>
<td>Revised STandards for Reporting Interventions in Clinical Trials of Acupuncture</td>
</tr>
<tr>
<td>TCM</td>
<td>Traditional Chinese Medicine</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual analog scale.</td>
</tr>
</tbody>
</table>
**List of publications**

**Paper I**

**Paper II**

**Paper III**

**Paper IV**
http://bmjopen.bmj.com/content/5/11/e009486.full
1. Introduction

1.1. Infantile colic

1.1.1. Definitions

Infantile colic is a poorly understood and probably painful condition of early childhood. It is classified according to the ROME III Paediatric working group (1) as one of the childhood functional gastrointestinal disorders (2). Crying peaks at 4–6 weeks of age and diminishes gradually towards 12 weeks (3). The clinical description and diagnostic criteria were first established by the American paediatrician Morris Arthur Wessel in 1954 (4): “Paroxystic uncontrollable crying and fussing in an otherwise healthy infant under three months of age, with more than three hours of crying per day in more than three days for more than three weeks.” This definition has governed clinical practice and clinical research up to now.

In 2006, the ROME III working group modified the criteria, extending the age of inclusion to four months, and reducing the criterion of three weeks to one week: (2). “Must include all of the following in infants from birth to 4 months of age: 1. Paroxysms of irritability, fussing, or crying that start and stop without obvious cause. 2. Episodes lasting 3 or more hours per day and occurring at least 3 days per week for at least 1 week. 3. No failure to thrive.” Ten years later, the Paediatric group of Rome IV argued that infantile colic is a behavioural syndrome, that most cases of colic probably represent the high part of normal developmental crying, and that there is no proof that the cause of crying is abdominal pain (3). In their 2017 systematic review and meta-analysis on fussing and crying durations, Wolke et al. (5) concurred with this. Normal crying patterns show wide variability, with averages of two hours over the first six weeks to one hour at three months (5, 6). However, infantile colic has the distinct characteristics of prolonged crying, unsoothable crying, apparently painful facial expressions, abdominal distension, leg withdrawing and increased crying levels at evening times (3, 6). For
clinical and research purposes, the ROME IV group advocated that a new and extended
definition should be used: “Diagnostic criteria for infant colic must for clinical purposes
include all of the following: 1. An infant who is <5 months of age when the symptoms start
and stop. 2. Recurrent and prolonged periods of infant crying, fussing, or irritability reported
by caregivers that occur without obvious cause and cannot be prevented or resolved by
caregivers. 3. No evidence of infant failure to thrive, fever, or illness. “Fussing” refers to
intermittent distressed vocalization and has been defined as [behaviour] that is not quite
crying but not awake and content either. Infants often fluctuate between crying and fussing, so
that the 2 symptoms are difficult to distinguish in practice. For clinical research purposes, a
diagnosis of infant colic must meet the preceding diagnostic criteria and also include both of
the following: 1. Caregiver reports infant has cried or fussed for 3 or more hours per day
during 3 or more days in 7 days in a telephone or face-to-face screening interview with a
researcher or clinician. 2. Total 24-hour crying plus fussing in the selected group of infants is
confirmed to be 3 hours or more when measured by at least one prospectively kept, 24-hour
behaviour diary” (3).

1.1.2. Prevalence and costs

Reporting on prevalence varies widely, reflecting differences in study design, diagnostic
criteria, populations screened, and each family’s perceptions. Retrospective studies tend to
estimate higher prevalence, reflecting recall bias (7). Retrospective studies have estimated the
occurrence rate at 8–40% (6), and prospective studies at 3–28% (7, 8). Of paediatric
consultations, 10–20% with infants aged two weeks to three months involve infantile colic
(9). The incidence does not differ between breast- and bottle-fed infants, between preterm and
term births, or between male and female infants (6). A retrospective review of all afebrile
infants evaluated at an emergency department because of excessive crying showed that only
5.1% of them had a serious underlying organic disease (10). In a cost burden analysis by
Morris et al. (11), infantile crying and sleeping problems were estimated to cost the British National Health Service £65 million annually. There is no such estimate for Norway.

1.1.3. Aetiology

The aetiology is unknown (3, 9, 12). It is likely to be multifactorial (3, 9, 12). Hypotheses are based on knowledge of infant pain perception, the function of the immature gastro-intestinal tract, including allergenic theories and changes in the gut microbiota, on theories of normal development, observations of parent–child interactions, and extrapolation from clinical trials of interventions.

Infants have a lower pain threshold than adults. They lack descending inhibitory control, and do not seem to be able to modulate their pain experiences (2, 3). Small pains can be perceived as major ones, and they cannot differentiate abdominal discomfort from serious pain. As colic is a functional disorder, pain is no longer a warning signal of peripheral disturbance or damage—as in acute pain models—but the illness itself. This neurophysiological knowledge should be taken into consideration in explanatory models of infantile colic (2, 3).

Food allergens, and in particular cow milk proteins, have long been established as possible aetiological factors associated with the incidence of infantile colic (12-15). Gut dysmotility/hypermotility have also been postulated to explain perceived abdominal pain in infants (6, 12, 15, 16). Infants’ gastro-intestinal microbiotic flora has been shown to be altered and this has been suggested as a possible important influencing factor in infantile colic (15, 17-20). In 2016, Camilleri et al. at the Mayo Clinic (9) added the concept of bile acid formation to the gut hypothesis: immaturity of bile acid formation and excretion mechanisms, alterations in gut motility, and changes in the microbiome combine to give rise to increased nociception of the gastro-intestinal tract in infants with colic.
Parent–child interaction associations with infantile colic have been proposed. Parental anxiety, disturbed infant–parental interactions, and family tension—although complex and controversial factors—have all been seen as possible contributing factors to the development and persistence of infantile colic (3, 6, 9, 12, 21).

1.1.4. Associated risks and consequences

Infantile colic is a self-limiting condition but stakes can be high, as parents tell of this being a very stressful and disrupting period in their family life. Among others, Landgren et al. have explored this subject in a qualitative article with the tell-tale title: “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” (22). Associations have been reported between infantile colic and early breast-feeding termination, and with maternal depression (6, 23). More disturbingly, there seem to be strong associations between incessant infant crying such as seen in infantile colic, and child abuse, child violence, and shaken baby syndrome (3, 24-26). From a neurophysiological development perspective, if infantile colic entails pain, long-term impacts may involve altered pain perception and sensitization (3).

1.1.5. Current treatment strategies and evidence

The basis of medicine is that aetiology defines treatment. However, as the aetiology of infantile colic is unknown, treatment suggestions are diverse and contested. None of the current treatment strategies appear sufficiently effective to warrant a general recommendation (3, 9, 12, 27). Treatment strategies can be divided roughly into six major patterns, with combinations of these often used: parental counselling; behavioural approaches/soothing techniques; dietary approaches; gut microbiotic changes; oral medications; and alternative/complementary therapies.
**Parental counselling.** Speaking for the ROME IV recommendations, Benninga et al. advised that as infantile colic is a functional disorder, parental counselling and support should be a mainstay of any treatment strategy (3). Some studies indicate a sharply reduced crying time with this approach (27).

**Behavioural approaches/soothing techniques.** Recommendations include changing feeding techniques, holding the baby during crying bouts, rocking the baby, driving the baby in a car, rubbing the baby’s abdomen, introducing pacifiers and changing scenery (12, 27).

**Dietary approaches.** Restricting cow milk-based formulas and reducing exposure to cow milk proteins in selected infants with a family history of cow milk allergenic susceptibility or atopy have been a mainstay of advice based on studies that show decreased infantile colic when restricting cow milk formulas (12, 15, 28, 29). However, this view is contested and the recommendation for infants with colic who otherwise thrive is not to change their diet (3, 27).

**Gut microbiotic changes:** The treatment of unfavourable microbiomes with coliform bacteria in infants with colic has been proposed (19, 30, 31). Several randomized controlled trials (RCTs) have shown the effects of administering industry-produced lactobacillus strains in infantile colic (17, 18, 31-33). However, critics argue that all of the positive studies were industry-sponsored, and there is still no consensus on the eventual benefits, given that other trials have shown no effect (3, 7, 9, 34, 35).

**Oral medications.** The use of herbal teas, sugar, dicyclomine, and cimetropium bromide has not shown persistent clinically relevant effects and studies have been prone to bias (7, 36).

**Alternative/complementary therapies:** Homeopathy has no place in the treatment of infantile colic (37-39). Chiropractic manipulation, extensively used by distressed parents, does not show clinical effects in unbiased scrutiny of systematic reviews and meta-analyses (40, 41). Acupuncture treatments are also widely used and practitioners tell stories of swift
effects (42), and small RCTs and non-RCTs support this possible favourable tendency (43-45). One systematic review by Raith et al. (46) on the effects of acupuncture on preterm and term infants suggested an effect on infantile colic on the basis of these studies. No meta-analyses have been conducted.

1.2. Acupuncture

1.2.1. The tradition of acupuncture

Acupuncture is part of the ancient medical system of traditional Chinese medicine (TCM). As with the Greco-Roman European medical tradition, this canon presents complex systems of thought and therapies reflecting the pre-industrial characteristics of nature reflected in the body, the inseparable connection with body and soul, vital “energies” within the body, and the idea that diseases result from imbalances of these vital energies that can be rebalanced in one therapeutic way or another (47-51). TCM developed in the society and within the cultural context of ancient China, and tradition is upheld in the sense that the classical medical textbooks of ancient times are still considered to carry the most profound medical truths. The most distinguished of all these, together with three other classic texts from that era, and the reference text for all text-books that were to come during the millennia, is “Huang Di Nei Jing, the Yellow Emperor’s Canon of Medicine”, of disputed origin between the Spring and Autumn period and the Warring States period, but compiled (52) in the Han Dynasty (206 BC–220 AC) (50, 51, 53-55).

TCM integrates Taoist and Confucian thought in its medical systems. In the beginning and as its basis is the concept of Tao, the unfathomable all-pervading “life force” in nature. From Tao springs the Yin and the Yang, the two opposing and balancing influences of all life, and also the opposing and balancing influences within the human body and soul. Illness and disease in their most basic ways reflect imbalances of Yin and Yang. In the body and soul
complex resides and flows the vital energy, Qi, whose obstruction or depletion results in excesses or deficiencies, which manifest as disease. There are five vital substances, the Jing, Shen, Yin Je, Blood, and Qi, which form a complex system with the five major organ systems: the kidneys, heart, spleen, liver, and lungs. Through all these systems are believed to run an interconnected flow of channels, “meridians”, which have deep branches in the vital organs, and superficial branches in the skin and outer layers of the body. Imbalances can be influenced and changed by stimulating these meridians in the superficial layers at specific points believed to have specific functions. These are the acupuncture points, on which the rationale of acupuncture treatment is based (49-51, 56).

Modern TCM has five traditions forming the basis for treatment recommendations. Herbal medicine, which can include mineral and animal parts in addition to botanicals, is by far the most important and complex element. There are more than 5000 items in its Materia Medica, and herbal medicine is the one governing tradition and therapy. The four others are medical massage (Tui Na), meditative exercise (Qi Gong), lifestyle and diet counselling, and acupuncture/moxibustion (50, 51).

Zhen Jui is the classical name of the acupuncture tradition. Acupuncture (Zhen: needling at acupuncture points) and moxibustion (Jui: burning or heating at acupuncture points) were integral and inseparable parts of TCM in ancient times (57). After the Chinese Communist Party’s ascent to power in 1949, and the introduction of acupuncture to the West after President Nixon’s opening of US–China relations in 1972 (58), acupuncture separated and became a therapy in its own right, especially in the West. Communist China redefined and reorganized TCM thought and texts to streamline a once very diverse and varied canon. This was also true for acupuncture texts, therapies, and organizations (55-57).

Acupuncture is primarily percutaneous needle acupuncture. Hand pressure (acupressure) is also used. Distinct acupuncture styles have developed regionally, such as Korean hand
acupuncture, Japanese superficial acupuncture, and French-style acupuncture. There are several microsystem acupuncture methods, such as scalp acupuncture and ear acupuncture. Today, Western segmental acupuncture, and techniques of laser- and electroacupuncture have developed and are in widespread use (51, 59, 60).

There are 365 classical acupuncture points (49, 51), which together with additional auxiliary points all have defined actions and properties. Alone or in combination, they can influence pain, lung function, blood deficiencies, urinary tract problems, depression, palpitations, anxiety and sleep, to name but a few conditions and symptoms. The indications are legion (61-64). Acupuncture in everyday practice is preferably used for pain problems because of the needle’s specific ability to change the flow of obstructed Qi in the meridians, which in TCM theory is the source of pain problems (51, 58, 65, 66). According to Professor Guang Ji, Vice-President of Shanghai University of Traditional Chinese Medicine and head of research, 90% of acupuncture treatments in contemporary China are for musculoskeletal pain and dysfunction (67).

The needles used today are sterile thin steel needles 0.15–0.40 mm in diameter and 15–120 mm long; they are inserted through the skin and into the soft tissues, preferably muscles, to obtain their effects (51, 59). How deep, how long the needles are retained, how many needles, and how to manipulate them once they are inserted are questions of lore and debate (60, 66, 68-70). Traditions vary, but there is a common recognition of the importance to elicit ‘de Qi’, the needle sensation, to obtain the desired effects. This feeling is described by the subject being acupuncture in multiple terms, as sore, throbbing, pressure-like, aching, or painful-like (51, 59, 60).
1.2.2. Neurophysiological mechanisms and the evidence base of acupuncture effects

The neurophysiological mechanisms of acupuncture have been researched and documented extensively. When using acupuncture needles penetrating the skin, there are local effects of direct vasoactive changes by the release of substances through tissue damage with both local inflammatory and anti-inflammatory effects (51, 58, 71, 72). Sensory signals from the activation of nociceptive A-delta and C fibres in addition to alpha beta sensory motor fibres in the periphery induce spinal effects by inhibiting signal transfer in the dorsal horn with possible influence on N-methyl-d-aspartate (NMDA) receptor activity in the dorsal spinal area (72, 73). There are well-documented spinal and brainstem effects of activating the pain-inhibiting nuclei and descending pathways from the brainstem to the spinal area, namely the gate-control mechanisms (56, 58, 71, 72, 74). Hypothalamic effects involve increased secretion of endorphins to the cerebro-spinal fluid and blood vessels during acupuncture treatment, and this is presumed to cause general analgesic effects (51, 71, 73-75). Multiple associations have been described between acupuncture stimuli in research laboratory settings and neuro-immunological, immunological, and hormonal changes (51, 71, 72, 74, 76-80). Functional magnetic resonance imaging (fMRI) laboratory studies show needle-specific cortical activation changes (51, 72, 73, 81, 82). Possible limbic effects by reduced signal intensity in structures that are central in maintaining and strengthening the cognitive-emotional experience of pain have been postulated (51, 73). Higher cortical effects, such as placebo effects, have been shown extensively in the activation of cortical expectation/reward systems and the conditioning of Pavlovian reflexes around the needle-insertion ritual (75, 83-85).

However, there are discrepancies between the observed physiological effects of acupuncture in laboratory settings, and the lack of clinical effects in clinical trials (71, 74, 82, 86). Except as treatment for pain and nausea (51, 71, 74, 75, 87-89), there is still no robust evidence of
acupuncture having the diversity of clinical effects that laboratory research results seem to indicate (51, 58, 71, 82, 90). This is the central enigma of acupuncture.

In clinical trials of chronic pain in adults, acupuncture treatments have repeatedly shown consistent results. The needle-specific standardized mean difference (SMD) of acupuncture treatments of chronic pain in adults is ~0.2 (87, 89-94). This corresponds to a Visual Analog Scale (VAS) difference between treatment and control of 5 mm on a 100-mm scale, or a difference of 7% between intervention and control treatments (88, 90). This effect size is of little clinical relevance (88, 90, 95-99).

There is a consistent and corresponding additional non-specific effect of 0.3 SMD for acupuncture (69, 71, 88, 92), irrespective of the mode of stimulation, type of chronic pain indication, and this is seemingly independent of acupuncture style, whether Western, TCM, or other. The only associations with better outcomes are the number of needles and frequency of treatment (69, 75). Acupuncture is known to have a powerful placebo effect, and this is thought to be behind its rather large non-specific effects. However, this effect is questioned by some because pain reduction is a patient-reported outcome (PRO) and prone to bias (69, 90, 100-104). Taken together, the specific and non-specific effects of acupuncture treatments of chronic pain give a medium effect size of SMD of 0.5, corresponding to 13 mm on a 100 mm VAS, or a difference of 20% between intervention and control treatments (88, 90). This can be interpreted as a clinically relevant effect (88, 92, 95-97, 99, 105), although some argue that for PROs such as pain, thresholds for clinically relevant effects should be higher (90, 95, 96, 98, 106). There is also debate on the specific effects of needle acupuncture for pain, isolated from its non-specific effects. The question is whether 0.2 SMD/5 mm on a VAS—the needle-specific effect—is enough to warrant a general recommendation of acupuncture treatment for pain because of the additional non-specific placebo effect of 0.3 SMD (71, 88, 90, 92, 107). There is no interdisciplinary consensus on this question. Acceptance of research-isolated
placebo effects and their inclusion in the total effect size to implement treatment regimens of complementary medicine are still disputed (71, 88, 90, 101, 102, 107-110).

1.2.3. Adverse effects of needle acupuncture

Modern acupuncture practice, with sterile needles and in experienced hands who know the anatomy, is considered safe (51, 72, 111). It is also considered safe for children (112). However, there are critics who argue that the data on acupuncture’s adverse effects are selective, and do not take into account non-systematic and anecdotal reports, and that the cumulative risks of adverse side effects such as pneumothorax and septic infections are under-estimated (113, 114). In addition, there is the still little-debated question of acupuncture needle insertion as a painful stimulus. Consenting adults take on many treatment regimens that are painful and poorly documented, with full autonomy and full consent. In children, especially small children who cannot consent to treatment, it is a different matter. In acupuncture, these are uncharted waters and in need of clarification (44, 62, 115-120).

1.2.4. Acupuncture as a painful stimulus

Skin-penetrating needle acupuncture is by definition a painful neurophysiological stimulus. It results in micro-tissue damage as a result of needle skin penetration into the superficial layers or deeper tissues. This activates, among others, afferent A delta and C nociceptive fibres that send pain signals to the spinal dorsal horn where the signals cross the midline, progress into the spinothalamic tract to the thalamus, and on to the sensory cortex areas in the classic 3-neuron pathway. It is a pain signal, and much of the explanation of the pain-inhibiting effects is secondary to this primary signal, as several central nervous pain-inhibiting mechanisms set in (51, 73, 74).

The Nordic Medical Acupuncture Congress II in Oslo, Norway on 6 September, 2014 specifically discussed acupuncture as a painful stimulus in an acupuncture research academic
panel session moderated by the author. The following are some verified quotes: “Generally, acupuncture is a pain stimulus” (Frauke Musial) (121); “Acupuncture is a therapeutic pain stimulus” (Thomas Lundeberg) (122); and with respect to a meta-analysis of fMRI studies of brain activation in pain areas following acupuncture: “Acupuncture is painful—what we see may be just pain activation in the brain” (Florian Beissner) (123).

There is little reporting of acupuncture pain as an adverse effect, except in paradoxically strong or persistent pain. A small RCT by Assefi et al. (124) on acupuncture for fibromyalgia in acupuncture-naïve patients reported discomfort at the needle site in 65% of patients after skin-penetrating needling. A much-quoted review by White (111) excluded pain overall as a significant adverse effect because of the difficulties in judging its severity. A review on the adverse effects of acupuncture in children by Adams et al. (112) did not discuss pain as an adverse effect other than in reported excessive cases. In RCTs of acupuncture for infantile colic, Landgren et al. (44, 45) systematically registered and reported crying time after needle insertion. In addition, there have been individual study reports of systematic pain assessment on acupuncture in children (115). Acupuncture needle insertion is acknowledged as being potentially painful in Western text-books on acupuncture in children, and attention has focused on minimizing this pain (62, 116, 125).

1.2.5. Acupuncture in children

Acupuncture in children, including infants, has gained acceptance in the West (12, 46, 126-130). If it involves percutaneous needle acupuncture, because neurophysiological it constitutes a pain stimulus (73, 121), it is potentially painful in children (44, 46, 66, 115). Ethical concerns regarding its use in small children without competence to give consent have been raised (117, 118, 131). There is limited scientific knowledge of the effect sizes in acupuncture treatment of children, including treatment for pain conditions (46, 58, 126). Of
the eight Cochrane systematic reviews on acupuncture for different children’s conditions published up to 2014, none were positive (132). A 2015 Norwegian RCT by Liodden et al. testing for postoperative nausea after tonsillectomy was negative (133). A 2016 RCT by Mitchell et al. testing laser acupuncture for pain reduction in heel lancing procedures was also negative (134). A 2015 laser-acupuncture RCT for neonatal abstinence syndrome by Raith et al. was positive (135). Published cases or qualitative studies on acupuncture in infants and small children among Western acupuncturists are very optimistic, and recommend the use of acupuncture (127, 136, 137).

There are no clinical guidelines on acupuncture in children (138), and practices vary substantially between countries. Individual text-book recommendations on paediatric conditions suitable for acupuncture treatment are extensive, from ear infections to autism (62) and from asthma to inflammatory bowel disease (125). Western-based text-books of acupuncture argue that its effects in children—especially small children—are swift and often stronger than in adults (116, 139-141). These text-book notions and recommendations do not in general refer to scientific studies or clinical trials, but are mostly based on references to the Chinese TCM tradition and on the authors’ personal views and clinical observations. There is little scientific knowledge regarding the nature and extent to which acupuncture, including acupuncture for pain conditions, is used in paediatric populations in modern-day China.

China-based English-language text-books consider herbal remedies the pivotal and primary treatment method in children (142), followed by Tui Na. Acupuncture is often mentioned as an adjunct or auxiliary treatment (143, 144). In a literature search conducted in January 2015, we did not find any English-language studies describing contemporary TCM clinical practices or clinical guidelines on the use of acupuncture for infants or small children from the People’s Republic of China, Taiwan, Singapore, or Hong Kong. Survey studies of the use of TCM in children have been reported from Singapore and Taiwan (145-147). Emphasis on TCM use in
children was on herbal treatment, acupuncture was most often used for injury and musculoskeletal disorders, and TCM use in general was reported to be low in infants and toddlers. We have not found any English-language studies or published abstracts on the views and attitudes of Chinese TCM practitioners towards acupuncture treatment for infants and small children.

1.2.6. Acupuncture for treating infantile colic

There is no TCM equivalent to the diagnosis of infantile colic. The closest we come to this Western diagnosis is the TCM term “night crying”. However, here the diagnostic criteria differ: the term obviously specifically describes crying during the night, which can last up to six months of age (148). The recommended treatment for this was Tui Na, medical massage (149). Recurrent abdominal pain in infants—a broad category consisting of a number of TCM syndromes—has different herbal treatments recommended as primary therapies (150).

All Western trials of acupuncture treatments for infantile colic have used percutaneous needle insertion (43-45, 136, 151). All Medline-indexed reported qualitative studies, case studies, and controlled trials are from Scandinavia (22, 43, 44, 136, 137, 151, 152). In particular, Kajsa Landgren of Lund University in Sweden has published extensive quantitative and qualitative research on the use of acupuncture in treating infantile colic (22, 42, 44, 45, 151-153). However, this does not mean that there is a purely Scandinavian focus. Landgren interviewed 23 paediatric acupuncturists from five European and four non-European countries at a major acupuncture conference in Germany. They reported widespread use and recommendations for acupuncture in treating infantile colic (137). TCM diagnostic orientation and differentiation, as well as the actual acupuncture points used and recommended, showed considerable variance. Points LI 4 (in the thenar muscles on the hand) and ST36 (in the proximal anterior fibularis muscle on the leg) were the most commonly recommended points.
In their text-book, the Norwegian physicians Oscar Heyerdahl and Nils Lystad recommend acupuncture for treating infantile colic (154), with ST36 being their main acupuncture point. Also recommended by the paediatric acupuncturist Julian Scott in his text-book “Acupuncture in Children” are the intra-articular extra Sifeng points on the palmar side of the proximal interphalangeal joints (62). Two RCTs from Landgren et al. and one non-RCT from Reinthal et al. showed statistically significant effects of standardized or semi-standardized percutaneous needle acupuncture of infantile colic (43-45, 151). The article “Management of infantile colic” in the BMJ suggests on the basis of the trials up to 2013 that acupuncture could be a possible intervention. However, the two papers from Landgren et al. (44, 151) on the same trial were interpreted as being two separate trials (12). Raith et al. (46) carried out their 2013 review on the use of acupuncture on preterm and term infants. This was based on one uncontrolled case series study of 913 cases (136), one non-RCT of 40 patients (43), and one non-blinding-validated RCT of 80 patients (44, 151). They concluded: “In summary current evidence suggests that acupuncture is safe, effective, and a cheap method to treat infantile colic.”
2. Objectives

The main aim of this research was to gain knowledge regarding the practice and the efficacy of acupuncture treatments of infantile colic.

The objectives were:

1: To develop and test in a pilot study a RCT research strategy that was able to deliver a tight and blinding-validated structure in a percutaneous needle acupuncture trial of infantile colic.

2: To test the efficacy of the standardised acupuncture treatment for infantile colic adopted by Norwegian general practitioners affiliated to the Norwegian Society of Medical Acupuncture in a blinding-validated multicentre randomised controlled trial.

3: To evaluate the existing randomised controlled trials testing the efficacy of acupuncture for infantile colic in a systematic review and meta-analysis.

4: To investigate in a qualitative field study the opinions of Chinese TCM practitioners in Shanghai concerning acupuncture treatments of infants and the use of this practice.
3. Methods and summary of results

3.1. Paper I


3.1.1. Method

The aim of the pilot study was to assess the feasibility of the proposed design of the subsequent RCT. The pilot study was scheduled with 12 randomised patients. We recruited nine patients during May to August 2009, two withdrew. The parents of the final seven infants were informed that they were participating in the pilot study and that it would be open to changes during the course of the study. The results would not be included in the main study.

The pilot study was conducted by the two project coordinators of the main study (HS and TS).

Study participants were recruited via information distributed to parents at maternity departments, well-baby clinics and at GPs’ offices. The patients fulfilled Wessel’s criteria (crying more than 3 h/day, 3 days/week, for more than 3 weeks), were born at full term (defined as 36 weeks or more) with birth weight 2500g or more and should be less than 3 months old at inclusion. There were otherwise no exclusion criteria.

Participants were randomised to active treatment or to no-acupuncture treatment control. The allocation to treatment or control was concealed in a numbered envelope. The crying registration form had previously been validated in a controlled trial of chiropractic spinal manipulation for infantile colic(155). Our trial would use a standardised bilateral needling of the acupuncture point ST 36. This point is situated over the proximal part of the Anterior Fibularis Muscle. A proposed neurophysiological mechanism is a beneficial effect on gut
dysmotility via parasympathetic vagal reflexes, as well as a centrally opioid-mediated pain inhibitory pathway(156).

The GP was alone in the treatment room with the infant during the intervention.

In the intervention group, an ethylene-oxidised sterile Seirin acupuncture-needle (0.20×15 mm) was inserted at ST36. The point was needled bilaterally to approximately 12 mm depth. The two needles were left inserted without manipulation for 30 s while the infant was lying on his/her back on the bench. The infant stayed in the treatment room for 3 min in total, and the parents were called for. An identical procedure, but without the needle insertions, was performed in the no-treatment control group. Intervention consisted of three treatments of 30 seconds duration during three consecutive workdays. The parents and the assistant were blinded to the intervention. There were no restrictions on the parents concerning what they could do to alleviate the infant’s symptoms during the trial observation period (four weeks), except further acupuncture treatment.

The parents filled in the crying registration form (Appendix 5) from the day after inclusion and through the intervention period until the day after the last treatment, as well as one and four weeks after the last treatment. The assistant conducted a telephone interview one and four weeks after the last intervention day (Appendix 6). The effectiveness of blinding was assessed after the first intervention and on day 33, four weeks after the last intervention.

3.1.2. Main results

The protocol was changed on several points as a consequence of the pilot study.

Changes concerning standardisation.

The original protocol did not specify what time of day to start the crying registration. During the pilot study, we became aware of this standardisation problem. We therefore determined that the crying registration for all participants should start at midnight two days before the
first intervention day. It also became apparent during the pilot study that we lacked a detailed and comprehensive procedure for the actual handling of the infant by the GP. We had agreed on the standardised needling technique, but differed in the handling of the child. Non-verbal communication by the practitioner is found to influence outcome in blinded acupuncture studies in adults(103). We consequently agreed on a detailed level on the non-needling aspect of the handling of the infant, and made a video presentation showing the three minutes handling time on the bench to further facilitate intervention standardisation in the main study of all participating doctors.

Changes concerning blinding.

The risk of visible needle marks and of post-insertion erythema in the acupuncture group is a possibility in all placebo-controlled trials of acupuncture treatment in infants. Explicit recommendations of reporting of blinding validation in acupuncture trials are not a part of the revised Standards for Reporting Interventions in Clinical Trials of Acupuncture guidelines of 2010- STRICTA (157), an adaption to the CONSORT guidelines. The original post-insertion blinding method was a three cm in diameter round marking made with a dark red water-resistant marker, centred in point ST36. This was in order to hide an eventual needle insertion mark and to hide the post-insertion erythema. Only adhesive dressing was considered too risky because of accidental or deliberate removing. The original protocol was set up with the water-resistant marking alone. During the pilot study, the mother of one of the infants in the acupuncture group thought she had seen a faint needle mark the day after the last intervention. We decided on the following changes in the blinding procedure (1): We made a point mark at ST36, with the dark red water-resistant marker, three mm in diameter. (2) Covering the point mark, we applied a waterproof, circular adhesive dressing, 24 mm in diameter. This was changed after each intervention in both groups. These changes led to additional advantages: the needle insertion was standardised on the same spot each time, the adhesive dressing
protected the dark red colour from fading and post-insertion erythema remained covered. The blinding was also kept if the adhesive dressing incidentally or purposely were removed.

**Figure 1** The ST36 point marking and inserted needle. The immediate post-insertion erythema is clearly visible in a diameter of 2.2 cm.

**Figure 2** The infant lying with inserted needles for 30 seconds without further manipulation.

**Figure 3** The 2.4 cm adhesive water-resistant adhesive dressing covering the point mark and erythema.
3.2. Paper II


The relevant items of the intervention period and follow-up, and the standardisation and blinding validation changes of the trial protocol were described in the previous method section paper I, the pilot study.

3.2.1. Method

The study was a prospective, blinding-validated, multi-centre, randomised controlled trial involving 13 GPs’ offices in Southern Norway. The trial was registered with Clinical Trial Registry Identifier NCT00907621. The data collection was approved by the Norwegian Social Science Data Services (reference 21490/2/JE) (Appendix 2). Ethical approval was given by the Regional Ethics Committee of South-Eastern Norway (reference S-08732b 2008/17889s) (Appendix 1) and the trial was carried out in accordance with the Helsinki Declaration. The parents of the infants gave informed consent (Appendix 4). The inclusion period was from September 2009 to December 2012. Two work-shops to secure standardisation, understanding of and compliance with protocol, and reporting were carried out with the participating GPs and the assistant personnel. One before start of the RCT and the second half-way in the trial period.

Participants: The participating doctors were all GP specialists with a minimum of 300 hours of acupuncture education and five years of practicing acupuncture. The assistant personnel were health secretaries working in the GPs’ practices. The patients fulfilled Wessel’s criteria and were born at full term. There were no exclusion criteria.
Randomisation, allocation concealment and blinding: The infants were randomised to active treatment or to no acupuncture control. Randomisation was done manually by two persons not otherwise involved in the study. Sealed, opaque, numbered envelopes were used. Randomization was closed until the start of the first intervention. The parents and assistants were blinded to the allocation.

Blinding validation: Blinding validation of the parents was done by two blinding validation questions, one immediately after the first intervention, “Do you think your child has received acupuncture or not?”, the second at week four, “Have you noticed any needle insertion marks?”.

Intervention: The interventions are described in the previous pilot study section.

Outcomes: The primary outcome was difference in changes in crying time in the registration period. Clinically relevant efficacy was defined as one-hour difference according to the pre-trial protocol. Secondary outcomes were differences in fulfilling Wessel’s infantile colic criteria of 180 minutes crying estimated from the crying diary, the parents’ assessment of the child’s condition on a Likert scale, and adverse effects, both as reported by the parents during the telephone interview. We also asked parents on additional therapies during the observation period.

Statistical power: We anticipated a standardized difference of 0.5 and a clinically relevant difference of one hour. With p 0.05 and 80% power, we needed to include 120 infants.

Reporting: The crying diary and telephone interviews are described in the previous pilot study section.

Data analysis: We used the software programs SPSS 19 and 20 for the statistical analyses. We used linear mixed models statistics for the analyses of the main outcome variable, and chi-square and Fisher’s exact tests for secondary outcomes of categorical data, and blinding
validation analyses. T-tests were used for the continuous data estimates at the various time points. Magne Thoresen, professor at the Department of Biostatistics, University of Oslo, supervised in the statistical analysis and carried out the baseline-adjusted time/effect analysis in the linear mixed model.

3.2.2. Main results

We did not reach the estimated 120 crying time diaries during the extended trial period. A total of 113 patients were recruited; 23 patients were excluded, and 90 randomized; 79 diaries and 84 interviews were analysed.

Blinding validation.

The primary blinding validation question “Do you think your child has received acupuncture or not?” was answered in 83 cases and showed a random distribution. In the acupuncture group 22 of the parents believed the child had acupuncture and 22 believed the child was in the control group. In the control group the numbers were 16 and 23 (p 0.41). The second blinding validation question, “Have you noticed any needle insertion marks?”, was answered in 38 cases: 17 in the acupuncture group and 17 in the control group answered “No”. Three in the acupuncture group and one in the control group answered “yes” (p 0.60). We thus consider the blinding as valid.

Primary end-point.

The primary end point of the trial was the difference in crying time changes between the acupuncture group and the control group. Linear mixed model main analyses of interaction time versus group for primary end point gave p 0.26. There was no statistically significant baseline-corrected difference for the whole period, (Table 1) or at any measured time period from baseline to the last measure after four weeks. There were no statistically significant changes over time in the interaction analyses between groups dividing into strata of more or
less than four hours crying at baseline (p= 0.20), the eventual influence of other types of treatment (p = 0.30), type of feeding (p=0.97), or concerning the primary blinding validation question (p= 0.98). Corrected for baseline differences, there was a small tendency in favour of the acupuncture group, with a non-significant baseline-corrected sum of mean differences in crying time during the assessment period of 13 minutes (95% CI -24 to + 51). The effect size for the intervention in this trial was estimated to be SMD 0.16.

**Table 1** Baseline-corrected mean overall difference in crying time reduction: The ST36 infantile colic acupuncture trial.

<table>
<thead>
<tr>
<th></th>
<th>Total mean difference (minutes)</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acupuncture Group/Control group</td>
<td>13,25</td>
<td>0.482</td>
<td>-23,99</td>
</tr>
</tbody>
</table>

**Figure 4** Plot diagram of changes in crying time: The ST36 infantile colic acupuncture trial.
Secondary end-points.

On day 6, nine out of 38 in the acupuncture group and 10 out of 41 in the control group fulfilled Wessel’s criteria of more than three hours’ crying per day (odds ratio 1.1, CI 0.4 – 3.2, p = 0.94). This lack of differences was repeated at day 12 (p 0.60) and day 33 (p 0.31). The parents’ evaluation showed a slight tendency in favour of acupuncture over time, with a statistically significant difference on day 33 of 0.51 (CI 0.04 – 0.99), (p 0.034). There were no serious adverse effects in the acupuncture or control group.

3.3. Paper III


3.3.1 Method

The protocol was registered at the University of York Centre for Reviews and Dissemination – PROSPERO 2015: CRD42015023253 (158). The study has been reported using PRISMA (159, 160) and PRISMA-IPD (161) recommendations.

Eligibility criteria

We included only full RCTs of acupuncture treatments for infantile colic. Participants were infants fulfilling Wessel’s criteria or Wessel’s modified criteria of infantile colic. There were no exclusion criteria. Intervention was percutaneous needle acupuncture treatment, 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. 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 (162, 163). We sought assistance from the Chinese Centre for Evidence-Based Medicine, Beijing University of Chinese Medicine to search Chinese language databases.

English language database search: Electronic scoping searches were conducted in Best Practice, UpToDate, Cochrane, and Prospero from inception to the search date. 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.

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. One methodological expert, Yang Shao 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, the thesis’ author, in Beijing in May 2016. Relevant articles according to the predefined eligibility criteria were translated and further discussed with me.

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.
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, statistical methods and outcome reporting would make an analysis based on aggregated data insufficient. So, we invited trialists of eligible trials to take part in a collaborative group and asked them 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 (158). All datasets were stored securely and pseudo-anonymously by identifier deletion. Kjetil Gundro Brurberg of the Norwegian Institute of Public Health, Division for Health Services, independent of the trialists, performed all analyses, based on the IPD from the included trials and carried out the meta-analysis.

Risk of bias

We used the Cochrane Collaboration’s tool for assessing risk of bias (164). 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. The quality of the blinding procedure was tested for each of the included studies by performing Pearson’s chi-squared tests and by calculating James and Bang blinding indices (165). We performed a sensitivity analysis in which studies assessed to be at high risk of bias were excluded.

Data synthesis

The data synthesis was carried out at the Norwegian institute of Public Health, by Kjetil
Gundro Brurberg at the Division of Health Services. Data from the included studies were analysed using a two-step approach. At the first step, he 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(166). 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, he modelled odds ratios (ORs) by logistic regression using the function GLM in R. The results were presented as ORs with 95% CIs and were adjusted for baseline differences in crying time. At the second step, he 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**

On the basis of papers trying to outline guidance in establishing clinical relevant endpoints in pain trials in adults and children, outlined in previous section 2.2, and after discussion in the Lund consensus meeting, we considered the minimally important difference in baseline-corrected crying time between acupuncture and control to be about 30 minutes, a number that
is roughly equivalent to an SMD of 0.3. The primary end-point was a baseline-corrected difference in crying time in minutes between intervention and control as measured during treatment one week after treatment ended and one month after treatment ended, with a 30 minutes difference in reduction in crying time between intervention and control groups considered as a clinically relevant effect. We changed “one 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. 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): 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. James ‘s blinding index were added post hoc to add information. The chi-squared test may indicate adequate blinding if P > 0.05, but the sensitivity becomes poor if both groups are fully unblinded. 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 and Bang’s blinding index is calculated for each intervention group separately, and reflects adequate blinding if it centres around 0 (165). We took all the different tests 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 I² 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.

**Subgroup and sensitivity analysis**

We performed sensitivity analyses based on the risk of bias in included studies. The results of the blinding validation tests were used to guide the risk of bias assessments in the blinding of outcome assessor domain.

**GRADE**

We assessed overall quality of evidence according to Grading of Recommendations Assessment, Development and Evaluation (GRADE) (167, 168).

**3.3.2. Main results**

We identified 384 English language and 24 Chinese language studies after removal of duplicates, 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) (43); one reported on data concerning feeding and stooling changes from the same study as reporting on crying time changes (151); and one was an open pilot study with seven patients and changes during the trial (169). Individual patient data were sought and obtained for all eligible RCTs (44, 45, 170). One trial (45) consisted of two active treatment groups receiving standardised and semi-individualised acupuncture,
respectively, and in accordance with the protocol of this trial(153), the two active arms were treated as separate comparisons by randomly splitting the control group. For two of the included studies, we reported previously unpublished data, i.e. the results of blinding validation in Landgren et al 2010(44) and the result of parental evaluation of effects in Landgren et al 2010(44) and Landgren et al 2017(45).
Figure 5 The PRISMA IPD flow diagram.

© Reproduced with permission of the PRISMA IPD Group, which encourages sharing and reuse for non commercial purposes
**Risk of bias and blinding validation:**

All included RCTs had adequate randomisation procedures reported and allocation concealment described. Acupuncturists were not blinded in any of the studies. Parents acted as outcome assessors in all studies, and a thorough blinding validation indicated that the parents in Landgren et al.’s study from 2010(44) seemed to be un-blinded to treatment allocation. In contrast, Landgren et al. in 2017(45) achieved blinding outcome assessment in the control group, but not in the acupuncture group, whereas Skjeie et al (170) were able to mask all parents irrespective of the group to which the infant was allocated.

**Figure 6** Risk of bias summary report
Table 2 Blinding indices. The IPD meta-analysis.

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Blinding index</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>James</td>
<td>Bang (Acupuncture)</td>
<td>Bang (Control)</td>
<td>Chi-squared test</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Landgren 2010</td>
<td>43</td>
<td>0.21⁺</td>
<td>0.77⁻</td>
<td>0.67⁻</td>
<td>0.87⁺</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>96</td>
<td>0.56⁻</td>
<td>0.44⁺</td>
<td>−0.08⁺</td>
<td>0.007⁺</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>96</td>
<td>0.51⁻</td>
<td>0.63⁻</td>
<td>−0.08⁺</td>
<td>&lt;0.0001⁻</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skjeie 2013</td>
<td>83</td>
<td>0.46⁻</td>
<td>0.00⁺</td>
<td>0.18⁻</td>
<td>0.55⁺</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

⁺ Not blinded for outcome assessors (parents).
⁻ Unsure blinding of outcome assessors.
⁺⁺ Blinded for outcome assessors.

End-points

Heterogeneity was negligible in all analysis.

Primary end point – Reduction in crying time: We did not detect important differences in crying time between acupuncture and no acupuncture control at any of the prespecified time periods. There was a statistically significant difference in mean crying time (MD −24.88 minutes/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 319 (SMD −0.13; 95% CI −0.35 to 0.09) when the study assessed as un-blinded was excluded in a sensitivity analysis.
Primary end-point: crying time

A Mean difference – MD (Mid-treatment, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Mean Difference IV, Random, 95% CI</th>
<th>Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landgren 2010</td>
<td>-55.30 [-94.62, -15.98]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>-20.60 [-61.97, 20.97]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>-9.90 [-45.06, 25.26]</td>
<td></td>
</tr>
<tr>
<td>Skjærov 2013</td>
<td>-12.00 [-62.41, 38.41]</td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI): -24.88 [-46.20, -3.57]

Heterogeneity: $I^2 = 32.33$, $p = 0.002$, $d = 3$, $P = 0.35$, $I^2 = 9$

Test overall effect: $Z = 2.29$ ($p = 0.02$)

B Standardised mean difference – SMD (Mid-treatment)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landgren 2010</td>
<td>-0.49 [-0.84, -0.14]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>-0.19 [-0.67, 0.30]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>-0.11 [-0.50, 0.28]</td>
<td></td>
</tr>
<tr>
<td>Skjærov 2013</td>
<td>-0.09 [-0.48, 0.29]</td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI): -0.23 [-0.42, -0.05]

Heterogeneity: $I^2 = 0.00$, $p = 0.30$, $d = 3$, $P = 0.35$, $I^2 = 9$

Test overall effect: $Z = 2.43$ ($p = 0.01$)

C Mean difference – MD (End of treatment, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Mean Difference IV, Random, 95% CI</th>
<th>Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landgren 2010</td>
<td>-10.90 [-58.02, 20.20]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>-9.80 [-51.85, 32.05]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>-10.20 [-45.36, 24.96]</td>
<td></td>
</tr>
<tr>
<td>Skjærov 2013</td>
<td>-3.10 [-54.34, 47.84]</td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI): -11.41 [-31.77, 9.85]

Heterogeneity: $I^2 = 0.00$, $p = 0.00$, $d = 3$, $P = 0.97$, $I^2 = 9$

Test overall effect: $Z = 1.10$ ($p = 0.27$)

D Standardised mean difference – SMD (End of treatment)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landgren 2010</td>
<td>-0.17 [-0.62, 0.30]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>-0.09 [-0.47, 0.25]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>-0.11 [-0.50, 0.27]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skjærov 2013</td>
<td>-0.02 [-0.41, 0.37]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI): -0.10 [-0.30, 0.10]

Heterogeneity: $I^2 = 0.00$, $p = 0.31$, $d = 3$, $P = 0.30$, $I^2 = 9$

Test overall effect: $Z = 1.03$ ($p = 0.28$)

E Mean difference – MD (Long-term follow-up, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Mean Difference IV, Random, 95% CI</th>
<th>Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skjærov 2013</td>
<td>-11.83 [-42.85, 39.19]</td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI): -11.83 [-42.85, 39.19]

Heterogeneity: $I^2 = 0.00$, $p = 0.00$, $d = 3$, $P = 0.96$, $I^2 = 9$

Test overall effect: $Z = 1.03$ ($p = 0.28$)

F Standardised mean difference – SMD (Long-term follow-up)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skjærov 2013</td>
<td>-0.90 [-0.48, 0.30]</td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI): -0.90 [-0.48, 0.30]
Primary end-point: Crying time with sensitivity analysis

A  Mean difference – MD (Mid-treatment, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Mean Difference IV, Random, 95% CI</th>
<th>Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landgren 2010</td>
<td>26.6%</td>
<td>-55.30 [-94.62, -15.98]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>24.1%</td>
<td>-20.50 [-61.97, 20.97]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>32.5%</td>
<td>-9.90 [-45.06, 25.28]</td>
<td></td>
</tr>
<tr>
<td>Skjeie 2013</td>
<td>16.8%</td>
<td>-12.00 [-62.41, 38.41]</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>-24.88 [-46.20, -3.57]</strong></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 42.33; \chi^2 = 3.29, \text{df} = 3 (P = 0.35); I^2 = 9\%$
Test for overall effect: $Z = 2.29 (P = 0.02)$

B  Sensitivity analysis. Mean difference – MD (Mid-treatment, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Mean Difference IV, Random, 95% CI</th>
<th>Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landgren 2010</td>
<td>0.0%</td>
<td>-55.30 [-94.62, -15.98]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>32.6%</td>
<td>-20.50 [-61.97, 20.97]</td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>45.3%</td>
<td>-9.90 [-45.06, 25.28]</td>
<td></td>
</tr>
<tr>
<td>Skjeie 2013</td>
<td>22.1%</td>
<td>-12.00 [-62.41, 38.41]</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>-13.82 [-37.50, 9.86]</strong></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 0.00; \chi^2 = 0.15, \text{df} = 2 (P = 0.93); I^2 = 0\%$
Test for overall effect: $Z = 1.14 (P = 0.25)$
Secondary end-point – Disappearance of colic: We found no 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 to 2.70).

Secondary end-point – Parental evaluation: Parents of the infants in the acupuncture groups were more likely to report that the colic had improved at the end of the treatment, with an OR of 3.03 (95%CI 1.56 to 5.89) for rating the condition as much improved and OR 2.67(95%CI 1.43 to 4.97) for improved.

Adverse effects: No major adverse effects were reported in the included trials. Crying during treatment was assessed by Landgren et al. in 2010(44) and 2017(45). 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 to 20.64). 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). The OR was 2.53 (95% CI 0.72 to 8.86).
Secondary end-points

A Parental report at the end of treatment

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IV, Random</td>
<td></td>
</tr>
<tr>
<td>Landgren 2010</td>
<td>12.5%</td>
<td>10.39</td>
<td>[1.67, 64.62]</td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>21.9%</td>
<td>1.78</td>
<td>[0.46, 6.86]</td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>25.5%</td>
<td>4.64</td>
<td>[1.34, 16.02]</td>
</tr>
<tr>
<td>Skjøde 2013</td>
<td>40.2%</td>
<td>2.11</td>
<td>[1.82, 5.45]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>100.0%</td>
<td>3.03</td>
<td>[1.56, 5.89]</td>
</tr>
</tbody>
</table>

Heterogeneity: $I^2 = 91\%$, $Q = 3.36, df = 3 (P = 0.34)$
Test for overall effect: $Z = 3.27 (P = 0.001)$

1.4.2 Improved

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IV, Random</td>
<td></td>
</tr>
<tr>
<td>Landgren 2010</td>
<td>26.7%</td>
<td>1.68</td>
<td>[0.45, 6.51]</td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>30.3%</td>
<td>3.60</td>
<td>[1.16, 11.10]</td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>26.8%</td>
<td>3.53</td>
<td>[1.06, 11.74]</td>
</tr>
<tr>
<td>Skjøde 2013</td>
<td>18.3%</td>
<td>2.19</td>
<td>[0.51, 9.38]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>100.0%</td>
<td>2.67</td>
<td>[1.43, 4.97]</td>
</tr>
</tbody>
</table>

Heterogeneity: $I^2 = 90\%$, $Q = 3.36, df = 3 (P = 0.48)$
Test for overall effect: $Z = 2.79 (P = 0.006)$

1.4.3 Worse

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IV, Random</td>
<td></td>
</tr>
<tr>
<td>Landgren 2010</td>
<td>100.0%</td>
<td>0.83</td>
<td>[0.22, 3.18]</td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>Not estimable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>Not estimable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skjøde 2013</td>
<td>Not estimable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>100.0%</td>
<td>0.83</td>
<td>[0.22, 3.18]</td>
</tr>
</tbody>
</table>

Heterogeneity: $I^2 = 90\%$, $Q = 3.36, df = 3 (P = 0.76)$
Test for overall effect: $Z = 2.79 (P = 0.079)$

B <180 min crying per day at the end of treatment

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IV, Random</td>
<td></td>
</tr>
<tr>
<td>Landgren 2010</td>
<td>31.9%</td>
<td>1.68</td>
<td>[0.62, 4.66]</td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>20.3%</td>
<td>2.85</td>
<td>[0.82, 9.62]</td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>18.7%</td>
<td>0.80</td>
<td>[0.22, 2.90]</td>
</tr>
<tr>
<td>Skjøde 2013</td>
<td>29.2%</td>
<td>1.30</td>
<td>[0.40, 3.94]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>100.0%</td>
<td>1.54</td>
<td>[0.88, 2.70]</td>
</tr>
</tbody>
</table>

Heterogeneity: $I^2 = 90\%$, $Q = 3.36, df = 3 (P = 0.58)$
Test for overall effect: $Z = 1.90 (P = 0.13)$

C Crying during treatment

- Landgren 2010 (No acupuncture)
- Landgren 2010 (Standard acupuncture)
- Landgren 2017 (Standard acupuncture)
- Landgren 2017 (Individualised acupuncture)
Table 3 Summary of findings – GRADE. The IPD meta-analysis

GRADE

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

<table>
<thead>
<tr>
<th>Timing</th>
<th>Average min. crying per day</th>
<th>SMD (95% CI)</th>
<th>No. of participants</th>
<th>Quality of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Acupuncture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid-treatment</td>
<td>193 min</td>
<td>25 min less</td>
<td>-0.23 (-0.42 to -0.05)</td>
<td>307 (3 studies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(46 less to 4 less)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>End of treatment</td>
<td>156 min</td>
<td>11 min less</td>
<td>-0.10 (-0.29 to 0.08)</td>
<td>304 (3 studies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(32 less to 9 more)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long-term follow-up</td>
<td>97 min</td>
<td>12 min less</td>
<td>-0.09 (-0.48 to 0.30)</td>
<td>79 (1 study)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(63 less to 39 more)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^A Wide confidence intervals (CI) and imprecision; ^B One study with few participants. SMD, standardised mean differences.

Summary of findings for secondary outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Absolute effect per 100 (95% CI)</th>
<th>Relative effect OR (95% CI)</th>
<th>No. of participants</th>
<th>Quality of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Acupuncture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No colic§</td>
<td>60</td>
<td>70</td>
<td>1.54 (0.88 to 2.70)</td>
<td>304 (3 studies)</td>
</tr>
<tr>
<td></td>
<td>(57 to 80)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Much improvement</td>
<td>26</td>
<td>52</td>
<td>3.03 (1.56 to 5.89)</td>
<td>264 (3 studies)</td>
</tr>
<tr>
<td></td>
<td>(35 to 67)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some improvement</td>
<td>65</td>
<td>83</td>
<td>2.67 (1.43 to 4.97)</td>
<td>264 (3 studies)</td>
</tr>
<tr>
<td></td>
<td>(73 to 90)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worsening</td>
<td>30</td>
<td>26</td>
<td>0.83 (0.22 to 3.18)</td>
<td>46 (1 study)</td>
</tr>
<tr>
<td></td>
<td>(9 to 58)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crying during treatment</td>
<td>37</td>
<td>81</td>
<td>7.5 (2.7 to 20.6)</td>
<td>81 (1 study)</td>
</tr>
<tr>
<td></td>
<td>(61 to 92)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

§ defined as >180 minutes of crying per day; ^A Wide confidence intervals (CI) and imprecision; ^B Not possible to estimate in two studies because of too few events.
3.4. Paper IV


3.4.1. Method

Study design

We chose a qualitative approach, with participating observations, interviews and literature searches. An extended, flexible approach was developed and a combination of qualitative methods was used, reflecting the validity threats and logistical challenges of conducting qualitative research on the attitudes towards and use of pediatric acupuncture among TCM clinicians in Shanghai, China. Subject knowledge matters, as described in Kvale and Brinkmann (171), guided the decision not to do surveys or rely on focus group interview (172) as the only sources of information. The study process was informed by “Qualitative Research Design” by J. Maxwell (173). Special emphasis was on validity threats: We used Maxwell’s eight-point checklist to strengthen validity as a guide when designing and carrying out the study and in the analysing process. His recommendations of several information gathering methods to ensure validity was central to our decisions (174):

Long-term involvement- we included four field visits over the course of fourteen months.

Rich data- we transcribed ad verbatim all field notes, quotes from informal conversations and formal interviews.

Respondent validation- we had two stages of informant validation.

Intervention- this was not appropriate in our study.

Searching for Discrepant Evidence and negative Cases- we did search for and reported discrepant evidence.

Triangulation-we used both participant observation and informal conversations, formal
interviews and textbook searches.

Numbers- we counted opinions, pages and patients, and reported it in a table.

Comparison- we had participant observations at three departments and at different times.

The main analyses of the transcribed field notes and interviews were informed by Thematic analysis (175), focused on manifest content, and is further described in the Data Analysis section below.

**Setting**

I conducted four field visits between February 2014 and March 2015 to Longhua Hospital, which is a teaching hospital of the Shanghai University of Traditional Chinese Medicine (SHUTCM) (figure 1). Longhua Hospital is a municipal TCM hospital with 2000 beds. The WHO Collaboration Centre at SHUTCM facilitated study access and granted permission to collect data at Longhua Hospital.

**Data collection**

1. The field visits: 3 days of participant observation in the pediatric department, 3 days of participant observation in the acupuncture department and 1 day of participant observation in the Tui na department. An interpreter who was either an English-speaking TCM doctor or an English–Chinese linguist was present at any time. We followed the daily routines and engaged in informal conversations with several acupuncturists, and pediatric and Tui na specialists. The interpreters themselves were also a rich source of information.

2. The focus group interview: A 90 min, semi-structured focus group interview using a predefined interview guide, not pilot-tested. The interview was conducted in English and without an interpreter, outside the workplace. The informants were three leading officials at acupuncture units in Shanghai, all of whom had clinical and administrative responsibilities. They were experienced acupuncturists.

3. The in depth interview: A 60 min, in-depth individual interview using a predefined
interview guide, not pilot-tested, with a non-acupuncture TCM practitioner who had academic credentials, including work and research experience in Western Europe.

4. The literature search: A search in the standard national textbooks used for teaching acupuncture, Tui na and herbal medicine at Shanghai University of Traditional Chinese Medicine, one book for each curriculum.

5. Informant validation: Two stages of informant validation on the main and final results. Main results by one central informant, and final results by two independent informants.

Data analysis

All data were used, including notes from participant observations, informal conversations, semi-structured interviews and textbook page analysis. The field notes were structured and systematised, including the participant observations and the information gathered from informal conversations. The focus group and in-depth interviews were recorded on two parallel, portable tape recorders and transcribed in verbatim by the first author. The data analyses were carried out using thematic analysis (175). The analyses were performed by hand and coding was primarily on semantic (manifest) content. The first author read through the texts several times to define broad categories from the interviews and field notes with subsequent coding and re-categorisation. The main categories and selected opinion statements were then sequenced into tables. The initial analysis was then reviewed by the second author before refinement and undergoing final organisation. The study was reported in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist (176).
3.4.2. Main results

We explored the use and recommendations for needle acupuncture on infants and small children by TCM practitioners in Shanghai. During 3 days of participant observations at the Longhua hospital outpatient acupuncture clinic, which receives on average 400 patients per day, we observed only two children and no infants. During formal interviews and informal conversations with TCM practitioners representing different specialties, the prevailing opinion was that needle acupuncture for infants and small children is considered painful, is inconvenient and is not indicated for routine clinical practice, except for certain neurological diseases.

We learned that TCM methods for children have traditionally focused on herbal medical treatments, and to a lesser extent, on Tui na. Acupuncture had historically been used for certain acute illnesses, which are now treated in Western medicine hospitals. The majority of acupuncturists, and all of the non-acupuncture TCM practitioners, had limited or no
experience treating children or infants with needle acupuncture and would not recommend it as a therapeutic option. The exceptions were the opinions of two TCM acupuncture officials.

Two themes emerged as reasons for the non-use of acupuncture in small children: 1 - internal TCM traditions and practices, which do not support the routine practice of needle acupuncture in infants and toddlers and 2 - external public health care system changes which increasingly limit the use of acupuncture and TCM practices in favour of Western medicine.

**Figure 10** Results in numbers. The Shanghai qualitative study.
4. Discussion

4.1. Development of the thesis

The main research question addressed in this thesis concerning the use of acupuncture treatments for infantile colic was “does the needle work?”

Initially, the answer to this question would be “yes”, because many times we had seen an apparently astonishing change from an intensely crying infant to a calm one during our clinical interventions. In parallel, a growing body of evidence of the effects of acupuncture for pain problems in adult populations was emerging. We planned to test our own practice with a pilot study followed by a multicentre RCT, and a qualitative study on the attitudes and beliefs of GPs with broad acupuncture experience.

The results of our main project, the multicentre RCT, took us by surprise, and made us change course. It is therefore necessary to explain here the rationale for the decisions we made. What we wanted to test was the efficacy component of this acupuncture practice, which we anticipated would approximate the needle effect, because the known anticipation/reward effects of acupuncture in adults might not occur in infants. We believed that if we could blind the outcome assessors—the parents—and standardize the treatment procedure, we would have a truly blinded acupuncture trial, where the efficacy of the needle would show up. If there was an effect, then this would be the needle effect, and we would have data on both the efficacy of acupuncture for infantile colic, and on the needle component of acupuncture treatment in this patient population, adding data to the empirical body of knowledge on the efficacy of acupuncture.

The pilot study was created to test whether our protocol for the main RCT could hold the line of standardization and the blinding of outcome assessors. We considered inadequate blinding to be the most important element of possible bias to be minimized (102, 104, 177). Indeed, the
pilot study showed certain weaknesses of the protocol (169). With this being an open study, we could change the protocol of the main RCT to secure standardization and blinding. When the baseline-corrected result of the main analysis of the RCT was produced in 2013, we had not reached the estimated sample size, although we had 79 final full outcome datasets. We were rather surprised by the results. The treatment-effect difference between acupuncture and no acupuncture control in a linear mixed model over time was not clinically relevant, and not statistically significant. However, it was remarkably similar to the needle-specific effect sizes of the large meta-analysis of acupuncture for chronic pain conditions using individual patient data (IPD) from over 17,000 adults (SMD 0.15–0.23), published in 2012 (88). Effect sizes of this magnitude in treating pain problems are normally not clinically relevant whether in adults (95, 96, 98) or children (97, 99, 105). We reasoned as follows.

1. We had to act on our data and adjust the thesis project.

2. The practice of needle acupuncture on infants is potentially painful, as shown in the 2010 study by Landgren et al. (44) on acupuncture for infantile colic.

3. We often had experience of short treatment crying episodes during our own previous practice on treating these infants with needles, but considered the apparent benefit of the intervention to outweigh this.

4. We knew that needle acupuncture involves a painful signal. What if our study reflected the real possible needle effect and the efficacy of treatment was indeed very low? Should this practice be recommended at all for infants, who cannot consent to treatment?

We could not make any firm conclusions on the basis of one small RCT, and with other controlled trials showing more positive results. We needed more data. Therefore, we abandoned the original final part of the thesis, namely the qualitative study of attitudes and beliefs among GPs who use acupuncture in everyday clinical practice. Two other published controlled studies of acupuncture in treating infantile colic both claimed positive results and
recommended the practice (43, 44), and we knew of a larger two-armed RCT under way (153, 178). We therefore changed course and, as the final quantitative part of the thesis, we decided to run a systematic review and meta-analysis of existing fully completed RCTs up to when the final RCT was published in 2016 (45). Differences in statistical methods and baselines convinced us that only an IPD-based meta-analysis with access to all raw data files would solve our problem. We also needed qualitative information on the origins and background of the initial recommendations of the practice of needle acupuncture on infants—whether this entailed new Western or original Chinese TCM-based practice—to complete our evaluation of the practice. Therefore, in parallel, we planned and carried out a field study in Shanghai in 2014–2015 to gain knowledge on this matter. With all this information combined, we could evaluate our own previous practices and available empirical data, and thereby have enough data on which to base recommendations.

4.2. Summary of main results

A pilot study of ST36 acupuncture for infantile colic

We improved the standardization of the intervention procedure and changed the blinding procedure as a result of the pilot study. Blinding validation questions were considered necessary. The pilot study thus led to important changes that were implemented in the final trial protocol for the subsequent RCT.

Acupuncture for infantile colic: a blinding-validated, randomized controlled multicentre trial in general practice

The blinding validation questions in the RCT showed a random distribution between intervention and control, with \( P = 0.41 \) for the acupuncture or no acupuncture question and \( P = 0.60 \) for the question of needle marks, indicating likely true blinding. We found no statistically significant difference in crying time reduction between the acupuncture and
control groups at any of the measured intervals, nor in the main linear mixed model analysis of differences in changes over time ($P = 0.26$). There was a tendency in favour of the acupuncture group, with a statistically non-significant total baseline-corrected mean of 13 minutes difference in crying time between the groups (95% confidence interval, CI = −24 to 51). This was not considered clinically relevant according to the protocol, whether or not reaching statistical significance.

**Acupuncture treatments for infantile colic: a systematic review and IPD meta-analysis of blinding-validated randomised controlled trials.**

We included three RCTs with data from 307 participants. Only one of these obtained a probable successful blinding of the outcome assessors in both the acupuncture and control groups. The mean difference in crying time between acupuncture intervention and no acupuncture control was −24.9 minutes (95% CI = −46.2 to −3.6; three trials) at mid-treatment, −11.4 minutes (95% CI = −31.8 to 9.0; three trials) at the end of treatment, and −11.8 minutes (95% CI = −62.9 to 39.2; one trial) at the 4-week follow-up. The corresponding 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), respectively. No results were clinically relevant according to our protocol. 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 (mean difference: −13.8 minutes, 95% CI = −37.5 to 9.9) and SMD of −0.13 (95% CI = −0.35 to 0.099). The registration of crying during treatment suggested that more infants cried during acupuncture treatment versus controls, indicating needle pain (odds ratio, OR = 7.7; 95% CI = 2.7 to 20.6; one trial). Using the GRADE system, we rated the findings as showing moderate quality evidence.
“Big needles, small bodies”. The absence of acupuncture treatment for infants in contemporary Shanghai: a qualitative study

The Longhua Hospital outpatient acupuncture clinic in Shanghai receives on average 400 consultations per day, including referrals from the paediatric department. During three days of participant observations at the hospital, we saw two children but no infants. During three days at the paediatric department and one day at the Tui Na department, we saw no referrals for acupuncture treatment. Formal interviews and informal conversations with acupuncturists and other TCM professionals revealed that acupuncture was neither routinely practised nor recommended for infants and small children. Percutaneous needle acupuncture was considered potentially painful for this young patient population.

4.3. Methodological considerations

4.3.1. Relevant methodological considerations in controlled clinical trials of acupuncture and pain

Blinding and blinding validation in acupuncture trials.

The Cochrane Handbook of Systematic Reviews of Interventions discusses bias due to inadequate blinding; Chapter 8 deals with assessing the risk of bias in included studies (164). It is argued that a lack of blinding results in an exaggerated estimated intervention effect of 9% in RCTs in empirical studies with both objective and subjective outcomes. Furthermore, subjective outcomes such as pain, crying, or days experiencing a common cold have been shown to be more influenced by lack of blinding than objective outcomes such as haemoglobin measures, or the incidences of coronary thrombosis or joint swelling. The assessment of blinding should take into account the type of outcome in the evaluation of risk of bias. In their meta-epidemiological study in 2008, Wood et al. found trials with subjective outcome estimates to be exaggerated when there was a lack of blinding with ORs of 0.75 (0.61 to 0.93) (104). Hróbjartsson et al. of the Cochrane Collaboration Centre in Copenhagen
published a systematic review in 2014 on bias caused by lack of blinding (102). They found “empirical evidence of pronounced bias due to lack of patient blinding in complementary/alternative randomized clinical trials with patient-reported outcomes”; 10 of the 12 trials involved acupuncture. The estimated average effect exaggeration was an SMD of 0.67 in the acupuncture trials. In their 2000 article on the statistics of blinding in clinical trials (177), Day and Altman noted that blinding patients to treatment is particularly important when the response criteria are subjective, such as alleviation of pain; furthermore, blind assessment of outcome might be more important than blinding the administration of the treatment. In her 2016 PhD dissertation “Acupuncture for postoperative morbidities in children and placebo by proxy” (75), Ingrid Liodden from Tromsø, Norway argued that placebo by proxy could play a significant part in subjective treatment outcomes in children. Placebo by proxy refers to placebo effects caused by a positive expectancy felt by other persons, for example, parents or caregivers. There is no item on blinding validation reporting in STRICTA (STandards for Reporting Interventions in Clinical Trials of Acupuncture of 2010) (157), which is an extension of the CONSORT (179) statement. Blinding validation, where possible, could increase trust in the results.

**Statistical significance and clinical relevance in trials of pain.**

Subjective treatment outcomes in trials, such as evaluations of pain or crying, are most often reported as patient-reported outcomes (PROs). A statistically significant effect is not necessarily a clinically relevant effect, and there have been several attempts in various clinical disciplines to reach consensus on what determines the minimal important difference (MID), the minimal important change (MIC), or the minimal clinically important difference (MCID) for pain and pain-related conditions (95-98, 105, 180). As a benchmark, Dworkin et al. set a 10–20% reduction of chronic pain in an anchor-based numerical rating scale as minimally
important, and 30% as moderately important (95). Furlan et al. of the Cochrane Group on back pain established a 30% difference on a VAS/Numerical Rating Scale (NRS) on pain as being clinically significant (96). This was partly based on Ostelo et al. in *Spine*, which proposed a MIC value of 15 mm on a VAS scale or a 30% improvement from baseline (98). Pach et al. (181) and Skonnord et al. (106) adhered to this threshold in trials of the use of acupuncture for treating back pain. On self-reported pain scores in children, von Baeyer suggested a 10–20% reduction or 10–20 mm on a VAS scale to be the smallest MID (99, 105). On pain in children with rheumatic disease, Dhanani et al. estimated a minimum of 8 mm on a 100 mm VAS to achieve meaningful improvement (97). Discussing acute pain in children, Powell et al. recommended 10 mm on a VAS scale as MID (180). We have found no papers that attempt to establish guidelines on MCID or MID measures in trials on pain or crying in infants, where parents would estimate the changes.

4.3.2. Papers I and II. The pilot study and the main RCT

An important methodological consideration was whether the pilot study, conducted by the first and second authors of this study in their own general practices, should be open or closed in testing the original protocol. Having a closed study adhering to a strict protocol would limit changes and experimentation, but would also add inclusion numbers to the main RCT. Having an open pilot study would mean that we could not use the eventual end-point data in the final analyses, but could open up to necessary changes, facilitating interchanging views and observations and comments from participants. We needed to secure standardization and blinding, and especially had to ensure that the blinding of outcome assessors was tested and adequate before starting the main RCT. Thus, the choice was easy, and we conducted an open pilot study.
We chose a standardized needle method (bilateral acupuncture point ST36, 30 seconds, ~12 mm deep, no manipulation) that was identical in the pilot study and in the main multicentre RCT. This was because we wanted to test our own practice, using our own recommended treatment formula from the Norwegian Society of Medical Acupuncture. This protocol had a seemingly valid TCM rationale (182-184), a possible neurophysiological explanation (156), and it was simple to administer. Few data existed on tested alternatives at the time, and arguments nationally or internationally for one or the other method were anecdotal (62).

There were arguments on standardized versus individualized acupuncture treatments, but there was little evidence for choosing one or the other until the large IPD study by MacPherson et al. in 2013 showed no difference between the various styles and choices of treatments for treating pain conditions in adults (69). After our first protocol draft was drawn up, Reinthal et al. (43) and Landgren et al. (44) published on acupuncture point LI 4.

However, we did not consider those study results sufficiently strong to change our proposed needle method. The short treatment period of three consecutive days was again chosen on pragmatic grounds. In the absence of guidelines, this was what we recommended and used, given that we had had positive experiences in treating cases of infantile colic. We were also aided by anecdotal references in the literature to needle acupuncture in children working faster, with fewer needles, and with a non-retaining method involving the quick insertion and withdrawal of needles (62).

The primary outcome—differences in crying time reduction between acupuncture and no acupuncture control—was in line with a larger body of data internationally on testing therapies for infantile colic, whether as a dichotomous 50% crying time reduction cut-off, or as a continuous mean reduction in crying time (17, 18, 33, 36).

The emphasis on an optimal blinding procedure of the parents as outcome assessors, and the testing and subsequent change of blinding procedure to an ink dot plus an adhesive tape were
a consequence of experiences with post-insertion erythema and parents’ notion in the pilot study of skin needle marks. We used two blinding validation questions. After the first treatment, we asked the parents: “Do you think your child has had acupuncture or not?” After four weeks, we asked: “Have you seen any needle marks?” These questions aimed to control for what we considered important: that parents could sense intervention changes in the child, and that they might be able to spot needle marks in spite of the blinding procedure. In retrospect, and after having learned more about blinding validation tests and questions (165) during the work on the thesis, the questions were a form of validation, but not an optimal one (see section 4.4.2, Limitations and Strengths in Papers I–IV).

Our main statistical estimate for the primary end-point was a linear mixed model to test differences between groups over time as one result. We decided on this to obtain an overall test of efficacy. We also reported on all end-points and eventual statistical significance/clinical relevance in a table and in the text. None of the results showed statistical significance given our small sample size and the surprisingly small SMD. However, our emphasis and thus primary objective concerned overall efficacy and this statistical model suited the purpose. We chose a baseline-corrected difference to correct for the problem of reduced or exaggerated results when not countering baseline differences in small trials. Professor Magne Thoresen at the Department of Biostatistics, University of Oslo was of invaluable help in advising us on choice of tests, and carried out the statistical analysis.

4.3.3. Paper III. The systematic review and IPD meta-analysis

The first methodological consideration in a systematic review is how to obtain the full existing dataset. Most systematic reviews are based on English-language sources. This is an inherent limitation, and possibly more so in trials of acupuncture, as the method and rationale originate in China. However, Chinese-language medical research databases have developed
considerably in recent years, and there are recommendations to include these in systematic reviews (162, 163, 185). Yan Shao at the Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine searched and interpreted the Chinese-language database to avoid omitting trials of infantile colic that we guessed would not be included on the basis of our Shanghai field study, but that we nevertheless could not take for granted. For a complete search on acupuncture treatments for infantile colic, we considered it necessary to include the Chinese-language databases.

The decision and subsequent rearrangements of changing from an ordinary meta-analysis to an IPD-based meta-analysis are described thus in Paper III.

“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.”

We chose a 30-minute MID in crying time reduction between acupuncture and no acupuncture control in the final protocol of our IPD meta-analysis. The notion of clinical relevance in trials of efficacy is particularly relevant in general practice, where a myriad of trials with statistical significance from a host of specialities often confuse more than inform the reader. More specifically, in trials with subjective outcomes such as pain, there is an emerging consensus that establishing clinically relevant effects (i.e., MID/MIC/MCID) is a prerequisite for relevant information on the effects of clinical trials to guide clinicians in their choice of treatment (95, 96, 105). Our 30-minute threshold emerged at the Lund University
consensus meeting as a pragmatic combination of available threshold data from the literature on pain conditions in children (97, 99, 105, 180), and from the clinical experience and suggestions of the trialists of the IPD.

The decision to apply a triage of blinding tests and evaluate the blinding or unblinding of the trials led to discussions of available tests and their pros and cons. As stated in our IPD-based article:

“The chi-squared test may indicate adequate blinding if p > 0.05, but the sensitivity becomes poor if both groups are fully un-blinded. 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. Bang’s blinding index is calculated for each intervention group separately, and reflects adequate blinding if it centres around 0. We took all the different tests into account before making any conclusions about the success of blinding.”

There is no gold standard of blinding test evaluation. We decided on these combinations on the basis of an informative article by Bang et al. (165) and because we had experience in using the chi-squared test in our own RCT. We decided to go further and aligned the James and Bang indices together in a synthesis with the common chi-squared method. This made sense for us in evaluating overall blinding test validity. Although a novel test combination, it adjusted for imbalances in the individual tests and added robustness to Bang’s index (165).

4.3.4. Paper IV: “Big needles, small bodies”. The Shanghai qualitative study

Following our non-significant and rather surprising RCT results, and reflecting on and discussing in the small research group consisting of my co-author Trygve Skonnord, and my two supervisors Arne Fetveit and Mette Brekke, I mused on the fact that, apart from my first clinical course in China in 1998 with a visit to a clinic that treated children with birth-related
neurological deficits, infants and small children had been remarkably absent from all my later participation at various acupuncture departments in Beijing and in Shanghai. In October 2013, Landgren had just published her qualitative paper on Western practitioners’ enthusiastic attitude towards and practice with needle acupuncture for infantile colic (42). “Maybe needle acupuncture is not a TCM practice in China, but a Western construct”, I said. “Go find out,” Professor Brekke replied.

I chose Shanghai because I had field knowledge and contacts there. Over the years, I had organized, administered, and supervised clinical acupuncture courses for Norwegian practitioners several times at Longhua Hospital in south Shanghai, in co-operation with Chinese TCM colleagues and friends at the International Education College at the Shanghai University of Traditional Chinese Medicine (SHUTCM). Field knowledge and local contacts are not inherent biases in qualitative research. On the contrary, they can be strengths and greatly facilitate the possibility of focused strategic samples of informants, and reduce mistrust (186).

Why did we choose several different qualitative methods? Subject knowledge matter (172) and field knowledge guided us to choose a flexible qualitative design (173) with several different research methods. This was an important methodological decision that started in the protocol planning stage and developed further during the study. There were validity threats that we had to take into consideration. We felt we could not rely on written and returned predefined questionnaires, but had to meet the clinicians in person at various places and times. This would also ensure rich data collection and comparisons. As there were no previous research data on the issues to be addressed that could guide us, we had to take a direct, explorative approach. We could have started with interviews and then developed a survey, but we chose a flexible approach so that we could make adjustments as more information became available. A survey risks the possibility of biased “official doctrine” answers, which we
needed to take into consideration. Focus group interviews run some of the same risks, getting “group-norm” answers (187) or “doctrinal” answers, as our focus group interview suggested. No ordinary clinicians wanted to participate in a planned focus group interview. Three leaders/assistant leaders of departments finally agreed to participate. This focus group interview strengthened our decision concerning broad and multiple informal information-gathering methods. Our research focus was on TCM clinicians practical use of and their attitudes towards acupuncture, one of the pillars of TCM, in a particular patient population. This can be a sensitive issue in China, and our findings would probably not be supported in official papers. Some of the divergent opinions in the focus group might reflect this.

Why use thematic analysis (175)? This was a pragmatic decision when we were faced with several formal and informal sets of data. Thematic analysis is an approach that is fairly straightforward and relatively easy to learn, and it is well suited for analysing data from several information gathering methods. Akin to content analysis (188, 189), it allows for coding of semantic (manifest) content in a realist approach, and the searching for main themes or patterns. It also allows for quantification in numbers. As an approach, thematic analysis amply suited our purpose.

4.4. Limitations and strengths in papers I–IV

4.4.1. Overall considerations

The main limitation involves the limited numbers of participants in all studies. The pilot study comprised seven infants; the main RCT failed to reach the goal of 120 participants and ended up with 90 and then 79 participants in the final efficacy analysis. There were 307 participants in the IPD-based meta-analysis, which is very limited for such an analysis. As a consequence, the confidence intervals in both the RCT and the IPD-based meta-analysis are wide, adding some uncertainty to the results presented. The qualitative study comprised only 14 informants,
one in-depth interview, and one focus group with three participants. Thus, reaching saturation—a widely used goal in qualitative analysis—was never an option. There were seven field days, which could reduce the internal validity of observations and informal information sources.

The main strength lies in the trends observed. In the IPD-based meta-analysis, the results of the blinding-validated and pilot-adjusted RCT are consistent with the combined baseline-adjusted results of other RCTs, using other point combinations and treatment times. The results showed no clinically important efficacy of acupuncture for infantile colic; the SMD values were in the same range, and there was no heterogeneity. The qualitative study helps reducing doubts on whether there might be other and better needle methods or systems of needle treatment rooted in TCM. According to our data obtained during the field study in Shanghai, needle acupuncture on infants is not a recommended TCM practice in China.

4.4.2. Papers I and II. The pilot study and the main RCT

The most important limitation is that we were underpowered. We failed to reach our sample size goal of 120 diaries. Our benchmark of 1-hour difference in crying time and an estimated SMD of 0.5 were ambitious and unfeasible given what we know now. However, at the time of our study, the report by Vickers et al. on chronic pain in adults showing needle-specific efficacy of SMD values of 0.15 to 0.23 (88) had not yet been published, and there were no reliable estimates of acupuncture treatment efficacy in children when we made the preliminary protocol in 2008. On the other hand, if we had to estimate sample sizes from what we know now, with an SMD of 0.10 at the end of treatment, we would have needed 1500 inclusions and the trial would never have started.

Another important limitation of the pilot study and the RCT is that our blinding validation of the outcome assessors (parents) could have been better. At the time of the protocol and the
pilot study, we were not aware of Bang’s article (165), of the James and Bang blinding indices, and of the discussion concerning how and when to put the blinding validation questions. We asked a “yes or no” question for acupuncture after the first intervention. In retrospect, we should have asked the question with the response options “yes/no/do not know” after the first intervention, after the last intervention, and at the last contact, as did Landgren et al. (45) in their 2017 study. Consequently, our question posed at the last contact (“Have you seen any needle marks”) reflected the actual unblinding of parents, as discussed in the pilot study (169). If our results had been positive, this would have lowered confidence in the results. As it turned out, we had no positive results on primary end-points in spite of the uncertainty regarding the blinding of outcome assessors.

A third important limitation is the way in which we checked the inclusion criteria. Both the pilot study and the RCT were undertaken in real-life clinical settings during normal working hours, where time and compliance by parents and participating GPs are limiting factors. To achieve enrolment in the study, we had to rely on the parents’ correct assessment of crying time during the interview. We did not screen the crying time diaries before enrolment: what the parents reported was what we accepted after close scrutiny. This constitutes a possible bias of recruiting infants who did not in fact fulfil the colic criterion of 180 minutes or more crying time. That said, in the analysis, the mean baseline crying time was 220 minutes in the acupuncture group and 212 minutes in the control group.

Other limitations are as follows. The GPs who conducted the interventions met and discussed with the parents during the inclusion interview but before randomization. Subsequently, the procedure was one of strict neutrality once the randomization decision had been revealed, and the GP took no further part in crying registration. Our choice of the method of percutaneous needle acupuncture, the acupuncture points, the depth, the needle retention, and the number of sessions deemed necessary for change was pragmatic, and not based on solid prior evidence
of particular efficacy or effectiveness. Such data did not and still do not exist. We did not register treatment-linked crying as an adverse effect, and we did not argue that any possible minor needle pain would outweigh what we thought was a considerable reduction in colic pain, or that we should assess it.

The major strength of this analysis is that a pilot study was conducted. We wanted to ensure that the protocol could hold the line of standardization and uniformity of intervention procedures, blinding procedures, and validation. Furthermore, the crying diary, although validated in another study (155), also worked in our decentralized setting and was understood by the parents. The protocol did not hold the line completely, and because of the decision in advance not to include the results of the pilot in the RCT numbers, we could exchange experiences, make changes, validate procedures, and improve the quality of the final RCT protocol.

4.4.3. Paper III. The systematic review and the IPD meta-analysis

As noted, the major limitation concerned the limited numbers. Few eligible trials have been conducted—three in all—and these were small studies, involving a total of 307 infants. These led to wide confidence intervals and added some uncertainty to the results. In the case of broad heterogeneity, this would have lowered confidence in the results to a considerable degree. As heterogeneity was negligible in all the end-points, this partly corrected for the small numbers, as reflected in the GRADE evaluation. Because only one trial assessed treatment pain in needle acupuncture versus no acupuncture control, there could be some doubt about whether this was a procedure that reflected normal treatment pain. However, the 2017 study by Landgren et al. (45) registered more treatment crying in infants randomized to the intervention with several needles than in infants randomized to one needle. Thus, although not statistically significant because of the small numbers, the OR of 2.53 (95% CI = 0.72 to
Another limitation concerned the non-uniformity of blinding test validations in the individual trials. The decision to combine chi-squared tests and the James and Bang blinding indices (165), and make decisions about outcome assessor blinding on the basis of the overall results of all three, represents a novel approach, one that to our knowledge has not been validated in comparable IPD-based meta-analyses. Interpretation of these results is complex because the timing of the blinding validation varied between the studies. In the 2010 study by Landgren et al. (44), parents were asked about the allocation of their children after the last treatment session. In the 2013 study by Skjeie et al. (170), parents were questioned about allocation after the first treatment session, and after four weeks, they were asked whether they had noticed any needle marks. In the 2017 study by Landgren et al. (45), parents were questioned about their children’s allocation after the second, third, and fourth treatment sessions, as well as at the follow-up. We decided to calculate blinding indexes obtained after the last treatment session to facilitate comparison with the results of Landgren et al. (44). We are aware of this limitation in standardizing the evaluation time and questions, and the consequent uncertainty in validation of the blinding tests. Concerning the most important blinding test validation by Landgren et al. (44), which led to their study being left out of a sensitivity analysis at mid-treatment, we are reasonably confident that this study was unblinded to the outcome assessor. Kjetil Gundro Brurberg at the Norwegian Institute of Public Health had access to all raw data files. When examining the spreadsheet with individual patient data from the trial (44), it became clear that some participants were asked at the last intervention day: “Do you think your child received acupuncture?” Although data were missing for one participant, they were available for 43 patients. Of 22 cases randomized to acupuncture, 18 sets of parents guessed correctly, one incorrectly, and three did not know. Of 21 randomized to the control, 16
guessed correctly, two incorrectly, and three did not know. If the efficacy of the trial was high, this could represent unblinding because of efficacy. However, as there was little efficacy, if any, these results point towards inadequate blinding.

We decided on 30 minutes, an approximate SMD of 0.3, and the difference in crying time between acupuncture and no acupuncture control as the MID, the clinically important effect in our primary end-point. Other choices could have led to other results, and there is no consensus in the literature on infants. It is a clear limitation that our choices were based mainly on studies on adults (95, 96, 98), on children older than infants (97, 99, 105), and on our own consensus discussions at the University of Lund in February 2017.

The major strength of this IPD-based meta-analysis was that we were able to include all eligible trials and receive the raw data files to reanalyse them independently at the Norwegian Institute of Public Health. Adjusted for baseline differences, there was negligible heterogeneity in the primary end-point. There was no clinically relevant efficacy, and when adjusted for unblinding of outcome assessors, there were no statistically significant effects at any estimation of the primary end-point. Meta-analyses of acupuncture studies are often marred by heterogeneity. In our study, there was none, in spite of the different insertion methods, acupuncture point choices, and the numbers and intervals between interventions. Another strength of our data is that we combined a Chinese-language with a fully updated English-language literature search. However, we found no RCTs from China. In light of our information from the Shanghai study, this would be because acupuncture is not performed on infants.

4.4.4. Paper IV. “Big needles, small bodies”: the Shanghai qualitative study

Again, our main limitations involved the numbers of participants and time available. When we decided to conduct the study in Shanghai, this would have to be carried out by me, in co-
operation with my contacts there, in-between attending my practice, and in the time that my part-time PhD allocated. In sum, I managed four visits over 14 months, with seven days of participant observations, one focus group interview, one in-depth individual interview, and one literature search. There were 14 informants in all, and two informant validations. None of this could have been achieved without the invaluable help of my friends Heng Lee and Han Chouping (the latter now Vice-Director of the International Education College at SHUTCM) who obtained the permissions and organized all matters.

More generally, the limitations of the study involved the challenges of conducting qualitative research in situations requiring interpreters, and in which both the researchers and informants had English as an additional language. We relied entirely on our Shanghai contacts to gain access and organize interviews with TCM practitioners and officials, and as such, the selection of informants could have been biased. The TCM setting in Shanghai might also differ in some ways from that in other cities and provinces in China. However, the centralized and uniform organization of education and practice of TCM in Shanghai and the People’s Republic of China should mitigate doubts concerning the external validity of our findings.

My own prejudices and standard of reflexivity in the processes of gathering information and analysing data, and the resulting internal validity, are also important considerations. I am a medical doctor and acupuncturist with 25 years of clinical experience and 15 years of teaching acupuncture and general principles in TCM. The publication of our RCT (170) might have influenced the perception of our informants’ views and attitudes and my subsequent coding and analysing processes.

The strengths of the study included the uniformity of participant observations and the information collected from the TCM clinicians, the variety of sources of information, the triangulation of methods, and the validation process (174). The only discrepant voices were
two officials who had both clinical and administrative responsibilities, and who regarded acupuncture as a universal method of treatment.

Qualitative research designs run the risk of biased and selective reporting. The use of a diversity of informants, a range of information-gathering methods, and a detailed description of the methodological aspects of the process might have partly counteracted such bias.

4.5. Ethical considerations

4.5.1. Ethical considerations of percutaneous needle acupuncture treatment in small children

Concerns have been expressed regarding the ethical aspects of acupuncture treatment in children (117, 118, 131), and more so the smaller and younger they are. Acupuncture by its nature is a neurophysiological pain stimulus (51, 73, 74, 121-123), and has been shown to be experienced as painful in children (44-46, 115). In contrast to adults and older children, infants lack the ability to provide informed consent. This should require a higher threshold for interventions. In “Principles of Biomedical Ethics” (120), Beauchamp and Childress state that reduced capacity to consent justifies lowered limits of acceptable risks (119). In 2012, the Ethics Council of the Norwegian Medical Association under its then chairman T. Markestad wrote a commentary in the Journal of the Norwegian Medical Association entitled “Physicians and alternative treatment”. The practice of acupuncture on infants was criticized specifically: “The Council regards it as especially worthy of criticism to recommend non-documented treatments of children, especially if it is possible that the treatment can give discomfort or adverse effects. Acupuncture can be painful, and the Council regards it as in violation of Ethical Rules For Physicians to recommend acupuncture on infants” (118) (my translation). Swedish law prohibits alternative treatments on children under 8 years of age (190). However, there are no such restrictions in the Norwegian Law on the Alternative Treatment of Disease (191).
4.5.2. The pilot study and the main RCT

These trials aimed to validate trial procedures and then test standardized percutaneous needle acupuncture on infants suffering from infantile colic. Percutaneous needle acupuncture is potentially painful and infants do not have the competence to give consent. These ethical considerations were duly discussed in our application to the Regional Ethics Committee. The argument was that infantile colic is likely a painful condition for the child that may last months, and there is no proven effective treatment available. Therefore, testing the eventual efficacy of a commonly used method in a controlled trial would outweigh the potential pain of the needling. The parents were requested to read and sign the consent forms on behalf of their child. The pilot study and the main RCT were approved by the Regional Ethics Committee (REK) of South-Eastern Norway (reference S-08732b 2008/17889). The data collection was approved by the Norwegian Social Science Data Services (NSD; reference 21490/2/JE).

The commentary from the chairman of the Ethics Council of the Norwegian Medical Association was published in 2012 (118) during our RCT and, as noted, specifically stated that acupuncture on infants was deemed unethical. It was therefore of some importance to clarify directly with the Council whether our trial would be regarded as being in violation of their previous resolution. The trial and the arguments for the justifications for carrying out the trial were debated as Case 4-2013 in the Council (Appendix 3). The resolution recommended continuation of the trial to clarify the possibility of treatment effects that could outweigh the adverse effects of the intervention. If such a treatment effect could not be verified in sound trials, the practice should be discontinued: “If it is a reasonable theoretical argument of an intervention having a significant effect, it is in full order, and of merit, to examine this in an approved and adequately designed clinical trial.” Furthermore: “If in reliable randomized studies with a study design that excludes a placebo effect ("bias") it is proven that the effect of acupuncture clearly outweighs the discomfort the treatment would involve, then the
treatment could be acceptable. If such an effect is not obtained in reliable studies, it is not acceptable to recommend or use such treatment” (my translation).

4.5.3. The systematic review and IPD meta-analysis

This study did not require formal ethical approval as it reviewed and combined published data from approved clinical RCTs.

4.5.4. “Big needles, small bodies”: the Shanghai qualitative study

We considered it important in the evaluation of acupuncture treatments for infantile colic to compare Western opinions with current opinions among Chinese TCM practitioners on the tradition and use of acupuncture in infants. As this was a field study of professionals in Shanghai, with no direct or indirect patient intervention, the ethical questions for the heads of hospital departments and those interviewed were straightforward and mainly concerned their informed consent to participate and the logistics of anonymity, described in the study article and also elaborated in the open review process. A formal request was submitted to the Regional Ethics Committee of South-Eastern Norway, and exemption from ethical approval was granted (reference 2014l/197) (Appendix 7). At the start of the study and as it developed, the heads or acting heads of departments at Longhua hospital in Shanghai and those interviewed all signed formal written consent forms (Appendix 8).

4.6. Discussion of results

We tested percutaneous needle acupuncture of infants as a treatment for infantile colic, and conducted a study on the opinions and practices of TCM practitioners in contemporary China concerning this approach. Emphasizing the needling aspect is important for two reasons. First, acupuncture covers a wide range of techniques. Although a literal translation of the term is “puncturing by a needle”, it also encompasses various non-penetrating techniques. Second,
our emphasis was on percutaneous needle intervention, that is, a nociceptive stimulus (51, 59, 74, 121-123). The benefit of such an intervention should outweigh its harm; especially among patients such as infants who cannot consent to treatment (119). The quintessential question is thus: does this treatment method have efficacy that outweighs its harm (118)? Although the harm might be limited, needle pain still constitutes harm. We argue that there was coherence between the main results of our RCT and the IPD-based meta-analysis of all reported RCTs. Heterogeneity was negligible in the primary end-points between Skjeie et al. in 2013 (170) and the results of the two other RCTs reported by Landgren et al. in 2010 (44) and 2017 (45). In the IPD-based meta-analysis, corrected for baseline imbalances and the effects of unblinding the outcome assessors, the differences in crying time reduction between acupuncture and no acupuncture control were neither clinically relevant nor statistically significant, and all had SMD values <0.15 at the three measured time points.

One of the mainstays of acupuncture as taught in China and the West is the idea of individualized, customized needling treatment for each patient as a result of the practitioner’s diagnosis of a TCM syndrome differentiated and based on a description of symptoms, pulse, and tongue diagnosis. This syndrome differentiation, which can take many forms, results in specifically tailored acupuncture point selections to reach a maximum needle effect in each patient. As each acupuncture point has specific effects according to TCM tradition, only individualized treatment in experienced hands can cure a particular ailment. Standard prescriptions might give approximate therapies, but cannot replace such treatments. However, there are no solid scientific data confirming this notion. On the contrary, in their evaluation of differences in efficacy between various modes and styles of acupuncture treatment for chronic pain conditions in 17,922 adults, MacPherson et al. found no differences in efficacy (69). Concerning children, there are no data of sufficient quality to verify individualized versus standardized needle point selection. Landgren’s study on Western practitioners’ choice of
method for infantile colic (42) covered a wide range of points and insertion techniques, with no consensus on their efficacy. Syndrome differentiation varied, not unexpectedly, given the individualized nature of TCM. In our IPD-based meta-analysis, there were no significant differences in primary end-points between the three different methods used: a standardized ST36 acupuncture point method for three days, with two needles each time; a standardized LI4 acupuncture point method twice weekly for three weeks, with one needle each time; and a standardized LI4 point with one needle, or semi-individualized acupuncture for up to five needles each time, twice weekly for two weeks. There are no guidelines in China on this issue, and our own study (192) suggested that needle acupuncture would not be recommended as a treatment for infants at all. Thus, we cannot argue that a lack of efficacy in primary end-points is a consequence of a suboptimal needle treatment method, acupuncture point selection, or the duration of treatment. Macpherson et al. reported a tendency to slightly better efficacy (69) with more needles and more treatment sessions. However, this cannot apply in cases of infantile colic on ethical grounds, knowing the pain quality of the intervention and in light of our own data on there being no difference between using one and up to five needles (45) and the data on overall treatment efficacy levels in our IPD-based meta-analysis.

Percutaneous needle acupuncture activates nociceptive A delta and C nerves. If the effect of a painful intervention outweighs the harm, the intervention might be justified in infants. The effectiveness of acupuncture treatments in chronic pain conditions of consenting adults (SMD 0.5) (88) might render the intervention both ethically justifiable and useful (88, 92). In infants, where efficacy when correcting for parents’ influence (193) approximates effectiveness because there are no direct anticipation/reward effects, the reasoning is different. SMDs for treating pain problems are in the range of 0.09–0.23, so if the intervention itself is painful, it would not be justifiable. If we calculate from an SMD of 0.10 at the end of treatment on a statistically non-significant estimate, and a 30-minute MID, the number needed to treat (NNT)
in Cochrane Handbook estimations would be roughly 26 infants treated for each one reaching this end-point.

Data on treatment pain are lacking in systematic reviews on acupuncture in children (112, 126, 132, 194) and this is not a required reporting item in the revised STRICTA guidelines that extend the CONSORT statement (157). This is puzzling. It may be that there is a bias with the expected benefits being of such magnitude that there is no relevance in reporting treatment pain in trials. This was at least the case in our RCT (158) and Landgren et al. (44, 45), stand out as exceptions to the rule. Data on the treatment of crying from Landgren et al. (45), monitoring only needle acupuncture interventions, gave an indication of the amount of treatment pain: “In total, 388 treatments were given. On 200 occasions the infant did not cry at all, on 157 occasions the infant cried up to 1 min, and on 31 occasions the infant cried for >1 min (mean 2.7 min”).

The parents valued the acupuncture treatment as being more effective than the no acupuncture control. The finding was convincing and statistically significant as shown in this secondary end-point. Parental evaluations of the effect gave a Much improved OR value on a Likert scale of 3.03 (95% CI = 1.56 to 5.89) and an Improved OR of 2.67 (95% CI = 1.43 to 4.97). How can this be explained, given the lack of efficacy in the primary end-point? A combined SMD of ≤0.15 is roughly equal to a VAS difference of ~3 mm on a 100-mm scale. Such a difference cannot be detected for a subjective outcome. Is there something else in the acupuncture treatment, some soothing, comforting, and balancing quality that is not registered in the reduction of crying time duration? These positive results from parental evaluations are not consistent with the lack of important differences in total crying times. One possible explanation relates to the type of crying. Not all crying is colic-related, 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 the crying time
evaluation. Another possible explanation is that acupuncture works by other means than reducing the crying time. However, observations that acupuncture was not associated with changes in the frequency of feeding, stooling, or sleeping compared with controls in Landgren et al. (151) do not support this hypothesis. Subjective outcomes such as “normalized stooling” were reported in the same study more frequently among parents in the acupuncture group. The tendency towards more positive results for subjective outcomes might of course be related to inadequate blinding. Our ad hoc sensitivity analysis based on blinded validation data did not suggest such a relationship, but it had a prerequisite that the blinding was sufficient in our RCT (170). As noted, the timing and framing of blinding validation questions were not sufficient, and our study could also have had unknown elements of unblinding. Moreover, we cannot be sure that existing blinding tests are sensitive enough to detect all relevant differences (165). 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 points (45), and small, non-significant improvements at the end of treatment, do not support the idea that the timing of blinding questioning is essential for the results reported here, nor that unblinding is likely because of the efficacy of intervention. The differences in parental evaluation of effects between the apparently unblinded study (44) with an OR of 10.4, and a partially blinded study (45) with an OR of 1.8, both using the same treatment method, suggest an influence of unblinding.

One interesting aspect concerning the efficacy found in our RCT (SMD 0.16) and in our IPD-based meta-analysis, SMD 0.23 (0.13 in the sensitivity analysis) mid-treatment and 0.10 at the end of treatment, is the striking similarities with the needle-specific effects found in the largest ever IPD-based meta-analysis of acupuncture on treating pain: 17,922 adult patients - SMD range 0.15–0.23 (88). This might be a coincidence, but it is unlikely. Infants have no anticipation/reward effects. If you can blind the outcome assessors or correct for unblinding,
the efficacy actually found will equal the needle efficacy. The notion in Western-based textbooks that the effects of acupuncture in children, especially small children, are swifter and often stronger than in adults (116, 139-141) is not true for infantile colic.

Finally, have we been testing for acupuncture treatment of a real pain problem? There are arguments that infantile colic is not abdominal pain but reflects the upper end of normal crying behaviour in infants. In their systematic review and meta-analysis of crying patterns in 8,690 infants, Wolke et al. (5) argued: “The rapid developmental change in fuss/cry duration has implications for treatment and interpretation of treatment studies. Colic is the extreme of normal fuss/cry behaviour, self-limiting, and, thus, the vast majority will spontaneously remit. Adequate management of fussing and crying in the first 3 months rather than treatment may be required.” Benninga et al. on behalf of ROME IV (3) were more direct: “Most cases of colic probably represent the upper end of the normal developmental ‘crying curve’ of healthy infants and there is no proof that the crying in such cases is caused by pain in the abdomen or any other part of the infant’s body”. In many ways, infantile colic remains an enigma and such discussion is beyond the scope of this thesis. Our concern has been on the effects of percutaneous needle acupuncture treatments for the condition; with the current data available, we must conclude that there is little, if any, efficacy.
5. Clinical implications

We have tested the efficacy and evaluated the practice of percutaneous needle acupuncture as a method for treating infantile colic. We did not find any clinically relevant efficacy. The positive recommendations for acupuncture in small children in Western paediatric acupuncture text-books (62, 116, 125) and the positive evaluations in the published literature on acupuncture in treating infantile colic (42-46, 136) are not supported by our results reported here. Our data indicate that percutaneous needle acupuncture on infants seems to be a Western practice, given that contemporary Chinese TCM practitioners normally do not use and do not recommend the method. Needle acupuncture is a nociceptive stimulus (51, 59, 74, 121-123), and infants cannot consent to treatment. Therefore, we do not recommend the use of acupuncture treatments for infantile colic. We recommend that acupuncture practitioners take these results into consideration, until data emerge indicating that the effects of such treatments outweigh their potential harm.
6. Future research

If clinical trials evaluating needle acupuncture efficacy in infants are to be carried out, it is important to elaborate on and predefine a clinically relevant efficacy threshold. Emphasis should be on securing blinding of the parents as outcome assessors, testing the efficacy of blinding with validation questions at intervals, and analysing them with a combination of predefined blinding tests. Treatment-related crying as an adverse effect should be monitored and evaluated quantitatively using evaluation forms. Another line of inquiry would be to explore more qualitative aspects of parents’ evaluation of effects (151) and consider other positive effects of acupuncture not registered in terms of the crying time. If parents are truly blinded, this could give valuable insights into the less explored and possibly positive effects of acupuncture on infants with colic. These effects could then be evaluated against efficacy and treatment pain.
7. Concluding remarks

I started the work that resulted in this thesis in 2009. We are now nearing the end of 2017. Much has changed in that time. What has not changed is the peculiar insistence among the most ardent supporters of acupuncture that it is a cure for all ills. Acupuncture has its place in Western medicine, but possibly to a lesser degree than many of us thought 10 years ago. I hope I have made a contribution to a better clinical evaluation of needle acupuncture practice in this very young patient population. Acupuncture treatments of pain conditions in consenting adults are likely to be a reasonable practice, but percutaneous needle acupuncture treatments for infantile colic are not.
List of references


52. Chen L-Y. Director of Medical History Department, Shanghai University of Traditional Chinese Medicine. Personal communication on the dating of the reference text “Huang Di Nei Jing”. 8 November, 2016.


67. Ji G. Vice-President of Shanghai University of Traditional Chinese Medicine, Shanghai, China. Scientific meeting, Acupuncture Research Group, Department of General Practice, University of Oslo, and Research Department, Shanghai University of


   http://nccam.nih.gov/grants/acupuncture/priorities?nav=upd


75. Liodden I. Acupuncture for Postoperative Morbidities in Children and Placebo by Proxy. PhD thesis. The National Research Center for Complementary and Alternative Medicine, Faculty of Health Sciences, UiT The Arctic University of Tromsø, Norway; 2016.


https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-7-7


153. Landgren K, Tiberg I, Hallström I. Standardized minimal acupuncture, individualized acupuncture, and no acupuncture for infantile colic: study protocol for a multicenter


178. Acupuncture in Infantile Colic – A Three Armed Randomized Multi Center Trial (ACU-COL) [Internet]. 2012. Available from:

180. Powell CV, Kelly AM, Williams A. Determining the minimum clinically significant
versus individualized acupuncture for chronic low back pain: a randomized controlled
184. Heyerdahl O, Lystad N. Tradisjonell kinesisk akupunktur; dens filosofi og praktiske
anvendelse i moderne medisin; lærebok 2. Oslo: Heyerdahl og Lystad ans; 2003.
resources contribute to Cochrane reviews. Journal of Alternative and Complementary
186. Maxwell JA. Qualitative Research Design. Thousand Oaks, California, USA: Sage
Publications Ltd.; 2013. pp. 87-120.
187. Bloor M, Frankland J, Thomas M, Robson K. Trends and uses of focus groups. Focus
188. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts,
procedures and measures to achieve trustworthiness. Nurse Education Today.
2004;24(2):105-12.


Appendix 1. REK ethics approval pilot and RCT (Norwegian)
S-08732b Effekt av akupunktur ved spedbarnskolikk [6.2008.2175]

Vi viser til brev oversendt via e-post 27.02.09 med svar på komiteens merknader, vedlagt revidert invitasjon til foreldre med samtykkeerklæring, revidert intervjuskjema og protokoll.

Komiteen tar svar på merknader til etterretning.

Komiteen har ingen merknader til verken informasjonskriv, intervjuskjema eller protokoll.

Vedtak
Prosjektet godkjennes.

Med vennlig hilsen

Tor Norseth (sign.)
Leder

Julianne Krohn-Hansen
Komitésekretær

Kopi:
- Trygve@skonnord.no
- Allmenmedisinsk Forskningsutvalg v/ Tove Rutle, rmrtove@online.no
Norsk samfunnsvitenskapelig datatjeneste AS
NORWEGIAN SOCIAL SCIENCE DATA SERVICES

Mette Brekke
Institutt for allmenn- og samfunnsmedisin
Universitetet i Oslo
Frederiks hus Kirkeveien 166
0450 OSLO

Vår dato: 03.04.2009
Vær ref: 21490 / 2 / IE

Deares dato: 
Deares ref: 

TILRÅDING AV BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 25.02.2009. Meldingen gjelder prosjektet:

21490
Behandlingsan认清: Eftfekt av akupunktur ved spedbarnshelselekk
Daglig ansvarlig: Mette Brekke

Personvernombudet har vurdert prosjektet, og finner at behandlingen av personopplysninger vil være reguleret av § 7-27 i personopplysningssaken. Personvernombudet tilår at prosjektet gjennomføres.

Personvernombudets tilrådning forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, eventuelle kommentarer samt personopplysningssaken/-helseregistren som forskriver. Behandlingen av personopplysninger kan settes i gang.


Vennlig hilsen

Vigdis Namtvedt Kvalheim

Janne Sigbjørnsen Eie

Kontaktperson: Janne Sigbjørnsen Eie tlf: 55 58 31 52

Vedlegg: Prosjektvurdering

Auditorisk kontor i Distrikt kontor/:
OSLO: NDS Universitetet i Oslo, Postboks 1005 Blindern, 0316 Oslo. Tlf: 47-22 85 52 11. nds@ub.uio.no
TRONDHEIM: NDS: Norges teknisk-naturvitenskapelige universitet, 7491 Trondheim. Tlf: 47-70 50 19 27. Kenneth.server@ntnu.no
TRONDHEIM: NDS: SIF. Universitetet i Trondheim, 9037 Trondheim. Tlf: 47-77 66 45 35. nnds@ub.uio.no
Personvernombudet for forskning

Prosjektvurdering - Kommentar

Formålet med prosjektet er å studere akupunktur på spedbarn med kolikk. De skal være fullbrune spedbarn og de skal oppfylle Wessels definisjon for spedbarnskolikk.

Studien skal utføres hos 10-15 allmennleger i Norsk Forening for Medisinsk Akupunktur, og pasientene rekrutteres fra fødeavdeling, helsestasjoner og fastleger i områdene rundt de involverte legekontorene som er spresd rundt om i Norge. 200 pasienter skal inkluderes. Halvparten av barna vil få akupunkturbehandling i form av bilateral punksjon av punktet Stomach 36 (ST36), mens resten er kontrollgruppe og får en "late-late"-behandling av punktet Tisso (LV55).

Foredrenerne og legesekretærer er blendet. Foreldrene skal foreskrive granskjemaet og vil også interviewes om deres vurdering av tilstanden ved ulike tidspunkt i forhold til behandlingen.

Pasientene rekrutteres via informasjon om studien til fødeavdeling, helsestasjon og fastleger. De som ønsker å delta tar kontakt med et av legekontorene som er med i studien. Ombudet forstår at de vil melder sin interesse til å få informasjonsskriv med full informasjon om studien. Foreldrene samtykker skriflig til deltakelse. Informasjonsskrivet vedlagt meldingskemaet finnes tilfredsstillende forutsatt at følgende endringer gjøres:

- Setningen "Opplysninger som allerede er innsamlet fra deg vil ikke bli slettet" må endres til f.eks. "Opplysninger som allerede er innsamlet vil da anonymiserses." Dersom noen trekker seg underveis har de krav på at innsamlede opplysninger anonymiserses.
- Det må tilføyes at Universitetet i Oslo v/ Mette Brekke er ansvarlig for studien.

Ombudet ber om at revidert informasjonsskriv ellersendes i god tid før utvalget kontaktes.

Det registreres sensitive opplysninger om helseforhold, i f.personopplysningsloven § 2 nr 8 e)

I tillegg til daglig ansvarlig Mette Brekke vil også Trygve Skonnord og Holger Skjøde ha tilgang til datamaterialet. Personvernombudet forutsetter at oppbevaring av datamaterialet på privat pc er i tråd med datasikkerhetsrutinene til Universitetet i Oslo.

Det er sendt søknad til Regional komité for medisinsk og helsefaglig forskningsetikk (REK). Ombudet forutsetter at prosjektet godkjennes av REK og ber om at kopi av godkjennelsen ellersendes.

Appendix 3. Ethical Council of the Norwegian Medical Association statement

– Case 42013 (Norwegian)
Spørsmål vedr. akupunktur og annen behandling av spedbarn som innebærer smerte


Saken ble behandlet av Rådet for legeetikk i møte den 6.2.2013.

Av protokollen fremgår:

Sak 4/2013 Spørsmål vedr. akupunktur og annen behandling av spedbarn som innebærer smerte
12/5032

Holger Skjeie har i brev av 7.12.12 spurt om det er tillatelig å prøve ut akupunkturbehandling mot kolikk hos spedbarn. Bakgrunnen for forespørselen er at Rådet i sitt vedtak om alternativ behandling i 2012 uttrykte at «Akupunktur kan være smertefullt, og rådet anser at det er i strid med Etiske regler for leger å anbefale akupunktur på spedbarns». (Gjengitt i Tidsskr Nor Legeforen 2012; 132:2409-10). Holger Skjeie viser til at mye av den behandlingen vi utsetter spedbarn for, medfører smerte og ubeheg. Han er i ferd med å avslutte en randomisert kontrollert studie om akupunkturbehandling ved spedbarnskolikk og spor om dette er i strid med rådets vedtak. Studien er godkjent av regional etisk forskningskomite.

Rådet vedtok å uttaie:

Det er riktig at spedbarn utsettes for mye som er plagsomt innenfor konvensionell medisin, f.eks. vaksinasjoner. Dette er uproblematisk så lenge det anses å være i barnets beste interesse. Det som er hele poengen med uttalelsen fra Rådet, er at det ikke er akseptabelt for leger å utsette noen for risiko eller plager med behandling som ikke er rimelig godt dokumentert. Her er det i prinsippet ingen forskjell mellom konvensionell og alternativ medisin. Hovedbudskapet i vedtaket er at vi forventer at behandling som viser seg ikke å ha effekt i pålitelige studier, ikke skal videreføres - det være seg konvensionell eller såkalt alternativ behandling.
En helt annen sak er vitenskapelige studier som har som formål å finne ut om en behandling er effektiv eller ikke. Dersom det er en rimelig teoretisk begrunnelse for at en behandling kan ha signifikant effekt, er det helt på sin plass, og fortjenestefullt, å undersøke dette i en godkjent og adekvat utformet klinisk studie. Det vises i den sammenheng til setningen i vedtaket forut for den som er stiert i forespørselen: «Rådet anser det som særlig kritikkverdig å anbefale ikke-dokumentert behandling til barn, særlig dersom det er mulig at behandlingen kan medføre plager eller bivirkninger» [Rådets understreknings]. I den aktuelle saken var det snakk om ikke-dokumentert behandling av spedbarn. Dersom det i pålitelige randomiserte studier med studiedesign som utelukker en placeboeffekt («bias») bevises at effekten av akupunktur klart oppveier de plagene behandlingen måtte medføre, kan behandlingen være akseptabel. Dersom slik effekt ikke oppnås i pålitelige studier, er det ikke akseptabelt å anbefale eller benytte slik behandling, jf. Etiske regler for leger Kap I, § 9, som bl.a. sier at «Legen må ikke gjøre bruk av eller anbefale metoder som savner grunnlag i vitenskapelige undersøkelser eller tilstrekkelig medisinsk erfaring».


RÅDET FOR LEGEETIKK

(Sign.)
Trond Markestad
Leder

Stine K. Tønsaker

Den norske legeforening • Postboks 1152 Sentrum • NO-0107 Oslo • legeforeningen@legeforeningen.no • Bygningsadresse: Akselt. 2 • www.legeforeningen.no • Telefon: +47 23 10 90 00 • Faks: +47 23 10 90 10 • Org.nr. NO 960 474 341 MVA • Bankgiro 5085 06 25180

125
Appendix 4. Parent consent form RCT (Norwegian)
Forespørsel om deltakelse i vitenskapelig studie

"Effekt av akupunktur ved spedbarnskolikk"

Bakgrunn og hensikt
Dette er en forespørsel til deg om å delta i et forskningsprosjekt som skal undersøke om og i hvilken grad akupunkturbehandling har effekt på spedbarnskolikk.

Spedbarnskolikk er en ufarlig, men svært plagsom lidelse for barn og foreldre. Det er ingen klare og entydige behandlingsanbefalinger for denne lidelsen.

Akupunkturbehandling av spedbarnskolikk er en hyppig brukt metode, og vi som driver med dette har inntrykk av at det ofte hjelper. Behandlingen er rask og enkel. Den er så godt som uten bivirkninger og er dermed helt ufarlig. Likevel mangler man god nok vitenskapelig dokumentasjon av behandlingens effekt og derfor ønsker vi nå å gjennomføre denne studien.

Studien utføres av erfarne leger i allmennpraksis som alle er medlemmer av Norsk Forening i Medisinsk Akupunktur. De har bred erfaring i akupunkturbehandling av forskjellige lidelser i allmennpraksis.

Hva innebærer studien?
Vi ønsker å sammenligne utviklingen av kolikkplager hos barn som får akupunktur med barn som ikke får slik behandling.

Under behandlingen er legen alene med barnet på et behandlingsrom i 3 minutter mens du venter på venterommet. Halvparten av barna vil få ekte akupunktur ved at de får en tynn akupunktturnål i øvre del av hver legg i 30 sekunder. Halvparten av barna vil få ”late-som” behandling, og vil altså ikke bli stukket.


For å registrere endringer som følge av behandlingen, trenger vi at du fyller ut et dagbokskjema over barnets gråt i 2 dager før første behandling og til og med dagen etter siste behandling. Vi ønsker også at du fyller ut dette skjemaet før én dag en uke etter siste behandling og 4 uker etter siste behandling.

Det eriktig for studien at du er så nøyde som mulig med å føre inn tidspunktene i dagboken og at du så summerer tiden med gråt for hvert døgn. Ta med gråtadgangen til første behandling slik at vi kan være sikre på at skjemaet føres på riktig måte.

Hjelpersonellet vil også stille deg noen spørsmål i forbindelse med behandlingene, samt 1 og 4 uker etter siste behandling. Det gjelder hva slags annen behandling du har forsøkt, hvor lenge barnet har vært plaget, barnets ernæring og om hvordan barnets kolikk har utviklet seg etter behandlingen. Vi vil også spørre om mulige bivirkninger av behandlingen.
Det er ingen restriksjoner under studien på hva du ellers måtte ønske å gjøre i forhold til barnets plagers, bortsett fra at du ikke kan gå til annen akupunktur før siste observasjonsdag er over (4 uker etter siste behandling).

Dersom barnet ditt blir sykt i løpet av observasjonstiden (4 uker), kan dere ikke lenger være med i studien. Vi ber derfor om å få beskjed om dette slik at vi kan registrere dette. Hvis du er usikker på om barnet er så sykt at det ikke kan være med lenger, kan du ta kontakt med legekontoret.

**Mulige fordeler, ulemper og alvorlige bivirkninger**

Det er 50 % sjans for at ditt barn får akupunktur, og dere er derfor ikke garantert at en eventuell effekt av slik behandling kommer deres barn til nytte. Du vil imidlertid være med å bidra til å øke kunnskapen om slik behandling hos fremtidige barn med kolikk.

Du må altså la barnet være alene med legen i noen minutter. Dette er nødvendig for å være sikker på at du ikke vet hva slags behandling barnet har fått, da det kan påvirke svarene som vi ønsker å undersøke.

Akupunkturbehandlingen kan av og til få barnet til å gråte, men er ellers en trygg behandling med svært små muligheter for skade/bivirkninger. Det kan av og til forekomme en liten blodning i form av en bloddråpe etter at nålen er tatt ut. Blødning under huden forekommer svært sjelden hos spedbarn.

**Hva skjer med informasjonen om barnet ditt?**


**Frivillig deltagelse**

Det er frivillig å delta i studien. Du kan når som helst og uten å oppgi noen grunn trekke deg fra studien uten at det får konsekvenser for din videre behandling. Du undertegner samtykkeerklæringen dersom du ønsker å delta.

Har du spørsmål til studien, ta kontakt med legesekretær eller lege på legekontoret hvor behandlingen skjer, eventuelt en av forskningslederne (se nedenfor).

**Personvern**

Opplysninger som registreres om barnet ditt er kjønn, alder, vekt og antall svangerskapsuker ved fødselen. Dessuten registreres varighet av kolikkene, hva slags behandling som er forsøkt og hva slags
ernæring barnet får. Videre registrerer vi resultatene av gråtedagboken og hvordan barnets kolikk har utviklet seg etter behandlingen, samt eventuelle bivirkninger.

Representanter fra kontrollmyndigheter i inn- og utland kan få utlevert studieopplysninger og gis innsyn i relevante deler av ditt barns journal. Formålet er å kontrollere at studieopplysningene stemmer overens med tilsvarende opplysninger i ditt barns journal. Alle som får innsyn i informasjon om barnet har taushetsplikt.

**Innsynsrett og oppbevaring av materiale**

Hvis du sier ja til å delta i studien, har du rett til å få innsyn i hvilke opplysninger som er registrert om barnet ditt. Du har videre rett til å få korrigert eventuelle feil i de opplysningene vi har registrert. Dersom du trekker deg fra studien, vil det ikke samles inn flere opplysninger eller mer materiale. Opplysninger som allerede er innsamlet fra deg vil ikke bli slettet.

**Finansiering**

Studien er finansiert av Allmennmedisinsk forskningsutvalg (AFU).

**Forsikring**

Du er forsikret i henhold til Helsepersonelloven gjennom legenes ansvarsforsikring.

**Veiledning**

Prosjektet er veiledet av Mette Brekke og Arne Fetveit ved Institutt for samfunns- og allmennmedisin ved Universitetet i Oslo.

**Informasjon om utfallet av studien**

Du har rett til å få informasjon om utfallet av studien. Ta enten kontakt med legen som behandlet deg eller prosjektlederne. Det kan ta 1-2 år før resultatene er ferdig analysert og foreligger.

Med vennlig hilsen

Prosjektledere for studien:

Trygve Skonnord  
Spesialist i allmennmedisin,  
Brår legekontor, Re i Vestfold.  
Tlf 41323232.

Holgeir Skjeie  
Spesialist i allmennmedisin,  
Hellemyr legekontor,  
Kristiansand i Vest-Agder.  
Tlf 91608790
Samtykke for deltakelse i studien

Ved underskrift på denne samtykkeerklæringen samtykkes det kun for at informasjon om prosjektet er mottatt og at du ønsker å delta.

Jeg, som verge for mitt barn som er nevnt under, har lest ovenstående informasjon og denne erklæringen, og samtykker til barnets deltagelse i studien.

Foresattes navn (med blokkbokstaver):

----------------------------------------------------------------------------------------------------------------

Barnets navn (med blokkbokstaver):

----------------------------------------------------------------------------------------------------------------

Barnets fødselsdato/nummer:

----------------------------------------------------------------------------------------------------------------

Dato:

--------------------------

Underskrift:

----------------------------------------------------------------------------------------------------------------

Bekreftelse på at informasjon er gitt deltakeren i studien
Jeg bekrer å ha gitt informasjon om studien

----------------------------------------------------------------------------------------------------------------

(Signert, rolle i studien, dato)
Appendix 5. Crying time diary form RCT (Norwegian)
"Effekt av akupunktur ved spedbarnskolikk"

**Dagbok for barn med spedbarnskolikk**  
Pasient nr.  

Før opp når barnet begynner og slutter å skrike i tidstabellen for hvert døgn.

Husk at hvert døgn begynner ved midnatt. Det er viktig at alle gjør det som i eksempelet under for hver dag som skal måles, slik at vi kan dobbeltsjekke.

Det er også veldig viktig at du selv legger sammen tiden barnet ditt har grått hvert døgn, og skriver opp summen, slik som i eksempelet nedenfor.

Når du legger sammen, oppgir du summen i timer og minutter, som i eksempelet under: 4 timer og 45 minutter.

**Eksempel:**

<table>
<thead>
<tr>
<th>Dato: 01.01.09</th>
<th>Timer per døgn: 4 timer, 45 min.</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>01.10-40</td>
</tr>
<tr>
<td>08</td>
<td>09</td>
</tr>
<tr>
<td>16.45-</td>
<td>17.15</td>
</tr>
</tbody>
</table>

De påfølgende to sider er selve registreringen.

Side 2 er dagene før og under selve behandlingen.

Side 3 er 1 uke og 1 måned etter behandlingen. Vi foreslår at du legger inn en alarm på mobiltelefonen kvelden før registreringen starter.

Dersom du lurer på noe underveis, kan du enten ta kontakt med legekontoret eller med forskningslederne.

Lykke til!
Dag 1 (dagen etter informasjon) Dato:__________

<table>
<thead>
<tr>
<th>Klokkeslett:</th>
<th>Timer per døgn:</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>01 02 03 04 05 06 07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>09 10 11 12 13 14 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17 18 19 20 21 22 23 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dag 2 Dato:__________

<table>
<thead>
<tr>
<th>Klokkeslett:</th>
<th>Timer per døgn:</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>01 02 03 04 05 06 07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>09 10 11 12 13 14 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17 18 19 20 21 22 23 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dag 3 (første behandlingsdag) Dato:__________

<table>
<thead>
<tr>
<th>Klokkeslett:</th>
<th>Timer per døgn:</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>01 02 03 04 05 06 07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>09 10 11 12 13 14 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17 18 19 20 21 22 23 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dag 4 (andre behandlingsdag) Dato:__________

<table>
<thead>
<tr>
<th>Klokkeslett:</th>
<th>Timer per døgn:</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>01 02 03 04 05 06 07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>09 10 11 12 13 14 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17 18 19 20 21 22 23 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dag 5 (tredje behandlingsdag) Dato:__________

<table>
<thead>
<tr>
<th>Klokkeslett:</th>
<th>Timer per døgn:</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>01 02 03 04 05 06 07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>09 10 11 12 13 14 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17 18 19 20 21 22 23 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dag 6 (dagen etter siste behandling) Dato:__________

<table>
<thead>
<tr>
<th>Klokkeslett:</th>
<th>Timer per døgn:</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
<th>__________</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>01 02 03 04 05 06 07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>09 10 11 12 13 14 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17 18 19 20 21 22 23 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Side 2 av 3

Pasient nr. __________

Gråtedagbok
Dag 12 (1 uke etter siste behandling) Dato:__________

Klokkeslett:  

<table>
<thead>
<tr>
<th>Timer per døgn:</th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>09</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

Dag 33 (4 uker etter siste behandling) Dato:__________

Klokkeslett:  

<table>
<thead>
<tr>
<th>Timer per døgn:</th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>09</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

Ferdig utfylt skjema sendes i vedlagte frankerte konvolutt så snart som mulig. Takk for deres deltagelse i studien!

Vennlig hilsen Trygve Skonnord og Holgeir Skjeie, prosjektledere
Appendix 6. Parent interview form RCT (Norwegian)
Intervjuskjema – "Effekt av akupunktur ved spedbarnskolikk"

Pasient nr. _______________________ Barnets alder (antall uker): ____________________ Kjønn: K □ M □
Antall svangerskapsuker ved fødsel: _______ Vekt ved fødsel: _______

Hvordan fikk du høre om studien?
Helsestasjon? □
Fødeavdelingen/ informasjon fra fødeavdelingen? □
Fastlege? □
Internett/ www.spedbarnskolikk.no? □
Massemedia? □
Annet? □ Ev. beskrivelse: ____________________________________________________________

Dato inklusjonsintervju (dag 0): ____________________________

Vurdert som friskt barn ved nyfødtundersøkelsen på sykehuset? Ja □ Nei □
Bortsett fra kolikken, vurderer du at barnet ditt er friskt? Ja □ Nei □
Har barnet vært plaget med mer enn tre timers gråt per dag i mer enn tre dager i uken i mer enn tre uker?
Ja □ Nei □
Hvis ja, Hvor lenge: ____________________________________________________________

Hva slags behandling har vært forsøkt?
________________________________________________________________________

Hva slags ernæring får barnet? (Morsmelk/tilllegg/begge) ____________________________

Hvis tillegg (morsmelkerstatning), hvilken type? ______________________________________
Hvor ofte gis det tillegg? __________________________________________________________

Dato første behandling (dag 3): ____________________________ Behandling kl. __________

Etter første behandling:
Spørsmål til foreldre: Tror du/dere at barnet har fått akupunktur eller ikke? Akupunktur □ Kontroll □

Dato andre behandling (dag 4): ____________________________ Behandling kl. __________

Side 1 av 3 Pasient nr. ____________________________ Intervjuskjema
Dato intervju/ tredje behandling dag 5: _____________________ Behandling kl. ____________
Foreldrenes rapport: Hvordan har barnets kolikk utviklet seg etter behandlingen?

1. ☐ Verre 
2. ☐ Ingen forandring 
3. ☐ Noe bedring 
4. ☐ Markert bedring 
5. ☐ Helt bra 

Eventuelle kommentarer:
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________

Dato intervju dag 12: _________________(Samme døgn som foreldrene skal fylle ut gråtskjema 24 timer)

Spørsmål 1:
Foreldrenes rapport: Hvordan har barnets kolikk utviklet seg etter behandlingen?

1. ☐ Verre 
2. ☐ Ingen forandring 
3. ☐ Noe bedring 
4. ☐ Markert bedring 
5. ☐ Helt bra 

Eventuelle kommentarer:
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________

Spørsmål 2:
Har barnet hatt noen ubehagelige reaksjoner på behandlingen? Ja ☐ Nei ☐ Vet ikke ☐
Hvis ja, hvilke? Beskrivelse:________________________________________________________________________

Så spøres uansett om foreldrene har svart ja, nei eller vet ikke:
Har barnet hatt noe av følgende, og i tilfelle noter beskrivelsen av størrelse/omfang:
Blåmerke eller blødning? Eventuell beskrivelse:________________________________________________________________________
Hevelse eller infeksjonstegn? Eventuell beskrivelse:________________________________________________________________________
Uvanlig uro? Eventuell beskrivelse:________________________________________________________________________
Bekymringsfull trøtthet? Eventuell beskrivelse:________________________________________________________________________
Uvanlige bevegelser eller rykninger? Eventuell beskrivelse:________________________________________________________________________
Annet? Eventuell beskrivelse:________________________________________________________________________

Side 2 av 3 
Pasient nr. ____________ Intervjuskjema
Dato intervju dag 33: ________________ (Samme døgn som foreldre skal fylle ut gråtskjema 24 timer)

Spørsmål 1:

Foreldrenes rapport: Hvordan har barnets kolikk utviklet seg etter behandlingen?

1. □ Verre
2. □ Ingen forandring
3. □ Noe bedring
4. □ Markert bedring
5. □ Helt bra

Eventuelle kommentarer:

________________________________________________________________________

Spørsmål 2:

Har det vært forsøkt annen slags behandling siden behandelingsstart? Ja □ Nei □

Hvis ja, hvilken eller hvilke! ____________________________________________________________________________________________

________________________________________________________________________

Spørsmål 3:

Har du sett noe som du mener er merker etter stikk i huden hos barnet? Ja □ Nei □

Side 3 av 3  Pasient nr. ___ Intervjuskjema
Appendix 7. REK-consideration on the qualitative study (Norwegian)
Mette Brekke
Universitetet i Oslo

2014/197 Akupunkturbehandling av spedbarn i Kina

Vi viser til søknad om forhåndsgodkjenning av ovennevnte forskningsprosjekt. Søknaden ble behandlet av Regional komité for medisinsk og helsefaglig forskningsetikk (REK sor-øst) i møtet 13.02.2014. Vurderingen er gjort med hjemmel i helseforskningsloven § 10, jf. forskningsetikklovens § 4.

Forskningsansvarlig: Universitet i Oslo
Prosjektleder: Mette Brekke

Prosjektskrivelse

Formålet med studien er å undersøke i hvilket omfang, grad og ved hvilke teknikker akupunktur blir benyttet i behandling av små- og spedbarn ved en pediatrisk poliklinikk i Kina. Et delmål er å undersøke om akupunktur brukes i behandling av kolikk.

Akupunkturbehandling ved spedbarnskolikk brukes i økende grad i Skandinavien til tross for at effekten er mangelfullt dokumentert. Det er heller ikke bred enighet om hvordan behandlingen skal gjennomføres. Akupunktur er opprinnelig en kinesisk behandlingsmetode, og i dette prosjektet ønsker man å få mer kunnskap om bruk av akupunktur på sped- og småbarn i Kina.

Man har i samarbeid med Shanghai University of Traditional Chinese Medicine fått anledning til å observere behandlingen som gis ved Long Hua pediatriske poliklinikk i Shanghai. Prosjektpersonal vil være tilstede sammen med en translatør som samarbeider med universitetet. Det er innhentet tillatelse fra ledelsen ved sykehuset. I søknaden vurderer man at observasjon som del i opplæring av studenter og helsespersonell er vanlig ved et universitetssykehus. Det skal tas notater fra observasjonen, men det vil ikke registreres personidentifiserbare opplysninger.

Det skal i tillegg gjøres fokusgruppeintervju eller dybdeintervju med legene som arbeider ved poliklinikken.

Komiteens vurdering

Prosjektet vil gi et innblikk i hvordan akupunkturbehandling gis til små barn i Kina, og hvordan kolikk behandles ved en poliklinikk i Shanghai. I prosjektsammenheng utsettes ikke pasientene for intervensjon eller behandling som medfører risiko eller belastning utover det som er å anse av ordinær behandling.

Prosjektet er helserelatert, men formålet slik det er beskrevet faller ikke inn under helseforskningslovens virkeområde.
Det er institusjonen der observasjonen skal gjennomføres som må gi tillatelse til å være til stede ”med annet formål enn å yte helsehjelp”. Forskningsansvarlig i Norge er ansvarlig for at prosjektet blir gjennomført på en forsvarlig måte og innenfor de gjeldende ordningene i Kina.

Vedtak

Prosjektet faller utenfor helseforskningslovens virkeområde, jf. § 2, og kan derfor gjennomføres uten godkjenning av REK. Det er institusjonens ansvar på å sørge for at prosjektet gjennomføres på en forsvarlig måte med hensyn til for eksempel regler for taushetsplikt og personvern.

Komiteens vedtak kan påklages til Den nasjonale forskningsetiske komité for medisin og helsefag, jf. helseforskningsloven § 10 tredje ledd og forvaltningsloven § 28. En eventuell klage sendes til REK Sørøst A. Klagefristen er tre uker fra mottak av dette brevet, jf. forvaltningsloven § 29.

Med vennlig hilsen

Knut Engedal
Professor dr. med.
Leder

Anne S. Kavli
Førstekonsulent

Kopi til: jeanette.magnus@medisin.uio.no; Universitetet i Oslo ved øverste administrative ledelse: universitetsdirektor@uio.no; h-skjeie@frisurf.no
Appendix 8. Information and consent form – participating observation – the qualitative study
Participating observation

In May 2014, at the Acupuncture Department in Long Hua Hospital in Shanghai, there will be conducted a Participating observation by Holgeir Skjeie on the subject of acupuncture treatment of infants. The participating observation will last for three work days.

The participating observation is part of Holgeir Skjeie’s PhD project, with research grant from the Norwegian Research Fund in General Practice, and based at The Department of General Practice, University of Oslo.

Chief supervisor is Professor Mette Brekke (Mette.Brekke@medisin.uio.no).

The research projects aim is to explore eventual effects of acupuncture treatment in infantile colic, and its rationale. The participating observation will focus on the treatment considerations and methods used by the TCM doctors and nurses. No tape recording or photos will be taken. The observations will be interpreted and analyzed in a qualitative context, and all references to any persons views or actions, will be anonymous.

The Acupuncture Department at Long Hua Hospital may at any time withdraw the consent.

Yours Sincerely,
Holgeir Skjeie, M.D, PhD-research fellow
Mette Brekke, M.D, Professor in Medicine

Consent Declaration.
On behalf of the Acupuncture Department at Long Hua Hospital, Shanghai, I consent to the information concerning the project is received and that the Acupuncture Department at Long Hua wish to participate.

Name (In block letters)
Position:

Date and signature
ORIGINAL ARTICLE

Acupuncture for infantile colic: A blinding-validated, randomized controlled multicentre trial in general practice

HOLGEIR SKJEIE, TRYGVE SKONNORD, ARNE FETVEIT & METTE BREKKE

Department of General Practice, Institute of Health and Society, University of Oslo, Norway

Abstract

Objective. Infantile colic is a painful condition in the first months of infancy. Acupuncture is used in Scandinavia as a treatment for infantile colic. A randomized controlled trial was carried out with the aim of testing the hypothesis that acupuncture treatment has a clinically relevant effect for this condition. Design. A prospective, blinding-validated, randomized controlled multicentre trial in general practice. Research assistants and parents were blinded. Setting. 13 GPs’ offices in Southern Norway. Intervention. Three days of bilateral needling of the acupuncture point ST36, with no treatment as control. Subjects. 113 patients were recruited; 23 patients were excluded, and 90 randomized; 79 diaries and 84 interviews were analysed. Main outcome measures. Difference in changes in crying time during the trial period between the intervention and control group. Results. The blinding validation questions showed a random distribution with p = 0.41 and 0.60, indicating true blinding. We found no statistically significant difference in crying time reduction between acupuncture and control group at any of the measured intervals, nor in the main analysis of differences in changes over time (p = 0.26). There was a tendency in favour of the acupuncture group, with a non-significant total baseline-corrected mean of 13 minutes (95% CI −24 to +51) difference in crying time between the groups. This was not considered clinically relevant, according to protocol. Conclusion. This trial of acupuncture treatment for infantile colic showed no statistically significant or clinically relevant effect. With the current evidence, the authors suggest that acupuncture for infantile colic should be restricted to clinical trials.

Key Words: Acupuncture, acupuncture point, general practice, infantile colic, Norway, randomized controlled trial

Introduction

Infantile colic is a painful and incompletely understood condition in the first months of infancy. The majority of studies of infantile colic have used the definition by Wessel et al: “Paroxystic uncontrollable crying and fussing in an otherwise healthy infant under three months of age, with more than three hours of crying per day in more than three days for more than three weeks.”[1]

Although infantile colic is a self-limiting condition, it is a severe strain on both the child and parents.[2] The aetiology is considered multifactorial. Possible mechanisms include physiological factors like painful intestinal contractions and altered gut motility, immaturity of gut function, lactose intolerance, food hypersensitivity, altered intestinal flora and gas, and psychological factors like inadequate mother–child interaction, anxiety, and infant temperament [3,4].

There is no consensus on treatment strategies for the condition, which include common strategies like hypoallergenic diet, soy formula, reduced stimulation, sucrose, and herbal tea [2,5]. Chiropractic manipulation has not shown effects in controlled studies.[6] Recently, administration of drops of specific Lactobacillus strains has shown promising results [7,8].

Acupuncture is a frequently used alternative treatment modality in Scandinavia [9] and is also used for infantile colic [10]. There is a well-founded concern for the ethics and the evidence concerning alternative treatment of paediatric conditions [11,12].

Acupuncture is an original Chinese treatment method using thin steel needles penetrating through
Acupuncture is used in Scandinavia as a treatment for infantile colic. This blinding-validated randomized controlled trial carried out in general practice found no statistically significant difference in duration of daily crying between a group subjected to acupuncture and a control group (p = 0.26). A non-significant 13-minute trend in favour of acupuncture was considered not to be clinically relevant.

the skin and into connective tissue and muscle fibres. The neurophysiologic basis for the observed effects, especially the pain-inhibiting effects, is relatively well understood.[13] Acupuncture is a safe procedure when used by trained practitioners, and the risk of serious adverse effects is low [14], in children also [15].

Two controlled trials of children with infantile colic treated with acupuncture have been published [16,17]. Both studies concluded that acupuncture significantly reduced crying and pain-related behaviour without noticeable adverse effects. Effect sizes were small, and there was no blinding validation.

General practitioners educated within the programmes of the Norwegian Society of Medical Acupuncture use a standardized bilateral needling of the point ST36 when treating infant colic. ST36 is located in the proximal part of anterior tibial muscle and is the acupuncture point considered most important for ailments of the gastro-intestinal apparatus in traditional Chinese medicine (TCM) [18,19]. A postulated neurophysiologic mechanism explains a beneficial effect on gut dysmotility by way of the parasympathetic vagal reflexes, as well as a centrally opioid-mediated pain inhibitory pathway [20].

We carried out a randomized controlled trial with the aim to test the hypothesis that such acupuncture treatment has an effect above no-treatment control in infantile colic. The main study was preceded by a pilot study [21].

Material and methods

Trial design, participants, and interventions

The study was a prospective, blinding-validated, multicentre, randomized controlled trial involving 13 GPs’ offices in Southern Norway. The trial was registered with Clinical Trial Registry Identifier NCT00907621.

The data collection was approved by the Norwegian Social Science Data Services (reference 21490/2/JE). Ethical approval was given by the Regional Ethics Committee of South-Eastern Norway (reference S-08732b 2008/17889s) and the trial was carried out in accordance with the Helsinki Declaration. The parents of the infants gave informed consent. The inclusion period was from September 2009 to December 2012. The participating doctors were all GP specialists with a minimum of 300 hours of acupuncture education and five years of practising acupuncture.

The patients fulfilled Wessel’s criteria and were born at full term. They were randomized to active treatment or to no-treatment control. The assistant instructed the parents on how to fill in the crying registration form, and the patient was given appointments with the same GP three, four, and five days after inclusion.

The GP was alone in the treatment room with the infant during the intervention. The GP made a mark, 3 mm in diameter, at the point ST36 bilaterally on all children, to hide the insertion mark. In the intervention group, an ethylene-oxidised sterile Seirin acupuncture-needle (0.20/H1100315 mm) was inserted at the acupuncture point ST36. The point was needled bilaterally to approximately 12 mm depth. The two needles were left inserted without manipulation for 30 seconds. The needles were then withdrawn and a waterproof circular adhesive dressing, to further hide the insertion area, was applied. An identical procedure, except for the needle insertions, was performed on each infant in the no-treatment control group. The same procedure was performed on days 4 and 5.

Outcomes

The primary outcome was difference in changes in crying time in the registration period. Clinically relevant effect was defined as one-hour difference according to the pre-trial protocol. Secondary outcomes were differences in fulfilling Wessel’s infantile colic criteria, the parents’ assessment of the child’s condition, and adverse effects.

Sample size, randomization, and blinding

We anticipated a standardized difference of 0.5 and a clinically relevant difference of one hour. With p < 0.05 and 80% power, we needed to include 120 infants. Randomization was done manually by two persons not otherwise involved in the study. Sealed, opaque, numbered envelopes were used. Randomization was closed until the start of the first intervention. The parents and assistants were blinded to the allocation. Blinding validation of the parents was...
done by two blinding validation questions, one immediately after the first intervention, the second at week four.

Statistical methods
We used the software programs SPSS 19 and 20 for the statistical analyses. We used linear mixed models statistics for the analyses of the main outcome variable, and chi-square and Fisher’s exact tests for secondary outcomes of categorical data, and blinding validation analyses.

Results
Participant flow and baseline data
A total of 113 patients were recruited; 23 patients were excluded, and 90 randomized; 79 diaries and 84 interviews were analysed. A flow-chart diagram is shown in Figure 1. For baseline data, see Table I.

Blinding validation
The primary blinding validation question “Do you think your child has received acupuncture or not?” was answered in 83 cases and showed a random distribution. In the acupuncture group 22 of the parents believed the child had acupuncture and 22 believed the child was in the control group. In the control group the numbers were 16 and 23 (p = 0.41). The second blinding validation question, “Have you noticed any needle insertion marks?”, was answered in 38 cases: 17 in the acupuncture group and 17 in the control group answered “No”. Three in the acupuncture group and one in the control group answered “yes” (p = 0.60). We thus consider the blinding as valid.

Figure 1. Trial flow chart: The ST36 infantile colic acupuncture trial.
Changes in crying time

The primary end point of the trial was the difference in crying time changes between the acupuncture group and the control group. Linear mixed model main analyses of interaction time versus group for primary end point gave \( p = 0.26 \). There was no statistically significant baseline-corrected difference for the whole period, (Table II) or at any measured time period from baseline to the last measure after four weeks (Table III and Figure 2). There were no statistically significant changes over time in the interaction analyses between groups dividing into strata of more or less than four hours crying at baseline \( (p = 0.20) \), the eventual influence of other types of treatment \( (p = 0.30) \), type of feeding \( (p = 0.97) \), or concerning the primary blinding validation question \( (p = 0.98) \). Corrected for baseline differences, there was a small tendency in favour of the acupuncture group, with a non-significant baseline-corrected sum of mean differences in crying time during the assessment period of 13 minutes (95% CI −24 to +51).

Table II. Baseline-corrected mean overall difference in crying time reduction*: The ST36 infantile colic acupuncture trial.

<table>
<thead>
<tr>
<th></th>
<th>Acupuncture group ( (n = 44) )</th>
<th>No treatment-control group ( (n = 40) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex-number of male/female</td>
<td>22/22</td>
<td>20/20</td>
</tr>
<tr>
<td>Mean age in weeks at inclusion (range)</td>
<td>6 (3–15)</td>
<td>6(3–9)</td>
</tr>
<tr>
<td>Mean gestation age in weeks at birth (range)</td>
<td>40 (36–42)</td>
<td>39 (36–42)</td>
</tr>
<tr>
<td>Mean birth weight in grams (range)</td>
<td>3590 (2580–4600)</td>
<td>3544 (2375–4410)</td>
</tr>
<tr>
<td>Feeding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>28</td>
<td>23</td>
</tr>
<tr>
<td>Supplement</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Both breast and supplement</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Pre-study interventions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chiropractic</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Milk free diet in mother</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Other (dimethichoner, herbal tea, malt extract, sugar water, massage)</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Several different treatments tried</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>No pre-study treatments</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Crying time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline crying time in minutes (range)</td>
<td>220 (80–394)</td>
<td>212 (0–353)</td>
</tr>
<tr>
<td>Number of children with crying time ≥ 4 hours</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>Number of children with crying time &lt; 4 hours</td>
<td>26</td>
<td>24</td>
</tr>
</tbody>
</table>

*The number of patients differ because of difference in numbers of interviews and reported crying diaries.

The effect size for the intervention in this trial was estimated to be 0.16.

Secondary outcomes

On day 6, nine out of 38 in the acupuncture group and 10 out of 41 in the control group fulfilled Wessel’s criteria of more than three hours’ crying per day (odds ratio 1.1, CI 0.4–3.2, \( p = 0.94 \)). This lack of differences was repeated at day 12 (\( p = 0.60 \)) and day 33 (\( p = 0.31 \)). The parents’ evaluation showed a slight tendency in favour of acupuncture over time, with a statistically significant difference on day 33 of 0.51 (CI 0.04–0.99), (\( p = 0.034 \)). There were no serious adverse effects in the acupuncture or control group.

Discussion

The trial showed a non-significant baseline-corrected mean difference of 13 minutes for the main outcome variable, changes in crying time, in favour of the acupuncture group. A linear mixed model of time versus group gave \( p = 0.26 \). This is not statistically significant or clinically relevant according to the protocol.

We failed to reach our sample size goal of 120 crying time diaries. With a baseline-corrected mean difference of 13 minutes in our trial and a standardized mean difference (SMD) or effect size of 0.16, we would have needed ten times as many participants, i.e. 1200 infants in a two-sample t-test for one
time measure and 800 in a mixed model analysing changes over time, in order to show a statistically significant effect.

Needle-specific acupuncture effect for various pain conditions in adults have in meta-analyses turned out to be small, ranging from 0.15 to 0.23 [22,23]. Clinical acupuncture effects in adults are divided into three parts: cortical anticipation/reward (placebo) effects, non-specific physiological effects, and needle-specific effects.

Infants can be considered to have no anticipation/reward effects regarding acupuncture treatment. The effects the parents observe are the true needle effects, on the condition that the parents are truly blinded, as was the case in the present trial. The effect size of 13 minutes and SMD 0.16 is similar to the needle effect size found in adults, and contradicts the often implied notion by practitioners and textbooks that acupuncture effects in small children are stronger and faster [24–26].

There is only one review regarding acupuncture effects on pain conditions in infants, with only four eligible RCTs [27]. The authors conclude that the data are sparse, and that acupuncture on infants should be limited to clinical trials. There are no estimations of needle-specific effect sizes.

Two previous trials investigate acupuncture treatment in infantile colic, by Reinthal et al. and Landgren et al. [16,17]. Both trials concluded that acupuncture for infantile colic may reduce crying intensity and crying duration. These trials had different primary endpoints, and used different acupuncture points and different insertion time and method. The median differences in reduction in crying time in favour of the acupuncture group over the assessment period were 19 minutes (n = 40) and 11 minutes (n = 80). The
corresponding result in our trial \((n = 79)\) was a mean difference of 13 minutes, also in favour of the acupuncture group. So the estimated effect sizes in these three trials are similar, and small. Our choice of acupuncture points and technique could be inferior to other more potent acupuncture approaches. There are as yet no trial data on infants to support this. In adults, however, individual patient data analysis of acupuncture for various chronic pain conditions, with over 18 000 patients in total, showed effect sizes that were similar across groups and with different treatment approaches [23].

Acupuncture treatment in children is considered a safe intervention [15,28], which was confirmed in the present trial. Adverse reactions were few and insignificant.

Acupuncture in children is a potentially painful treatment [27], and this concern is especially important in small children without competence of consent [29]. Nevertheless, painful or potentially painful interventions in small children are used when the benefit clearly outweighs the harm. This consideration should also apply to acupuncture treatments. Acupuncture has been shown to work consistently in pain conditions in adults [23,30]. The needle-specific effects are small, and the combination of needle-physiological effects and cortical anticipation/reward effects contributes to improvement in several pain conditions. Small children have no established anticipation/reward systems; it is the real effect of the needle that defines the relevance of acupuncture treatment. It would be unethical to treat in response to the parents’ hope of improvement if the effect size does not outweigh the potential pain inflicted on the child. Our trial could not prove such a justification.

Conclusion

Our trial of acupuncture for infantile colic showed no statistically significant difference in reduction of crying time between the acupuncture and the control group. There was a non-significant baseline-corrected mean overall difference in crying time reduction of 13 minutes in favour of the acupuncture group, which is not considered clinically relevant. With the current evidence, we suggest that acupuncture for infantile colic should be restricted to clinical trials.

Acknowledgements

The authors would like to thank Magne Thoresen, professor at the Department of Biostatistics, University of Oslo, for his invaluable help in the statistical analyses of this trial.

Ethical approval

Regional Ethics Committee of South-Eastern Norway, reference, S-08732 2008/17889s.

Funding

The study was financed through a grant from the Norwegian Research Fund in General Practice (AMFF).

Trial Registry

Clinical Trial Registry Identifier: NCT00907621.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Acupuncture for infantile colic


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

Holgeir Skjeie¹, Trygve Skonnord¹, Mette Brekke¹, Atle Klovning¹, Arne Fetveit¹, Kajsa Landgren², Inger Kristensson Hallström², Kjetil Gundro Brurberg³ ⁴

¹ Department of General Practice, Institute of Health and Society, University of Oslo, P.O. Box 1130 Blindern, N-0318 Oslo, Norway

² Faculty of Medicine, Department of Health Sciences, Lund University, P.O. Box 157, Lund SE-22100, Sweden

³ Norwegian Institute of Public Health, Division for Health Services, P.O.Box 4404 Nydalen, N-0403 Oslo, Norway

⁴ Western Norway University for Applied Sciences, Centre for Evidence Based Practice, P.O.Box 7030, N-5020 Bergen, Norway

Corresponding author: Holgeir Skjeie. Department of General Practice, Institute of Health and Society, University of Oslo, P.O. Box 1130 Blindern, N-0318 Oslo, Norway.
E-mail: holgeir.skjeie@medisin.uio.no

Word Count: 5,092

References: 69

Short running title: IPD meta-analysis of acupuncture for infantile colic.
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 endpoints were crying time at mid-treatment, at the end of treatment, and at a 1-month follow-up. A 30-minute 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 minutes [95% confidence interval, CI –46.2 to –3.6; three trials] at mid-treatment, –11.4 minutes [95% CI –31.8 to 9.0; three trials] at the end of treatment, and –11.8 minutes [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 minutes [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 to 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

MeSH Keywords: Acupuncture. Infant. Colic. Review. Meta-analysis.
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.
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 three months of age, with more than three hours of crying per day in more than three days for more than three 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 fibers. 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 to 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
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 use 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] [Appendix 1]. The unabridged protocol is included as [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 minutes 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 [Appendix 3]

Electronic scoping searches were conducted in Best Practice, UpToDate, Cochrane [CDSR, DARE, and HTA], and Prospero from inception to the search date. 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 [HS and TS] and one methodological expert [AK], 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 [AK] 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 [MB].

**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 AK, 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 痉气 [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 [YS] 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 [HS] in Beijing in May 2016.
Relevant articles according to the predefined eligibility criteria were translated and further discussed with HS. 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, KGB 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. KGB, 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 [TS, AK] 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-blindness 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 postintervention measurements. Data from all measurement points were included in a single model. The postintervention 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 a 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 a 10-20% reduction or 10-20mm 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 minutes, 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 minutes difference in reduction in crying time between intervention and control groups was considered as a clinically relevant effect. We changed “one 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 to 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 [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 [AK, KGB] 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.[32-34] The characteristics of the included studies are described in detail in Table I. 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 2010 and the result of parental evaluation of effects in Landgren et al 2010 and Landgren et al 2017.

**Risk of bias and blinding validation**

The risk of bias summary is presented in Figure 2. Detailed risks of bias assessments are available in Appendix 4. Briefly, all included RCTs had adequate randomisation procedures reported and allocation concealment described [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 II] 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 2010 were asked about allocation beliefs after the last treatment session. In
Skjeie et al 2013, parents were questioned about allocation beliefs after the first treatment session, and after four weeks they were asked if they had noticed any needle marks. Landgren et al 2017 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 2010, 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 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 [Appendix 1 and 2]. Table III 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 prespecified time periods [Figure 3]. There was a statistically significant difference in mean crying time [MD –24.88 minutes/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 to 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 the end of the treatment [Figure 5], with an OR of 3.03 [95%CI 1.56 to 5.89] for rating the condition as much improved and OR 2.67 [95%CI 1.43 to 4.97] for improved. The odds ratio for worsening was only available from the Landgren 2010 trial [OR 0.83; 95% CI 0.22 to 3.18].

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 to 20.64], although the majority stopped crying within seconds²⁴ [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 to 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 seconds, 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 minutes 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 contradicts 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] 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 are 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

The authors would like to thank Yan Shao [YS] at the Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine for carrying out the Chinese language literature search and the invaluable discussion of the search results with HS.
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.

Funding

The study was financed through a grant from the Norwegian Research Fund in General Practice [AMFF]. The sponsor had no role in the design, analysis, or interpretation of findings, nor in the manuscript.

Conflict of interest

The manuscript has been read and approved by the authors who meet the International Committee of Medical Journal Editors [ICMJE] criteria for authorship. No other persons have contributed significantly to the manuscript’s preparation. There are no conflicts of interest. All authors have completed the ICMJE Conflict of interest form.
References


Table I: Characteristics of included studies

Table II: Blinding indices

Table III: Summary of findings – GRADE

Figure 1: PRISMA IPD Flow Diagram

Figure 2: Risk of bias summary report

Figure 3: Primary end-point

Figure 4: Sensitivity analysis

Figure 5: Secondary end-points

Appendix 1: Prospero pdf mars-17 IPD-meta

Appendix 2: Final full protocol after Lund consensus meeting

Appendix 3: Search profile 30.01.17

Appendix 4: Risk of bbias table- full report

Appendix 5: PRISMA-2009-Checklist MS-Word
Table 1. Characteristics of included studies.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Landgren et al., 2010 Sweden</th>
<th>Landgren et al., 2017 Sweden</th>
<th>Skjeie et al., 2013 Norway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td>RCT</td>
<td>Multicentre RCT</td>
<td>Multicentre RCT</td>
</tr>
<tr>
<td>Participants</td>
<td>Healthy infants 2–8 weeks old, born after week 36, with appropriate weight gain, fulfilling the modified Wessel’s criteria of crying/fussing ≥3 hours/day for ≥3 days a week. Exclusion criterion: medications other than dimethicone or Lactobacillus reuteri</td>
<td>Healthy infants, 2–8 weeks old, born after week 36, with appropriate weight gain, crying/ fussing ≥3 hours/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</td>
<td>Healthy infants born at full term and &lt;3 months of age at inclusion. Filling Wessel’s criteria of paroxystic uncontrollable crying/fussing ≥3 hr/day for ≥3 days a week in ≥3 weeks. No exclusion criteria</td>
</tr>
<tr>
<td>Outcomes</td>
<td>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</td>
<td>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</td>
<td>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</td>
</tr>
</tbody>
</table>
### Table II. Blinding index with three different blinding tests.

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>James [Acupuncture]</th>
<th>Bang [Control]</th>
<th>Chi-squared test P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landgren 2010</td>
<td>43</td>
<td>0.21*</td>
<td>0.77</td>
<td>0.67*</td>
</tr>
<tr>
<td>Landgren 2017 [1]</td>
<td>96</td>
<td>0.56*</td>
<td>0.44</td>
<td>-0.08*</td>
</tr>
<tr>
<td>Landgren 2017 [2]</td>
<td>96</td>
<td>0.51*</td>
<td>0.63</td>
<td>-0.08*</td>
</tr>
<tr>
<td>Skjeie 2013</td>
<td>83</td>
<td>0.46*</td>
<td>0.00*</td>
<td>0.18*</td>
</tr>
</tbody>
</table>

* Not blinded for outcome assessors [parents].
*+ Unsure blinding of outcome assessors.
*+ Blinded for outcome assessors.
### Table III. Summary of findings for primary outcome: differences in crying time.

<table>
<thead>
<tr>
<th>Timing</th>
<th>Average min. crying per day</th>
<th>SMD</th>
<th>No. of participants</th>
<th>Quality of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Acupuncture</td>
<td>[95% CI]</td>
<td></td>
</tr>
<tr>
<td>Mid- treatment</td>
<td>193 min</td>
<td>25 min less [46 less to 4 less]</td>
<td>−0.23 [−0.42 to −0.05]</td>
<td>307 [3 studies]</td>
</tr>
<tr>
<td>End of treatment</td>
<td>156 min</td>
<td>11 min less [32 less to 9 more]</td>
<td>−0.10 [−0.29 to 0.08]</td>
<td>304 [3 studies]</td>
</tr>
<tr>
<td>Long-term follow-up</td>
<td>97 min</td>
<td>12 min less [63 less to 39 more]</td>
<td>−0.09 [−0.48 to 0.30]</td>
<td>79 [1 study]</td>
</tr>
</tbody>
</table>

<sup>A</sup> Wide confidence intervals [CI] and imprecision; <sup>B</sup> One study with few participants. SMD, standardised mean differences.

### Summary of findings for secondary outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Absolute effect per 100 [95% CI]</th>
<th>Relative effect OR [95% CI]</th>
<th>No. of participants</th>
<th>Quality of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crying during treatment</td>
<td>37</td>
<td>81 [61 to 92]</td>
<td>81</td>
<td>⊕⊕⊕⊙ Moderate&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td>Worsening</td>
<td>30</td>
<td>26 [9 to 58]</td>
<td>46</td>
<td>⊕⊕⊕⊙ Low&lt;sup&gt;ABA&lt;/sup&gt;</td>
</tr>
<tr>
<td>Some improvement</td>
<td>65</td>
<td>83 [73 to 90]</td>
<td>264</td>
<td>⊕⊕⊕⊙ Moderate&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td>Much improvement</td>
<td>26</td>
<td>52 [35 to 67]</td>
<td>264</td>
<td>⊕⊕⊕⊙ Moderate&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td>No colic&lt;sup&gt;§&lt;/sup&gt;</td>
<td>60</td>
<td>70 [57 to 80]</td>
<td>304</td>
<td>⊕⊕⊕⊙ Moderate&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>§</sup> defined as >180 minutes of crying per day; <sup>ABA</sup> Wide confidence intervals [CI] and imprecision; <sup>A</sup> Not possible to estimate in two studies because of too few events.
The PRISMA IPD flow diagram

© Reproduced with permission of the PRISMA IPD Group, which encourages sharing and reuse for non commercial purposes
<table>
<thead>
<tr>
<th></th>
<th>Landgren 2010</th>
<th>Landgren 2017</th>
<th>Skjoie 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence generation</td>
<td>+ + ?</td>
<td>+ + ?</td>
<td>+ +</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td>+ + ?</td>
<td>+ + ?</td>
<td>+ +</td>
</tr>
<tr>
<td>Blinding of outcome assessors</td>
<td>+ + ?</td>
<td>+ + ?</td>
<td>+ +</td>
</tr>
<tr>
<td>Blinding of outcome data</td>
<td>+ + ?</td>
<td>+ + ?</td>
<td>+ +</td>
</tr>
<tr>
<td>Incomplete outcome reporting</td>
<td>+ + ?</td>
<td>+ + ?</td>
<td>+ +</td>
</tr>
<tr>
<td>Selective outcome reporting</td>
<td>+ + ?</td>
<td>+ + ?</td>
<td>+ +</td>
</tr>
<tr>
<td>Other sources of bias</td>
<td>+ + ?</td>
<td>+ + ?</td>
<td>+ +</td>
</tr>
</tbody>
</table>
Primary end-point: crying time

### A Mean difference – MD (Mid-treatment, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Mean Difference</th>
<th>95% CI</th>
<th>Mean Difference</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lansgreen 2010</td>
<td>-58.90 (64.90, -53.90)</td>
<td></td>
<td>-58.90 (64.90, -53.90)</td>
<td></td>
</tr>
<tr>
<td>Lansgreen 2017 (Y)</td>
<td>-30.50 (41.97, -20.70)</td>
<td></td>
<td>-30.50 (41.97, -20.70)</td>
<td></td>
</tr>
<tr>
<td>Lansgreen 2017 (O)</td>
<td>-6.00 (45.20, 20.00)</td>
<td></td>
<td>-6.00 (45.20, 20.00)</td>
<td></td>
</tr>
<tr>
<td>Shave 2013</td>
<td>-12.00 (62.41, 38.41)</td>
<td></td>
<td>-12.00 (62.41, 38.41)</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>-24.80 (-48.36, -0.07)</td>
<td></td>
<td>-24.80 (-48.36, -0.07)</td>
<td></td>
</tr>
<tr>
<td><em>P</em> for overall effect: Z = 2.20 (P = 0.02)</td>
<td></td>
<td></td>
<td><em>P</em> for overall effect: Z = 2.20 (P = 0.02)</td>
<td></td>
</tr>
</tbody>
</table>

### B Standardised mean difference – SMD (Mid-treatment)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Std. Mean Difference</th>
<th>95% CI</th>
<th>Std. Mean Difference</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lansgreen 2010</td>
<td>-4.60 (-8.64, -0.56)</td>
<td></td>
<td>-4.60 (-8.64, -0.56)</td>
<td></td>
</tr>
<tr>
<td>Lansgreen 2017 (Y)</td>
<td>-1.10 (-5.97, 0.18)</td>
<td></td>
<td>-1.10 (-5.97, 0.18)</td>
<td></td>
</tr>
<tr>
<td>Lansgreen 2017 (O)</td>
<td>-6.11 (-10.50, -0.28)</td>
<td></td>
<td>-6.11 (-10.50, -0.28)</td>
<td></td>
</tr>
<tr>
<td>Shave 2013</td>
<td>-0.08 (-3.48, 3.30)</td>
<td></td>
<td>-0.08 (-3.48, 3.30)</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>-4.23 (-4.82, -3.65)</td>
<td></td>
<td>-4.23 (-4.82, -3.65)</td>
<td></td>
</tr>
<tr>
<td><em>P</em> for overall effect: Z = 2.43 (P = 0.01)</td>
<td></td>
<td></td>
<td><em>P</em> for overall effect: Z = 2.43 (P = 0.01)</td>
<td></td>
</tr>
</tbody>
</table>

### C Mean difference – MD (End of treatment, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Mean Difference</th>
<th>95% CI</th>
<th>Mean Difference</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lansgreen 2010</td>
<td>-14.90 (16.60, -3.20)</td>
<td></td>
<td>-14.90 (16.60, -3.20)</td>
<td></td>
</tr>
<tr>
<td>Lansgreen 2017 (Y)</td>
<td>-5.80 (11.85, 22.00)</td>
<td></td>
<td>-5.80 (11.85, 22.00)</td>
<td></td>
</tr>
<tr>
<td>Lansgreen 2017 (O)</td>
<td>-12.10 (44.96, 20.96)</td>
<td></td>
<td>-12.10 (44.96, 20.96)</td>
<td></td>
</tr>
<tr>
<td>Shave 2013</td>
<td>-0.10 (0.06, 0.44)</td>
<td></td>
<td>-0.10 (0.06, 0.44)</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>-14.64 (15.15, -4.14)</td>
<td></td>
<td>-14.64 (15.15, -4.14)</td>
<td></td>
</tr>
<tr>
<td><em>P</em> for overall effect: Z = 1.50 (P = 0.27)</td>
<td></td>
<td></td>
<td><em>P</em> for overall effect: Z = 1.50 (P = 0.27)</td>
<td></td>
</tr>
</tbody>
</table>

### D Standardised mean difference – SMD (End of treatment)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Std. Mean Difference</th>
<th>95% CI</th>
<th>Std. Mean Difference</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lansgreen 2010</td>
<td>-1.17 (-5.52, 3.15)</td>
<td></td>
<td>-1.17 (-5.52, 3.15)</td>
<td></td>
</tr>
<tr>
<td>Lansgreen 2017 (Y)</td>
<td>-0.09 (-5.67, 5.26)</td>
<td></td>
<td>-0.09 (-5.67, 5.26)</td>
<td></td>
</tr>
<tr>
<td>Lansgreen 2017 (O)</td>
<td>-6.11 (-5.83, 2.27)</td>
<td></td>
<td>-6.11 (-5.83, 2.27)</td>
<td></td>
</tr>
<tr>
<td>Shave 2013</td>
<td>-0.82 (-4.84, 3.25)</td>
<td></td>
<td>-0.82 (-4.84, 3.25)</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>-0.10 (-4.29, 4.08)</td>
<td></td>
<td>-0.10 (-4.29, 4.08)</td>
<td></td>
</tr>
<tr>
<td><em>P</em> for overall effect: Z = 1.05 (P = 0.29)</td>
<td></td>
<td></td>
<td><em>P</em> for overall effect: Z = 1.05 (P = 0.29)</td>
<td></td>
</tr>
</tbody>
</table>

### E Mean difference – MD (Long-term follow-up, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Mean Difference</th>
<th>95% CI</th>
<th>Mean Difference</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shave 2013</td>
<td>-1.85 (-9.05, 5.36)</td>
<td></td>
<td>-1.85 (-9.05, 5.36)</td>
<td></td>
</tr>
</tbody>
</table>

### F Standardised mean difference – SMD (Long-term follow-up)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Std. Mean Difference</th>
<th>95% CI</th>
<th>Std. Mean Difference</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shave 2013</td>
<td>-0.09 (-0.43, 0.25)</td>
<td></td>
<td>-0.09 (-0.43, 0.25)</td>
<td></td>
</tr>
</tbody>
</table>
Primary end-point: Crying time with sensitivity analysis

A  Mean difference – MD (Mid-treatment, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landgren 2010</td>
<td>26.6%</td>
<td>-55.30 [-84.62, -15.98]</td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>24.1%</td>
<td>-20.50 [-41.97, 20.97]</td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>32.5%</td>
<td>-9.90 [-45.06, 25.26]</td>
</tr>
<tr>
<td>Skjønne 2013</td>
<td>16.8%</td>
<td>-12.00 [-62.41, 38.41]</td>
</tr>
</tbody>
</table>

Total (95% CI) 100.0% -24.88 [-46.20, -3.57]

Heterogeneity: \( \tau^2 = 42.33; \chi^2 = 3.29, \text{df} = 3 (P = 0.35); I^2 = 9\%

Test for overall effect: \( Z = 2.29 \) (\( P = 0.02 \))

B  Sensitivity analysis. Mean difference – MD (Mid-treatment, min/day)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landgren 2010</td>
<td>0.0%</td>
<td>-55.30 [-84.62, -15.98]</td>
</tr>
<tr>
<td>Landgren 2017 (1)</td>
<td>32.6%</td>
<td>-20.50 [-41.97, 20.97]</td>
</tr>
<tr>
<td>Landgren 2017 (2)</td>
<td>45.3%</td>
<td>-9.90 [-45.06, 25.26]</td>
</tr>
<tr>
<td>Skjønne 2013</td>
<td>22.1%</td>
<td>-12.00 [-62.41, 38.41]</td>
</tr>
</tbody>
</table>

Total (95% CI) 100.0% -13.82 [-37.50, 9.86]

Heterogeneity: \( \tau^2 = 0.00; \chi^2 = 0.15, \text{df} = 2 (P = 0.93); I^2 = 0\%

Test for overall effect: \( Z = 1.14 \) (\( P = 0.25 \))
Secondary end-points

A Parental report at the end of treatment

B <180 min crying per day at the end of treatment

C Crying during treatment
ABSTRACT

Objective: To explore contemporary practices and clinical recommendations regarding the use of acupuncture for infants by Traditional Chinese Medicine (TCM) practitioners in Shanghai.

Design: A qualitative study consisting of four field visits between February 2014 and March 2015. Data was collected using participant observation, focus group interview, in-depth interview, textbook page analysis and informant validation.

Participants: 14 Shanghainese professionals, including interpreters and TCM practitioners, of which seven were acupuncturists.

Setting: The Longhua Hospital (paediatric, acupuncture and Tui na departments) in southern Shanghai and the campus of the Shanghai University of Traditional Chinese Medicine.

Results: The Longhua Hospital outpatient acupuncture clinic receives 400 consultations on average per day. Children, including patients from the paediatric department, are referred to this clinic. During 3 days of participant observations at this department, we saw two children. No infants. During 3 days at the paediatric department and 1 day at the Tui na department we saw no referrals. Formal interviews and informal conversations with acupuncturists and other TCM professionals revealed that acupuncture was neither routinely practiced nor recommended for infants and small children. Acupuncture was considered potentially painful for this young patient population. Alternative treatment options such as herbal treatments or medical massage were widely available and preferred. Western medical diagnostics and treatment were also used, recommended, and trusted.

Conclusions: Acupuncture for infants is not a preferred therapeutic method among TCM practitioners working in contemporary Shanghai. Acupuncture on broad indications in infants appears to be a Western practice with little basis in TCM modern-day practice.

BACKGROUND

Acupuncture is a part of Traditional Chinese Medicine (TCM). There are three main pillars of TCM: Herbal medicine, acupuncture-moxibustion (needling and heating) and Tui na (medical massage). Herbal medicine is the mainstay of TCM in China, while acupuncture-moxibustion and Tui na are regarded as auxiliary and complementary therapies. Acupuncture is used widely in Western Europe, North America and Australia.1–6 The treatment principles of acupuncture are relatively straightforward, especially for treatment of various pain conditions in adults and the mechanisms for the neurophysiological effects of acupuncture are fairly well understood.7 In acupuncture, thin steel needles are penetrated through the skin and into connective tissue and muscle fibres to elicit effects. Compared to no treatment or treatment-as-usual, specific needle effects are small, with standardised mean differences (SMD) ranging from 0.15 to 0.23. The overall effectiveness is larger, with a SMD of about 0.5 SMD.8 This larger overall effect is attributed to acupuncture being a particularly good placebo irrespective of treatment approach or practitioner style or experience.9
Acupuncture in children, including infants, has also gained acceptance in the West. However, there is scant evidence for treatment effects, no clinical guidelines exist and practices vary substantially between countries. Individual textbook recommendations cover the whole range of paediatric indications from ear infections to autism and from asthma to inflammatory bowel disease. There are concerns regarding the ethical aspects of this treatment, as it is a potentially painful method. In contrast to adults and older children, infants lack the ability to provide informed consent and this requires a higher threshold for interventions. The relatively large placebo effects of acupuncture through the patient’s belief in the treatment being effective, deemed important in the treatment of adults, would arguably not occur in this young patient population. Acupuncture in experienced hands is considered safe. It is also, with experienced practitioners, considered safe in children. Yet the scientific evidence on the effectiveness of acupuncture treatment for young children is sparse or non-existent, and randomised controlled trials regarding the efficacy of acupuncture for pain conditions in infants are few with conflicting results.

Western-based textbooks of acupuncture argue that acupuncture effects in infants and small children are swift, and often stronger than in adults, contradicting the few existing randomised controlled studies. These textbook notions seem to be based on references to TCM tradition and on the authors’ personal views or clinical observations rather than evidence from randomised, controlled clinical trials. Published case reports and qualitative studies are also generally supportive of acupuncture in infants and small children, and recommend acupuncture for broad indications in paediatric populations. There is little scientific knowledge regarding the nature and extent to which acupuncture is used for pain conditions in paediatric populations in modern-day China. Chinese-English language textbooks consider herbal remedies as the primary treatment method in children, followed by Tui na, although acupuncture is often mentioned as an adjunct treatment. A literature search of major databases conducted by the first author (Cochrane, MEDLINE, EMBASE, AMED, Maternity and Infant Care, Global Health, PsycINFO, Anthropology Plus, Sociological Abstracts, ISI-Web of knowledge, IBSS, AIO, BASE and Wanfang Data, including conference proceedings from academic conferences in China) identified no English-language studies describing contemporary TCM clinical practices or clinical guidelines on the use of acupuncture for infants or small children from the People’s Republic of China, Taiwan, Singapore and Hong Kong. Survey studies of TCM in children have been reported from Singapore and Taiwan, but we found no English-language studies or published abstracts on TCM practitioners’ views and attitudes toward acupuncture treatment for infants and small children.

The purpose of this study was to investigate current opinions and clinical practices regarding needle acupuncture for infants among TCM clinicians in Shanghai.

METHOD

We chose a qualitative approach, with participating observations, interviews and literature searches. An extended, flexible approach was developed and a combination of qualitative methods was used, reflecting the validity threats and logistical challenges of conducting qualitative research on the attitudes towards and use of paediatric acupuncture among TCM clinicians in Shanghai, China. Subject knowledge matters, as described in Kvale and Brinkmann, guided the decision not to do surveys or rely on focus group interviews as the only sources of information. The study process was informed by ‘Qualitative Research Design’ by Maxwell. Special emphasis was on validity threats: We used Maxwell’s eight-point checklist to strengthen validity as a guide when designing and carrying out the study and in the analysing process. His recommendations of several information gathering methods to ensure validity was central to our decisions. Long-term involvement—we included four field visits over the course of 14 months. Rich data—we transcribed ad verbatim all field notes, quotes from informal conversations and formal interviews. Respondent validation—we had two stages of informant validation. Intervention—this was not appropriate in our study. Searching for discrepant evidence and negative cases—we did search for and do report discrepant evidence.

Triangulation—we used both participant observation and informal conversations, formal interviews and textbook searches. Numbers—we counted opinions, pages and patients, and report it in a table. Comparison—we had participant observations at three departments and at different times. The main analyses of the transcribed field notes and interviews were informed by Thematic analysis, focused on manifest content, and is further described in the Data Analysis section.

Setting

The first author conducted four field visits between February 2014 and March 2015 to Longhua Hospital, which is a teaching hospital of the Shanghai University of Traditional Chinese Medicine (SHUTCM) (figure 1). Longhua Hospital is a municipal TCM hospital with 2000 beds. The acupuncture department is an outpatient clinic with 42 beds, staffed by 10 doctors and 20 interns who receive an average of 400 consultations per day. The paediatric department receives about 130 outpatient consultations per day, with an average of 10 inpatients admitted daily to the unit at any given time. There is also a small Tui na department which runs an outpatient service, staffed by five doctors who receive approximately 50 consultations per day. The WHO Collaboration Centre at SHUTCM facilitated study
access and granted permission to collect data at Longhua Hospital. The first author is a male general practitioner and acupuncturist with 25 years of experience, with an established working relationship with the WHO Collaboration Centre.

Data collection

The field visits consisted of the following:

1. Three days of participant observation in the paediatric department, 3 days of participant observation in the acupuncture department and 1 day of participant observation in the Tui na department. An interpreter who was either an English-speaking TCM doctor or an English–Chinese linguist was present at any time. We followed the daily routines and engaged in informal conversations with several acupuncturists, and paediatric and Tui na specialists. The interpreters themselves were also a rich source of information.

2. A 90 min, semistructured focus group interview using a predefined interview guide, not pilot-tested. The interview was conducted in English and without an interpreter, outside the workplace. The informants were three leading officials at acupuncture units in Shanghai, all of whom had clinical and administrative responsibilities. They were experienced acupuncturists. No parallel field notes were taken during the interview, and transcripts were not returned. We did not succeed in recruiting regular clinicians for this interview, because they felt they lacked sufficient knowledge of and had minimal experience with, paediatric acupuncture. Five experienced acupuncture clinicians were contacted in succession, all declined for the same reasons. Only then did we decide to go up one level and invite leading officials. We had to take into consideration the probability of group norm answers, but this would also give valuable information and expand the triangulation.

3. A 60 min, in-depth individual interview using a predefined interview guide, not pilot-tested, with a non-acupuncture TCM practitioner who had academic credentials, including work and research experience in Western Europe. The interview was conducted in Chinese with an English–Chinese interpreter outside the workplace. No parallel field notes were taken during the interview and transcripts were not returned.

4. A search in the standard national textbooks used for teaching acupuncture, Tui na and herbal medicine at Shanghai University of Traditional Chinese Medicine, one book for each curriculum. These textbooks are compulsory nation-wide. They form part of the basis for the various 5-year TCM teaching programmes to qualify as herbalist, acupuncturist or Tui na practitioner in the People’s Republic of China. The search was assisted by a linguist fluent in Chinese and English. We counted section pages and case pages to gauge the factual emphasis on using TCM therapeutic methods for paediatric populations in the national textbooks.

5. Two stages of informant validation on the main and final results. Main results by one central informant, and final results by two independent informants. We presented main and final results personally in written form on a paper, the first time with five major preliminary findings, the second time with the Results section of the study. The validity of the findings was then discussed, minor ambiguities were corrected and the Results section was then accepted as valid in the opinion of these informants.

A total of 14 informants were included during the field study. Informants had diverse clinical backgrounds representing several TCM specialty areas, including seven acupuncturists (figure 2). There were six women and eight men between 35 and 60 years of age, with a

Figure 1  Study flow chart.
minimum of 10 years clinical experience. To maintain anonymity, further details on the participants are not disclosed.

**Data analysis**

All data were used, including notes from participant observations, informal conversations, semistructured interviews and textbook page analysis. The field notes were structured and systematised, including the participant observations and the information gathered from informal conversations. The focus group and in-depth interviews were recorded on two parallel, portable tape recorders and transcribed in verbatim by the first author. The sound files were transferred to a USB storage pen and securely stored. The original sound tracks on the tape recorders were erased.

The data analyses were carried out using thematic analysis. The analyses were performed by hand and coding was primarily on semantic (manifest) content. We did not seek saturation. The first author read through the texts several times to define broad categories from the interviews and field notes with subsequent coding and recategorisation. The main categories and selected opinion statements were then sequenced into tables. We did not use a coding tree. The initial analysis was then reviewed by the second author before refinement and undergoing final organisation. All quotes have been retained in their original form.

The study is reported in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist.

**Ethics**

The Regional Ethical Committee of Southeastern Norway, REK South East granted this project an exemption from formal approval owing to the lack of patient data (Ref: 2014/197/REK South East). The informants who participated in formal interviews were informed of the purpose of the study and they signed a written consent form. The head or acting head of the three departments at the Longhua hospital were informed of

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Results in numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Patients observed to have acupuncture during 3 days in Longhua Acupuncture department (400 consultations a day, only out-patients)</td>
<td></td>
</tr>
<tr>
<td>Infants</td>
<td>0</td>
</tr>
<tr>
<td>Children</td>
<td>2</td>
</tr>
<tr>
<td>II. The Shanghai TCM acupuncturists (7 informants) Do they have experience in needle acupuncture on infants?</td>
<td></td>
</tr>
<tr>
<td>Very little or no experience</td>
<td>5</td>
</tr>
<tr>
<td>Some experience</td>
<td>2</td>
</tr>
<tr>
<td>Regular or daily experience</td>
<td>0</td>
</tr>
<tr>
<td>III. All Shanghai TCM professionals (12 informants) Do they recommend needle acupuncture on infants?</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>9</td>
</tr>
<tr>
<td>Yes, but only as the last choice</td>
<td>3</td>
</tr>
<tr>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td>IV. Amount specifically on paediatric conditions in Standard national textbook in undergraduate Traditional Chinese Medicine (TCM) studies, Shanghai.</td>
<td></td>
</tr>
<tr>
<td>Herbal medicine:</td>
<td></td>
</tr>
<tr>
<td>Special textbook on paediatric herbal medicine:</td>
<td>140 pages</td>
</tr>
<tr>
<td>Tui na/medical massage:</td>
<td>57 of 248 pages</td>
</tr>
<tr>
<td>Paediatric section of general textbook:</td>
<td>6 of 328 pages</td>
</tr>
<tr>
<td>Acupuncture-moxibustion:</td>
<td></td>
</tr>
<tr>
<td>No paediatric section in general textbook.</td>
<td></td>
</tr>
<tr>
<td>Conditions relating to children:</td>
<td></td>
</tr>
<tr>
<td>TCM, Traditional Chinese Medicine.</td>
<td>140 pages</td>
</tr>
</tbody>
</table>
the purpose of the study and they signed a written consent form.

RESULTS

The absence of children

During the 3-day field visit at the acupuncture department at Longhua TCM hospital, we moved freely between the consultation rooms, waiting rooms and reception area. In total, we saw two children. No infants (table 1-I).

‘Nobody does paediatric acupuncture’. Lack of experience

During informal conversations with clinicians at the acupuncture department, we were informed that children and infants were very seldom referred for treatment, and thus, the clinicians stated they lacked experience with needle acupuncture treatment on children. Several of the acupuncturists reported they had treated only one or two children past year. None of the clinicians had treated infants. As one TCM acupuncturist described, “Not so many children here. I treated two patients last year. Six-year olds. Facial paralysis and poor concentration. No infants.” During the formal focus group interview with the leading officials of acupuncture units in Shanghai, two of three reported some past experience in treating children, including infants. However, they did not provide information on whether this occurred on a regular basis (table 1-II).

‘Not convenient’. Lack of recommendations

Formal interviews and informal conversations with TCM clinicians revealed that acupuncture for infants and a small children was not recommended. The focus group interview confirmed this finding, and the officials from the acupuncture clinics stated that acupuncture therapy for infants and small children was considered ‘the last choice’ (table 1-III). One TCM acupuncturist stated, “Never used acupuncture, especially with small children. How could you? Not convenient. Big needles, small bodies. We use Tui na and herbs.” A TCM paediatrician expressed “Acupuncture is considered invasive. We are reluctant to do that on infants.” One exception to this consensus included neurological conditions such as cerebral palsy, neurological birth defects and recently, attention deficit hyperactivity disorder, which is regarded as a neurological disease. For these conditions, needle acupuncture was considered a possible treatment method.

‘No classic, no modern’. Lack of textbooks and training

In interviews and informal conversations, TCM clinicians reported a lack of specialised textbooks and training in performing needle acupuncture on children. In contrast, textbooks and specialised courses exist for paediatric herbal medicine and paediatric Tui na. Traditional and modern teaching in acupuncture does not distinguish between acupuncture point combinations for adults versus children. As one TCM acupuncturist explained, “Never been taught pediatric acupuncture in MD China or postgraduate education. No special points, no special treatments, no special meridians, no special organ systems.”

A 5-year programme of integrated Western and TCM coursework is required for MD China, which qualifies for certification as a TCM herbalist, acupuncturist or Tui na specialist. A search of the national textbooks at SHUTCM Shanghai University of Traditional Chinese Medicine revealed the following page counts and specific chapter dedicated to children per textbook:

- Moxibustion: 6 of 328 pages, no paediatric chapter.

In the treatment chapters, children are mentioned on page 254 (enuresis), malnutrition (page 255), polio sequelae (page 255), mumps (page 258), convulsions (page 274) and otitis media (page 286) (table 1-IV).

‘For now, zero’. The lack of a strong research base

During the focus group interview, we explored the informants’ knowledge regarding research on acupuncture treatment for infants and small children. Despite being familiar with numerous studies on herbal treatments and Tui na in paediatric samples, they reported no knowledge of ongoing or existing research on needle acupuncture in infants and small children.

‘Of course it is painful’. The possibility of pain

Pain as a consequence of needle acupuncture was discussed with all the TCM clinicians, including paediatricians, Tui na specialist and acupuncturists. Prevailing attitudes among clinicians was that acupuncture was painful for infants and small children. The potential for pain was stated as a primary reason for the overall limited use of needle acupuncture in infants. For example, one TCM acupuncturist stated, “We have very few children. No infants. It is hard to make good acupuncture treatment. And it is painful.” Two of the leading acupuncture officials diverged in opinion, however, stating that needle acupuncture in children could be painless when performed in the right way with the proper method. One TCM official reported, “There are some special methods, though, for inserting the needle without pain. Too many adults and children are just scared of the needles.”

‘Eighty percent of acupuncture on children is CP or other neurology’. Indications for acupuncture

Acupuncturists and other TCM clinicians expressed the clear opinion that needle acupuncture was indicated for very few paediatric conditions in general. Indications included neurological diseases in which other treatment options were limited. However, two of the TCM
acupuncturists/hospital officials diverged in opinion by recommending broader indications for children, including asthma, malnutrition and digestive problems. Specifically concerning the infant population, there was broad consensus that needle acupuncture was not a treatment option in modern TCM. A Tui na specialist stated, “Acupuncture has such a long history, it never separated kids and adults. But you mentioned under one year, but this only for these acute problems, like convulsions, high fever, emergency. Then they may use needles.”

‘TCM paediatrics is herbs’. The availability of TCM alternatives: Herbal medicine and Tui na

Historically, several TCM treatments have been available for paediatric patients. In particular, treatments with herbal medicine and Tui na have been specifically developed for children. For contemporary Shanghai TCM practitioners working with children, treatment is largely focused on herbal medicine in the form of orally-administered drugs or herbal paste applied to acupuncture points. As explained by a TCM herbalist, “We focus on herbs. And some Tui na. Always like that. If acupuncture points, we use moxibustion.” This approach to treatment was confirmed during the participant observations at the TCM paediatrics department. The vast majority of paediatric outpatients were prescribed herbal treatment combinations, and some were recommended Tui na. If acupuncture comprised the treatment, the specific recommendation included herbal paste, often heated, as a point application. We did not observe needle acupuncture treatment, nor did we observe any recommendations provided for acupuncture during the field visit to the paediatrics department. Penetrating needle acupuncture on infants was not considered a treatment option for most TCM paediatric conditions.

‘Now parents take children to the Western Children Hospitals’. Trust in Western paediatrics

There are four governmental Children’s Hospitals in Shanghai. These hospitals are large Western medical hospitals with inpatient wards and specialised outpatient clinics staffed solely with medical doctors and other clinical staff educated in Western medicine. The vast majority of paediatric cases in Shanghai, particularly infants and toddlers, are treated at these hospitals. These hospitals provide treatment for a range of illnesses, from minor ailments to chronic or life-threatening conditions. Western medical hospitals are widely regarded as the first choice by Shanghai parents seeking assessment, consultation and treatment for their children. The TCM professionals themselves provide recommendations and referrals to the hospitals for infants and small children, as indicated by a TCM paediatrician, “They have better diagnostic equipment at Western hospitals, for example the pediatric ultrasound, for abdominal diagnostics. We do not have that.” A TCM herbalist stated, “Now parents take children to the Western Children Hospitals. If needed, herbs in addition. Acupuncture is never in their minds”.

‘The whole TCM world is shrinking in China, or at least in Shanghai’. The money and the system

During the field study at Long Hua hospital, there was a general sense that traditional TCM practice was diminishing in importance and facing challenges from Western medicine, even in the TCM hospitals. The ratio of Western to TCM hospitals in Shanghai is seven to one. By law, all hospitals are required to have a TCM department. Long Hua is an integrated hospital, such that TCM and Western medical approaches to diagnostics and treatment are combined. The TCM clinicians at Long Hua confirmed that TCM is losing ground to Western medicine. In the paediatric department, for example, the majority of the outpatient consultations we observed included a Western medical prescription. For example, the inpatients, many of whom had respiratory tract infections and/or asthma exacerbations, received antibiotics and/or corticosteroid medication. The clinicians reported that traditional Chinese medicine departments were being reduced in terms of beds and staff, and that entire departments in other TCM hospitals had merged or been significantly reduced in size. Financial reasons were identified to explain this trend. In brief, the hospitals are financed through governmental funding as well as patient fees for procedures, medicines and consultation fees. Although basic equipment is free, the more advanced and consequently, more Western equipment is more expensive. All hospitals, including TCM and Western, increase their earnings by focusing on Western medical therapeutics. As such, the general trend in TCM hospitals is toward greater utilisation of Western medical treatment and procedures. An anonymous informant stated, “One surgeon can earn for the hospital as much as a hundred TCM doctors. TCM herbal medicines is cheap little money.”

‘It is a vicious cycle’. The slipping confidence

Several of the TCM practitioners reported changing attitudes towards healthcare, with increasingly assertive parents and a more critical, consumer-driven approach towards treatment and healthcare professionals, especially for treatment of children. The TCM practitioners were reluctant to advise against parents’ wishes and they were similarly reluctant to do anything that could elicit unpleasant reactions in the child, like crying. Acupuncture was generally considered painful, and in turn, TCM clinicians were increasingly hesitant to recommend acupuncture for treatment in children. This trend perpetuated a cycle which limited experience, limited expertise and lowered confidence. One TCM practitioner reported, “It has also two sides, because at first the parents are reluctant to go, but doctors they need experience. Since they have less and less patients,
they don’t have that expertise. They don’t have confidence. It is a vicious cycle.”

‘I have one cerebral palsy patient, 6 months old, who tolerates needles with no pains’—the divergent point of view

The formal focus group interview differed in many aspects from the individual, in-depth interviews and the informal conversations with clinicians. Specifically, opinions voiced during the focus group differed with regard to tradition, experience, indications for acupuncture and the question of whether acupuncture was a painful intervention. Among two of the three hospital officials, there was a general reluctance to limit the possibility of acupuncture as a universal treatment option. This viewpoint contrasted with TCM practitioners, among whom there was broad agreement regarding the limitations of needle acupuncture in infants and small children.

DISCUSSION
Main findings

This qualitative study explored the use and recommendations for needle acupuncture on infants and small children by TCM practitioners in Shanghai. During 3 days of participant observations at the Longhua hospital outpatient acupuncture clinic, which receives on average 400 patients per day, we observed only two children and no infants. During formal interviews and informal conversations with TCM practitioners representing different specialties, the prevailing opinion was that needle acupuncture for infants and small children is considered painful, is inconvenient and is not indicated for routine clinical practice, except for certain neurological diseases. We learned that TCM methods for children have traditionally focused on herbal medical treatments, and to a lesser extent, on Tui na. Acupuncture had historically been used for certain acute illnesses, which are now treated in Western hospitals. The majority of acupuncturists, and all of the non-acupuncture TCM practitioners, had limited or no experience treating children or infants with needle acupuncture and would not recommend it as a therapeutic option. The exceptions were the opinions of two TCM acupuncture officials. The reason for the non-use of acupuncture in small children seemed to revolve around two main themes (1) internal TCM traditions and practices, which do not support the routine practice of needle acupuncture in infants and toddlers and (2) external system changes which increasingly limit the use of acupuncture specifically and TCM practices in general.

Internal TCM tradition and practice

The most important reasons for the lack of acupuncture in infants and small children attribute to TCM traditions and practices. Specifically, herbal medical treatments or Tui na are widely available and universally recommended as a first-choice TCM treatment for paediatric populations.

External system changes

These were opinioned in informal conversations throughout the study period, and by several of the informants, and are not taken from official sources or literature. The topic might be a sensitive issue. System changes within the general healthcare system in Shanghai over the past years are also relevant in explaining the lack of paediatric acupuncture. TCM departments and practices, including acupuncture, are losing ground to Western medicine, with widespread consequences. Even original TCM hospitals and institutions are losing credentials and resources to Western diagnostics and treatment, which offers greater financial compensation than traditional TCM practices. This is also true for paediatrics.

Strength and weaknesses

Strengths of the study include the uniformity of the participants and the information collected from the TCM clinicians, the variety of sources of information, the triangulation of methods and the validation process. The discrepant voices were two of the officials who had both clinical and administrative responsibilities, who regarded acupuncture as a universal method of treatment. Limitations of the study involved the challenges of conducting qualitative research in situations requiring interpreters, or in which both the researchers and informants have English as their second language. We relied entirely on our Shanghai contacts to gain access and organise interviews with TCM practitioners and officials, and as such, the selection of informants may be biased. It is worth noting that the situation in Shanghai might differ substantially from TCM in other cities and provinces in China. However, the centralised and uniform organisation of education and the practice of TCM in Shanghai and the People’s Republic of China, arguably strengthens external validity. An important consideration is the first author’s prejudices and standard of reflexivity in the information gathering and analysing process. He is a medical doctor, acupuncturist, with 25 years of clinical experience and 15 years of teaching acupuncture and general principles in TCM. His main field of interest is paediatric acupuncture. He has in recently published a blinding-validated multicenter randomised controlled study on the effect of acupuncture on infantile colic that showed no clinically relevant effect of the intervention. Qualitative research designs run the risk of biased and selective reporting. The use of a diversity of informants, several information gathering methods and a detailed description of the methodological aspects of the process might, however, counteract such bias.

Implications for practice

This study adds to the ongoing discussion on the evidence, utility and limitations of acupuncture in children, as well as in medicine in general. Contrary to our observations in Shanghai, acupuncturists in the Western
world encourage acupuncture for children, including infants. One rationale underlying this trend is that acupuncture is widely considered in the West to be an integral part of TCM tradition, for patients of all ages. According to our study, Western beliefs that acupuncture is routinely indicated and recommended for infants and small children within TCM are unfounded. Such beliefs may appear to be a Western interpretation, and are not based on actual Chinese modern-day practices or therapeutic recommendations. Recent controlled trials investigating the efficacy of acupuncture treatment in the infant population yield conflicting results, and little clear empirical evidence is available to support the use of acupuncture. As such, it appears that the rise of infant acupuncture in the West could be partly attributable to Western acupuncturists’ own clinical observations and theoretical beliefs. In line with the principles of biomedical ethics, however, this study underscores that this rationale is insufficient to recommend a potentially painful treatment for infants and small children who lack the competence to provide informed consent.

CONCLUSION
Acupuncture for infants and small children is neither routinely practiced nor recommended by TCM clinicians working in Shanghai. It is generally considered a potentially painful therapeutic method. Alternative TCM treatments are widely available and preferred by TCM practitioners in Shanghai, including herbal treatments and medical massage (Tui na). Needle acupuncture on broad indications in infants appears to be a Western practice, with little basis in Chinese TCM contemporary practice.

Acknowledgements The authors would like to thank the clinicians at Longhua hospital and the teachers and translators at Shanghai University of Traditional Chinese Medicine for their invaluable cooperation and support.

Contributors HS and MB developed the original idea, developed the flexible study process and wrote the initial protocol. HS carried out the field study in Shanghai, transcribed the interviews and field notes and carried out the initial analysing. MB reanalysed the field notes and interviews. HS wrote the initial draft of the article. MB reviewed and structured the article. HS is the guarantor of the study.

Funding The study was financed through a grant from the Norwegian Research Fund for General Practice.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES
'Big needles, small bodies'—the absence of acupuncture treatment for infants in contemporary Shanghai: a qualitative study

Holgeir Skjeie and Mette Brekke

BMJ Open 2015 5:
doi: 10.1136/bmjopen-2015-009486

Updated information and services can be found at:
http://bmjopen.bmj.com/content/5/11/e009486

References
This article cites 31 articles, 7 of which you can access for free at:
http://bmjopen.bmj.com/content/5/11/e009486#BIBL

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

Complementary medicine (87)
Paediatrics (344)
Qualitative research (367)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/