Author's response to reviews

Title: How large are the non-specific effects of acupuncture? A meta-analysis of randomized controlled trials

Authors:

  Klaus Linde (klaus.linde@lrz.tum.de)
  Karin Niemann (karin.niemann@gmx.net)
  Antonius Schneider (antoniusschneider@lrz.tum.de)
  Karin Meissner (karin.meissner@med.uni-muenchen.de)

Version: 2 Date: 23 September 2010

Author's response to reviews: see over
Dear editors,

Enclosed we submit the revised version of our manuscript

**Non-specific, specific and total effects of acupuncture - a meta-analysis of randomized controlled trials**

now titled

**How large are the non-specific effects of acupuncture? A meta-analysis of randomized controlled trials**

Please find the details regarding our revision below.

All authors have seen and approved the revised manuscript.

Sincerely,

Karin Meissner & Klaus Linde

**Responses to the comments of the reviewers**

**General response to all reviewers**

The primary aim of our analysis was to estimate the size of non-specific effects associated with (sham) acupuncture interventions. This was stated explicitly at the end of the introduction and the discussion mainly addressed the problems related to this issue. However, it seems that our title and the detailed presentation of the secondary results on specific and total effects obscured our primary goal. We apologize that we were not clear enough and probably over-stated our secondary findings. Therefore, we have chosen a new title and revised the manuscript considerably focussing more clearly on non-specific effects. We are aware that a focus on non-specific effects is unusual but we believe that the basic hypothesis that not all sham intervention seem to be the same is a quite fundamental issue for the evaluation of specific effects for all non-pharmacological interventions. We hope that the aim of the manuscript is now clearer.
We tried to address comments of the reviewers as far as possible, however, in our view most comments focussed on methodological problems mainly related to the estimation of specific effects and total effects. Therefore, our changes in the manuscript and our responses to the comments might not always directly take up all recommendations. We hope that are changes are satisfactory.

We did another change in the manuscript which was not requested: we substituted fixed effects analyses by a subgroup analysis comparing larger (> 100 patients per comparison) and smaller trials. This change was based on a new key publication (and subsequent communication with its' first author) published in the BMJ after our initial submission (Nuesch et al. BMJ 2010;341:c3515).

Responses to the single comments are listed below.

Responses to reviewer Brent Bauer

Comment (C): Because the thrust of the paper hinges on the concepts of "specific" and "non-specific" effects, clarity would be enhanced by a more careful explanation of these terms. One or two examples would be particularly helpful early on to help the reader place the findings in context.
Response (R): We have expanded the introduction and added examples.

C: I would like to see the dates of the reviewed studies included in the abstract.
R: The date of the last search has been added.

C: Page 4 - states "all available randomized trials" were reviewed. Does this include Chinese language articles as well?
R: We did not have a priori language exclusion criteria but we did not systematically search Chinese databases. We have deleted the “all” in the introduction.

C: In the first paragraph of the Discussion, the authors state "...size of these effects is likely to be overestimated due to bias". Why? This question is touched upon later in the discussion but leading the discussion with a statement such as this requires immediate justification - or else consider moving the hypothesis regarding bias closer to the discussion of the funnel plot results, etc.
R: We have deleted this phrase from the discussion.

Responses to reviewer Charlie Xue

Comment (C): 1. There are more than "point location and skin penetration" in determining the appropriateness of acupuncture, it would be more beneficial to the broader audience (such as traditional acupuncturists and medical practitioners) to benefit from this review if the included studies were also assessed with respect to the quality by using more comprehensive tools that have been developed and widely used in assessing quality of the studies involving acupuncture.
Response (R): We agree that there is clearly more than point location and skin penetration in determining the appropriateness of acupuncture. We have re-written this part in the introduction. Given our (hopefully now more clear) focus on non-specific effects we do not think that an assessment of the appropriateness of acupuncture is of major relevance for our paper. Furthermore, we do not think that there are validated widely used tools for that purpose. One of us (KL) has been involved in several attempts for such tools, but the results were not convincing at all. While there is some agreement on some very basic issues acupuncturists often disagree on important details. In our tables we try to give at least basic information on treatment details. We hope that for the purpose of assessing non-specific effects this is acceptable as sufficient.
C: 2. With regards to the conclusions, as the majority of the quality studies were in pain management, it may be more appropriate to have this review to focus on pain only as the applicability of these findings to other conditions seem to have little relevance due to the rather narrow range of trials that were included in this study. The conclusions are too general and as besides pain, the base for such conclusion is rather limited.

R: We agree that regarding specific and total effects this comment is justified. We believe that the comment is less relevant to the issue of non-specific effects. The discussion has been revised, however, we do not focus on the pain issue a lot.

Minor Essential Revisions
C: 1. Page 2 Abstract: Methods line 3, do you mean extracted instead of "abstracted"?
R: corrected

C: 2. Page 4, Methods line 3, should be from instead of "form"
R: In attempt to shorten the manuscript (as the introduction was expanded) the paragraph was deleted.

C: 3. Page 6: 2nd last line those >0.7 instead of <0.7
R: Corrected

C: 4. Page 9, under Meta-analysis non specific effects: line 1: For the comparison of sham acupuncture with NO acupuncture, no was missing
R: corrected

Responses to reviewer Girdhar Agarwal
Comment (C): This manuscript deals with the systematic review of randomized controlled trials (RCTs). The authors have followed most of the methodological guidelines for the conduct of the systematic reviews. Although, they have not addressed one of the major issue, namely “control of bias” in this review.
Response (R): Our review has addressed major bias issues in detail. We only included RCTs, assessed adequacy of concealment, drop-out rates, and whether a primary outcome measure was pre-defined. Furthermore, we assessed small study bias. To be more clear we re-named what we called lower and higher quality into higher and lower risk of bias. (see also our response to Reviewer Jaime Peters, comment 5)

C: 1. Definition of groups. I am unclear about “Sham Intervention” group (point 4, p. 4). Where was insertion of needle done for this group? What points were chosen? Are these arbitrary? If yes, will not be harmful for the subjects?
R: We now describe the basic principles of sham acupuncture interventions in the introduction. In the table summarizing the included studies the column “intervention details” indicates what type of sham intervention has been used (with or without skin penetration, stimulation at non-indicated or outside of acupuncture points, skin penetration yes/no). It is difficult to go into further details in our paper. In many trials needles are inserted superficially at non-acupuncture area, but specifics vary considerably (For a review see reference 16 in our manuscript). Acupuncture needles are very thin, so risks are very minor, but obviously the procedure (whether acupuncture or sham acupuncture) is not completely risk free. As in any trials participants are informed about potential risks.
C: 2. Measuring effect size. Participants could be either completely untreated or receive treatment and were allowed “rescue medication” (p. 4). This will affect the outcome measure. How the effect of intervention (acupuncture) will be identified from that of “rescue medication”?
R: Trials were included only if all groups (acupuncture, sham, no acupuncture) received the same basic treatment. It is impossible and ethically unacceptable in many conditions to keep trial participants completely untreated. Obviously, it can happen that, for example, patients in the no treatment/no acupuncture group used more rescue medication. This could lead to an under-estimation of group differences. This problem is addressed in the discussion section. Furthermore, we did a sensitivity analysis for trials with more and less intense co-interventions.

C: 3. Study Selection (p.5) The search strategy and data extraction should be more explicitly specified.
R: The appendix gives the detailed search strategies for Medline and Embase. This is in accordance with PRISMA recommendations. Our description of the data extraction is similar to that in most other systematic review. We are uncertain what more should be written.

C: 4. Methodological quality assessment How was this done so that the studies with high & low quality can be identified?
R: For general issues see our response to the general comment above. For identifying studies: The methods section explicitly states when a trial is considered low quality (now high risk of bias) and when high quality (now low risk of bias). In the table summarizing the included studies it is easily possible to see which studies meet the criteria by looking on the columns giving the percentage of drop-outs and concealment of allocation. (see also our response to Reviewer Jaime Peters, comment 5).

C: 5. Heterogeneity This is the most difficult task to handle. Personally, I think it is better to draw conclusions in subgroups than combining the results over large heterogeneous groups. This is more important here as the effect size measure is not same across all the studies. Also different outcomes are measured in different studies. How is “standardized mean difference” defined? Is it some kind of coefficient of variation?
R: We believe that by focusing more clearly on non-specific effects the problem of heterogeneity is at least a bit smaller now. Still, in the abstract and the results section we now first report the findings for subgroups. However, subgrouping the studies does not resolve the problem. Apart from short term studies \( (I^2 = 0\%)\), heterogeneity is moderate (47%) among the chronic pain and substantial (67%) among the other studies. For specific and total effects it is substantial in all subgroups. Even if we go to the level of single conditions heterogeneity is often substantial. Reasons for heterogeneity seem to be very complex in our case...
We now explain SMD (mean difference/divided by pooled standard deviation) in the methods section.

C: 6. Meta-analysis specific effects (p.9) Heterogeneity coefficient \( (I^2 = 0.83)\) was very high. In this light, the pooled effect size will have no clear meaning. The funnel plot was also highly asymmetrical (fig. 3 B)
R: We agree, in principle. However, heterogeneity for specific effects is not reduced by subgrouping studies according to condition indicating that the main reasons for heterogeneity probably are somewhere else.

7. Meta-analysis total effects. Here also \( I^2 = 0.80 \) is very high. again the pooled effect size will have no meaningful interpretation.
R: see response to C5.
8. Meta-regression (p. 10). Actually meta-regression is also one method of addressing the heterogeneity. So the same comments apply as in 5.
R: See response to C5.

9. In view of comments 5-8, I would suggest pooling the results over subgroups, which are nearly homogenous, rather than over large heterogeneous group.
R: see response to C5

Responses to reviewer Jaime Peters

Comment (C) 1. I have some concerns that for half of the included RCTs (17 out of 32) there was no clearly defined main outcome measure, and so the outcome measures chosen by the authors for inclusion into the analyses were done so where “two reviewers independently chose the outcome considered most important”. The sensitivity analyses show that there are some differences between the pooled estimate for those having a pre-defined outcome and those not (even though the p-value suggests this is not statistically significant). I believe this is an important potential source of bias in the methods used by the authors, and at the very least this should be acknowledged. Furthermore, the authors have not reported any information on the number of disagreements and this would likely be useful for the reader to know, in addition to the author’s beliefs on the importance of this as a potential source of bias in their study.

R: We now included a statement reporting that we had two disagreements when selecting outcomes. Our sensitivity analysis (see former table 2, now appendix table 3) includes one analysis using other outcomes. The results are very similar. For our main comparison (sham vs. no treatment) trials with a predefined main outcome measure even show (albeit non-significant) larger differences (as for other indicators of better quality). So far we did not address the main outcome issue in more detail in our revision but we hope that the reviewer agrees with us that with our focus on non-specific effects this does not seem a major concern.

C 2. The authors reviewed and included 3-arm trials only: sham vs no treatment vs acupuncture. Further data may be available from 2-arm trials, say for sham vs acupuncture. There should be some discussion as to why these other data were not included, as they could help to inform the question of interest to the authors. Inclusion of these additional data could be undertaken in an MTC and it would be useful for readers to understand why this wasn’t done by the authors, or whether they considered this approach at all? In fact, a neater analysis would be to undertake a MTC of the 3-arm trials, rather than reporting 3 pairwise comparisons, although undertaking meta-regression analyses would be more complicated and so I can see why such an approach would not have been pursued by the authors.

R: We would love to do such a network review (and we tried) but this was impossible with the resources we had. We hope to do this in the near future (in case we are able to get funds).

3. As the authors themselves note, there is a great deal of heterogeneity between study estimates meta-analysed in this paper. This is not unexpected given the heterogeneity in outcomes, clinical categories, quality, intensity of intervention etc. However, there needs to be more emphasis of the heterogeneity with respect to the pooled estimates the authors report. For instance, except for reporting the I2 value, there is no mention of heterogeneity in the abstract.
Furthermore, under the section “Heterogeneity as main limitation” in the Discussion there is only mention of the fact that heterogeneity was expected and that subgroup analyses were undertaken. There is no discussion of the heterogeneity in the results
and how this may impact on the interpretation of these results, i.e. the fact that heterogeneity is the main limitation.

R: Again we hope that the shift in the presentation of our manuscript explains why we did (do) not go into further details regarding the discussion of heterogeneity. Furthermore, the second phrase in our results read (and still reads): “The included studies varied strongly regarding patients, interventions, outcome measures, methodological quality and effect sizes reported.” While this does not include the word (statistical) heterogeneity we believe that this is a clear indication that our studies were heterogeneous on several levels.

4. A large number of subgroup and sensitivity analyses have been undertaken. There is therefore concern for issues of multiple testing, and statistically significant p-values being found solely by chance as so many comparisons have been undertaken. The authors should make it explicit to readers that these subgroup/sensitivity analyses are exploratory/hypothesis-generating rather than them being seen as providing answers, particularly as for some subgroup analyses subgroups consist of <5 studies. Furthermore, there are some results that may not necessarily be as expected (e.g. Trials in which the sham intervention involved skin penetration yielded significantly smaller effects over no acupuncture groups than trials which used nonpenetrating sham techniques from Subgroup and sensitivity analyses), yet not much weight can be put on these analyses because so many have been undertaken and there is so much heterogeneity.

R: The methods sections includes a statement (“For exploratory analyses we defined further subgroups”) stating explicitly that subgroup analyses were exploratory. We included a phrase that sensitivity analyses were performed to check the robustness of results. We deleted the separate section on subgroup and sensitivity analyses from the results, put the table with these analyses into the appendix, and only mentioned some subgroup findings in the remaining results sections (repeating the these analyses are exploratory).

5. It is interesting that quality was based on just two features: randomization concealment and drop-out. A number of quality checklists are available in the literature and so I wonder why the authors did not use one of these. Also, what is the reasoning for 15% drop-out? Why not 10% or 20%?

R: Indeed in the beginning we planned to use the Cochrane risk of bias tool. However, for our main comparison (sham vs. no treatment/no acupuncture) blinding is not discriminative as all comparisons are unblinded. We then decided to follow the simple approach used by Madsen et al. (ref. 23) in their review on pain trials. We now explicitly indicate this.

6. Small study bias is investigated and reported by the authors. As with the great deal of heterogeneity identified, this could be emphasised more in the paper. E.g. noting in the abstract that the acupuncture vs no acupuncture also had evidence of small study effects.

R: For our main comparison there is no indication of small study bias (if anything a trend to larger non-specific effects in bigger studies). While we shortened the text on specific and total effects we tried to be a bit more explicit about biases associated with sample size and risk of bias.

7. Second paragraph in “Data synthesis and analysis”: SMDs # -0.4 were considered small effects, those between -0.41 and -0.7 moderate and those < -0.7 large effects [20]. Should this be: SMDs # -0.4 were considered small effects, those between -0.41 and -0.7 moderate and those > -0.7 large effects [20]?

R: corrected
R: correct reference (Deeks et al.) now given

9. Second paragraph in “Data synthesis and analysis”: State that exploratory analyses with and without skin penetration are for the sham group.
R: done

10. In “Description of included studies”: Be explicit that the main results are those based on the 32 RCTs where a continuous outcome could be obtained.
R: The phrase “The main analyses are based on the 32 trials reporting data on a continuous outcome” has been added at the beginning of the “results – meta-analysis” section.

11. First line of “Meta-analysis non-specific effects”: should be sham acupuncture with no acupuncture.
R: corrected

12. Within each paragraph for each of the three comparisons (sham vs no, acupuncture vs sham and acupuncture vs no), the sentences describing the number of SMDs as being above 0.7, between 0.4 and 0.7 and smaller than 0.4 should actually refer to above -0.7, between -0.4 and -0.7 and smaller than -0.4.
R: corrected

13. Second sentence of “Meta-analysis specific effects”: Replace “There were no significant (p=0.71)” with “There were no statistically significant (p=0.71)”.
R: done

14. In “Meta-analysis specific effects” point out that the small study bias explains why fixed and random effects results so different.
R: see second part of the response to all reviewers above

15. Why is appendix figure 2 not in the main report with the same plot for the two previous comparisons?
R: the reason was to shorten the manuscript. Given the new focus also the figure on specific effects is put into the appendix.