Reviewer’s report

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

Version: 1 Date: 31 August 2010

Reviewer: Jaime Peters

Reviewer’s report:

Minor essential revisions

General comments

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.

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.

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.

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.

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%?

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.

Specific points
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]?


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

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.

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

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.
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)”.

14. In “Meta-analysis specific effects” point out that the small study bias explains why fixed and random effects results so different.

15. Why is appendix figure 2 not in the main report with the same plot for the two previous comparisons?

Which journal?: Appropriate or potentially appropriate for BMC Medicine: an article of importance in its field

What next?: Accept for publication in BMC Medicine after minor essential revisions

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests