Author's response to reviews

Title: Acupuncture for osteoarthritis of the knee: A pilot study for an open parallel-arm randomised controlled trial

Authors:

Harriet Lansdown (harrietlansdown@gmail.com)
Katie Howard (katiedrawoh@hotmail.co.uk)
Stephen Brealey (sb143@york.ac.uk)
Hugh MacPherson (hm18@york.ac.uk)

Version: 2 Date: 25 June 2009

Author's response to reviews: see over
RESPONSE TO REVIEWERS.

Reviewer One.

1. We have renamed the manuscript with “pain and osteoarthritis of the knee”. This change is also reflected in the first line of the abstract.

2. In the Introduction, we have expanded the focus of the paper to “Knee pain and osteoarthritis (OA) of the knee”

3. “and less physical activity” has been deleted.

4. “deteriorating” has been deleted

5. “high quality” has been moved to be prior to “meta-analyses”

6. We have changed the wording from “comparing acupuncture to usual care” to “comparing acupuncture to usual care alone”. We hope our design is also clear from the start of our Methods section where we state: “We conducted a pragmatic parallel two-armed randomized control trial (RCT) comparing ‘acupuncture and usual GP care’ to ‘usual GP care’ alone.”

7. The reviewer states that there is “no long term benefit from acupuncture”, however we believe the evidence is inconclusive, so we have replaced “insufficient evidence” with “inconclusive evidence”.

8. Sentence now deleted.

9. Agree, and have added, “appropriate validated outcome measures” as an outcome

10. There is a conflict here with another of the reviewers who wants “GP” inserted everywhere. On balance we agree with this reviewer, and have deleted “GP” as a rider to “usual care” throughout.

11. We disagree with the reviewer and quote the mean differences on meta-analyses of pooled data below for the three reviews that we referenced, all showing confidence intervals not crossing zero.

<table>
<thead>
<tr>
<th>Study</th>
<th>Timeframe</th>
<th>Outcome</th>
<th>Effect Size</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kwon et al 2006</td>
<td>Short term (pain)</td>
<td>SMD = 0.24</td>
<td>95% CI 0.01 to 0.47</td>
<td></td>
</tr>
<tr>
<td>Manheimer et al 2007</td>
<td>Short term (pain)</td>
<td>SMD = 0.35 95% CI 0.15 to 0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long term (pain)</td>
<td>SMD = 0.13 95% CI 0.01 to 0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White et al 2007</td>
<td>Short term (pain)</td>
<td>WMD = 1.54, 95% CI 0.49 to 2.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long term (pain)</td>
<td>WMD = 0.54, 95% CI 0.05 to 1.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12. Our selection was based on self-reports of knee pain. One finding from this study was that we need to screen to ensure a minimum level of severity at entry to the trial, a point now raised in the Discussion.

13. We have added that the outcomes were collected using “patient self-reported outcome measure data” and “using the Likert-based” measures.

14. We have experience of collecting adequate self-reported resource use data covering a period of three months, however accept that the next nine months (from 3 to 12 post
baseline) will require three monthly questionnaires, a recommendation that has been added to the Discussion.

15. We accept that the eagerness of respondents will be helpful for adherence and response rate, a reason why we were happy to recruit these patients. We have added to the discussion. “We also note that by selecting the first patients to respond to our invitations, we recruited the more eager volunteers, with possibly higher adherence and follow-up rates than may be present in the primary care population.”

16. We have added that the questionnaire data was collected by post, driven by computerised mailings. This means the Hawthorne effect will be minimised regarding self-reports.

17. Only the baseline scores were used when evaluating outcomes in the regression model. This has been added to the statistics section of the Methods.

18. Table 3 and 4 now combined.

19. We would like to retain both WOMAC stiffness and WOMAC Global, as the study was not powered to find an effect, more we wanted to explore outcomes and seek to identify appropriate measures for a fully powered trial.

20. We have identified as a weakness our poor follow efforts, and in the Discussion have stressed the need for better follow-up in a large scale trial.

21. We agree that there will be a weakening of effect if there is unplanned cross-over by individual patients, but this reflects the pragmatic design of the trial and what might happen in routine clinical practice. As we allowed for this to occur in the pilot, no additional factoring in is required in the sample size calculation.

22. We accept there are different ways to establish a sample size, however our advice has been to retain the current sample size analysis, and suggest that a fully powered study could present the results as defined by treatment responders in order that the analysis would be more meaningful for the clinicians.

23. Very few trials are powered on the basis of health economics, and when they are, the size of the trial tends to be unfeasibly large. We are basing our sample size calculation on current practice and not powering the sample size on the EQ-5D.

24. We accept that the trial design does not answer the question as to whether there is a specific effect of acupuncture. We regard the question about overall benefit as more relevant to patients, providers and policy makers, a point we have made in our discussion. We have added, “such as expectations of acupuncture” as an example of a non-specific effect.

**Reviewer Two**

1. We agree that the primary outcomes from this pilot are to establish the potential recruitment rate, attendance levels for acupuncture treatment, loss to follow up and the sample size for a full-scale trial. We have added establishing appropriate validated outcome measures as another outcome, which we have added throughout. As advised we have refocused and reordered to stress the key findings are these process-oriented ones. Resulting changes have occurred in The Abstract, Introduction, Methods and Results and Discussion.

2. We have contradictory advice here, in that Reviewer One has advised us to remove the “GP” from “usual GP care”, whereas Reviewer Two advises its retention. On balance we
have adopted reviewer One’s point that the full range of usual care interventions go beyond those that are GP initiated, and might for example include self-referrals for massage or over-the-counter medications. We agree that capturing the “usual care” in the full scale trial will be important, and have added a point to the Discussion about self-reports on usual care being collected at three month intervals in the full scale trial.

3. We have revised the discussion regarding whether there are longer-term benefits.

Minor improvements:

a. We have revised the Abstract adding with focus on the (now) five “objectives”

b. We have added a sentence about core treatments (advice, etc), to the Discussion.

c. We have added to our existing comments on the need for better follow-up, with the edited sentence: “For a full scale trial, it is recommended that a more systematic approach to collecting patient outcome data be put in place, including follow-up reminders and possibly financial incentives in the form of a voucher or payment to patients.”

d. Corrected this reference and expanded on the protocol for treatment.

e. We have added this point about top-up sessions within the discussion.

f. Now clarified that it is the WOMAC global variance used in the sample size calculation.

g. We accept that a large cohort might be better than a trial if one is seeking representative data. However we have suggested collecting qualitative data, which could be nested within a future trial.

h. The two graphs currently have the full range of WOMAC scale points, See Figure 2, range 0 to 9, and Figure 3, range 0 to 40.

i. See point 2. above where we have addressed the labelling of usual care vs. usual GP care.

j. We have merged Tables 2 and 3 as suggested.

**Reviewer Three**

1. There are conflicting opinions about the importance of the clinical outcomes in this study, which was not designed with the power to detect differences. We prefer to follow the advice of Reviewer Two, and de-emphasise the clinical outcomes, and therefore will leave the effect size out of the Abstract.

2. To clarify, we have added “A regression model, with baseline outcome measure as the covariate, …” to the Methods, removing reference to analysis of co-variance.

3. There are differences between SF-36 sub-scales scores, however as discussed above, this study was not powered to detect differences and we believe further details of statistical differences here would detract from the purpose of the study. Given that there was no mention of the outcomes for SF-36, we have edited the last sentence on Outcomes to read: “No significant differences were observed on the OKS scale or the domains of the SF-36.”
4. Again, with regard to the suggestion regarding effect sizes, we would like to reinforce the point made by Reviewer Two that the objectives of this study were process-oriented rather than outcome-oriented.

5. We did not use clinical change that we found in our study as a basis for the sample size calculation, as it was based on too small a sample, and a different sample may well have given us a different clinical difference. We believe it is the clinically important difference that should inform the sample size calculation, not a clinical difference found in our small-scale study. Pilot studies are needed to determine the standard deviation (or variance) and it is these data, along with the minimally important clinical difference, that are required for estimating the sample size. All data needed to calculate the sample size are provided within the paper.

6. The rationale for using WOMAC global scale for the sample size calculation has been added to the Results section.

7. In the discussion where we compare results with other studies we have included reference to the Witt et al 2006 paper.

8. The Reinhold reference has been added in the Introduction.

9. We have included name of software (STATA) used for the randomisation.

10. We again wish to follow the recommendation of Reviewer Two, who does not want this paper to focus on the outcomes, rather the process. So we have not presented measures of variance etc. The formatting can be changed if required by Biomed Central.

11. The formatting can be changed if required by Biomed Central.

12. We have Tables 3 and 4, now merged, has characteristics relabelled as outcomes. The measure of variance is given as standard deviation.