Reviewer's report

Title: Exploring integrative medicine for back and neck pain - a pragmatic randomised clinical pilot trial

Version: 1 Date: 28 February 2009

Reviewer: Karen Sherman

Reviewer's report:

This is a pragmatic trial of integrative medicine versus usual care for spine pain. The authors found that both groups improved, which is not surprising for a study of spine pain, but they found no differences between groups. Although the results appear straightforward, I believe the authors try to over interpret their data by making too much of the within group differences when one of the groups shows a significant change and the other does not. I do not believe such comparisons are valid scientifically.

• Major Compulsory Revisions

1. Please revise the statistical analysis section so that it is written more concisely and is easier to understand. I think the material on scanning and double data entry can be omitted. I don’t understand the sentence: “The data structure with repeated measurements within patients was accounted for in the analysis? Please revise it. I don’t understand the sentence on the power calculations. What was the effect size on which the study was powered?

2. In the results section, mean differences within (or between groups) of less than 10 or so would not be clinically important, so some of the changes you described would not be of interest. I notice that in general, the percentages are not very different at follow-up for the two groups. You also have substantially more people in the IM group for data analysis (n=36) than in the CT group (n=27). Thus, I am not at all surprised that you find some within group differences that are significant in the IM group, but not in the CT group. However, you never discuss that as an explanation for why you get significant differences in one group versus another. This problem is exacerbated when you limit analyses to the chronic group.

3. Where are the 12 week data? Did you analyze them at all? While I am not suggesting a table with them, they should be mentioned if you analyzed them.

4. Under “additional findings”, the second sentence “Sensitivity testing with last-observations carried forward …” is hard for me to understand. Please reword it.

4. The discussion section is much too long and needs focus. For example, you don’t need to talk about the rational for the study (you should have done so in the introduction), the lack of a placebo control and other issues that are generic for pragmatic trials.
The second paragraph on page 13 is hard for me to understand. Your data do show a greater fraction of persons at follow-up in both groups trying CM than at baseline for the chronic groups and a greater fraction of the Conventional group using CT — although the numerator is similar at both timepoints. Please delete (or revise this paragraph).

The paragraph on sample sizes could be shortened and worded more clearly. Are you implying that the study was underpowered because you anticipated larger effects than you saw? Populations of primary care patients with spine pain traditionally exhibit lots of variability, so that should be built into estimates. Moreover, you can power a study for clinically important differences, if you know what those are. I would delete the paragraph on trying to get more homogenous groups because I don’t think this is a useful strategy.

The paragraph on blinding could be deleted and the paragraph on “washout effects”. Those could be replaced with a short and focused paragraph on the strengths and limitations of the study.

5. In the conclusions section, I would take out the sentence about the within group changes because I think it over interprets the data.

• Minor Essential Revisions

1. the word “data” is plural and the authors need to change that throughout the text.
2. In the abstract – I would remove the comments pertinent to within group differences.
3. On page 5, 2nd paragraph – am I correct that in order to be selected, a prospective study participant needed to have visited a general practitioner at least 6 times in two weeks? (3 visits per week for 2 weeks each)? If so, this seems like an excessive requirement as I am unclear what the GP would do with them for so many visits. If not, please correct the error.
4. What was the primary outcome measure? Were any of the measures besides the SF-36 validated?

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Not suitable for publication unless extensively edited

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

Declaration of competing interests:

I have no competing interests.