Author's response to reviews

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

Authors:

Tobias Sundberg (tobias.sundberg@ki.se)
Max Petzold (max.petzold@nhv.se)
Per Wandell (per.wandell@ki.se)
Anna Ryden (anna.ryden@medicine.gu.se)
Torkel Falkenberg (torkel.falkenberg@ki.se)

Version: 2 Date: 28 May 2009

Author's response to reviews: see over
Dear Iratxe Puebla,

Thank you for the opportunity to revise and clarify our manuscript entitled “Exploring integrative medicine for back and neck pain – a pragmatic randomized clinical pilot trial” [MS: 7919351342550535].

Please find our replies below in a point-by-point format addressing each of the four reviewers’ concerns and the editorial comments. All comments and replies have been numbered for easier reference.

Analysing the four referees’ reports we found that there were somewhat different suggestions on how to revise and improve the manuscript and the dissemination of our findings. We feel that the overall line of reconsiderations proposed by Dr McDonough was very much in agreement with our intentions with this paper. We have hence tried to revise the manuscript with this in mind at the same time addressing the many valuable comments given by Drs Endres, Sherman and Poitras.

We hope that the manuscript is now more concise and clear and look forward to your response. Thank you.

Sincerely yours,

Tobias Sundberg
Max Petzold
Per Wändell
Anna Rydén
Torkel Falkenberg

---

**Editor’s comments:** Iratxe Puebla

1) Consent- Please document in the Methods section of the revised version of the manuscript the type of consent sought from the participants.

   ➔ Reply: All participating patients gave oral and written informed consent. We have revised the methods section accordingly, please see pages 5-6.

2) Trial registration? I am unable to find the record for the NTC00565942 number under the clinicaltrials.gov registry, can I ask you to please confirm that this is the correct number, and provide a link to the record
for the trial under that database?
→ Reply: The correct trial registration number is NCT00565942, and the link to the record for the trial in the online database can be found here: http://clinicaltrials.gov/ct2/show/NCT00565942. The trial registration number can be found at the end of the abstract; please see page 3.

Reviewer: Suzanne McDonough

Reviewer's report:

Major Compulsory Revisions

1) Thank you for asking me to review this interesting paper which appears to be a follow on from the piece of work published in 2007 by this group (Sundberg et al, 2007). The 2007 paper described the development and implementation of an integrative medicine model in primary care. I would have expected that the next stage of answering this research question would be to test the feasibility of using the model compared to current care, however the aim of the current paper, identified on page 4 of the introduction, was instead ‘to explore the general effectiveness of the developed IM model compared to conventional primary care management’. I think this aim should be revised and that the authors should present this as a pilot study which tests the feasibility of an IM programme compared to conventional care.

→ Reply: Thank you. We agree and the aim has been revised accordingly, please see the last paragraph on page 4: “…the aim of this pilot study was to explore the feasibility of conducting a randomised clinical trial to compare the effectiveness of the developed IM model with conventional primary care management…”

2) Specific objectives might then be 1. define recruitment and retention rates to the trial for both arms. 2. collect data in order to calculate the power for a main trial (and move the emphasis away from describing results in an underpowered study which is too small to be robust). 3. define the IM’s used in the IM arm. 4. Test the consent and assessment procedures and collection of outcome measures etc. 5. Make a crude estimate of the size effect in both arms and concentrate on the clinical importance of the changes between the groups as opposed to statistical tests.

→ Reply: We agree with this line of reasoning and have specified the specific objectives as follows (please see last line of page 4 and top of page 5): “The specific objectives included analytical exploration of recruitment and retention rates, characteristics of patients and health care provision, statistical and clinical outcome differences between IM and conventional care and to collect data to adequately power a future full-scale trial. Additional objectives were to test the feasibility of assessment, consent and data collection procedures.”
3) The discussion could then be written around the lessons learned from the pilot and how this information may be used to define a main trial in which effectiveness can be robustly tested. Please see Kennedy et al (2008) for an example of a feasibility write up.

Reply: We have totally revised the discussion trying to detail lessons learned from the pilot trial including the implications this may have for the design of a future main trial. Please see the discussion section starting at page 14.

4) I don’t think it is appropriate that the team carried out sub group inferential analyses, the sample was already unpowered before you split the groups, and so it very unlikely to be robust. It might be more useful to look at clinically important differences?

Reply: We can see the point and have removed the inferential sub-group analyses for the chronic patients.

5) The team have not published a detailed protocol for this pilot RCT so it is not possible to compare their assumptions in this trial with a planned protocol. However I am surprised that the team would not have predicted large variability in the outcome measures given the heterogeneity of their sample. Also given that this detail in not in your 2007 paper or a published protocol you need to include much more detail on how you decided the size of the current study (page 10), taking into account the high degree of variability you would expect from such a heterogeneous population.

Reply: Since we had no previous data for our specific study population the sample size for the pilot trial was based on a hypothesis of disability scores, i.e. one of the core outcomes prioritised by the IM team. The hypothesis included that 0-10 ratings of disability would be about equal in the two groups at baseline, around 5. We then hypothesised that there should be a mean improvement of at least 2 points in the integrative medicine group vs. 1 point in the conventional care group in order for the integrative medicine model to be clinically advisable at follow-up after 16 weeks. Applying 80% power, significance level of 5% and assuming a standard deviation of 1.5 gave a sample size of 36 patients per treatment arm (n=72) (STATA software). The sample size was boot to n=80, which we reasoned sufficient to give an estimate of proportion for this type of explorative pilot study.

In retrospect the variability of disability scores were underrated. This was a valuable lesson that needs to be considered in a main trail. We have detailed the above issue in the methods, results and discussion; please see the last paragraphs at pages 10, 12 and 17.
6) I think there are other really interesting results that I would like to see the team draw out in this paper. Page 6, the drop out rates were quite high (esp in the usual care group) over such a short follow up period, how might they address this in a main trial? In your discussion how do these drop out rates compare to CM or conventional care trials in neck and back pain?

Reply: The patients lost to follow-up were relatively large in this study population reaching 25% (9/36) for conventional care and 18% (8/44) for IM care after 16 weeks. We have discussed this in the revised manuscript (please see pages 14-16). An interesting point possibly related to this fact is that the current study employed postal questionnaires administered outside of the primary care units in order to collect data. This strategy was chosen for several reasons, primarily to decrease “the burden” of the primary care units and the GPs to participate in the trial, but also to ensure that the patients would not feel stressed having to complete the questionnaires at the GP’s office when other patients most certainly would be waiting in line, and thirdly to make sure that the patients could be certain that the answers provided would not be read by the providers of care potentially inflicting on the relationship between the patients and the care givers.

We have seen that other back/neck pain studies may have chosen different data collection strategies possibly contributing to higher retention rates. Such strategies might have employed filling in questionnaires on-site at the providing care units or by conducting data collection and follow-ups via telephone interviews. These strategies may be considered in a main trial given that the available resources allow for such engagement of personnel.

It further appears that some studies with low drop out rates, i.e. high retention rates, seem to have recruited patients through advertising or via different health related registers, business or insurance companies. We decided against such an “external advertising/recruitment” approach reasoning that patients enrolled through these types of recruitment strategies may display other characteristics than those actually seeking help in Swedish primary care for back/neck pain. E.g. it may be reasoned that patients having actively applied to be part of a study through an external advertising campaign may be more motivated to comply with different “study assignments” such as data collection procedures in order to stay in the trial. If so that could potentially reduce interpretation of results and generalisability of findings (external validity) in relation to the target population. A study population of “regular patients” may on the other hand be relatively less motivated to comply with different study assignments or to stay in a study over time since their initial motivation was “just to get care”. We have added the above line of reasoning in the discussion.

7) -which CMs were chosen, reasons for choice, were multiple CMs used?

Reply: There are at least two interesting perspectives or areas relating to this comment. First, in relation to the IM model, which “CMs”, i.e.
complementary therapies, were chosen to be part of the IM model and why? This we have discussed in a previous paper (Sundberg et al 2007*), but in short relates to the utilisation rates of these therapies in Sweden, to the use of these therapies for back/neck pain and to the emerging evidence base for these therapies in back/neck pain management. The complementary therapies were Swedish massage therapy, manipulative therapy/naprapathy, shiatsu, acupuncture and qigong.

Secondly, which complementary therapies were chosen to be part of the individual treatment plans devised by the IM team managing patients in the clinical trial? We found that patients randomised to the IM arm received about 7 complementary treatments over a ten week period divided as follows: Swedish massage (1.5), manipulative therapy/naprapathy (1.8), shiatsu (2.8), acupuncture (0.3) and qigong (1.0). Please see the new Table 4.

Related to this we can also see that there were an about equal distribution of the number of treatment sessions between the two main CT categories, i.e. western/manipulative body based complementary therapies (i.e. Swedish massage and manipulative therapy/naprapathy, 3.3 treatments on average) and so called eastern/energy based complementary therapies (i.e. shiatsu and acupuncture, 3.1 treatments on average), please see Table 4.

The low number of acupuncture sessions was related to the difficulties recruiting an acupuncturist to participate in the trial and that acupuncture was the complementary treatment option most often opted out by the patients. The IM team compensated for this by providing shiatsu, a therapeutic approach that to a large extent shares the philosophy of Chinese medicine, instead of acupuncture. Details on the above have been added in the results (please see page 13), in the new tables 4 and 5, and the discussion (page 18).


---

8) -need to also support why you expect to see moderate to large effects between groups when many studies which compare between two active forms of treatment e.g. conventional treatments and CMs show only small to moderate effects.
Reply: Please see reply 5 above relating to our hypothesis for sample size estimation and lesson learned.

---

9) -Highlight the excellent recruitment rates, why do you think this was as often trials in primary care struggle to get GPs to refer into trials. This would be a useful discussion point.
Reply: The relatively high recruitment rates may relate to the fact that we made an effort to facilitate participation for the primary care units and the GPs, e.g. that they only had to screen regular primary care patients. Further
the main GP of the IM team had previous professional experience providing complementary therapies in terms of e.g. Swedish massage therapy and acupuncture/traditional Chinese medicine in addition to being family physician with a special interest in pain management. This was known to several of the collaborating GPs and may have increased their trust in the project and willingness to refer patients. A dual trained GP could possibly also explain the rationale for the study to his GP colleagues in other terms compared to if a complementary therapist or academic administrator would have done this. A fact that may also have contributed to the relatively high recruitment rate.

We have expanded this reasoning in the discussion about recruitment and retention, please see pages 14-15.

10) -finally I think a future trial needs to explore costs and if you have collected data from which crude costs could be estimated for this study it might help inform how you would do this for a main trial. It would be interesting for example to get an idea of how much additional treatment/CMs the integrative medicine group received, this needs to be offset against the levels of difference in clinical outcomes between the groups. It may also be useful to look at side effects if prescription medication or invasive interventions were lower in the IM group.

⇒ Reply: The costs for distributing the added number of complementary therapies in the IM trial have been estimated and provided in the results section (please see page 13). We have also discussed the estimated costs and in relation to health policy implications (please see page 17). We conclude the need for more extensive exploration of costs and cost effectiveness relating to different scenarios (page 16 first paragraph, and pages 17 and 19).

• Minor Essential Revisions

11) This is a very well written and presented paper. I only have one correction. Page 11, line 16th, change advert to adverse.

⇒ Reply: Thank you, this has now been corrected.


Level of interest: An article of importance in its field

Quality of written English: Acceptable
**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:** I declare that I have no competing interests

---

**Reviewer:** Heinz G. Endres

**Reviewer's report:**

Sundberg et al. present a pilot trial consisting of what is described as a “pragmatic randomised clinical trial” on the treatment of both acute and neck and back pain patients, conducted at four primary care units in suburban Stockholm.

1) The stated objective of the pilot trial was to explore the general clinical effectiveness of conventional care combined with complementary therapies in an integrative medicine (IM) model, compared to conventional care alone. This is an unconventionally defined objective, since the purpose of pilot trials is not to test the hypothesis of the subsequent main trial (effectiveness of IM compared to conventional care), but rather to obtain missing basic data, e.g. to estimate the differences in treatment outcomes that may realistically be expected, for the purpose of calculating the sample size for the main study. Since no such data are yet available in the literature the authors were indeed unable to calculate sample size with any degree of confidence. However, a pilot trial could conceivably have other objectives as well. Therefore the authors should define the ultimate object of the pilot trial more clearly.

_reply:_ Thank you. We agree and this is also in agreement with Dr McDonough's comments above. We have redefined the general aim and specific objectives, please see pages 4 and 5 and also our replies to Dr McDonough’s comments number 1 and 2.

---

2) Eighty patients with acute or chronic neck and back pain were recruited. Since it was not clear whether IM is equally effective for acute and chronic pain, both types of pain were included in the study. This should be more clearly pointed out.

_reply:_ We agree and have tried to make it more clear in the background, please see the second paragraph on page 4: “It is currently unknown if sub-acute to chronic back pain, for which conventional medical care is often costly and of limited benefit, respond differently to IM.” In the methods section we
add to this by specifying the chosen duration in the inclusion criteria, please see page 5.

3) The 80 patients were randomised by a computer-generated procedure. Nevertheless, only 36 patients were allocated to the conventional primary care arm, while 44 were allocated to the IM arm. This is not a distribution that would normally be expected from a centralised randomisation procedure. The authors should explain this difference.

Reply: We utilised a computer generated simple randomisation procedure without blocking or stratification, giving equal probability for each participant to receive either conventional care or integrative medicine (this is now stated on page 6). This resulted in somewhat unequal number of patients in the pilot trial. Hence, in retrospect a blocking procedure might have been more suitable for this trial. This has been added as a limitation of the study, please see page 19.

4) Outcome data were collected 16 weeks after randomisation. After 16 weeks there were 9 (25%) drop outs in the conventional arm and 8 (18%) in the IM arm. In the “Statistical analysis” part the authors state that all patients were kept in their assigned groups for the final analysis, in accordance with the ITT principle. However, the ITT principle also requires that all randomised patients be included in the statistical analysis. This the authors do not do. Rather, all patients lost to follow-up were excluded from the final analysis. The problem with this approach is that drop outs and other protocol violations are not simply events by chance, but are related to the respective treatment. If drop outs are excluded from the statistical analysis, the patients who remain in the study (who will generally be the more satisfied ones) will result in massive distortion of study results toward more positive treatment effects. If the intended objective of the pilot trial (which the authors still need to define) were in fact the estimated magnitude of the treatment effect, then such a distortion of the results would be of major importance.

Reply: Thank you for letting us clarify this. In accordance with your valuable comments about the ITT principle we do agree that there may be advantages of including all randomised patients in the statistical analysis. However, we also understand that imputation of missing data, in order to be able to conduct a complete ITT analysis when there are missing cases, may be challenging and controversial regarding the multitude of different strategies and procedures that can be found on how to select or define the “correct value” for imputation.

We hence reasoned that it may be more suitable to present the primary results based on those patients with both baseline and follow-up data available. However, considering the possible effects of the missing patients we did actually do an additional “ITT-analysis” as well (although we, perhaps misleadingly, called it “sensitivity testing”) where the last observed measure
was imputed for all missing data. This imputation procedure hence implied a “no improvement” strategy to balance for possibly positive distortion of excluding cases with missing data. The results of this “sensitivity testing” did not change the previous main conclusions, i.e. the general lack of statistically significant findings between groups. This was previously stated in the results section as “Sensitivity testing with last-observation-carried-forward imputation showed that the general lack of significant results between groups was not sensitive for excluding cases with missing data.”. We apologise if our previous writing on this was unclear and have tried to update the text about this in the results and the discussion section. Please see page 10 (second paragraph) and page 12 (first paragraph).

5) But even if the objective of the pilot trial were something entirely different, at the very least the very positive interpretation of the outcomes in the “Results” section (“statistically significant improvements within both groups of care at follow-up after 16 weeks”) should be corrected. This means that the authors need to either re-analyse their data to include all randomised patients, or they must revise the central conclusion of their pilot trial concerning what are described as statistically significant improvements.
  ➔ Reply: In agreement with Dr McDonough's comments above (please see e.g. replies 1, 2 and 4) we have changed the focus away from inferential statistics within and between groups towards identifying clinical important differences between groups. Please also see reply 4 directly above about ITT-analysis.

6) A very large number of diverse outcomes are included in this study. There is nothing wrong with this. However, the authors should clearly establish at the outset which of these many outcome variables is intended to serve as the primary outcome in the subsequent main trial.
  ➔ Reply: Having changed the aim and specific objectives to include exploring the clinical importance of outcome differences between groups (please also see reply 5 above) we have now specified self rated disability as the main outcome measure. This has been stated in the methods, please see page 7.

7) The authors state that they used a p-value of <0.006 for the Bonferroni correction, which corresponds to the testing of 8 variables for the same patient pool. However, in Table 2 there are 12 variables that are tested, and one would therefore expect a corrected p-value of <0.004.
  ➔ Reply: The Bonferroni correction was based on explorative analysis of the eight core “IM tailored” variables defined and prioritised by the IM team (ADL disability, stress, wellbeing, days with pain, use of prescription and non-prescription analgesics, use of conventional care and use of complementary therapies outside of study). The eight SF-36 domain variables were included as an additional (or “secondary”) outcome. The previous Table 2 showed the four numerical variables ADL disability, stress, wellbeing and days with pain, together with the eight domains of the SF-36, i.e. 12 variables in the same
We realise that this may have been confusing and unclear to the readers of the first version of the manuscript.

In accordance with Dr McDonough’s comments the general aim and specific objectives of the trial have been revised to focus more on the feasibility of implementing integrative medicine and identification of clinically important differences between groups (as opposed to statistically significant differences within and between groups). In agreement with this we have also removed the inferential statistical analysis for the sub group of chronic patients (please see reply 4 to Dr McDonough’s comments). Taken together with the definition of one main outcome measure, the new aims and objectives have markedly reduced the number of statistical analysis performed and thus, in agreement with Dr Poitras’ comment number 14 below, we reason that the Bonferroni correction may no longer be necessary. The methods section has been revised to reflect this, please see the statistical procedures section starting on page 9.

8) Moreover, stating a p-value of 0.000 in Tables 2 and 3 is nonsensical, even if SPSS does give this value. What SPSS means by this is that p < 0.001. The “0.000’ p-values in these tables must be corrected.
   ➔ Reply: Thank you, this has been removed and is no longer applicable. Tables 2 and 3 have been revised to reflect the new objectives.

9) For the comparison of baseline variables, it is not really necessary to give p-values, since one would not expect a p-value of <0.05 after randomisation in any case (even though it could occur after randomisation in rare cases).
   ➔ Reply: To increase the clarity of our tables we agree with you and have removed the p-values for the comparison of variables at baseline for all tables. However, we have also learned that different authors have different opinions about this, see e.g. Dr Poitras’ comment number 16 below.

10) In the “Statistical analysis” section, the authors state that the dichotomous variables were analysed by logistic regression. It is absolutely essential to indicate whether these analyses were univariate or multivariate. Also, every reader will justifiably expect a precise description of the model underlying the logistic regression (definition of outcome variable and independent variables). This information is not apparent in Table 3.
   ➔ Thank you. We utilised univariate logistic regression with group allocation as independent variable. The use of prescription analgesics, the use of non-prescription analgesics, the use of conventional care and the use of complementary therapies were dependent variables in the respective analyses. The manuscript has been revised accordingly to clarify this, please see last paragraph on page 9 and top of page 10 as well as the revised Table 3.
11) In summary, the authors need to specify the following:
• what is the actual objective of their pilot trial
  ➔ Reply: Please see reply 1 above.

12) • which variable is intended to serve as the primary outcome measure in the subsequent main trial
  ➔ Repy: Please see reply 6 above.

13) • what are the treatment outcomes in the two treatment arms when drop outs are included in the final analysis in accordance with the ITT principle.
  ➔ Reply: Please see reply 4 above.

14) The authors should also address or correct all of the above-mentioned other issues in their manuscript.
  ➔ Reply: Please see replies 1-10 above.

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

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

Reviewer: Karen Sherman

Reviewer’s report:

1) 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.
  ➔ Reply: Thank you for pointing this out. In agreement with Dr McDonough’s comments we have revised the aim and specific objectives to focus more on
clinically important differences between groups. Please see replies 1 and 2 to Dr McDonough’s comments above.

• Major Compulsory Revisions

2) Please revise the statistical analysis section so that it is written more concisely and is easier to understand.
   ➔ Reply: The statistical analysis section has been revised to be more clear and easy to understand, please see the statistical procedures section starting at page 9.

3) I think the material on scanning and double data entry can be omitted.
   ➔ Reply: These measures are quality assurance procedures that are highly valued in clinical research of the Swedish health care system. We hence opt to keep this information in the manuscript.

4) I don’t understand the sentence: “The data structure with repeated measurements within patients was accounted for in the analysis? Please revise it.
   ➔ Reply: The sentence has been revised, i.e. “For the analysis of difference in change over time between intervention and control-group a mixed model was applied to account for repeated values within patients.”, please see last sentence, top paragraph, page 10.

5) I don’t understand the sentence on the power calculations. What was the effect size on which the study was powered?
   ➔ Reply: Please see reply 5 to Dr McDonough’s comments above.

6) 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.
   ➔ Reply: The aim and objectives of the trial have now been revised to focus more on exploring clinical important differences between groups (please see e.g. the replies to Dr McDonough’s comments 1 and 2 above). The manuscript has been extensively revised in this matter, please see the results and discussion sections dealing with statistical and clinical differences between groups, pages 11-12 and 16-17).
7) 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.

Reply: Thank you for this input, we have now added a part of the discussion to this, please see last paragraph page 16 and top of page 17. In accordance with the revision of the general aim and the specific objectives the sub group analyses for chronic patients is no longer applicable.

8) 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.

Reply: We did not analyse the 12 week data utilising inferential statistics for this paper. The 12 week data has not been included in the revised tables.

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

Reply: We see that this writing may have been unclear. Please see the reply to Dr Endres’ comment number 4 above.

10) 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.

Reply: The discussion section has been completely revised mainly in accordance with Dr McDonoughs suggestions. Please see pages 14-19.

11) 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).

Reply: Thank you, this paragraph has been deleted. Please also see the reply to comment 10.

12) 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.

⇒ Reply: We have tried to phrase this more clearly. Please see the reply to Dr McDonoughs comment number 5 above.

13) Moreover, you can power a study for clinically important differences, if you know what those are.

⇒ Reply: Please see reply 12.

14) I would delete the paragraph on trying to get more homogenous groups because I don’t think this is a useful strategy.

⇒ Reply: This has been deleted.

15) 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.

⇒ Reply: The paragraphs on blinding and washout effects have been deleted and the discussion section has been revised to include a more focused paragraph on strengths and limitations of the study. Please see last paragraph on page 18.

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

⇒ Reply: This sentence is deleted and the conclusion has been revised. Please see page 19.

17) • Minor Essential Revisions

the word “data” is plural and the authors need to change that throughout the text.

⇒ Reply: This has been corrected.

18) In the abstract – I would remove the comments pertinent to within group differences.

⇒ Reply: The abstract has been revised.

19) 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.
20) What was the primary outcome measure?
Reply: Self-rated disability, please see the revised methods section at page 7 as well as the reply to Dr Endres' comment number 6 above.

21) Were any of the measures besides the SF-36 validated?
Reply: Self-rated disability was slightly modified from the reliable, valid and responsive Bournemouth questionnaire for back/neck pain. This info has been added to the methods section, please see page 8. Further, the stress and wellbeing scales were considered face valid and chosen to reflect common question areas targeted by CT providers in clinical practice. The number of days in pain during a fixed period of time has been valued an important indicator of pain persistence, and based on group consensus and pre-testing of the selected outcome measures we stipulated an appropriate recall period of two weeks. The same time frame was chosen for assessing the face valid outcomes of self reported use (yes/no) of prescription and non-prescription analgesics, conventional and complementary care. We have revised the methods section to inform about this, please see page 8.

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.

Reviewer: Stéphane Poitras

Reviewer's report:
Major Compulsory Revisions

1- The authors state that the study is justified because of a lack of knowledge regarding the efficacy of IM in chronic pain patients, since it has been demonstrated ineffective in acute patients. However, the inclusion criterion is two weeks pain duration. This inclusion criterion
does not fit the study's purpose. This did not appear to have a major impact, since most patients suffered from chronic pain. However, the 2 weeks criterion should be justified.

Reply: The study aimed to include patients with sub-acute to chronic back/neck pain. The manuscript has been revised to more clearly point this out, please also see the reply to Dr Endres' comment number 2 above.

---

2- It is not clear if the allocation was concealed to the research coordinator. This should be clarified

Reply: The allocation was concealed to the research coordinator until baseline data had been collected. The methods section has been revised to clarify this, please see page 6.

---

3- Who assessed patients for their eligibility?

Reply: The GPs screened the patients and then the research coordinator verified the eligibility over telephone. The methods section has been revised to clarify this, please see page 6.

---

4- I'm a bit puzzled by the power analysis. On page 9, it is stated: "The sample size for achieving a power of 80% in detecting a moderate to large difference between groups in favour for IM care was determined based on a pilot trial hypothesis of disability scores". However, on page 13 "Larger sample sizes may therefore be recommended to adequately power future clinical trials of the developed IM model targeting back/neck pain management." Also, "Having no previously reported IM clinical trial data for estimating appropriate sample size our hypothesis underlying the estimation appears partly impaired." However, using the MCID of the SF36 and it's SD, the authors can easily calculate the power of the study, according to the number of subjects who completed the study.

Reply: The section about power analysis and sample size has been revised, please see page 10. Please also see the reply to Dr McDonough’s comment number 5 above.

---

5- What was the time interval between baseline data and allocation?

Reply: Allocation was conducted as soon as the patients' written informed consent and baseline questionnaires had been received, please see page 6. This was generally within one to two weeks due to logistical reasons; hence we only included patients with sub-acute to chronic back/neck pain as mentioned above (comment 1).

---

7- The types of conventional care received in the control group should be described. General information was given on page 6, but more detailed information should be provided.

Reply: The methods, results and discussion sections has been revised
with more details in regards to this information, please see pages 6, 13 and
18.

8- On a note related to comment 7, most interventions included in IM are
eroutinely used by PTs, at least here in NA. Since subjects in the control
group were allowed to receive PT, it is difficult to assess how much they
were different regarding therapy received.
🪘 Reply: In reference to the pragmatic design both groups were allowed to
receive PT, and we agree that it may be difficult to assess precisely how
much the groups differed regarding therapy received. It was estimated that
about 26 % of the conventional care treatment plans included a written
referral to PT (please see results on page 13). Data on the amount of
conventional care (including PT) and complementary therapies (outside of
study) that was received were collected by questionnaires and did not display
any statistically significant differences between the groups (please see results
in Table 3, use of conventional and complementary care).

As an additional piece of information (we have not detailed this in the
manuscript), PTs in Sweden are not allowed to practice the full scope of
complementary therapies according to the Swedish health services act.
Although some PTs have special interest and additional training in
orthopaedic manual therapy procedures, and some work with “western
acupuncture” (please note that the practice of traditional Chinese
medicine/acupuncture is not legally possible to registered PTs or any other
licensed health care providers in Sweden), most PTs likely work with physical
training related activities and rehabilitation.

9- The average number of sessions per week appears to be at 1. This
seems to be relatively small and should be justified. Also, the lack of
difference between groups could be related to the lack of intensity of
sessions per week. This should be discussed.
🪘 Reply: There was no pre-set rule or justification for a specific number of
treatments per week. Following the pragmatic design the study rather set out
to explore what were the typical characteristics of integrative care. The results
showed that the average number of treatment sessions was about 7 over a
maximum period of 12 weeks (average was ten weeks). In accordance with
the pragmatic trial design the treatments were administered at the suitability of
the providers and patients. Treatments could e.g. be delivered more
frequently in the beginning of the treatment period and then spread out. The
possible contribution of more intense treatment attention in relation to the
observed effects have been added to the discussion, please see pages 16-17.

10- The psychometric proprieties and the justification of many of the
outcomes selected have not been provided. Namely:
- disability of activities in daily living due to back/neck pain, stress and
  wellbeing measured by 11 points numerical scales (0-10) where 0
  indicated no and 10 indicated maximum levels respectively;
- number of days with back/neck pain over the last two weeks (0-14);
- self reported use (yes/no) over the last two weeks for prescription and non-prescription analgesics, conventional care and CTs

Reply: Please see reply 21 to Dr Shermans comments above.

11- The outcome measures should be classified primary and secondary. I would suggest the SF36 as the primary outcome. This can then be used to calculate the power of the study (comment 4).

Reply: The general aim and specific objective have been revised and self-rated disability is specified as the main explorative outcome measure. Please see page 7.

12- I do not understand the reason behind: "In order to standardise the administration of the questionnaires patients were matched as far as possible between the randomised groups"

Reply: We see that this information may be confusing and have deleted this sentence in the revised manuscript. What we meant was that when patients enrolled in the IM arm were finished with the complementary treatments in the trial, the follow-up questionnaires was sent concurrently to those patients in the IM arm and those allocated to conventional care arm during the same time period.

13- The authors did not prohibit the use of CT during the study in the control group. To get a better understanding of possible contamination, the exposition of the control group to CT during the study period should be detailed.

Reply: We collected data on self reported use of CTs out of study for both arms and have reported this in the results section. Please see Table 3. Please also see the reply to comment 8 above.

14- The statistical analysis of change in a group is unnecessary. Amount of change in a group can be explained by several factors (placebo, natural recovery). We are more interested in the difference of change between groups. Thus, only the between group comparisons are necessary. (UC vs IM column (baseline and change)). This limits the number of analyses performed. The Bonferroni correction therefore becomes unnecessary, too harsh at the level of significance. A level of 0.05 can be used. This last column shows the lack of difference between the two groups.

Reply: Thank you. We agree and the manuscript has been revised accordingly. Please see the methods section pages 9-10, as well as our reply to Dr Endres’ comment number 7 above.

15- Why were non-parametric analyses used?
We used both non-parametric and parametric analysis depending on data type, i.e. changes in “days with pain” was analysed by the two sample t-test. Please see the methods section pages 9-10.

16- Statistical analyses at a level of 0.05 should be performed to assess the differences in baseline characteristics between groups (table 1). At a minimum, a p-value should be provided. There appears to be differences.

Reply: The statistical differences (p values) in baseline characteristics were previously given in Table 2 and 3, i.e. there were no statistically significant differences for any variables. However, in line with Dr Endres’ comments about the presentation of results we have removed this information to increase the clarity of the tables. Please also see our reply 9 to Dr Endres’ comments above.

Minor Essential Revisions

17) Data on the recruitment of patients (page 6) should be put at the beginning of the results section.

Reply: Corrected, please see the results section starting on page 11.

18) Data and analyses on drop-outs (page 6 and 14) should be put in the results section.

Reply: Corrected, please see the results section stating on page 11.

Discretionary Revisions

None

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

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests: I declare that I have no competing interests