Reviewer’s report

Title: An update of stabilisation exercises for low back pain: a systematic review with meta-analysis

Version: 2 Date: 28 July 2014

Reviewer: Luciana Macedo

Reviewer’s report:

The study is relevant as presents an update of a previously published review, although, the authors missed other relevant reviews in the area that were published after the review they aim to update.

Major compulsory revisions:

1. I think that the authors should not have pooled all of the studies together regardless of two things: the duration of pain (acute or chronic) and the treatment comparison. The later is very important and incorrect. Studies that compared stabilization exercises with a control group were pooled together with studies comparing stabilization exercises with other exercises. The literature is very clear that exercises have a small effect (thus stabilization exercises versus control or minimal intervention should give significant results). But the literature suggests that there is no difference between different types of exercises. Thus pooling everything together is not only heterogeneous due to fundamental differences between studies but also does not allow for a full interpretation of the results. For instance, stabilization versus control could have been clinically significant, but we don’t know this as this analysis alone was not conducted. I believe that only the subgroup analysis presented in the study should be included.

Minor Essential Revisions:

1. The authors should perform a systematic review of systematic reviews and make sure that they have not missed any given that many systematic reviews in the topic have been published. Including Macedo et al; Phys Ther 89:1; pg 9-25
2. The agreement in study selection was very low. This is surprising as normally the agreement tend to be very high, especially when there is a defined inclusion criteria. I think it would be good if the authors provided information where the disagreements occurred: type of exercise, population, etc?
3. Given the controversy in using quality scores to pooled data, why did you do this? It is still not clear to me why you would choose this approach.
4. You mentioned on the results “That left a total 18 publications for inclusion [25–28,57,62–75]”. (this is actually 19 references).
5. Conclusion: “there is a trend of worse fear avoidance scores” This is confusing, I did not understand what you meant with this statement.
6. I think you might need to better define the interventions that were included as
part of stabilization exercises. For instance the studies by Marshall et al and Xueqiang don’t explicitly say that TrA or MF were of interest and thus, it is left to the reader’s imagination to whether these exercises actually contained the isolation of the deep trunk muscles. It is very possible that it did not and that stabilization exercises that deals only with the global muscles (as suggested by Stu McGill) could have been used.

7. Why would a pilot study not be eligible for inclusion in the review? Why did you exclude this study?

8. You mentioned that three studies included a package of treatment that included stabilization. How can you assess the effect of the intervention within this package? Why did you decide to keep these studies in the review?

9. I found that there is a lot of descriptive information about the studies on the text of the results. I think that this would be better represented on a table with duration of symptoms, type of comparisons, quality of the trials, etc.

10. Given that you had studies including populations with different duration of pain: acute, subacute and chronic, I think it is probably better to separate these studies. This is what is recommended given the better prognosis of acute and subacute patients and the fact that exercise is not often recommended for this group.

11. How was clinical significance assessed? You do not mention this on your methods or results and it is left to the reader’s imagination to guess this.

12. It is odd to me that statistical heterogeneity was presented in the same sentence as clinical significance. This may cause one to think that they are related. Maybe this needs to be made clearer.

13. At long term there was no significant statistical or clinical difference; should read: At long term there was no clinical or statistical significance.

14. I² does provide some information on the strength of the results but I don’t think it should be the sole measure to evaluate the robustness of the analysis. This could even be a chance finding from small trials or something like that. Other things need to be taken in consideration. I think that the meaning of I² has been a little overused in the study. You should consider using GRADE or any other method to evaluate the strength of the results. At this point you used only I² to evaluate the strength of the evidence, whether other methods are more comprehensive.

15. Why did you include the results of the FABQ in your manuscript since this was not in your methods?

16. I think that supplement 1 could be made smaller and simpler and presented as part of the manuscript, especially if the manuscript text can be reduced.

17. Also, I don’t see the need for supplement 2 given that the PEDro scores a readily available on the PEDro database.
**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 am on the final steps to conclude a Cochrane review of same topic. "Motor control exercise for non specific low back pain". I guess this give me a complete understanding of the manuscript produced here.