Reviewer's report

Title: Preventive physiotherapy interventions for back care in children and adolescents: A meta-analysis

Version: 3 Date: 16 April 2012

Reviewer: Wolfgang Viechtbauer

Reviewer's report:

The manuscript provides the results from a meta-analysis of studies examining the effectiveness of preventive interventions for back care in children and adolescents. The goals of the meta-analysis are clear and well-defined. In general, the methods and techniques used are well-described and conform to standard meta-analytic methodology. Given that this appears to be the first meta-analysis on this topic, the manuscript has the potential to make a useful contribution to the literature. In general, I am supportive of the publication of this manuscript. However, before I would consider the manuscript ready for publication, I think some revisions are warranted.

Major Compulsory Revisions

--------------------------

I have two main concerns with this manuscript, both pertaining to the evidence base on which the conclusions are based.

1) Many of the effect sizes (SMDs) of the studies are quite large and some are just outright extreme. In fact, I have never seen effects as large as 13 or even 22 in any study. Accordingly, the mean effects that are found are also rather large and larger than usually observed in other meta-analyses. For example, Lipsey and Wilson (1993; The efficacy of psychological, educational, and behavioral treatment. American Psychologist, 48, 1181-1209) provide a comprehensive overview of the effects observed for a wide variety of different types of interventions and the largest mean effect size observed in hundreds of meta-analyses was 1.51. Three out of the four mean effects observed in the present meta-analysis are this large and one is even much larger. Many of the estimated effects based on the moderator analyses (i.e., in subgroups of studies) are even bigger.

It may very well be the case that some of the included studies achieved such large effects, but these findings need to be double-checked and the presence of these large effects needs to be discussed in the paper. Is there something peculiar about these studies that created such large effects?

Moreover, the robustness of the conclusions of the meta-analysis need to be checked. What happens when these extreme values are excluded? In particular, what happens with the mean effects and what happens with the results from the
moderator analyses? In fact, many of the moderators turn out to have a statistically significant relationship with the size of the effects. In my experience, this is somewhat unusual. I would assume that many of these significant findings are simply a result of the extreme effects falling into one or the other level of a categorical moderator (or at one end of the distribution for a quantitative moderator). If that is the case, then essentially many of the findings are determined by one or two studies. I would be hesitant to draw strong conclusions based on such findings.

2) My second concern also applies to the evidence base of this meta-analysis. I think it should be noted somewhere in the text that there are really only a few groups of researchers that have been conducting these studies. For example, two researchers from Belgium are the first/second authors of 10 (i.e. over half of the) studies (explaining why so many of the studies come from Belgium; it is really just one group that is conducting all of the studies there).

An implicit assumption underlying the analyses is that the 19 studies are independent. However, given the small number of research groups producing these 19 studies, this assumption is questionable. Accounting for the hierarchical dependence in these data is difficult (due to the low number of studies overall) and would require multilevel methods. However, at the very least, I think the authors should discuss this issue somewhere in the text as a potential limitation of the evidence base.

A few additional things that I think the authors should definitely address:

p. 15: Not finding a significant difference between published and unpublished studies does not provide sufficient evidence to "discard publication bias". I think the authors should examine/present funnel plots and examine the relationship between the precision of the estimates (e.g., based on the sample sizes) and the size of the effects (as a means of testing for funnel plot asymmetry). I would not conduct the Egger regression test with this data, because the way the sampling variances (and hence, the standard errors) of the SMD values are calculated will automatically induce a correlation between the SMDs and the standard errors (especially since many of the SMD values are so large). However, based on Peters et al. (2006; Comparison of two methods to detect publication bias in meta-analysis. Journal of the American Medical Association, 295, 676-680), one can use the (inverse) sample size for such tests.

Finally, the manuscript still needs some major language editing, ideally by a native speaker.

Minor Essential Revisions
-------------------------

Some additional issues/points:

p. 5: Why do the authors expect the sex of the participants to influence the results?
p. 7-8: Were studies with more than two treatment arms excluded? Or was it just coincidence that all of the studies included happened to only have two treatment arms?

p. 12: Instead of writing 10.53%, it would be clearer to state that 2 out of the 19 studies were unpublished. Also, the authors should describe what methods were used to examine the presence of publication bias (this becomes clear later on in the manuscript, but the methods should already be described here).

p. 13: When reporting the averages about the study characteristics (average number of weeks, mean intensity, etc.), some measure of variability (e.g., SD, range, interquartile range) should also be provided.

p. 16-18: How was the % of variance explained calculated?

Discretionary Revisions

p. 10: If 19 studies were included, then "37% of the studies" really just means that 7 articles were coded by two reviewers. In my opinion, all of the articles should have been coded by two reviewers (I think with 19 studies, this is not unreasonable). Therefore, the ICCs and kappa values are based on only 7 observations, which is really a rather low number for calculating these measures of agreement. It's now a bit late to change this, so there is probably little that can be done.

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 declare that I have no competing interests.