Reviewer’s report

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

Version: 2 Date: 28 August 2014

Reviewer: Neil O’Connell

Reviewer’s report:

This systematic review is an addition to an area that has been reviewed a fair amount already. However a rigorous up to date review is welcome given the persistent popularity of the intervention. In my view the manuscript in its present form requires some substantial revision before I would recommend it for publication with more methodological detail and justification needed in a number of areas, not least on the plan for evidence synthesis and the presentation of effect sizes. I have also asked for some clarifications regarding certain methodological choices (see below).

Compulsory revisions/ Clarifications

Abstract

Intro - I would reconsider the sentence regarding the lack of evidence to support its use as it is ambiguous. A lack of evidence or a lack of positive evidence? There is a fair amount of evidence about - hence your review. If a lack of supportive evidence would that not preclude the question of your review?

Results – When presenting percentage effect sizes specify what it is a percentage of. When mentioning clinical importance please specify the threshold. When mentioning significance state that it is statistical significance.

Conclusion – What is meant by “person specific”. Do you mean individually tailored? If so this is not explicitly specified in the inclusion criteria later and is unlikely to be true of group exercise studies.

Background

NSLBP - is not really defined as without cause, but without a cause that we can reliably identify.

The 62% statistic from Hestbaek for long term problems would benefit from some qualification. A much smaller % might be expected to have serious disabling chronic problems.

What is meant by the median duration of survival time?

Do you mean improvement or growth of the evidence base?

In justifying the review it would be worth also referring to the review by Bystrom et al (2013 Spine 38:6) and its critical summary by the CRD http://www.crd.york.ac.uk/crdweb/ShowRecord.asp?LinkFrom=OAI&ID=12013023228

Why would meta-analysis be a secondary aim - it is a critical tool in achieving the
primary aim. Were you not interested in comparisons with no treatment or usual care? If not it might be useful to explain why.

Methods

The search strategy seems quite limited. It is possible that a broad selection of terms might have been more sensitive. EMBASE or SCOPUS were not searched - could the authors comment on whether they may have missed any trials that way? Was any search made for unpublished studies or of the grey literature?

Study selection - were patients with radiculopathy included?

Study selection - why have a single author screen the titles and abstracts? Can the authors comment on possible reasons for the low level of agreement relating to full text screening?

Page 8 line 156 - correction - authors were contacted.

Were authors contacted where data required for meta-analysis were not available?

Quality Assessment - correction - in a number of places “were” should be replaced by “was”.

Can the authors comment on the decision to use PEDRO scores from the database rather than quality assessing the papers themselves with 2 independent reviewers. This feels like a shortcut.

Were studies scoring <6/10 considered to be of low quality? The use of these cut-offs is common but questionable - for example you have a “high quality” paper that reported insufficient data to analyse. So it is at high risk of bias for selective outcome reporting.

Statistical Analysis – why were results converted from their original scales? How was this done for disability scales such as the RMDQ? Why not use the standardised mean difference? This section requires substantially more detail. What was the established a priori plan for the evidence synthesis? Which comparisons were planned? With regards heterogeneity was the Chi squared test also used to assess the statistical significance of the heterogeneity? How was it decided whether trials were “sufficiently homogenous”. Was any subgroup analysis planned? Was any investigation of small study effects made?

Results

It is odd to reject papers in an update of a systematic review that were included in the previous versions. More detail on how the participants in those studies no longer met the criteria are needed.

Study quality and bias - correction “…failed to assess baseline comparability”

Be clear when using the word significant whether you are referring to statistical or clinical significance.

There needs to be a clarification of the scales - do they represent the scales all converted to 0-100 or some form of percentage as implied in the abstract. If a percentage what is it a percentage of? Baseline score? Post treatment control group score? What threshold are you using for clinical significance? Without
these details it is not possible to meaningfully interpret the results. They appear in the discussion but need to be more clear and precise, presented in the methods and stated whether they were established a priori.

Forest plots - it would be useful to see the plots for the primary analyses rather than just those for subgroup analyses. It would also be useful to see the number of participants clearly displayed in the forest plots.

Were the subgroup analyses pre-planned or post hoc? It is vital that this is made clear.

Discussion:

The most likely explanation for the high heterogeneity is arguably the different comparisons being made between trials - this should be acknowledged.

You state here that 2 authors independently extracted data but in the methods one author did this with another author checking. This needs to be consistent across the paper.

Discretionary revisions

Title

This could be clearer - what is the specific question - effectiveness? Compared with?

The discussion is quite strident in suggesting the increase in fear avoidance scores is the fault of the core stability paradigm. While I have sympathy with the theoretical suggestion this should be substantially toned down as there was no significant difference and the difference may also be clinically meaningless if it is there at all.

**Level of interest:** An article of importance in its field

**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'