Author’s response to reviews

Title: Effects of aerobic exercise on pain and disability in patients with non-specific chronic low back pain: a systematic review protocol

Authors:

Irlei dos Santos (irlei_santos@yahoo.com.br)
Adriana Lunardi (adrianalunardi@hotmail.com)
Naiane Oliveira (naiateixeira@hotmail.com)
Matheus Almeida (mathewalmeida@hotmail.com)
Leonardo Costa (lcos3060@gmail.com)

Version: 1 Date: 12 Dec 2018

Author’s response to reviews:

11th December 2018

Hyun Kang

Editor of Systematic Reviews

Dear Hyun Kang,

Thank you for providing us with the reviewers’ comments on our manuscript entitled “Effects of aerobic exercise on pain and disability in patients with non-specific chronic low back pain: a systematic review protocol”. We appreciate the comments and attention of the reviewers. We provided a point-by-point response for each of the comments as well as revised the manuscript accordingly.

Yours sincerely,

Irlei Santos (on behalf of all co-authors)
Reviewer #1:

I am pleased to get the opportunity to read and comment on this protocol. This is an interesting subject and very important to elucidate the existing knowledge on the field by a systematic review.

I have some comments and suggestions before finalization of the protocol.

Response: Thank you very much for your kind and insightful comments.

TITLE: Are you not assessing the effect instead of the efficacy?

Response: Thank you for this. We have changed the term efficacy to effects as we aim to get outcomes from clinical research. We have changed the title accordingly.

INTRODUCTION: consider rephrasing: "the most commonly prevalent..." to "the most prevalent...".

Response: Thank you. We have removed “commonly” as requested.

In the last paragraph on the first page, your state the rationale for, why aerobic exercise can improve outcome among patients with chronic back pain. This paragraph could be more precise in stating, what do we know from other cohorts, what is the effect and why should this work on the patient group that you are examining?

Response: Excellent suggestion, we have included more information on the effects of aerobic exercises in patients with low back pain from observational studies.
METHODS:

PICO: Why are you planning to exclude studies with inadequate randomization process - this is not common practice. Instead, you would classify these studies as having a high risk of bias and base your primary analysis on the trials with overall low risk of bias.

Response: We agree with you. We have removed this exclusion criteria, but we make it clearer that non-randomized controlled trials will not be included.

Why are you only interested in mixed gender trials? You may describe the heterogeneity of studies and limit the conclusions because of heterogeneity of the patients. You may risk excluding valuable knowledge.

Response: No, we are not. Gender is not an exclusion criteria. Most (if not all) clinical trials for back pain included mixed gender. In the case we found a trial which recruited only male or female gender, it will be included. We have removed “from both genders” of the sentence to avoid confusion.

Do you plan to include cross-over studies - and if so, what will an required “wash-out” period be as minimum?

Response: We are quite confident that there is no cross-over studies for patients with chronic back pain using clinical outcomes (such as pain and/or disability). It is very unlikely that someone would conduct a cross-over trial with this population. Back pain responds quite well with non-pharmacological interventions and an extremely large wash-out period would be needed for patients to experience a recurrence of symptoms.

Regarding the outcomes: how are pain intensity and disability going to be defined?

Response: Pain intensity are usually measured by likert-type scales such as visual analogue scales or numerical rating scales. Some larger questionnaires, such as the McGill Pain questionnaire has an 11-item Likert scale for measuring pain intensity.
There are a wide range of questionnaires for measuring disability associated with low back pain. The most common one are the Ronald Morris Disability Questionnaire, the Oswestry Disability Index and the Quebec Questionnaire. It is possible that some trials might use measures such as the Patient-Specific Function Scale. We intend to include any validated measure.

Would you consider only one primary outcome? Maybe QoL is more important to the patients - as it may contain pain and disability. …

Response: We would like to consider only pain intensity as the primary outcome. The newest core outcome set for back pain [https://www.ncbi.nlm.nih.gov/pubmed/29194127](https://www.ncbi.nlm.nih.gov/pubmed/29194127) indicates pain intensity, physical functioning (i.e. disability) and quality of life as the most important measures for patients with low back pain. Although quality of life is also considered as a core outcome, the number of trials that measure this outcome is smaller compared with trials that measures pain or disability. For these reasons, we would like to keep pain intensity as the primary outcome and quality of life, disability, kinesiophobia and return to work as secondary outcomes.

Search: you should consider also include studies regardless of the publication status, and then perform searches in trial registries and conference proceedings, to find unpublished trial results. Publication bias is one of the most important bias in systematic reviews. If only published material is included, there is a risk of overestimating the intervention effect. Therefore, attempts to limit the publication bias should be achieved.

Response: We agree with you. We will also search for ongoing or unpublished data using clinical trial registries. We have included this information in the method section.

Consider having two reviewers for both screening phases.

Response: We agree with you. We have included this information as suggested.

Data extraction: consider specifying exactly what characteristics (patients and interventions) you would like to extract.
Response: We have included this information as requested.

Risk of bias assessment: I am not familiar with the PEDro score, but the domains seem very appropriate and sufficient to be used instead of the Cochrane Handbook's tool for risk of bias assessment. Though, I have a major concern, that you will use other's assessments and not uniformly apply your own assessment and judgement for all included trials. Moreover, I am very concerned, that a cut off at 6 points is far to optimistic. In the Cochrane's tool, only when all domains are judged as being of low risk of bias, the overall judgement will be low - and ONLY these trials, will be of low risk of bias. As you mentioned in you preface, one limitation (important) to your review and meta-analysis is, that it can only be as good, as the included studies. By judging not-low-risk-of-bias-trials, as being low, you will introduce bias to your own review and results. It is far better to be completely transparent, and plan to perform your primary analysis on the trials with overall low risk of bias. If no trials are of overall low (score 10), then it is obvious, that this area needs more good clinical trials, and you can still perform secondary analyses, on all trials despite risk of bias. Alternatively, of you expect none to be of low risk of bias, you can in the protocol specify a category of overall "lower" risk of bias. Then you should specify, which items you would accept as not being met (getting a "no" answer). In this area, blinding is obviously impossible, and then these items could be allowed to have a "no", but you would still require the outcome assessors to be blinded.

Response: Thank you for these comments.

Firstly, all trials on PEDro are assessed by highly trained raters. In order to be a PEDro rater, someone has to do the PEDro online training and then has to rate 5 clinical trials. The maximum score for this exam is 55 points (i.e. 5 trials x 11 items) and this rater has to achieve at least 50 points. All trials on PEDro are rated by two of these raters and, in the case of discrepancies, arbitration is done by a third and very experienced rater. Our position is that getting the scores directly from PEDro are actually more reliable than by doing ourselves (although most of authors of this manuscript are PEDro raters anyway; the senior author of this protocol has rated more than 5 thousand trials).

Secondly, it is impossible for an exercise trial to get a score above 8/10 in this case as both the trial therapist and patients are impossible to blind. In this case we agree with you that 6/10 will be overly optimistic. Therefore we will follow your first option of running our primary analysis.
using only low risk of bias trials (i.e. 8/10) and then run secondary analysis regardless of risk of bias.

Statistics: If you plan to use two primary outcomes, you should adjust for statistical significance level to account for multiple testing and the chance of a type I error.

Have considered included subgroup analyses (with test of interaction) - on e.g. overall low versus overall high risk of bias studies?

Response: We decided to have only pain intensity as the primary outcome.

You should also consider including a sequential method (e.g. Trial Sequential Analysis) to account for random errors due to small sample sizes and multiple testing (if. the imprecision in the GRADE evaluation). Attempts to decrease bias and decrease random errors will improve your accuracy. Please see the proposed reference articles.1-3


Response: We were unaware of this type of analysis. Thank you for introducing it to us! We will use it as suggested.
Reviewer #2:

I think this will be a very interesting review! Some comments for your consideration.

Kinesiophobia is an unusual outcome for chronic pain patients - perhaps something more commonly used in physiotherapy? Many trials and non-experimental studies in chronic pain use the outcomes from IMMPACT - I apologize if I have missed it but I don't see any references to that set of documents in your protocol. Might you consider also looking at antalgia as a related outcome to kinesiophobia?

Response: Thank you for your comments. Up to a few years ago, most researchers from the low back pain field used IMMPACT, but very recently a group of experts in back pain developed their own core outcome domains and core outcome instruments (please see citation on the outcome’s section). Kinesiophobia is an extremely common outcome in the field of musculoskeletal pain. Many patients develop fear of moving due to musculoskeletal conditions (such as back pain) and this construct is highly associated with chronicity. Since then, many trials started using it as an outcome or as a mediator for more important outcomes such as pain and/or disability. As far as we know, most (if not all) clinical trials in the field of back pain, does not measure antalgia, but kinesiophobia.

You say: "If the outcomes are evaluated by different scales, they will be converted to a common scale ranging from 0 to 100." Most of the measures in studies of chronic pain patients generate ordinal data that is treated as interval data (you likely know this!) - in many cases though, it would be inappropriate to convert the scales to something as granular as a 0-100. Doing so may well render the outcome either inaccurate or increase the magnitude of the skew that is common in these measures. Please be careful with this - I suggest that you consider only converting 0-10 - VAS-type scales as a result.

Response: We agree with you. We decided to use standardized mean differences instead. We think this would be the best decision to avoid problems related to precision.

The paragraph relating to the GRADE approach - lines 17-30 following the GRADE table is very confusing - please consider rewording. As well, your synthesis section is limited to a discussion of GRADE and related procedures. If the measures are heterogeneous - as I imagine they will be
- you will need to do some sort of narrative synthesis to add to whatever meta-analysis you perform. I would suggest that the narrative component will add significantly to your statistical work.

Response: We have reworded lines 17-30 as requested. We also included a sentence on narrative synthesis as we do expect a large level of heterogeneity.