PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	THE INFLUENCE OF ALLOCATION CONCEALMENT AND INTENTION-TO-TREAT ANALYSIS ON TREATMENT EFFECTS OF PHYSICAL THERAPY INTERVENTIONS IN LOW BACK PAIN RANDOMIZED CONTROLLED TRIALS: A PROTOCOL OF A META-EPIDEMIOLOGICAL STUDY
AUTHORS	Almeida, Matheus; Saragiotto, Bruno; Maher, Chris; Costa, Leonardo

VERSION 1 – REVIEW

REVIEWER	Hala Nassif, PhD Publicis Health, France
REVIEW RETURNED	06-May-2017

GENERAL COMMENTS	The aim of this study is to investigate the influence of allocation concealment and the use of intention-to-treat analysis on estimates of the treatment effects of physical therapy interventions in low back pain clinical trials from meta-analyses. The authors of the study state that biased results from RCTs can lead to inadequate clinical decision making and consequently affects patients. It would be interesting if the authors could briefly state some examples of inadequate clinical decision making to clarify this point further.

REVIEWER	Juan Alfonso Andrade Ortega Complejo Hospitalario de Jaén (España-Spain)
REVIEW RETURNED	07-May-2017

GENERAL COMMENTS	The authors pose a very interesting topic in order to improve the methodological quality when dealing with physical therapy interventions in low back pain randomized controlled trials, so they should be congratulated. However, I would like to raise some concerns: CINAHL (Cumulative Index to Nursing and Allied Health Literature) is a very important database in the fields of Physical Therapy in addition to Nursing and other disciplines. Will it be considered? When carrying out the meta-regression, why “sample size” (< 100 per arm, >= 100 per arm) and “heterogeneity assessment” (I2) will be categorized. I think the regression outcomes would be strengthened when quantitative variables are used.

	Finally, intention-to-treat analysis is particularly suitable for superiority trials, but not when equivalence and non-inferiority trials are carried out. Will the authors keep in mind this?

REVIEWER	Greta Castellini IRCCS Istituto Ortopedico Galeazzi Italy
REVIEW RETURNED	15-May-2017

GENERAL COMMENTS	The authors have presented a topic, which can inspire health professionals dedicated to rehabilitation to improve the conduction of a clinical trial. Underling the effect of systematic bias such as the allocation concealment and the adequate use of the intention to treat analysis on the effect of an intervention can emphasise the importance of adequate perform a RCT in terms of conduction, transparency, and reporting. The manuscript is well written, however there are major concerns which I have related to (1) the unit of the analysis the authors have selected (original RCTs? Meta-analysis? Or systematic review?) (2) Authors' choice of using PEDro scale and (3) explanation of the meta-regression analysis. 1. Going through the whole manuscript, I had difficulties to understand which the unit of analysis is. In some lines, you referred to meta-analysis, in other lines to systematic reviews and even to randomized controlled trials. The aim of the manuscript is quite clear but I did not perfectly get the selection process of the unit of the analysis (meta-analysis?) and from where the authors extracted details about allocation concealment and intention to treat analysis. I have reported the lines where the reader might be confused: Line 137: "we include meta-analysis" Line 151: "Filters related to meta-analyses and systematic reviews". It is not clear whether you have used the "PubMed" or another database filter for meta-analysis/study design, just the free search term or the MESH term. Moreover, in the abstract you have mentioned just meta-analysis and not systematic review. Please clarify the search strategy. Line 160: "we will extract the information….. for each RCT". Here, you reported that RCT would be the unit of analysis from which extract the information regarding the effect of the intervention, the allocation concealment and the intention to treat analysis. However, in Line 172 the unit of the analysis is already changed: "two reviewers will independently extract data from selected systematic reviews". In this sentence seemed that systematic review would be selected and it is the study from which you would select the data. Please make the unit of analysis consistent throughout the text. Line 176: "Study characteristics". Are you referring to study characteristics of systematic reviews or meta-analyses. I suppose this step will not be clear until the unit of analysis is defined. If the interest is to select original RCTs from the systematic reviews, it is not actually necessary to summarize the characteristics of the systematic review but just the RCTs'. On the other hand, if the aim is to focus on meta-analysis included in the systematic review then it might have more sense to report review's characteristics. Line 186: "Individual study data will be retrieved from the met-analyses included in our study". Here, unit of analysis of the study: meta-analysis.

	I think it is not clear which is the systematic process you are going to follow to select studies and extract data throughout the manuscript. I suggest to re-write the section “identification and selection of studies” letting the reader better understand the selection process. For instance, “We are going to select all the meta-analysis of RCT included in published systematic review and from published meta-analysis alone on the effect of physical therapy interventions in adult patients with non-specific low back pain… on the outcome pain or disability…. “ Or, “we are going to select all the original RCTs included in the systematic review with a meta-analysis….” ? Please remember that a systematic review can have or not a meta-analysis but a meta-analysis can exist also without being included in a systematic review. In data extraction you should explain the following issues: Once selected the meta-analyses (if you decided to start form the meta-analyses), will you extract all the RCTs that are included in it and look for the full text? Then, are you going to search the RCT’s title in the Pedro database? … if the title is not included in the Pedro database, what are you going to do? If the RCT on the Pedro database is currently being rated, what are you going to do? Retrieve the full text and assess it? This information is missing. I think is necessary to consider all situations you can encounter and explain how to manage them. Moreover, what if a systematic review included more than one meta-analysis assessing pain or disability, are you considering both or just one meta-analysis for each systematic review? 2. The second remark referred to the use of the PEDro scale for the assessment of the methodological quality. The author decided to extract details about the allocation concealment and the intention to treat analysis form the assessment given by the PEDro database. Please give a reason why you have selected the PEDro scale as assessment tool. Generally, the Risk of bias tool of the Cochrane Collaboration give a most exhaustive information on the likelihood of the risk of bias and not just the presence or not of the domain in the manuscript . It might be useful to report in “the risk of bias assessment section” how the judgment of the PEDro works and how each single domain is rated (yes/no). In line 166 you stated “when the score of domain bias form an included article is not available at PEDro database, two review authors will independently assess it…”. This statement revealed the presence of two different types of bias assessment. On one hand, the Pedro score performed by external and anonymous assessors and reported on the PEDro database, on the other hand review authors (I imagine authors of this manuscript) assessing the RCTs. Who is going to perform the assessment? This method you chose can be valid and equal to the one on the PEDro database only if the assessor is blinded to the scope of the manuscript otherwise the two assessments types cannot be consider the same. 3. Data analysis. You are going to re-analyse the original meta-analysis and create a new one combining a pooled effect size obtained by all the RCTs included in the original meta-analysis with adequate allocation concealment and ITT analysis AND another pooled effect size obtained by the RCTs not having adequate allocation and ITT performed. Is it correct? The effect obtained will indicate if the treatment effect has been modified by the presence or absence of these domains.

	However, what are you going to do when a RCT has been rated, for instance, as NO for the allocation and YES for the intention to treat analysis? Are you investigating separately the two domains? I do not clearly get the reason why you have decided to investigate the presence of the association between bias and effect of the intervention from the meta-analysis included in the systematic review and not from the original RCTs included in the same SR. Theoretically, allocation concealment and intention to treat analysis might affect first the RCT results and then the meta-analysis pooled effect. Minor comments: Line 86-91. These paragraph seems superfluous. The concept of what is the meta-epidemiological studies should be not stressed twice (see line 92-93). Line 201: add be after will. Line 205: I suggest explaining better why you are investigating the effect of sample size and heterogeneity as covariates in the meta-regression

REVIEWER	Iosief Abraha Health Planning Service, Regional Health Authority of Umbria, Department of Epidemiology, 06124 Perugia, Italy
REVIEW RETURNED	17-May-2017

GENERAL COMMENTS	In this protocol the authors aim to assess the bias related to inadequate allocation or lack of intention-to-treat using a meta-epidemiological approach. Overall, the proposal is interesting and it is original in the context of physical therapy. The following are the major issues that the authors will need to address. a) it is not clear how the trials will be classified. To perform this authors will need to provided classifications based on the adequacy or not of allocation concealment (eg, adequate, inadequate, unclear) and the presence or lack of intention-to-treat. This classification will drive the review/meta-analysis (MA) categories. b) classification based on the intention-to-treat can be used in different ways and authors will have to clarify this. Authors should be aware that the case of intention to treat is often interlinked with the post-randomisation exclusions. They can use exclusively on the reported exclusions as performed in the meta-epidemiological study they cited (BMJ 2009;339:b3244.) but a potential misclassification may occur as trials might perform exclusions without reporting them. Alternatively they can base their ITT categorization on reporting as performed in another meta-epidemiological study (BMJ 2015;350:h2445). c) Authors should be aware that deviation from intention-to-treat or modified intention to treat are consistently present in MA and trials (http://dx.doi.org/10.1016/j.jclinepi.2016.11.012; BMJ 2010;340:c2697; http://dx.doi.org/10.1186/1745-6215-12-58). Hence, they will need to clarify how will consider studies with these characteristics. d) I am not a statistician but the paragraph on analysis should be correctly referenced. Authors may have a glance to the following paper that used continuous data BMJ 2012;344:e813. In addition, are they going to assess publication bias? Minor issues

	a) Page 5; line 100-101. The statement is not entirely correct: Regarding the reference number 16: I think that in their IPD assessment Tierney and Stewart found that the experimental intervention had a positive effect when exclusions were performed, compared with the effect measured when a true intention to treat analysis was done. In addition, Melander et al (BMJ 2003;326:1171-3) found that drug treatment had a favorable effect when a per protocol analysis was used instead of an intention to treat analysis.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Hala Nassif, PhD
Institution and Country: Publicis Health, France
Competing Interests: None declared

Comment: The aim of this study is to investigate the influence of allocation concealment and the use of intention-to-treat analysis on estimates of the treatment effects of physical therapy interventions in low back pain clinical trials from meta-analyses.
-The authors of the study state that biased results from RCTs can lead to inadequate clinical decision making and consequently affects patients. It would be interesting if the authors could briefly state some examples of inadequate clinical decision making to clarify this point further.

A: Thank you for your suggestion. We have added an example explaining how biased results from RCTs could lead to inadequate clinical decision making. Please see below:
“Clinicians may select interventions for their patients based on trials results that are inflated, and so these are not a very reliable guide to treatment selection. For example, a systematic review about the effectiveness of low level laser therapy for chronic LBP found a significantly greater reduction in pain in response to laser therapy compared to placebo. However, this finding was based on a meta-analysis with three clinical trials that did not conceal allocation or use an ITT analysis, so this result could be an overestimate of the true effect of laser. If this premise of overestimated effect is true, clinicians might in good faith select interventions that might not help their patients.”

Reviewer: 2

Reviewer Name: Juan Alfonso Andrade Ortega
Institution and Country: Complejo Hospitalario de Jaén (España-Spain)
Competing Interests: None declared

Comment: The authors pose a very interesting topic in order to improve the methodological quality when dealing with physical therapy interventions in low back pain randomised controlled trials, so they should be congratulated. However, I would like to raise some concerns:
- CINAHL (Cumulative Index to Nursing and Allied Health Literature) is a very important database in the fields of Physical Therapy in addition to Nursing and other disciplines. Will it be considered?

A- Thank you for this suggestion. We have added CINAHL as a database in the methods section.

Comment: When carrying out the meta-regression, why “sample size” (< 100 per arm, >= 100 per arm) and “heterogeneity assessment” (I2) will be categorized. I think the regression outcomes would be strengthened when quantitative variables are used.

A: Thank you for your suggestion. We firstly decided to dichotomise the variables because we think the interpretation of findings would be easier. For example, with the covariate ‘sample size’ categorized in the regression model (< 100 per arm, >= 100 per arm), we could interpret that studies with small sample sizes (<100 per arm) would be associated with higher magnitude of effect sizes. However, we agree that we should consider these variables as continuous variables as you proposed. We also believe that the regression model would be strengthened in this way.

Comment: Finally, intention-to-treat analysis is particularly suitable for superiority trials, but not when equivalence and non-inferiority trials are carried out. Will the authors keep in mind this?

A: Thank you for your comment. The main advantage of ITT is that it preserves the randomisation which is important in all trials.
Furthemore, as our comparisons of interest are physical therapy intervention versus placebo or no intervention, we suppose that all included trials in our study would be categorized as superiority trials. Furthermore, we believe that most of physical therapy trials, unfortunately, did not report the type of trial (superiority, noninferiority or equivalency).

Reviewer: 3
Reviewer Name: Greta Castellini
Institution and Country: IRCCS Istituto Ortopedico Galeazzi, Italy
Competing Interests: None declared
Comments attached.

Comment: The authors have presented a topic, which can inspire health professionals dedicated to rehabilitation to improve the conduction of a clinical trial. Underling the effect of systematic bias such as the allocation concealment and the adequate use of the intention to treat analysis on the effect of an intervention can emphasise the importance of adequate perform a RCT in terms of conduction, transparency, and reporting. The manuscript is well written, however there are major concerns which I have related to (1) the unit of the analysis the authors have selected (original RCTs? Meta-analysis? Or systematic review?) (2) Authors’ choice of using PEDro scale and (3) explanation of the meta-regression analysis.
- Going through the whole manuscript, I had difficulties to understand which the unit of analysis is. In some lines, you referred to meta-analysis, in other lines to systematic reviews and even to randomized controlled trials. The aim of the manuscript is quite clear but I did not perfectly get the selection process of the unit of the analysis (meta-analysis?) and from where the authors extracted details about allocation concealment and intention to treat analysis. I have reported the lines where the reader might be confused:

A- Thank you for your comment and suggestions. We agree that we used the terms about unit of analysis interchangeably. The unit of analysis in our study is randomised controlled trials included in systematic reviews with meta-analysis evaluating treatment of non-specific low back pain. We make this clear across the whole manuscript now. We also amended in the manuscript where we extracted data to make it clearer. Please see our comments in the following topics.

Comment: Line 137: “we include meta-analysis” Line 151: “Filters related to meta-analyses and systematic reviews”. It is not clear whether you have used the “PubMed” or another database filter for meta-analysis/study design, just the free search term or the MESH term. Moreover, in the abstract you have mentioned just meta-analysis and not systematic review. Please clarify the search strategy.

A: With regards to the search strategies, as described in Appendix 2 from the manuscript, we used the terms 'systematic review' and 'meta-analysis' as free search terms. We decided to use the term 'meta-analysis' in the search strategy, because some authors use this term to define the study design (instead of systematic review). So, in this way, we try not to miss any potentially eligible study.

Comment: Line 160: "we will extract the information….. for each RCT". Here, you reported that RCT would be the unit of analysis from which extract the information regarding the effect of the intervention, the allocation concealment and the intention to treat analysis. However, in Line 172 the unit of the analysis is already changed: "two reviewers will independently extract data from selected systematic reviews". In this sentence seemed that systematic review would be selected and it is the study from which you would select the data. Please make the unit of analysis consistent throughout the text. Line 176: "Study characteristics". Are you referring to study characteristics of systematic reviews or meta-analyses. I suppose this step will not be clear until the unit of analysis is defined. If the interest is to select original RCTs from the systematic reviews, it is not actually necessary to summarize the characteristics of the systematic review but just the RCTs'. On the other hand, if the aim is to focus on meta-analysis included in the systematic review then it might have more sense to report review's characteristics. Line 186: "Individual study data will be retrieved from the met-analyses included in our study". Here, unit of analysis of the study: meta-analysis.

A- Thank you for your suggestion. As stated above, our unit of analysis will be randomised controlled trials included in systematic reviews with meta-analysis. So, we will extract data (risk of bias assessment; characteristics of the studies; results) from the original RCTs included in the systematic reviews. We believe that the manuscript is clear now.

Comment: I think it is not clear which is the systematic process you are going to follow to select studies and extract data throughout the manuscript. I suggest to re-write the section "identification and selection of studies" letting the reader better understand the selection process. For instance, "We are going to select all the meta-analysis of RCT included in published systematic review and from published meta-analysis alone on the effect of physical therapy interventions in adult patients with non-specific low back pain… on the outcome pain or disability…. " Or, "we are going to select all the original RCTs included in the systematic review with a meta-analysis…." ? Please remember that a systematic review can have or not a meta-analysis but a meta-analysis can exist also without being included in a systematic review.

A- Thank you for this suggestion. We agree that the section 'identification and selection of studies' should be re-written to make it clearer. Please see revised text below:
"All RCTs included in systematic reviews with meta-analysis evaluating physical therapy treatment in adults with non-specific low back pain that included pain or disability (as continuous variables) as the main outcomes will be included in this study. Non-specific low back pain is defined as low back pain not attributed to a specific pathology, such as nerve root compromise or serious spinal pathology. The physical therapy intervention will be compared with placebo or no intervention. We will consider all possible meta-analyses from the same systematic review, since a systematic review may contain more than one meta-analysis (e.g different outcomes)."

Comment: In data extraction you should explain the following issues: Once selected the meta-analyses (if you decided to start form the meta-analyses), will you extract all the RCTs that are included in it and look for the full text? Then, are you going to search the RCT's title in the Pedro database? … if the title is not included in the Pedro database, what are you going to do? If the RCT on the Pedro database is currently being rated, what are you going to do? Retrieve the full text and assess it? This information is missing. I think is necessary to consider all situations you can encounter and explain how to manage them.

A- Thank you for your comment. We will extract data from all RCTs included in the selected systematic reviews. After the process of identification and selection, we will have to retrieve the full texts from these RCTs to extract the data. Two authors from our study are raters of PEDro database, so it will be possible to get the full texts that are indexed on PEDro. It is important to state that about 92% of physical therapy RCTs is indexed in PEDro database1. For those RCTs not indexed in PEDro, we will make all efforts to get these full texts (searching other databases, contact authors).
About domain bias assessments, if the article is not available in the PEDro database or is currently being rated, two blinded assessors (please see in the further comment our explanation about PEDro assessment) will rate the article following the same recommendations of PEDro scale. We added this information on the manuscript.
1 Michaleff ZA, Costa LO, Moseley AM, Maher CG, Elkins MR, Herbert RD, Sherrington C. CENTRAL, PEDro, PubMed, and EMBASE are the most comprehensive databases indexing randomized controlled trials of physical therapy interventions. Phys Ther. 2011 Feb;91(2):190-7. doi: 10.2522/ptj.20100116.

Comment: Moreover, what if a systematic review included more than one meta-analysis assessing pain or disability, are you considering both or just one metaanalysis for each systematic review?

A- We intend to consider all possible meta-analyses from a systematic review, since a systematic review may contain one meta-analysis about pain intensity and other about disability. We have now added this information in the revised text.

Comment: The second remark referred to the use of the PEDro scale for the assessment of the methodological quality. The author decided to extract details about the allocation concealment and the intention to treat analysis form the assessment given by the PEDro database. Please give a reason why you have selected the PEDro scale as assessment tool. Generally, the Risk of bias tool of the Cochrane Collaboration give a most exhaustive information on the likelihood of the risk of bias and not just the presence or not of the domain in the manuscript. It might be useful to report in “the risk of bias assessment section” how the judgment of the PEDro works and how each single domain is rated (yes/no). In line 166 you stated “when the score of domain bias form an included article is not available at PEDro database, two review authors will independently assess it…”. This statement revealed the presence of two different types of bias assessment. On one hand, the Pedro score performed by external and anonymous assessors and reported on the PEDro database, on the other hand review authors (I imagine authors of this manuscript) assessing the RCTs. Who is going to perform the assessment? This method you chose can be valid and equal to the one on the PEDro database only if the assessor is blinded to the scope of the manuscript otherwise the two assessments types cannot be consider the same.

A- We understand that the Risk of Bias tool from the Cochrane Collaboration is a very useful tool. However, we decided to use the PEDro scale due to the following reasons: 1) PEDro scale has high reliability for individual ratings and consensus ratings and can be used as a continuous scale for measuring the methodological quality of trials1,2; 2) PEDro scale is strongly correlated (0.83; 95% CI 0.76 to 0.88) with the Risk of Bias from the Cochrane Collaboration3; and 3) Feasibility: as two authors from this study are raters from the PEDro database, we have access to download the score of methodological quality for the potentially included RCTs in our study. As you suggested, we added this information to the revised manuscript.
About the judgment of the PEDro scale, we added the following sentence “Each domain will be rated as ‘yes’, when the criterion is clearly satisfied, or ‘no’ when the criterion is not satisfied or the information is unclear in the text.”
With regards to your consideration about the assessment of domains bias from the RCTs not scored by member of the PEDro database.

We agree that assessors should be blinded to the scope of the manuscript. In our department, we have some researchers that are trained raters of the PEDro database and are not involved in the study. So, we will ask them to assess the domains, when necessary. We have now added this information to the text.

1- Maher, C.G., Sherrington, C., Herbert, R.D., Moseley, A.M., and Elkins, M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther. 2003; 83: 713–721
2- Shiwa, S.R., Costa, L.C.M., Moseley, A.M., Lopes, A.D., Ruggero, C.R., Sato, T.O. et al. Reproducibility of the Portuguese version of the PEDro scale. Cad Saúde Pública. 2011; 27: 2063–2068
3- Yamato TP, Maher C, Koes B, Moseley A. The PEDro scale had acceptably high convergent validity, construct validity, and interrater reliability in evaluating methodological quality of pharmaceutical trials. J Clin Epidemiol. 2017 Mar 11. pii: S0895-4356(16)30338-9. doi: 10.1016/j.jclinepi.2017.03.002

Comment: Data analysis: You are going to re-analyse the original meta-analysis and create a new one combining a pooled effect size obtained by all the RCTs included in the original meta-analysis with adequate allocation concealment and ITT analysis AND another pooled effect size obtained by the RCTs not having adequate allocation and ITT performed. Is it correct? The effect obtained will indicate if the treatment effect has been modified by the presence or absence of these domains.

A- Yes, it is correct. This is exactly our objective.

Comment: However, what are you going to do when a RCT has been rated, for instance, as NO for the allocation and YES for the intention to treat analysis? Are you investigating separately the two domains?

A- Thank you for your question. We intend to investigate and present data for the two domains separately. We have added a sentence in the manuscript to make it clearer.

Comment: I do not clearly get the reason why you have decided to investigate the presence of the association between bias and effect of the intervention from the meta-analysis included in the systematic review and not from the original RCTs included in the same SR. Theoretically, allocation concealment and intention to treat analysis might affect first the RCT results and then the meta-analysis pooled effect.

A- Thank you for your question. We decided to investigate the association between bias and effect of the intervention from a collection of meta-analyses, because it will bring more robust information. We could investigate this association in original RCTs included in the same systematic review (like a sensitivity analysis), as you mentioned. However, individual meta-analyses may inaccurately estimate these effects.

Minor comments
- Line 86-91. These paragraphs seems superfluous. The concept of what is the meta-epidemiological studies should be not stressed twice (see line 92-93).

A- Thank you for your suggestion. We modified the paragraph. Please see below:
“Meta-epidemiological studies are designed to understand the impact of study level characteristics (i.e., methodological quality, study design) in randomised clinical trials (RCTs), investigating the association between these specific study characteristics and the intervention effect estimates from collections of meta-analyses. A collection of meta-analyses is necessary, since individual meta-analyses may inaccurately estimate these effects.”

Comment: Line 201: add be after will

A- Thank you for your correction.

Comment: Line 205: I suggest explaining better why you are investigating the effect of sample size and heterogeneity as covariates in the meta-regression.

A- Thank you for your suggestion. We decided to investigate the effect of sequence generation and sample size as covariates in the meta-regression, since these variables are most consistently associated with treatment effect estimates1. We added this information in the text
Please see on the manuscript that we decided to remove the heterogeneity assessment as a covariate in the meta-regression. We believe that the variables in the meta-regression need to be evaluated at trial level and heterogeneity comes from the pooled estimates.
1Dechartres A, Trinquart L, Faber T, Ravaud P. Empirical evaluation of which trial characteristics are associated with treatment effect estimates. J Clin Epidemiol. 2016 Sep;77:24-37. doi: 10.1016/j.jclinepi.2016.04.005. Epub 2016 Apr 29.

Reviewer: 4
Reviewer Name: Iosief Abraha
Institution and Country: Health Planning Service, Regional Health Authority of Umbria, Department of Epidemiology, 06124 Perugia, Italy
Competing Interests: None declared

Comment:
- In this protocol the authors aim to assess the bias related to inadequate allocation or lack of intention-to-treat using a meta-epidemiological approach. Overall, the proposal is interesting and it is original in the context of physical therapy.
The following are the major issues that the authors will need to address.
- It is not clear how the trials will be classified. To perform this, authors will need to provided classifications based on the adequacy or not of allocation concealment (eg, adequate, inadequate, unclear) and the presence or lack of intention-to-treat. This classification will drive the review/meta-analysis (MA) categories.

A- Thank you for your question. We agree that this information should be clearer in the text. The risk of bias of included trials will be extracted from PEDro database. In PEDro scale, each domain is rated as ‘yes’, if it is considered adequate, or ‘no’ if it is considered inadequate or ‘unclear’. When the score of domain bias from an included article is not available at PEDro database (the article may not be indexed in PEDro database; or the article may be in process to be rated), two blinded assessors will independently assess it, following the same recommendations stated above. We have now added this information in the revised text.

Comment: Classification based on the intention-to-treat can be used in different ways and authors will have to clarify this. Authors should be aware that the case of intention to treat is often interlinked with the post-randomisation exclusions. They can use exclusively on the reported exclusions as performed in the meta-epidemiological study they cited (BMJ 2009;339:b3244.) but a potential misclassification may occur as trials might perform exclusions without reporting them. Alternatively, they can base their ITT categorization on reporting as performed in another meta-epidemiological study (BMJ 2015;350:h2445).

A- We agree that ITT analysis can be classified in different ways, and the number of trials reporting the use of modified ITT (as the paper you cited) has increased over the time.

However, the two main principles of ITT analysis are: 1-) all participants should be analyzed in groups in which they were originally randomised (ie, grouping subjects according to their initial random allocation); and 2-) all randomized participants must be included in the analysis (that focus in the post-randomisation exclusions).
Loss to follow-up does not prevent intention to treat analysis, nor does the failure to use an imputation technique (eg, last known value carried forward, multiple imputation) imply that the analysis was not by intention to treat. However, removal of data for any subjects because they do not start or comply with the treatment program would violate the intention to treat principle, because the subjects were still available for follow-up measurement. Complete follow-up of all subjects does not guarantee that an intention to treat analysis has been performed (eg, research authors could swap equal numbers of subjects between the groups, perhaps due to low adherence, and performed a per protocol analysis). So even when there is no loss to follow-up, articles need either to use the term “intention to treat”, state that all subjects received treatment or control conditions as allocated, or state that subjects were analysed according to their initial group allocation. Those three features are considered by the PEDro scale in order to score a ‘yes’ for the item related to ITT analysis.

Comment: Authors should be aware that deviation from intention-to-treat or modified intention to treat are consistently present in MA and trials (http://dx.doi.org/10.1016/j.jclinepi.2016.11.012; BMJ 2010;340:c2697; http://dx.doi.org/10.1186/1745-6215-12-58). Hence, they will need to clarify how will consider studies with these characteristics.

A- We agree with you, this is exactly our aim: to compare the treatment effects in trials that did or did not perform intention to treat analysis.

Comment: I am not a statistician but the paragraph on analysis should be correctly referenced. Authors may have a glance to the following paper that used continuous data BMJ 2012;344:e813. In addition, are they going to assess publication bias?

A- Thank you for your suggestion. We added the following references in the section about data analysis:
- Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-60.
- Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002;21(11):1513-24
About publication bias, we did not plan to evaluate it. We intend to focus on evaluating if selection and attrition bias influence the effect size of physical therapy interventions. We believe that to assess the presence of publication bias in low back pain trials is very important, and maybe it would be a topic for a further study.

Minor issues
Comment: Page 5; line 100-101. The statement is not entirely correct: Regarding the reference number 16: I think that in their IPD assessment Tierney and Stewart found that the experimental intervention had a positive effect when exclusions were performed, compared with the effect measured when a true intention to treat analysis was done. In addition, Melander et al (BMJ 2003;326:1171-3) found that drug treatment had a favorable effect when a per protocol analysis was used instead of an intention to treat analysis.

A- Thank you for your comment. We stated the influence of ITT analysis is still unclear in the literature, because the results from the studies are not consistent with the hypothesis that studies that did not perform ITT analysis tend to overestimate the effect of the experimental intervention. We believe the attrition bias varied between studies because of the different methods and definitions used and different clinical areas addressed.

In Melander et al 2003, they found that a per-protocol analysis could result in large overestimate of effects compared with the intention to treat analysis. However, in Nuesch et al 2009 (reference number 14), the difference in the pooled effect size between trials with and without exclusion was not statistically significant (−0.13, 95% CI −0.29 to 0.04, P=0.13). Moreover, in Tierney and Stewart 2005, the results did not demonstrate a clear tendency that ITT analysis altered the results more in favor of either treatment or control. They stated in Results section “There was no clear indication that the exclusion of patients altered the results more in one direction than another (t = 1.537, P = 0.13). Mostly the differences between the HR were small, with 70% changing by 1 to 10%, but in 17% of trials the differences ranged from 11 to as much as 35%”… “In contrast, comparing the 14 pooled meta-analysis results, there was a tendency for the HR for ‘included’ patients to be more in favor of the research treatment than the HR based on all patients (t = 2.401, P = 0.03, Figure 3). This was irrespective of whether the overall effect was in favor of this treatment or not. These differences tended to be small, between 1 and 5%, and generally might not have altered the interpretation of the results.”

VERSION 2 – REVIEW

REVIEWER	Juan Alfonso Andrade Ortega Complejo Hospitalario de Jaén Spain
REVIEW RETURNED	25-Jun-2017

GENERAL COMMENTS	In my opinion, the authors have satisfied most requirements

REVIEWER	Greta Castellini IRCCS Istituto Ortopedico Galeazzi, Università degli Studi di Milano, Italy
REVIEW RETURNED	30-Jun-2017

GENERAL COMMENTS	The authors satisfied all my request. However, some minor points need to be considered: 1. Abstract, line 27. I suggest to add characteristics after the word “methodological” or change it into methods. 2. Page 6. Line 142. I appreciate how you have clarified this paragraph however, what are you going to do if you find a meta-analysis including just one RCT? I suppose that this meta-analysis would be excluded from the sample. Since it can happen, please report what you are going to do. 3. Page 7, line 159. When you referred to full-texts, you mean systematic review from which you extracted the meta-analyses. Is it correct? If yes, please explicit the terms systematic review because the reader can be confused. 4. Page 7, line 161: I think that the point where full-texts of RCT are searching is missed. Please add a sentence in line 161. Something like: “After the process of identification and selection, we will retrieve the full texts of the systematic review in order to look for the meta-analyses of our interest. Once we selected the systematic reviews with the meta-analyses to be included in our study, we look for the full texts of the RCT included in the chosen meta-analyses in order to extract their original data……”.

	5. Page 8. Line 197. Why are you extracting also the study design type? I think it is superfluous since all trials you want to include are RCT. 6. Page 10, line 231: add be after will.

REVIEWER	Iosief Abraha Health Planning Service, Regional Health Authority of Umbria, Department of Epidemiology, 06124 Perugia, Italy
REVIEW RETURNED	01-Jul-2017

GENERAL COMMENTS	I have two major concerns: I. Regarding the reporting or performing intention-to-treat the authors provided a response that such as “when there is no loss to follow-up, articles need either to use the term “intention to treat”, state that all subjects received treatment or control conditions as allocated, or state that subjects were analysed according to their initial group allocation” These conditions are clear but not exhaustive. In other words authors should consider (i) the reporting of ITT(as they cited ITT or “all subjects received treatment…”) and (ii) whether post-randomisation exclusion occurred, and clearly state which scenarios will be considered “yes”, or “no” (“unclear” might be another category that they may consider where necessary). Please note that these two issues are different but interlinked enough to influence the author judgement. The following are some examples of scenarios in which a clarification is required: a) A trial might NOT report the term ”intention-to-treat” and may NOT provide any statement like “that all subjects received treatment or control conditions as allocated” but there is no apparent exclusions in the analysis. How will authors classify these types of trials b) Another trial DO report the term ”intention-to-treat” or DO provide a statement like“that all subjects received treatment or control conditions as allocated” but post-randomization exclusion occur . How will authors classify these types of trials c) Another trial may state “all randomly assigned patients who received at least one dose of study medication” will be included in analysis but without excluding subjects. How will authors classify these types of trials. d) Another trial may state “analysis was performed by modified intention-to-treat” and no exclusions are reported. How will authors classify these types of trials In this regard authors will need to clarify their ITT categorization. II. There is a need to remark in the introduction that deviation from ITT are common in trials and reviews (J Clin Epidemiol. 2017 Apr;84:37-46. Authors report correctly the references regarding the meta-epidemiological studies that evaluated the influence of allocation concealment, however, I suggest to consider the combined effect of several meta-epidemiological studies (Health Technol Assess 2012;16:1-82.). Regarding ITT, I think they will need to cite and acknowledge the meta-epidemiological studies that evaluated deviation from ITT (BMJ. 2015; 350: h2445) and the modified ITT (J Clin Epidemiol. 2016 Apr;72:66-74). These two meta-epidemiological studies have different approaches an reach different conclusions.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Reviewer Name: Juan Alfonso Andrade Ortega
Institution and Country: Complejo Hospitalario de Jaén, Spain
Competing Interests: None declared
Comment:In my opinion, the authors have satisfied most requirements

Reviewer: 3
Reviewer Name: Greta Castellini
Institution and Country: IRCCS Istituto Ortopedico Galeazzi, Università degli Studi di Milano, Italy
Competing Interests: None declared

The authors satisfied all my request. However, some minor points need to be considered:

1. Abstract, line 27. I suggest to add characteristics after the word “methodological” or change it into methods.

A- Thank you for the suggestion. We have added the word ‘characteristics’ in the text.

2. Page 6. Line 142. I appreciate how you have clarified this paragraph however, what are you going to do if you find a meta-analysis including just one RCT? I suppose that this meta-analysis would be excluded from the sample. Since it can happen, please report what you are going to do.

A- Thank you for the comment. We agree that the meta-analyses with only one RCT will be not included in our study. We have added this information in the text.

3. Page 7, line 159. When you referred to full-texts, you mean systematic review from which you extracted the meta-analyses. Is it correct? If yes, please explicit the terms systematic review because the reader can be confused.

A- Thank you for the suggestion. We have added the term ‘systematic review’ in the text to make it clearer.

4. Page 7, line 161: I think that the point where full-texts of RCT are searching is missed. Please add a sentence in line 161. Something like: “After the process of identification and selection, we will retrieve the full texts of the systematic review in order to look for the meta-analyses of our interest. Once we selected the systematic reviews with the meta-analyses to be included in our study, we look for the full texts of the RCT included in the chosen meta-analyses in order to extract their original data……”.

A- Thank you for the suggestion. We have added a sentence in the text to make clear the process of identification and collection of the full-text of systematic reviews and RCTs.

5. Page 8. Line 197. Why are you extracting also the study design type? I think it is superfluous since all trials you want to include are RCT.

A- We agree that to extract the study design is not necessary since all trials will be RCT. We have removed this information from the text.

6. Page 10, line 231: add be after will.
A- Thank you for the correction.

Reviewer: 4
Reviewer Name: Iosief Abraha
Institution and Country: Health Planning Service, Regional Health Authority of Umbria, Department o Epidemiology, 06124 Perugia, Italy
Competing Interests: None declared

I have two major concerns:
Comment I. Regarding the reporting or performing intention-to-treat the authors provided a response that such as “when there is no loss to follow-up, articles need either to use the term “intention to treat”, state that all subjects received treatment or control conditions as allocated, or state that subjects were analysed according to their initial group allocation” These conditions are clear but not exhaustive. In other words authors should consider (i) the reporting of ITT (as they cited ITT or “all subjects received treatment…”) and (ii) whether post-randomisation exclusion occurred, and clearly state which scenarios will be considered “yes”, or “no” (“unclear” might be another category that they may consider where necessary).
Please note that these two issues are different but interlinked enough to influence the author judgement.
The following are some examples of scenarios in which a clarification is required:

A- Thank you for this suggestion. We believe it is important to clearly describe our method of classification regarding the use of ITT or not. We have answered the questions below based on item 9 from the PEDro scale, which was the validated risk of bias tool we chose for this study. Although the research team of this study includes two trained PEDro raters and one developer of the PEDro scale, we have asked the assistance of A/Prof Anne Moseley (director of the PEDro database) to double check our answers in order to prevent any misunderstanding by the readers.

Comment a) A trial might NOT report the term ”intention-to-treat” and may NOT provide any statement like “that all subjects received treatment or control conditions as allocated” but there is no apparent exclusions in the analysis. How will authors classify these types of trials?

A- In this case, we would rate as ‘no’; unless the trial clearly states that all participants were analysed in the groups in which they were originally allocated. Sometimes, it can be clear that there were no exclusions in the analysis, but it does not guarantee that participants were analysed in the groups that they were originally allocated to. For example, research authors could swap equal numbers of subjects between the groups, due to treatment cross-over, and perform a per-protocol analysis.

Comment b) Another trial DO report the term ”intention-to-treat” or DO provide a statement like“that all subjects received treatment or control conditions as allocated” but post-randomization exclusion occur. How will authors classify these types of trials?

A- In this case, we would rate as ‘yes’ since the authors stated that analysis was based on ITT; but it should be clear that post-randomization exclusions were not related to receiving (or not) the treatment. There are some trials that excluded patients after randomization if the authors realize that a participant is not eligible for the trial (for example Hancock 20071, Garcia 20172). In these specific cases, we intent to classify as yes.

Comment c) Another trial may state “all randomly assigned patients who received at least one dose of study medication” will be included in analysis but without excluding subjects. How will authors classify these types of trials?

A- In this case, we would rate as ‘no’ since authors excluded patients that did not receive at least one dose of medication.

Comment d) Another trial may state “analysis was performed by modified intention-to-treat” and no exclusions are reported. How will authors classify these types of trials?

A- Trials often report the use of modified intention-to-treat analysis when there were deviations from the original ITT approach, so if just modified ITT analysis is stated without further information we would rate as ‘no’. However, if it is clear that there were no exclusions and ALL subjects were analysed according to their initial group allocation, it will be classified as ‘yes’.

Comment: In this regard authors will need to clarify their ITT categorization.

A- We have added more information in the text to clarify the classification of ITT analysis that will be used in our study. Please see below:
“Trials will be classified as ‘yes’ for ITT analysis, if they use the term ‘intention to treat’, and it is clear that all subjects received treatment or control conditions as allocated, or that subjects were analysed according to their initial group allocation. When there are post-randomisation exclusions, a trial will be rated as ‘no’ if the exclusion is on the basis of not receiving allocated treatment. There are some trials that exclude patients after randomization if authors subsequentially realize that the participant is not eligible for the trial.32,33 In this specific case, the trial will be classified as ‘yes’. Trials will be classified as ‘no’, if they did not mention any intention to treat approach or reported the use of a modified intention to treat approach. However, if it is clear that there were no exclusions and all subjects were analyzed according to their initial group allocation, it will be classified as ‘yes’.

Comment II. There is a need to remark in the introduction that deviation from ITT are common in trials and reviews (J Clin Epidemiol. 2017 Apr;84:37-46).

A- Thank you for the suggestion. We have added this information in the text and reference cited.

Comment III. Authors report correctly the references regarding the meta-epidemiological studies that evaluated the influence of allocation concealment, however, I suggest to consider the combined effect of several meta-epidemiological studies (Health Technol Assess 2012;16:1-82.).

A- Thank you for the suggestion. We have modified the information in the text, adding the results of the combined analysis of meta-epidemiological studies about allocation concealment.

Comment IV. Regarding ITT, I think they will need to cite and acknowledge the meta-epidemiological studies that evaluated deviation from ITT (BMJ. 2015; 350: h2445) and the modified ITT (J Clin Epidemiol. 2016 Apr;72:66-74). These two meta-epidemiological studies have different approaches and reach different conclusions.

A- Thank you for the suggestion. We have added the information in the text addressing the content (finding and different approaches for the classification of ITT analysis) from the two suggested meta-epidemiological studies.

REFERENCES
1. Hancock MJ, Maher CG, Latimer J, et al. Assessment of diclofenac or spinal manipulative therapy, or both, in addition to recommended first-line treatment for acute low back pain: a randomised controlled trial. Lancet 2007;370(9599):1638-43.
2. Garcia AN, Costa L, Hancock MJ, et al. McKenzie Method of Mechanical Diagnosis and Therapy was slightly more effective than placebo for pain, but not for disability, in patients with chronic non-specific low back pain: a randomised placebo controlled trial with short and longer term follow-up. Br J Sports Med 2017.

VERSION 3 – REVIEW

REVIEWER	Greta Castellini 1. Unit of Clinical Epidemiology, IRCCS Istituto Ortopedico Galeazzi, Milan, Italy; 2. Department of Biomedical Sciences for Health, University of Milan, Milan, Italy
REVIEW RETURNED	24-Aug-2017

GENERAL COMMENTS	The authors have fully satisfied my request.

REVIEWER	Iosief Abraha Regional Health Authority of Umbria Perugia Italy
REVIEW RETURNED	02-Aug-2017

GENERAL COMMENTS	The responses provided by the authors were satisfactory