## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Physiotherapy rehabilitation for Whiplash Associated Disorder II: a systematic review and meta-analysis of Randomised Controlled Trials |
| **AUTHORS** | Rushton A, Wright C, Heneghan N, Eveleigh G, Calvert M, Freemantle N |

## VERSION 1 - REVIEW

| | |
|---|---|
| **REVIEWER** | AP Verhagen, Dept of General Practice, ErasmusMC, Rotterdam. The Netherlands |
| **REVIEW RETURNED** | 03/08/2011 |

| | |
|---|---|
| **THE STUDY** | I do think they need advice from a statistician, because I have some doubts about the analysis |
| **RESULTS & CONCLUSIONS** | I do believe that the methods, especially the risk of bias assessment and the pooling, leeded to results that are not credible. |
| **REPORTING & ETHICS** | None |
| **GENERAL COMMENTS** | This paper is a systematic review and meta-analysis on physiotherapy rehabilitation for whiplash associated disorders II (WAD-II). The paper is nicely written, according to the guidelines of the Cochrane Collaboration and follows the PRISMA statement.<br><br>General comments<br>1. I miss a lot of studies included in other systematic reviews on whiplash, and it is unclear to me why these studies are not included here. Please explain the exclusions more clearly.<br>2. The manuscript is rather lengthy and can be written more concise. Especially the tables are really lengthy and contain a lot of unnecessary information.<br><br>Specific comments<br>1. Title. The title does not cover the content of the manuscript. It is not regarding physiotherapy rehabilitation only as also studies evaluating acupuncture are included. Also it does not evaluate WAD-II only (which is impossible because studies do not mention often this information specifically), but on WAD 0-III. Please change the title accordingly.<br>2. Introduction, page 8. In the introduction is stated that WAD can be regarded as a distinct group within the broader non-specific neck pain population. This might be true, but is not the full story as other studies do not differentiate between the two because they do not consider WAD a distinct group. Please discuss the matter in a broader perspective.<br>3. Method, eligibility criteria, page 10. As stated before, please delete the II from WAD-II as a selection criterion, because you did not select only studies including WAD-II. |

| | |
|---|---|
| | 4. Method, eligibility criteria, page 10. Please add pain as an outcome measure. This is missing in the abstract and method, but appeared to be the main outcome measure in the results.<br>5. Method, risk of bias assessment, page 12. I think the authors do not fully understand the risk of bias (RoB) assessment. First, they do not actually use the tool from the Back Review Group, as described. Second, the tool does not distinguish between assessment of a study protocol or the full paper, but distinguishes between outcome measures used. Having also a study protocol available means that you can assess selective outcome reporting, but more information about RoB can often be presented in the protocol as compared to the study publication, but should be considered as one. So RoB assessment within studies is useless and wrong. Third, it seems that all studies are regarded as having high RoB because one or more items are assessed as high RoB. This is extremely strict and unusual. Furthermore it is unclear when studies score high RoB on item 6 (other potential threats to validity) as almost all studies seem to score negative on this item. This part of the method should be revised majorly, because it does not follow the Cochrane guidelines at all.<br>6. Results. All information on WAD-II or else can be shortened or deleted at all, because it varies between studies and one cannot make subgroups of studies based on this information as far as I am concerned.<br>7. Results, page 18. When the authors still think that all studies are of high RoB they really should reconsider whether to pool the results at all, as "when you pool rubbish, you get rubbish as a result".<br>8. Results, page 18. I have sincere doubts whether you can actually pool the studies in the two groups as the authors suggest, because pooling taping, exercise and acupuncture does not seem very clinically relevant. I think there is room to pool exercise studies as long as the authors do not separate the different WAD categories.<br>9. Results, page 19. The first paragraphs seems like vote counting. Please do only mention the pooled results (or any other summary measure).<br>10. Conclusion. I think the conclusion is way too strict and do not follow the conclusions properly. Also the authors did not perform an analysis according to the GRADE approach (as recommended by the Cochrane Back Review Group), therefore any mention about the quality of the evidence cannot be made.<br>11. Tables. Please provide in table 3 the actual data on the outcomes, only presenting p-values is not informative. Please drastically shorten this table too. Please delete table 5 because it contains redundant information.<br>12. Figures. Please mention not only the outcome measure of the figure, but also the measurement timing (e.g. short-term). Please also ask some statistical help, because in my opinion one cannot pool means and standard deviations together with mean changes (as in Scholten-Peeters et al see figure 2), but I am not sure. Also, when all studies report pain on a VAS (as I consider they did), please calculate WMD instead of SMD now. |

## VERSION 1 – AUTHOR RESPONSE

Point to address from BMJ Open reviewer

I do think they need advice from a statistician, because I have some doubts about the analysis.

One of the authors of the paper is a statistician and two other authors have considerable expertise in data analysis, particular systematic reviews and trials.

I do believe that the methods, especially the risk of bias assessment and the pooling, leeded to results that are not credible.

1] The risk of bias assessment was rigorous as justified in our response to specific comments below, and we hope that our response supports the credibility of our results and conclusions.
2] The pooling of results, when evaluated critically within the context of the assessment of high risk of bias did enable an overview of the evidence to be presented, to advance understanding.

This paper is a systematic review and meta-analysis on physiotherapy rehabilitation for whiplash associated disorders II (WAD-II). The paper is nicely written, according to the guidelines of the Cochrane Collaboration and follows the PRISMA statement.

Thank you for this positive feedback regarding the structure and format of the paper.

General comments

1. I miss a lot of studies included in other systematic reviews on whiplash, and it is unclear to me why these studies are not included here. Please explain the exclusions more clearly.

The authors agree that some trials included in other reviews are missing from this review. This is due to our focus to trials that identify inclusion of the Whiplash Associated Disorder II (WADII) classification within their population. Other reviews have drawn conclusions that are difficult to apply in clinical practice owing to inclusion of broader populations without attention to the difference between WAD classifications.
As stated on page 8, previous systematic reviews focused on a range of classifications, interventions, and inclusion of broader neck pain populations.
Your systematic review (reference 17) focused on WAD classifications I and II and management strategies broader than physiotherapy.
Our specific research question and pre-specified inclusion / exclusion criteria were focused on the management of WAD II patients who are those seen most commonly for conservative treatment by physiotherapists, thereby representing the individuality of this review.
The inclusion and exclusion criteria are detailed within the eligibility criteria (page 9) and inclusion / exclusion criteria (Table 2, page 12).
Figure 1 details the exclusions and the 23 articles excluded as they did not detail inclusion of a WADII population.

2. The manuscript is rather lengthy and can be written more concise. Especially the tables are really lengthy and contain a lot of unnecessary information.

We have reviewed the manuscript and made minor changes wherever possible to ensure that the manuscript is concise. Table 5 has been deleted.

Specific comments

1. Title. The title does not cover the content of the manuscript. It is not regarding physiotherapy rehabilitation only as also studies evaluating acupuncture are included. Also it does not evaluate WAD-II only (which is impossible because studies do not mention often this information specifically), but on WAD 0-III. Please change the title accordingly.

1] The objective of this review was to focus on WADII and therefore the title has been maintained. We believe that the review is justified in drawing conclusions re the WADII population as all trials

included this population. We have also attempted to obtain specific data from authors re WAD classification to inform our conclusions.
Figure 1 and 'study selection' (page 17) detail the categories of trials in line with the point that you raise re many trials not evaluating effectiveness in a WADII population only. This is one of the points that the review highlights and we have strengthened this in the discussion (page 25).
2] Acupuncture as an intervention is within the scope of physiotherapy practice as a conservative management intervention in many countries and is therefore included.

2. Introduction, page 8. In the introduction is stated that WAD can be regarded as a distinct group within the broader non-specific neck pain population. This might be true, but is not the full story as other studies do not differentiate between the two because they do not consider WAD a distinct group. Please discuss the matter in a broader perspective.

The text has been amended to reflect this broader perspective as suggested (page 7), and refers to your recent paper addressing this issue (reference 14).

3. Method, eligibility criteria, page 10. As stated before, please delete the II from WAD-II as a selection criterion, because you did not select only studies including WAD-II.

We have not deleted the WADII distinction as explained above. We included only trials that included WADII participants.
We have added a point page 10 to clarify this and to strengthen the point that some trials also included other classifications.
A sentence has been added into the discussion (page 25) regarding the problems of combining multiple classifications within one population that would in practice, receive different management interventions.

4. Method, eligibility criteria, page 10. Please add pain as an outcome measure. This is missing in the abstract and method, but appeared to be the main outcome measure in the results.

Our inclusion criteria for outcome measures are written as broad rather than defining individual measures. We chose to include all domains within the WHO definition of function, disability, and health as inclusion criteria. This definition includes pain within the 'sensory functions and pain' level 1 classification. It was one of the main outcomes discussed from the results as it was one of the few measures used consistently across trials to enable meta-analysis.
The detail has been amended on pages 4 and 10 to provide greater clarity. Content under 'data items' (page 14) does explain this further and specifically mention pain.

5. Method, risk of bias assessment, page 12. I think the authors do not fully understand the risk of bias (RoB) assessment. First, they do not actually use the tool from the Back Review Group, as described. Second, the tool does not distinguish between assessment of a study protocol or the full paper, but distinguishes between outcome measures used. Having also a study protocol available means that you can assess selective outcome reporting, but more information about RoB can often be presented in the protocol as compared to the study publication, but should be considered as one. So RoB assessment within studies is useless and wrong. Third, it seems that all studies are regarded as having high RoB because one or more items are assessed as high RoB. This is extremely strict and unusual. Furthermore it is unclear when studies score high RoB on item 6 (other potential threats to validity) as almost all studies seem to score negative on this item. This part of the method should be revised majorly, because it does not follow the Cochrane guidelines at all.

The authors do not describe and have not used the tool from the Back Review Group but have used the Cochrane risk of bias assessment tool as advocated by the Cochrane handbook 2011 and PRISMA 2009 (see page 14).
The tool is focused to the trial report, but evaluation of points 5a and 5b (re selective outcome

reporting, Table 4, page 20) are dependent on previous publication of a study protocol to enable evaluation as a low risk of bias. The protocol and trial report are therefore considered as one.
The interpretation of high risk of bias overall based on ≥1 item evaluated as high risk of bias is in line with the advocated use of the tool by Cochrane 2011. This is strict but not unusual and many authors are demanding further rigour in evaluating quality of trials as we support. We have added in the rigour of our evaluation as a point on page 14 and supported it with a paper that we have in press at present (reference 30).
The reasons for trials scoring high risk of bias on item 6 are all detailed in Table 4 (page 20).
We have not made changes to this method as it is in line with use of the Cochrane risk of bias assessment tool 2011. All decisions are clearly reported and justified to enable evaluation by the reader.
We have added the PRISMA reference in on page 14 to further support our use of the Cochrane risk of bias assessment tool.

6. Results. All information on WAD-II or else can be shortened or deleted at all, because it varies between studies and one cannot make subgroups of studies based on this information as far as I am concerned.

Further to our response above, our focus to the WADII population has been maintained. The categorisation of trials (page 17) makes explicit the different populations included in the review to inform the analysis and future debate.

7. Results, page 18. When the authors still think that all studies are of high RoB they really should reconsider whether to pool the results at all, as "when you pool rubbish, you get rubbish as a result".

Results of the meta-analysis are reported in the context of the overall risk of bias as detailed on page 15. The authors believe that the pooling of data adds new information in this context to inform readers. This justification is detailed on page 22. The discussion carefully explores findings in this context (page 24).

8. Results, page 18. I have sincere doubts whether you can actually pool the studies in the two groups as the authors suggest, because pooling taping, exercise and acupuncture does not seem very clinically relevant. I think there is room to pool exercise studies as long as the authors do not separate the different WAD categories.

The authors believe that the pooling into two groups of specific physiotherapy interventions or a multimodal active intervention is justified (page 18) as these represent two distinct approaches in clinical physiotherapy management. It is acknowledged that the interventions within each grouping are broad, but the authors argue that the analysis does add to current analysis and debate. Heterogeneity is subsequently explored within the discussion (page 25) and as it was anticipated, it was accounted for by using the random effects model.

9. Results, page 19. The first paragraphs seems like vote counting. Please do only mention the pooled results (or any other summary measure).

The authors are unsure of this point as the content (page 23) provides clarity of the interpretation of the Forest plots to ensure that the reader possesses all relevant information.

10. Conclusion. I think the conclusion is way too strict and do not follow the conclusions properly. Also the authors did not perform an analysis according to the GRADE approach (as recommended by the Cochrane Back Review Group), therefore any mention about the quality of the evidence cannot be made.

The conclusion is intended to be strict and clear based on the findings of the review. We believe

that this is justified based upon our disappointing findings to ensure that the key messages from the paper are clear. To strengthen the justification for the strength of the conclusions we have used the GRADE analysis as suggested (page 26).

11. Tables. Please provide in table 3 the actual data on the outcomes, only presenting p-values is not informative. Please drastically shorten this table too. Please delete table 5 because it contains redundant information.

The focus of our review is to identify evidence of comparative effectiveness across interventions. The 'main results' column of Table 3 (at end of text) therefore focused on the statistical significance of any difference between groups. The Forest plots then present standardised data on key outcomes to enable comparison.
It has not been possible to shorten Table 3 as all data is required to support interpretation of the review. This content had already been discriminated as far as possible. The online publication of BMJ Open facilitates the transparency of information in the table.
Table 5 has been deleted as the data can be found in Table 3 as suggested.

12. Figures. Please mention not only the outcome measure of the figure, but also the measurement timing (e.g. short-term). Please also ask some statistical help, because in my opinion one cannot pool means and standard deviations together with mean changes (as in Scholten-Peeters et al see figure 2), but I am not sure. Also, when all studies report pain on a VAS (as I consider they did), please calculate WMD instead of SMD now.

Clarification of 'short-term' has been added to the titles of the figures.
A combination of final value scores and change from baseline scores can be pooled as stated (page 15). The reference support of the Cochrane handbook has been added to support this point (section 9.4.5.2 Cochrane handbook).
The Cochrane handbook 2011 (section 9.2.3.1) advises against using the terminology of WMD, and recommends use of MD. We have then calculated SMD as pain has been measured in a variety of ways (numerical rating scales and VAS, 0-10 and 0-100 scales).