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)	Factors associated with return-to-work among people on work absence due to long-term neck or back pain: a narrative systematic review
AUTHORS	Rashid, Mamunur; Kristofferzon, Marja-Leena; Nilsson, Annika; Heiden, Marina

VERSION 1 - REVIEW

REVIEWER	Achim Elfering University of Bern, Switzerland
REVIEW RETURNED	01-Dec-2016

GENERAL COMMENTS	The ms reports a review on factors associated with return-to-work among people on work absence due to long-term neck/shoulder or back pain. 10 Studies were included in the review. The current study has several methodological weaknesses:
	1) Out of the 10 studies thet were included only 2 studies addressed pain in the neck, no study addressed pain in shoulders. There is no rationale to treat the review to refer to neck pain or shoulder pain.
	2) Studies with WA due to acute pain were intended to be excluded by including only studies with WA > = 2 weeks. However, refers WA > = 2 weeks to a mean or minimum value? The most used time-based category, hover reads acute = duration of pain less than 1 month, subacute = duration up to 3 months, chronic = duration of pain more than 3 months. Therefore, WA > = 2 weeks in this study failed to exclude individuals with acute pain.
	3) It is unclear whether the 10 studies that were included were included in previous reviews. Therefore, it is unclear whether the current review adds new information to the research field.

REVIEWER	Daniela Milani
	Universidade Estadual do Centro-Oeste (UNICENTRO)
	Brazil
REVIEW RETURNED	04-Jan-2017

GENERAL COMMENTS	Thanks for allowing me the opportunity to review this interesting paper.
	The authors appear to have done a clear/transparent study and reported this in the same manner.
	There are not many comments and my considerations are related

especially on the abstract.

I also was not able to see/access the complete supplementary reporting on the PRISMA checklist (not sure if the chacklist was not submitted or if it was not demanded to the authors when they submit). I've tryed to follow and identify the items of the study on the PRISMA checklist by myself and have found the majority of it on the report. However, it would be easier if the authors had submited this.

Abstract:

Make clear that the observational studies included were the longitudinal (prospective) ones.

In the Results session of the abstract it says: "seven studies fulfilled the methodological standard to be included...". When we read only the abstract we think that only these seven studies made part of the report. However, the entire report says that ten papers made part of this report. As related in the full report, the three other studies were included by the additional search through reference list and citations, of the included studies. So, it is important to relate this in the abstract. At least to say that made part of the entire report the ten studies instead only the seven ones.

It is also important to specify in the abstract what the authors considered "long term neck/shoulder or back pain": ≥15 days (two weeks). This is clear in the full report, but is missing in the abstract.

I suggest rewrite the conclusion and/or the results of the abstract. Both sessions are saying much the same thing when they could use the space to provide other more information.

REVIEWER	Dr Darnel Murgatroyd
	John Walsh Centre for Rehabilitation Research
	The University of Sydney
	The Kolling Institute
	North Shore Hospital
	St Leonards
	Sydney, NSW
	Australia
REVIEW RETURNED	01-Feb-2017

GENERAL COMMENTS	Thank you for the opportunity to review this paper.
	Summary
	The authors have conducted a literature review to investigate the prognostic factors for Return To Work (RTW) in people with chronic neck/shoulder or back pain. They summarise the results from a number of papers but the manuscript is difficult to read and requires further revision prior to it being ready for publication. The methodology and results require further explanation, which could assist the reader to interpret the findings. It is unclear what new information this study adds in the absence of addressing these issues. Please see comments below.

Abstract

1. The Abstract reflects the content of the article but needs

- revising based on comments for each section. It is unclear whether the authors conducted a literature review or systematic review due to the wording (e.g. systematically summarise, systematic literature search), this should be explicit in the abstract.
- The methods section should clearly define what 'guidelines of prognostic studies' was used to assess the results. For example, it is unclear from the methods how the authors determined 'sufficient' and 'insufficient or inconclusive'.

Introduction

3. In the introduction the authors refer to Work Absence (WA) and RTW, it is not clear whether they are interested in one or both measures. This could be assisted by a definition of both, either in the introduction or methods as they are not the same thing and the authors at times use the terms interchangeably, which is confusing for the reader.

Methods

- 4. The methods section requires extensive expansion to illustrate the study design, methods, statistical analysis. It is difficult to follow, and there is a lack of definitions.
- 5. The authors only refer to three databases, there is also Embase, Informit, Web of Science. Why were these databases not included.
- 6. The search strategy is also very short, were all terms explored? For example, why were the terms 'cohort studies, follow up studies, longitudinal studies' not used. Did the authors consult a medical librarian to assist them in developing the search strategies? If so, it should be reported.
- 7. The results should not be reported in the methods section, paragraph 1 of 'Selection process should be moved to Results. To report the results comprehensively the authors should include in Figure 1, which investigators (and how many) reviewed the studies at each period. The reasons for excluding studies should be presented and summarised either in the results or figure and put in each section (i.e. main reason excluded based on title and abstract, full text). A full list of reasons for exclusions should be attached to the Appendix for those reviewed and excluded based on full text (n=44).
- 8. For clarity list the inclusion and exclusion criteria in dot points or a table to assist the reader.
- Last sentence of paragraph 2 under 'Selection process' should be moved to Results.
- 10. Were the references and included studies of relevant 'reviews and meta-analysis' checked to see if they should be included in this review?
- 11. Definitions should be included for chronic/long-term back, neck/shoulder pain, and WA and RTW, particularly for RTW as there is no standardised definition. Did the authors use the study definitions?
- 12. The quality assessment methodology for systematic reviews of prognostic studies is not standardised and can be controversial, particularly if scoring is used. The authors need to elaborate and clearly articulate the reasons behind

- why they selected this method (in the 'Quality assessment' section). What was the rationale for scoring, setting the response rate limits, why only 'age, gender, prior WA and co-morbidity' as potential confounders? What was considered 'appropriate statistical analysis'? Just to have references for these decisions is insufficient; the authors need to explain their reasons.
- 13. There needs to be further explanation in the data synthesis section about how the results from the quality assessment were interpreted 'we categorised the factors into categories' what categories? This is partly presented below in 'Methodological quality' this section needs to move up to the 'Data synthesis' section. Why were 5-12 points considered 'adequate or good'? This is not the same as 'sufficient' and what is written in the Abstract, it needs to be consistent.
- 14. Importantly, why were low quality studies excluded and only the results of adequate or good quality studies included? Generally, all studies are included, then the quality evidence is weighted according to the quality of the study. The authors need to carefully articulate their reasons here or preferably include all studies (of all quality).

Results

- 15. There needs to be a short summary statement or paragraph at the beginning here prior to presenting Table 1. The characteristics of the studies should be summarised for the reader.
- 16. Each factor needs to be defined prior to being summarised. For example, how were recovery beliefs defined in the studies, what is the accepted definition, is there one, did the studies meet that. Health-related factors is very general, what were the measures used for quality of life and health transition. Summarise this for the reader.
- 17. Table 1 requires some additional clarification. The length of back and/or neck/shoulder pain should be defined. In sample size/drop-out why N/A? Why is there a column for non-significant factors that is largely blank. Work status needs to be defined, in the text and the table. How did each study define this and present in the table (e.g. RTW Yes/No, or further details such as duration, full/part-time hours/duties). How were the significant factors measured, what measure was used, this needs to be included.
- 18. There needs to be additional presentation of the results in terms of how the authors graded/synthesised the evidence in a table, to show which studies showed what factors as being significant, not significant, scores, ranking etc. There are many published reviews to guide in this regard.

Discussion

- 19. Commentary for the Discussion is limited as the authors need to address the issues in the Methods and Results first, which may alter the presentation of their findings and any discussion.
- 20. Parts of the Discussion need to be shifted to the Results (e.g. the sentences beginning 'In the two studies included.....recovery beliefs were assessed using 3

- items.....' in the paragraph 'Recovery beliefs', here the authors are summarising the results from the papers. Similarly in 'Workplace factors' some of the results appear to be reported here.
- 21. The section entitled 'Methodological considerations....' In particular the first two paragraphs of this section would be better placed in the Methods. The last paragraph beginning 'As effect sizes...' fits into the Results.
- 22. Again, it is unclear in 'Strengths and limitations' why low quality studies that fit the inclusion criteria were excluded from the analysis.
- 23. The Discussion will require further thought and analysis of the results, particularly if the above sections are addressed.

VERSION 1 – AUTHOR RESPONSE

Comments by Reviewer 1:

1. Out of the 10 studies thet were included only 2 studies addressed pain in the neck, no study addressed pain in shoulders. There is no rationale to treat the review to refer to neck pain or shoulder pain.

Answer: It is true that our literature search, although aiming for populations with pain in the neck/shoulder or back, did not find relevant studies of people with shoulder pain. Therefore, we have removed "shoulder" from the title and conclusion.

2. Studies with WA due to acute pain were intended to be excluded by including only studies with WA > = 2 weeks. However, refers WA > = 2 weeks to a mean or minimum value? The most used timebased category, hover reads acute = duration of pain less than 1 month, subacute = duration up to 3 months, chronic = duration of pain more than 3 months. Therefore, WA > = 2 weeks in this study failed to exclude individuals with acute pain.

Answer: With respect to work absence, inclusion criteria were a minimum of 2 weeks. This limit was set to avoid inclusion of studies on prognostic factors related to acute trauma, since they might differ substantially from prognostic factors related to long-term pain. We agree with the reviewer that definitions are not consistent across studies, and have added to table 1 the information about pain duration that was reported in the studies. We have also added text in the methods (page 5, line 3-6) to address this.

3. It is unclear whether the 10 studies that were included were included in previous reviews. Therefore, it is unclear whether the current review adds new information to the research field.

Answer: The 10 studies have been included in previous reviews. However, none of these reviews addressed the natural course of return to work among people with long-term pain in the neck/shoulder or back. We have also added the text to the discussion (page 19, line 2-6).

Comments by Reviewer 2:

There are not many comments and my considerations are related especially on the abstract. I also was not able to see/access the complete supplementary reporting on the PRISMA checklist (not sure if the chacklist was not submitted or if it was not demanded to the authors when they submit). I've tryed to follow and identify the items of the study on the PRISMA checklist by myself and have found the majority of it on the report. However, it would be easier if the authors had submited this.

Answer: A completed PRISMA checklist is now submitted as supplementary material.

Abstract:

Make clear that the observational studies included were the longitudinal (prospective) ones.

Answer: It is now specified in the abstract that nine of the ten studies were prospective cohort studies, and one was retrospective.

In the Results session of the abstract it says: "seven studies fulfilled the methodological standard to be included...". When we read only the abstract we think that only these seven studies made part of the report. However, the entire report says that ten papers made part of this report. As related in the full report, the three other studies were included by the additional search through reference list and citations, of the included studies. So, it is important to relate this in the abstract. At least to say that made part of the entire report the ten studies instead only the seven ones.

Answer: We realize that table 1 may have given the impression that all 10 studies were included in the data synthesis. This was not the case, as 3 studies were excluded due to having low quality (score ≤ lowest tertile of quality scores). In light of comment 14 by reviewer 3, and given the similarity in quality scores between most studies, we have now included all 10 studies in the data synthesis. As a result, the abstract has been revised.

It is also important to specify in the abstract what the authors considered "long term neck/shoulder or back pain": ≥15 days (two weeks). This is clear in the full report, but is missing in the abstract.

Answer: The information has been added to the abstract.

I suggest rewrite the conclusion and/or the results of the abstract. Both sessions are saying much the same thing when they could use the space to provide other more information.

Answer: Due to the inclusion of all 10 studies in the data synthesis, the abstract has been revised.

Comments by Reviewer-3

Abstract

1. The Abstract reflects the content of the article but needs revising based on comments for each section. It is unclear whether the authors conducted a literature review or systematic review due to the wording (e.g. systematically summarise, systematic literature search), this should be explicit in the abstract.

Answer: Following the recommendation by the editor (comment #6), the review has been explicitly named as a narrative systematic review in the abstract.

2. The methods section should clearly define what 'guidelines of prognostic studies' was used to assess the results. For example, it is unclear from the methods how the authors determined 'sufficient' and 'insufficient or inconclusive'.

Answer: The name of the guidelines for quality assessment is now stated in the abstract. Due to the inclusion of all 10 studies in the data synthesis, the abstract has been revised.

Introduction

3. In the introduction the authors refer to Work Absence (WA) and RTW, it is not clear whether they are interested in one or both measures. This could be assisted by a definition of both, either in the introduction or methods as they are not the same thing and the authors at times use the terms interchangeably, which is confusing for the reader.

Answer: We agree with the reviewer that work absence (WA) and return to work (RTW) is not the same thing. For clarification, we have added the definition of WA and RTW used in the study under the methods section (page 5, line 3-6 and page 5, line 23-27).

Methods

4. The methods section requires extensive expansion to illustrate the study design, methods, statistical analysis. It is difficult to follow, and there is a lack of definitions.

Answer: Definitions of neck/shoulder and back pain, work absence and return to work have been added in the methods section, and the selection of studies, quality assessment and data synthesis have been described in more detail.

5. The authors only refer to three databases, there is also Embase, Informit, Web of Science. Why were these databases not included.

Answer: We agree that there are many literature databases available. Through our search in Medline (Web of science), CINAHL and PsycInfo, 57% of the identified articles were duplicates. Therefore, we did not expect additional databases to contribute significantly to the findings.

6. The search strategy is also very short, were all terms explored? For example, why were the terms 'cohort studies, follow up studies, longitudinal studies' not used. Did the authors consult a medical librarian to assist them in developing the search strategies? If so, it should be reported.

Answer: During the literature search, a medical librarian was consulted. She has now been added to the acknowledgements section of the manuscript. The terms used in the search were defined based on the ideas behind the PICO model [1]. We selected suitable keywords for P (population), I (intervention), O (outcome), while C (comparison) was excluded since comparative studies were not the focus for this study (see table below). Keywords pertaining to the design of the study were not used as that might cause important studies to be missed.

P (population)	P (population)	I (intervention)	O (outcome)
neck	pain	factor*	return* to work
back	ache	prognos*	return-to-work
shoulder			job re-entry
lumbar			work absence
spine			work ability
spinal			ability to work

- 1. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. (2012) *Qualitative health research* 22: 1435-43.
- 7. The results should not be reported in the methods section, paragraph 1 of 'Selection process should be moved to Results. To report the results comprehensively the authors should include in Figure 1, which investigators (and how many) reviewed the studies at each period. The reasons for excluding studies should be presented and summarised either in the results or figure and put in each section (i.e. main reason excluded based on title and abstract, full text). A full list of reasons for exclusions should be attached to the Appendix for those reviewed and excluded based on full text (n=44).

Answer: Changes have been made accordingly, and a full list of reasons for exclusions is available in appendix 2.

8. For clarity list the inclusion and exclusion criteria in dot points or a table to assist the reader.

Answer: Changes have been made accordingly.

9. Last sentence of paragraph 2 under 'Selection process' should be moved to Results.

Answer: Changes have been made accordingly.

10. Were the references and included studies of relevant 'reviews and meta-analysis' checked to see if they should be included in this review?

Answer: Yes. In the citation search, reviews and meta-analyses were identified and checked for relevant included studies and references. Text has been added to the methods (page 4, line 16-19) to clarify this.

11. Definitions should be included for chronic/long-term back, neck/shoulder pain, and WA and RTW, particularly for RTW as there is no standardised definition. Did the authors use the study definitions?

Answer: The definitions requested have been added in the Selection of studies methods section.

12. The quality assessment methodology for systematic reviews of prognostic studies is not standardized and can be controversial, particularly if scoring is used. The authors need to elaborate and clearly articulate the reasons behind why they selected this method (in the 'Quality assessment' section). What was the rationale for scoring, setting the response rate limits, why only 'age, gender, prior WA and co-morbidity' as potential confounders? What was considered 'appropriate statistical analysis'? Just to have references for these decisions is insufficient; the authors need to explain their reasons.

Answer: We agree that quality assessment of prognostic studies is yet to be standardized. According to the work by Hayden et al [1], adequate quality assessment includes judgements of 6 areas of potential biases: study population, study attrition, measurement of prognostic factors, measurement of outcomes, measurement of and controlling for confounding variables, and analysis approaches. Following these recommendations, we have quality-assessed each of the areas. No weighting was used, as we did not consider any area of potential bias to be more important than another. Text has been added to the *Quality assessment* section to explain the scoring and the reasons behind calculating a quality sum score of each study.

- 1. Hayden JA, Cote P, Bombardier C. (2006) Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 144: 427-37.
- 13. There needs to be further explanation in the data synthesis section about how the results from the quality assessment were interpreted 'we categorised the factors into categories' what categories? This is partly presented below in 'Methodological quality' this section needs to move up to the 'Data synthesis' section. Why were 5-12 points considered 'adequate or good'? This is not the same as 'sufficient' and what is written in the Abstract, it needs to be consistent.

Answer: The data synthesis has been explained in more detail. Considering your next comment, and given the similarity in quality scores between most studies, we have now included all 10 studies in the data synthesis. Thus, the results have been revised throughout the manuscript.

14. Importantly, why were low quality studies excluded and only the results of adequate or good quality studies included? Generally, all studies are included, then the quality evidence is weighted according to the quality of the study. The authors need to carefully articulate their reasons here or preferably include all studies (of all quality).

Answer: In the revised manuscript, all studies have been included.

Results

15. There needs to be a short summary statement or paragraph at the beginning here prior to presenting Table 1. The characteristics of the studies should be summarised for the reader.

Answer: A short summary has been included at the beginning of the results section.

16. Each factor needs to be defined prior to being summarised. For example, how were recovery beliefs defined in the studies, what is the accepted definition, is there one, did the studies meet that. Health-related factors is very general, what were the measures used for quality of life and health transition. Summarise this for the reader.

Answer: In the revised manuscript, detailed information is given about each factor, and how it was measured, in the results section and in table 4.

17. Table 1 requires some additional clarification. The length of back and/or neck/shoulder pain should be defined. In sample size/drop-out why N/A? Why is there a column for nonsignificant factors that is largely blank. Work status needs to be defined, in the text and the table. How did each study define this and present in the table (e.g. RTW Yes/No, or further details such as duration, full/part-time – hours/duties). How were the significant factors measured, what measure was used, this needs to be included.

Answer: Table 1 (now table 2) has been changed according to the suggestions. N/A is short for "not available" (see table text), and used when the information was not presented in the article. The reason that the column for nonsignificant factors is largely blank is that nonsignificant factors were seldom reported, and sometimes disregarded before fitting the prediction model. Table 4 contains information about how the factors were measured in the studies.

18. There needs to be additional presentation of the results in terms of how the authors graded/synthesised the evidence in a table, to show which studies showed what factors as being significant, not significant, scores, ranking etc. There are many published reviews to guide in this regard.

Answer: In the revised manuscript, table 4 describes how the evidence was synthesized.

Discussion

19. Commentary for the Discussion is limited as the authors need to address the issues in the Methods and Results first, which may alter the presentation of their findings and any discussion.

Answer: The discussion has been revised according to previous comments.

20. Parts of the Discussion need to be shifted to the Results (e.g. the sentences beginning 'In the two studies included.....recovery beliefs were assessed using 3 items.....' in the paragraph 'Recovery beliefs', here the authors are summarising the results from the papers. Similarly in 'Workplace factors' some of the results appear to be reported here.

Answer: Changes have been made accordingly.

21. The section entitled 'Methodological considerations....' In particular the first two paragraphs of this section would be better placed in the Methods. The last paragraph beginning 'As effect sizes...' fits into the Results.

Answer: The methods, results and discussion sections have been revised.

22. Again, it is unclear in 'Strengths and limitations' why low quality studies that fit the inclusion criteria were excluded from the analysis.

Answer: In the revised manuscript, all studies have been included. As a result, the *Strengths and limitations* section has been revised.

23. The Discussion will require further thought and analysis of the results, particularly if the above sections are addressed.

Answer: The discussion has been revised according to previous comments.

VERSION 2 – REVIEW

REVIEWER	Prof. Dr. Achim Elfering University of Bern, Switzerland	
REVIEW RETURNED	D 04-Apr-2017	
GENERAL COMMENTS	All my points were addressed sufficiently. The author(s) did a good job in the revision.	

REVIEWER	Dr Darnel Murgatroyd John Walsh Centre for Rehabilitation Research, The University of
	Sydney, Kolling Institute, Sydney, NSW, Australia
REVIEW RETURNED	03-Apr-2017

GENERAL COMMENTS	Thank you for the opportunity to review this paper again. The authors have done an excellent job in addressing the reviewer comments, thank you. The remaining comments are minor: Introduction 1. The first mention of the word 'musculoskeletal disorders' shortened to 'MSDs' should be in capitals for the first letters. 2. Return to Work does not require hyphens. 3. The second use of MSDs should include an 's' Methods 4. Is there a definition of 'long-term' neck/shoulder or back pain. The authors provide a time frame definition of work absence, and state in 'Selection of studies' that they wished to avoid acute injury or trauma, but is long-term pain defined as longer than 2 weeks or as pain that was not caused by an acute injury or trauma requiring the time of work/work absence (i.e. the pain has been present for some time prior to the latest work absence). 5. The exclusion criterion 'mixed or wrong population' needs to be explained, defined, or perhaps is not necessary if the inclusion criteria is explicit. 6. The data synthesis (and abstract) state that 'we labelled each category according to the aspects covered by the factors in the category.' This statement is difficult to interpret, do the authors mean each category was labelled according to the significant

VERSION 2 – AUTHOR RESPONSE

Comments by Reviewer 3:

Introduction

1. The first mention of the word 'musculoskeletal disorders' shortened to 'MSDs' should be in capitals for the first letters.

Answer: Changes have been made accordingly.

2. Return to Work does not require hyphens.

Answer: The hyphens have been removed.

3. The second use of MSDs should include an 's'

Answer: Changes have been made accordingly.

Methods

4. Is there a definition of 'long-term' neck/shoulder or back pain? The authors provide a time frame definition of work absence, and state in 'Selection of studies' that they wished to avoid acute injury or trauma, but is long-term pain defined as longer than 2 weeks or as pain that was not caused by an acute injury or trauma requiring the time of work/work absence (i.e. the pain has been present for some time prior to the latest work absence).

Answer: Long-term neck/shoulder or back pain was defined as pain that was not attributed to acute injury or trauma requiring at least 2 weeks part-time or full-time absenteeism from work. This has been clarified in the *Selection of studies* section.

5. The exclusion criterion 'mixed or wrong population' needs to be explained, defined, or perhaps is not necessary if the inclusion criteria is explicit.

Answer: The wording "mixed or wrong population" has been rephrased to indicate populations who partly consisted of subjects fitting our inclusion criteria but for whom the results were not reported specifically.

6. The data synthesis (and abstract) state that 'we labelled each category according to the aspects covered by the factors in the category.' This statement is difficult to interpret, do the authors mean each category was labelled according to the significant findings/factors in the studies – consider rephrasing.

Answer: The statement has been rephrased.

Comments by Reviewer 1:

All my points were addressed sufficiently. The author(s) did a good job in the revision.

Answer: Thank you!

VERSION 3 – REVIEW

REVIEWER	Dr Darnel Murgatroyd
	John Walsh Centre for Rehabilitation Research, The University of
	Sydney, Kolling Institute, Sydney, NSW, Australia
REVIEW RETURNED	13-Apr-2017

GENERAL COMMENTS The authors have addressed all additional comments, thank you. I

have nothing further.