PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients
AUTHORS	Andrew Beswick, Vikki Wylde, Rachael Gooberman-Hill, Ashley Blom and Paul Dieppe

VERSION 1 - REVIEW

REVIEWER	Changhai Ding, Associate Professor University of Tasmania, Australia No conflicts of interests
REVIEW RETURNED	31/10/2011

GENERAL COMMENTS	The study was aimed to estimate the proportion of people reporting long-term moderate or severe pain after total hip or knee replacement for the treatment of osteoarthritis. Why did you select studies with 3 months' to 5 years' follow-up? Is this time period "long-term" as you defined? You excluded 129 studies with >5 years' follow-up which may be real "long-term". The uncertain outcome includes patients who "had revision surgery": how many patients had this only in 3 months to 5 years? did the authors explore the reason why had "revision surgery"? Was it due to increased pain? If it was, this should be listed as "an unfavourable outcome". The uncertain outcome includes patients who "were not followed up": All studies were prospective, so why were patients not followed up? "The WOMAC outcome used to define a poor pain outcome as a gain of less than 10 points on the 100 point scale": was this a total WOMAC score or a subscale score? If it was a total score (sum of 5 subscale score), why did it come to 100 points? "a gain of less than 10 points" should be written as "an improvement of less than 10 points" (in the results and table 1) "Baker and colleagues followed up 9417 patients with primary total hip replacement…" in page 8: hip or knee? Page 9: please add "uncertain" before "15.7%". Hid did you estimate "7 to 23%" or "10 to 34%" of patients experienced pain after hip or knee replacement?

REVIEWER	Steven J Kamper

	Research Fellow The George Institute, University of Sydney Australia No competing interests
REVIEW RETURNED	14/11/2011

GENERAL COMMENTS	General comments • The authors conducted a review investigating the prevalence of pain after joint replacement surgery in patients with hip and knee OA. • For the most part the study appears well conducted and is well-written, but some further work will improve the utility and interpretability of the information provided. • Methodological quality assessment of the included studies must be carried out. Specific comments Introduction 1. Pg 3; The introduction is generally clear and concise. I would suggest inclusion of some data from the existing literature to give the reader an impression of the size of the problem. Methods 2. Pg 3, line 55; to say that Cochrane methods have been used for this review is misleading, there are several, important ways in which the authors deviate from the Cochrane protocols (quality assessment, duplicate data extraction, no contact of authors for further information etc). This reference should be removed. 3. Pg 4, line 20; are the pain VAS and self-appraisal disease-specific measures? My impression was that they are generic measures. This leads to the question of why generic health measures are excluded, please provide justification for this decision. 4. Pg 4, line 39; I have difficulty understanding the repeated use of the term 'generally unselected' patients. All studies will have inclusion and exclusion and exclusion criteria, so all are selected to some extent. This being the case the authors need to specify what inclusion/exclusion criteria were and were not acceptable. This has further implications for selection based on study type (Pg 5, line 11), it does not seem supportable to exclude due to study design based on the premise that certain desins select a certain type of patient. Either the patients in the study (regardless of design) meet the a-priori inclusion criteria or they do not and studies will be included or not based on that criteria. 5. Pg 4, line 45; please confirm that limitation of the follow-up period was specified a-priori. 6. Pg 4, line; were title and abstract lists screened independently by two authors? 7. Pg 5, line; 8; please clarify whether retrospective studies were included or excluded. 8. Pg 5, line 49; while the decision to collect information according to proportions of people with pain is reasonable, please provide information regarding the mean pain course after surgery. Given that this information is readily available this should not be too difficult and would provide worthwhile information to the readers, particularly when summarised in a meta-analysis. It would potentially provide a large amount of information given that 83 studies excluded form analysis for not providing these data. 9. Pg 6, line 6; please conduct a study quality assessment, it provides critical information for interpretation of results. Results

	10. Pg ; Please put the results from each included study in some meaningful order. The most useful order would be according to study quality, alternatively sample size (this recommendation also holds for Figure 2). 11. Pg 6, line 59; were the patients included in the analysis from this study the same or different to those excluded? 12. Pg 8, line; 45; should this say 'knee' rather than 'hip'. 13. Pg 9, line 6; this illustrates a major problem with the selected analysis. By categorising people as improved or not according to the criteria used in each individual study, the authors make comparison between studies (and hence synthesis) extremely problematic. For example12.9% of patients had >40/100 knee pain in one study and 21.7% of patients had any pain in another, how is the reader to interpret these data together? 14. Pg 9, line 17 onwards; here the need for quality assessment is crucial, without knowing which estimates are more likely to reliable, the reader cannot judge where in the large range 'the truth' is most likely to lie. Discussion 15. Pg 10, line 38; the justification for not contacting authors of included studies to provide data is weak. The Cochrane handbook recommends it and it should be undertaken. 16. Pg 11, line 8; is it possible that some of these issues regarding heterogeneity could be explored with the extra power inclusion of RCTs would provide? Appendix 1 MOOSE checklist (sections to be amended) 17. Please change 'Background' to 'Introduction'. 18. Hypothesis statement; please outline where (what) this is. 19. Methods of addressing articles published …; not stated. 20. Description of any contact…; justification is weak. 21. Assessment of study quality…; this is not an assessment of study quality. 22. Assessment of hetereogenity; there is no such assessment. 23. Results of sensitivity…; not completed. 24. Quantitative assessment of bias…; not completed. 25. Assessment of quality…; insufficient.

VERSION 1 – AUTHOR RESPONSE

We thank the reviewers for their thorough review of our article. As requested we have provided a Quality analysis using relevant sections of the Cochrane risk of bias table and addressed the issues raised by the reviewers.

Specific issues raised by reviewers

The study was aimed to estimate the proportion of people reporting long-term moderate or severe pain after total hip or knee replacement for the treatment of osteoarthritis. Why did you select studies with 3 months' to 5 years' follow-up? Is this time period "long-term" as you defined? You excluded 129 studies with >5 years' follow-up which may be real "long-term".

This was an a priori decision based on evidence about a plateau in outcome after joint replacement and evidence about long-term revision. In the study of Bachmeier et al. 2001 (now described in introduction Para 4) pain levels reached a plateau by 3 months after hip and knee replacement. We chose 5 years as the upper limit because studies with longer term follow up generally concern implant failure. This does not reflect a sudden change in implant stability and need for revision – the study of Roberts 2007 (now cited in Methods Study selection para 3) showed that implant survival at 5 years is about 97.5% and at ten years about 94.4%. However as described in the Methods section it was our

perception that orthopaedic studies with over 5 years of follow up generally concern technical issues whereas studies with less than 5 years duration are more likely to investigate pain and function as primary outcomes.

The uncertain outcome includes patients who “had revision surgery”: how many patients had this only in 3 months to 5 years? did the authors explore the reason why had “revision surgery”? Was it due to increased pain? If it was, this should be listed as “an unfavourable outcome”.

In 6 studies there was either an explicit report of revisions or a less specific report of re-operation. These patients were not followed up with disease specific pain measures and we felt it more appropriate to include them in the uncertain group. The number of revisions across studies was small and it would not have an impact on overall results.

The uncertain outcome includes patients who “were not followed up”: All studies were prospective, so why were patients not followed up?

Some patients were lost to follow up. Those who “were not followed up” are the patients with revision surgery, contra-lateral surgery or dislocation who were not followed up with questionnaires. We have made alterations to the text (Methods Data synthesis and analysis) and attempted to make it clearer that these patients were not approached for a questionnaire study follow up.

“The WOMAC outcome used to define a poor pain outcome as a gain of less than 10 points on the 100 point scale”: was this a total WOMAC score or a subscale score? If it was a total score (sum of 5 subscale score), why did it come to 100 points? “a gain of less than 10 points” should be written as “an improvement of less than 10 points” (in the results and table 1)

Jones and colleagues reported that “Each WOMAC subscale score was transformed to a range of 0 (worst) to 100 (best)”. We used their 0-100 WOMAC pain subscale scores (not a total WOMAC score). Other authors also reported WOMAC pain converted to a 0-100 scale.
Changes made to text and table as requested. Also pain changed to 100 point pain scale to clarify that this was only the WOMAC pain dimension.
We took care to ensure all outcomes related to pain in the operated joint.

“Baker and colleagues followed up 9417 patients with primary total hip replacement…” in page 8: hip or knee?

Mistake now corrected: changed to knee

Page 9: please add “uncertain” before “15.7%”.

Uncertain added to text

Hid did you estimate “7 to 23%” or “10 to 34%” of patients experienced pain after hip or knee replacement?

We calculated these by estimating that a similar proportion of those lost to follow up for whatever

reason will have had an unfavourable outcome as those followed up. This is probably an underestimate and we have altered the text and hopefully made this clearer (Results Overview Total hip replacement Para 3). There is additional reporting of study quality in this section.

General comments
• The authors conducted a review investigating the prevalence of pain after joint replacement surgery in patients with hip and knee OA.
• For the most part the study appears well conducted and is well-written, but some further work will improve the utility and interpretability of the information provided.
• Methodological quality assessment of the included studies must be carried out.

We have now included as Appendix 3 a “Risk of bias table”. This is not identical to the Cochrane Risk of Bias table as each individual study does not represent a comparison of methods. We have used this table to address further issues raised by the Reviewers and feel this has improved the article considerably.

Specific comments
Introduction
1. Pg 3; The introduction is generally clear and concise. I would suggest inclusion of some data from the existing literature to give the reader an impression of the size of the problem.

Thank you for this suggestion; we have now included a description of the studies by Woolhead et al. and Judge et al. which indicate something of the scope of the problem.

Methods
2. Pg 3, line 55; to say that Cochrane methods have been used for this review is misleading, there are several, important ways in which the authors deviate from the Cochrane protocols (quality assessment, duplicate data extraction, no contact of authors for further information etc). This reference should be removed.

Thank you, we agree that our process does not completely adhere to Cochrane protocols, the reference to the Cochrane Handbook has been removed from Methods para 1 but retained in relation to the newly added risk of bias table.

3. Pg 4, line 20; are the pain VAS and self-appraisal disease-specific measures? My impression was that they are generic measures. This leads to the question of why generic health measures are excluded, please provide justification for this decision.

All outcomes included in the review related to pain in the operated hip or knee. We were careful not to include total WOMAC scores. We did not include generic scores as the older populations included in studies of joint replacement will have other health problems causing pain (including pain in non-operated joints) - Discussion Para 9.

4. Pg 4, line 39; I have difficulty understanding the repeated use of the term ‘generally unselected’ patients. All studies will have inclusion and exclusion and exclusion criteria, so all are selected to some extent. This being the case the authors need to specify what inclusion/exclusion criteria were and were not acceptable. This has further implications for selection based on study type (Pg 5, line

11), it does not seem supportable to exclude due to study design based on the premise that certain desins select a certain type of patient. Either the patients in the study (regardless of design) meet the a-priori inclusion criteria or they do not and studies will be included or not based on that criteria.

The reviewer raises an important point about how we make studies' selection criteria transparent to the readers. The criteria for exclusion of patients from individual studies are now summarised in the Risk of bias table (Appendix 3).We aimed to review studies of patients who were representative of the joint replacement population and have now made this clearer in the Methods Study selection section para 1. Hopefully this supports our attempt to include populations representative of the hip and knee replacement population. "Generally" has been removed where used.

5. Pg 4, line 45; please confirm that limitation of the follow-up period was specified a-priori.

As described above we chose the range of 3 months to 5 years in order to reflect the period when recovery can be considered maximal and not affected by increased incidence of joint loosening and revision over longer periods. Appropriate referencing now included.
We have now confirmed that our decision was made a priori in Methods Study selection para 3.

6. Pg 4, line; were title and abstract lists screened independently by two authors?

We have made clear that titles and abstracts were checked independently for eligibility by 2 reviewers in Methods Study selection para 4.

7. Pg 5, line; 8; please clarify whether retrospective studies were included or excluded.

Included studies were prospective in design. We have now made this clear.

8. Pg 5, line 49; while the decision to collect information according to proportions of people with pain is reasonable, please provide information regarding the mean pain course after surgery. Given that this information is readily available this should not be too difficult and would provide worthwhile information to the readers, particularly when summarised in a meta-analysis. It would potentially provide a large amount of information given that 83 studies excluded form analysis for not providing these data.

We chose not to summarise mean WOMAC scores in our study. The distribution of WOMAC scores after surgery is highly skewed and we feel that an analysis such as that we have used is more meaningful in the context of individual patient outcomes.
Bachmeier and colleagues in their study tracked changes in WOMAC pain in a cohort study. They reported means and standard deviations at 3, 6, 9 and 12 months after surgery. This is an important paper and potentially could be replicated using systematic review methods. However this was not our aim, we always aimed to report unfavourable (moderate to severe pain) in a way meaningful to patients and health care providers. Sometimes this is referred to as "unacceptable" pain but we did not use this term here.

9. Pg 6, line 6; please conduct a study quality assessment, it provides critical information for interpretation of results.

We have now added the requested risk of bias table (based on Cochrane method) which we have modified to be appropriate for observational studies (Appendix 3).

Results

10. Pg ; Please put the results from each included study in some meaningful order. The most useful order would be according to study quality, alternatively sample size (this recommendation also holds for Figure 2).

Thank you for this helpful suggestion. We have now ordered the studies more meaningfully based on 1) number of centres studied and 2) losses to follow up (i.e. a marker of study representativeness/ quality).

11. Pg 6, line 59; were the patients included in the analysis from this study the same or different to those excluded?

The patients without WOMAC data were similar with regard to age and sex and we have now reported this in the text.

12. Pg 8, line; 45; should this say ‘knee’ rather than ‘hip’

We are sorry for this error, we have changed this to ‘knee’.

13. Pg 9, line 6; this illustrates a major problem with the selected analysis. By categorising people as improved or not according to the criteria used in each individual study, the authors make comparison between studies (and hence synthesis) extremely problematic. For example12.9% of patients had >40/100 knee pain in one study and 21.7% of patients had any pain in another, how is the reader to interpret these data together?

Pain is considered a core outcome that should be assessed after orthopaedic surgery. However, there is no agreement in the research literature as to which outcome measures should be used or which aspects of the pain experience should be assessed. We have added a reference to this effect (MacKichan et al. 2009).
We have attempted to extract an outcome that reflects moderate or severe pain as observed in clinical practice. While this may fall short of acceptability for inclusion in formal meta-analyses we believe the individual studies have described an unfavourable pain group with moderate or severe pain.
To emphasise that the unfavourable pain outcomes authors have reported and that we have extracted are of potential clinical relevance we have summarised them in a new Appendix 4. This does not reflect new data but serves to emphasise that they do reflect an important outcome for patients.

14. Pg 9, line 17 onwards; here the need for quality assessment is crucial, without knowing which estimates are more likely to reliable, the reader cannot judge where in the large range ‘the truth’ is most likely to lie.

It was always our aim to identify and include only prospective studies with patient reported pain outcome measures in populations representative of the joint replacement population. However adding

the risk of bias assessment within these studies identifies another tier of quality which is clearly important. We have now made quality assessments and considered these at length particularly in the Methods and Results sections.

Discussion

15. Pg 10, line 38; the justification for not contacting authors of included studies to provide data is weak. The Cochrane handbook recommends it and it should be undertaken.

A study of pain outcomes after joint replacement including all studies that have measured WOMAC outcomes would be a considerable undertaking. As we are not able to perform such a review, we suggested that study authors should perform and publish appropriate analyses of data. We have changed the wording of Methods Data extraction para 1 to emphasise the difficulties associated with acquiring data from authors as described by Mullan et al.

In the context of randomised controlled trials it is potentially possible to analyse data on every person that has ever been randomised to receive a treatment or to a control group. However issues of representativeness of study populations and external validity are important, particularly in non-drug interventions. A fundamental problem in presenting an overview of treatment outcomes based on longitudinal studies is that only a proportion of the total number of people exposed to the intervention will be included in studies.

Longitudinal studies are only ever going to include a proportion of patients receiving an intervention, at best a proportion of patients who are representative of the total population. Our approach to review of longitudinal studies was to identify only those studies reporting appropriate pain outcomes (proportions with different pain levels) in populations representative of those receiving total hip or knee replacement.

16. Pg 11, line 8; is it possible that some of these issues regarding heterogeneity could be explored with the extra power inclusion of RCTs would provide?

We were hopeful that we would be able to present summary values reflecting an overall estimate of the proportion of patients with moderate or severe long term pain after total hip or knee replacement. The advice from a statistician with experience in systematic reviews and meta-analyses was that without information on the correlations between different pain measures we should not attempt to generate a pooled estimate.

Inclusion of further studies less representative of the hip and knee replacement population might add a few studies with appropriate data but would also add a new tier of heterogeneity relating to evaluated treatment effectiveness and trial inclusion criteria.

Appendix 1 MOOSE checklist (sections to be amended)

17. Please change ‘Background’ to ‘Introduction’.

Requested change made

18. Hypothesis statement; please outline where (what) this is.

The rationale behind the study is now identified in the MOOSE statement

19. Methods of addressing articles published …; not stated.

All of the studies were published in English.

20. Description of any contact…; justification is weak.

We have made attempts to justify this in our revisions in the article (Methods/ Data extraction).

21. Assessment of study quality…; this is not an assessment of study quality.

Information on this is now added to MOOSE statement and text as described above.

22. Assessment of hetereogenity; there is no such assessment.

Quality as a source of heterogeneity is addressed and described in the MOOSE statement.

23. Results of sensitivity…; not completed.

Text added on problem associated with range of outcome measures

24. Quantitative assessment of bias…; not completed.

This is now completed to reflect additional consideration of quality as requested by reviewers.

25. Assessment of quality…; insufficient.

Details based on extensive revisions relating to quality assessment are described.

>From the managing editor:
Please provide a structured abstract that either meets our criteria (http://bmjopen.bmj.com/site/about/guidelines.xhtml#research) or those described by the PRISMA checklist: (background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.).

Structured abstract replaced with version meeting PRISMA checklist criteria.

In the contributorship statement please include that all authors revised the ms and approved the final version of the ms (as long as this is true).

All authors saw the reviewers' comments and provided comments. This is confirmed in the Author's contributions.

Reviewer SJK insists that you go back and assess the quality of the included studies (ideally using

the Cochrane Risk of Bias tool). We understand that this is quite a lot to ask, but think it's reasonable and would make the study much stronger. If you can provide strong reasons for not doing so we may reconsider this request.

In the revised manuscript we have attempted to address the issues relating to study quality described by reviewers.

VERSION 2 – REVIEW

REVIEWER	Changhai Ding, Associate Professor University of Tasmania, Australia No conflicts of interests
REVIEW RETURNED	06/01/2012

The reviewer completed the checklist but made no further comments.

REVIEWER	Steven J Kamper Research Fellow The George Institute, University of Sydney Australia
REVIEW RETURNED	09/01/2012

GENERAL COMMENTS	Thank you to the authors for their substantial efforts in responding to my revisions. I have only a very few comments to make and do not insist on any further changes, I am happy for the editor to decide whether they need to to be addressed. 1. Results section: please check that the Appendix numbers stated in the text correspond to the correct Appendices. 2. Point 4 from the original review; I am still not entirely convinced by the authors' justification for excluding RCTs. Surely it is possible that some RCTs in the area recruited consecutive cases of knee or hip OA patients awaiting replacement? If this were the case then would not these patients be representative of the population? 3. Point 8 from the original review; I accept that recording mean pain scores post operation was not the primary aim of the review but only raised this point due to the possibility of providing information that might be interesting to readers with relatively little additional work. 4. Point 15 from the original review; I am still unconvinced by the strength of the argument put forward by the authors to justify not contacting study authors.