PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Comparative efficacy and safety of surgical and invasive treatments for adults with degenerative lumbar spinal stenosis: protocol for a network meta-analysis and systematic review
AUTHORS	Chen, Lingxiao; Ferreira, Paulo; Beckenkamp, Paula; Ferreira, Manuela

VERSION 1 - REVIEW

REVIEWER	Gordon Guyatt McMaster University
REVIEW RETURNED	23-Jul-2018

GENERAL COMMENTS	This study plans to assess the efficacy and safety of surgical and invasive procedures for adults with degenerative lumbar spinal stenosis, which is a valuable work to provide an overall picture on the surgical and invasive treatment for degenerative lumbar spinal stenosis. However, we have some concerns on the present version. ABSTARCT section – To provide a more informative abstract, the authors should consider to provide a summary including: 1) the databases you are planning to search and if there are any restrictions; 2) the criteria of risk of bias assessment you will use; 3) assessment of the certainty of evidence. Types of interventions –The control group will include no treatment, usual care, sham operation, another active opinion or a combination of approaches. Will these constitute a single node in the network? If so, the authors should address in the discussion the implications for heterogeneity and transitivity. Outcome measures – The authors are grouping the time points into less than 6 months, 6-12 months, and more than 12 months; it is not clear what “other time points” mean? It is also not clear whether authors will perform network meta-analysis in the three time points separately? For all outcomes in which investigators may use more than one instrument for one construct, how will the authors decide which instrument to use in their analysis. Search strategy – Is there any restriction like language and publication status? Risk of bias assessment – For efficacy outcomes, the authors plan to include only studies with a low risk of bias in the item “random sequence generation”; if we understand correctly, this should move to an eligibility criterion. Data analysis – The first time the authors mention NRS it should be spelled out in full. Data analysis – It seems that the authors will, for mortality outcomes, be including randomized and non-randomized studies in the same analysis. It seems to us that the NRS estimates are very likely to be biased: physicians and patients choose between

	treatments on the basis of patients' characteristics, and those characteristics are very likely to be associated with mortality. Thus, including unbiased assessment from randomized trials with biased assessments from observational studies is problematic. Pairwise meta-analyses – The authors should specify what software and method they will use to do pairwise meta-analysis? Assessment of the transitivity assumption – The authors should specify the variables they will consider when addressing the transitivity assumptions. Exploring sources of heterogeneity or inconsistency – The authors should specify prior assumptions for subgroup analysis in terms of direction of effect. - In outcome measures section, the authors group outcomes into three categories, but here they say they will perform subgroup analysis, so will they include all durations in primary analysis, then perform subgroup analysis? Rating the certainty/quality of the evidence – The approach the authors cite for rating quality/certainty of evidence is designed for randomized trials. How will they address certainty/quality of evidence for observational studies? And if they do (misguidedly in our view) combine observational and RCTs, how will they assess certainty of evidence then?

REVIEWER	Hwanhee Hong Duke University
REVIEW RETURNED	17-Oct-2018

GENERAL COMMENTS	Note that my review focuses on statistical issues. 1. Types of studies: Will cross-over and cluster randomized trials contribute to only safety outcomes? 2. Network meta-analysis a. Reference 47 is about Bayesian network meta-analysis methods, not about frequentist methods. b. The second sentence should be "The heterogeneity parameter is assumed the same for each intervention." c. I presume that the analysis will be done via mvmeta package in Stata (your reference 48). Does this package have an option for the same heterogeneity parameter for all interventions (as the second sentence stated)? It would be great to cite a reference discussing frequentist network meta-analysis under this assumption. Note that reference 49 discusses Bayesian network meta-analyses. d. What do you mean by prediction interval plot? Reference 50 does not explain prediction interval plot. e. For safety outcomes, both randomized controlled trials and observational studies will be included. Observational studies (are usually considered to) tend to yield somewhat biased results relative to randomized controlled trials. Will the analysis account for the variability across different sources (i.e., different types of studies)?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Gordon Guyatt

Institution and Country: McMaster University

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

This study plans to assess the efficacy and safety of surgical and invasive procedures for adults with degenerative lumbar spinal stenosis, which is a valuable work to provide an overall picture on the surgical and invasive treatment for degenerative lumbar spinal stenosis. However, we have some concerns on the present version.

1. ABSTARCT section – To provide a more informative abstract, the authors should consider to provide a summary including: 1) the databases you are planning to search and if there are any restrictions; 2) the criteria of risk of bias assessment you will use; 3) assessment of the certainty of evidence.

Answer (A): Thank you for your suggestions. We have added the relevant information to the abstract section based on your suggestions (page 2, line 10-13 and 19-22).

We will include randomised controlled trials assessing surgical and invasive treatments for adults with degenerative lumbar spinal stenosis. We will search AMED, CINAHL, EMBASE, the Cochrane Library and MEDLINE. Only English studies will be included and no restriction will be set for publication status.

Two reviewers will independently select studies, extract data and assess the risk of bias (Revised Cochrane risk-of-bias tool for randomized trials) of included studies. The quality of the evidence will be evaluated through the Grading of Recommendations Assessment, Development and Evaluation framework.

2. Types of interventions –The control group will include no treatment, usual care, sham operation, another active opinion or a combination of approaches. Will these constitute a single node in the network? If so, the authors should address in the discussion the implications for heterogeneity and transitivity.

A: We will separate the types of control groups into different nodes. However, if we have insufficient studies, we will combine no treatment and usual care into one node to make full use of the data. We have added relevant sentences in the article (page 5, line 27-30).

The interventions in comparison groups will be treated as different nodes. However, if we have insufficient studies to connect different interventions, we will combine no treatment and usual care into one node to make full use of the data.

3. Outcome measures – The authors are grouping the time points into less than 6 months, 6-12 months, and more than 12 months; it is not clear what "other time points" mean? It is also not clear whether authors will perform network meta-analysis in the three time points separately?

A: 'Other time points' mean different time points in the long-term follow-up, e.g. 1 year, 2 years, 5 years. We will perform network meta-analysis in the three time points separately. Although we will choose the data which is closest to 12-month as the long-term outcome, we plan to perform a subgroup analysis based on these different time points in the long-term follow-up assessment, given

there might be differences in treatment responses at those time points. We have revised all relevant sentences (page 6, line 4-8).

We will perform network meta-analysis in the three time points separately. For studies which report outcomes in multiple time points, data closest to the 6 and 12-month follow-up time will be included in the main analyses. For different time points in long-term follow-up assessment (e.g., 1 year, 2 years, 5 years), subgroup analyses will be performed.

4. For all outcomes in which investigators may use more than one instrument for one construct, how will the authors decide which instrument to use in their analysis.

A: We have revised the relevant content based on the previous reference (page 6, line 14-16 and 25-26; page 7, line 1-2).

For the outcome of pain, visual analog scale will be used as the first choice and numeric rating scale as the second choice. For the outcome of physical function, ODI will be used as the first choice, Roland Morris disability questionnaire as the second choice and core outcome measures index as the third choice. For the outcome of health related quality of life, EQ-5D will be used as the first choice, followed by SF-36, SF-12 and Nottingham health profile.

5. Search strategy – Is there any restriction like language and publication status?

A: Only English studies will be included as stated in the methods section. There will be no restrictions based on publication status. We have revised the relevant sentences (page 7, line 28-29).

Only English studies will be included and no restriction will be set for publication status.

6. Risk of bias assessment – For efficacy outcomes, the authors plan to include only studies with a low risk of bias in the item “random sequence generation”; if we understand correctly, this should move to an eligibility criterion.

A: You are correct. We have moved this sentence to the eligibility criteria section of the methods (page 7, line 18-19). In addition, we have changed the risk of bias assessment tool into RoB 2 tool considering it is a new recommended tool from Cochrane Collaboration (page 10, line 6-20).

To decrease bias, we excluded studies with a high risk of bias in the domain risk of bias arising from the randomization process.

Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) will be used to evaluate the risk of bias of included randomized parallel-group trials. The tool is comprised of five domains: 1. bias arising from the randomization process; 2. bias due to deviations from intended interventions; 3. bias due to missing outcome data; 4. bias due to missing outcome data; 5. bias in selection of the reported result. Each domain includes several signaling questions which elicit information relevant to an assessment of risk of bias. The answer option for each signaling question is: Yes, probably yes, probably no, no and no information. Based on the answers of all signaling questions in one domain, we rated the domain as low risk of bias, some concerns or high risk of bias. Finally, we got the overall risk-of-bias judgement as low risk of bias, some concerns or high risk of bias considering the risk-of-bias judgement in five domains.

For cluster-randomized trials, one more domain should be considered: bias arising from identification or recruitment of individual participants within clusters. For cross-over trials, analysis issues in cross-over trials should be additionally considered.

7. Data analysis – The first time the authors mention NRS it should be spelled out in full.

A: We have corrected it based on your suggestion.

8. Data analysis – It seems that the authors will, for mortality outcomes, be including randomized and non-randomized studies in the same analysis. It seems to us that the NRS estimates are very likely to be biased: physicians and patients choose between treatments on the basis of patients' characteristics, and those characteristics are very likely to be associated with mortality. Thus, including unbiased assessment from randomized trials with biased assessments from observational studies is problematic.

A: Thank you for your comment. We have decided to include only randomised controlled trials based on your comments. We have revised the manuscript accordingly.

9. Pairwise meta-analyses – The authors should specify what software and method they will use to do pairwise meta-analysis?

A: We will use random-effect model with DerSimonian and Laird inverse-variance method for the pairwise meta-analyses and have added this information on page 11, line 2-3.

We will perform traditional pair-wise meta-analyses through random-effect model with DerSimonian and Laird inverse-variance method for every direct comparison.

10. Assessment of the transitivity assumption – The authors should specify the variables they will consider when addressing the transitivity assumptions.

A: According to previous literature, age, gender, education level, baseline physical function, smoking habit, BMI, comorbidities and previous treatment will be considered when addressing the transitivity assumptions. We have added this information on page 11, line 9-12.

The potential baseline effect modifiers (age, gender, education level, baseline physical function, smoking habit, BMI, comorbidities and previous treatment) will be assessed to confirm they are similar among different comparisons before we perform network meta-analyses

11. Exploring sources of heterogeneity or inconsistency – The authors should specify prior assumptions for subgroup analysis in terms of direction of effect.

A: We have added this information on page 12, line 1-11.

1. Single level spinal stenosis versus multiple levels, the assumption is that patients with multiple levels spinal stenosis might have poorer physical function and higher all-cause mortality than patients with single level; 2. Duration of follow-up for long-term assessment (e.g., 1 year, 2 years and 5 years), the assumption is that patients who received injection therapies might have poorer physical function and higher all-cause mortality in longer duration of follow-up than patients who received surgical therapies; 3. Patients with versus patients without degenerative spondylolisthesis, the assumption is that patients with degenerative spondylolisthesis might have poorer physical function and higher all-cause mortality than patients without; 4. Type of disease: central, foraminal or lateral, the assumption is that patients with central lumbar spinal stenosis might have poorer physical function and higher all-cause mortality than patients with foraminal or lateral.

12. In the outcome measures section, the authors group outcomes into three categories, but here they say they will perform subgroup analysis, so will they include all durations in primary analysis, then perform subgroup analysis?

A: We will perform these subgroup and meta-regression analyses under the three categories except for the analysis on the duration of follow-up for long-term assessment. In this subgroup analysis, we will only focus on the long-term outcomes and categorise into 1 year, 2 years and 5 years. We have revised the relevant sentences (page 11, line 28-30; page 12, line 1).

For two primary outcomes (physical function and all-cause mortality), subgroup analyses and meta-regression will be performed under the three time categories (short-term, mid-term, and long-term) except for the analysis on the duration of follow-up for long-term assessment.

13. Rating the certainty/quality of the evidence – The approach the authors cite for rating quality/certainty of evidence is designed for randomized trials. How will they address certainty/quality of evidence for observational studies? And if they do (misguidedly in our view) combine observational and RCTs, how will they assess certainty of evidence then?

A: Thank you for your comment. Based on your previous comments, we have changed the inclusion criteria as including randomized controlled trials only.

Reviewer: 2

Reviewer Name: Hwanhee Hong

Institution and Country: Duke University

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Note that my review focuses on statistical issues.

1. Types of studies: Will cross-over and cluster randomized trials contribute to only safety outcomes?

A: Cross-over and cluster randomized trials will be included in both efficacy and safety outcomes. We have revised the relevant sentences on page 7, line 14-15.

Only randomised controlled trials (RCT), which includes parallel, cross-over and cluster trials, will be included.

2. Network meta-analysis

a. Reference 47 is about Bayesian network meta-analysis methods, not about frequentist methods.

A: We have changed the reference 47 into one based on frequentist – White IR. Network meta-analysis. Stata Journal. 2015;15(4):951-85.

b. The second sentence should be "The heterogeneity parameter is assumed the same for each intervention."

A: We have revised the sentence accordingly (page 10, line 16-17).

c. I presume that the analysis will be done via mvmeta package in Stata (your reference 48). Does this package have an option for the same heterogeneity parameter for all interventions (as the second sentence stated)? It would be great to cite a reference discussing frequentist network meta-analysis under this assumption. Note that reference 49 discusses Bayesian network meta-analyses.

A: We plan to use network commands (via mvmeta package) in Stata to run the analysis. This package has the default option for the same heterogeneity parameter for all interventions. At the start of the output, it shows that the parameter is estimated using REML (Restricted Maximum Likelihood) and a single tau or tau2 is reported. We have changed the reference 49 into one based on frequentist – White IR. Network meta-analysis. Stata Journal. 2015;15(4):951-85.

d. What do you mean by prediction interval plot? Reference 50 does not explain prediction interval plot.

A: The Prediction interval plot could facilitate heterogeneity assessment for each combined comparison. The interval includes the true intervention effects in future studies and the extent of between-study variation. We have changed reference 50 into two more relevant references.

Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ (Clinical research ed). 2011;342:d549.

Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society Series A, (Statistics in Society). 2009;172(1):137-59.

e. For safety outcomes, both randomized controlled trials and observational studies will be included. Observational studies (are usually considered to) tend to yield somewhat biased results relative to randomized controlled trials. Will the analysis account for the variability across different sources (i.e., different types of studies)?

A: We have revised the inclusion criteria to limit our inclusion to randomized controlled trials only.

VERSION 2 – REVIEW

REVIEWER	Gordon Guyatt McMaster University
REVIEW RETURNED	06-Dec-2018

GENERAL COMMENTS	The authors have addressed all our concern and taken all our advice.

REVIEWER	Hwanhee Hong Duke University
REVIEW RETURNED	21-Dec-2018

GENERAL COMMENTS	Authors revised the protocol according to reviewer's comments. I have one minor comment. In the subgroup analyses section, authors listed assumptions for individual subgroup analyses. However, in fact they are hypotheses, not assumptions.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Gordon Guyatt

Institution and Country: McMaster University

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

The authors have addressed all our concern and taken all our advice.

Answer (A): Thank you so much.

Reviewer: 2

Reviewer Name: Hwanhee Hong

Institution and Country: Duke University

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

Authors revised the protocol according to reviewer’s comments. I have one minor comment. In the subgroup analyses section, authors listed assumptions for individual subgroup analyses. However, in fact they are hypotheses, not assumptions.

A: Thank you for your suggestion. We have revised corresponding sentences using the word “hypothesis” (Page 12, Lines 2, 5, 8 and 11).