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)	Quality of life and objective functional impairment in lumbar spinal stenosis: A protocol for a systematic review and meta-analysis of moderators
AUTHORS	Ferretti, Fabio; Coluccia, Anna; Gusinu, Roberto; Gualtieri, Giacomo; Muzii, Vitaliano; Pozza, Andrea

VERSION 1 – REVIEW

REVIEWER	Martin N. Stienen Stanford University, CA (USA)
REVIEW RETURNED	25-Jun-2019

GENERAL COMMENTS	The rationale of the review and meta-analysis is given. The methods described in the article appear very sound; the analysis is registered and quality and risk of bias of each included study is assessed. I congratulate the authors to their well-planned research and wish them luck that they will be able to include a satisfactory number of high-quality studies in order to obtain meaningful results. Please find some aspects below that the authors may find helpful: - Why do the authors confine their methods to the QoL measures included in table 1? There are further, frequently used measures (e.g. SF-12, EQ5D, COMI, ...) that are used frequently for degenerative lumbar spine patients. - Prior research indicated the increased use of "objective" functional tests as outcome measures, supplementing the use of PROMS (typically questionnaire-based) and adding a further dimension to the comprehensive outcome evaluation (for review, see Spine J. 2019 Jul;19(7):1276-1293. doi: 10.1016/j.spinee.2019.02.014). I would like to propose the authors consider objective outcome measures in their review, as well, as those were shown to be less biased by depressive comorbidity, compared to PROMs (compare e.g. Spine J. 2017 Jun;17(6):807-813. doi: 10.1016/j.spinee.2016.12.004).

REVIEWER	Ulrike Held Dept. of Biostatistics, Epidemiology, Biostatlstics and Prevention Institute, University of Zurich, Switzerland
REVIEW RETURNED	29-Aug-2019

GENERAL COMMENTS	Review for the paper Impairment correlates in lumbar spinal stenosis: perceived health status and depression. A protocol of a systematic review and meta-analysis Ferretti and co-authors presented a systematic review protocol on

	the comparison of health-related quality of life between patients with lumbar spinal stenosis (LSS) and healthy controls, while accounting for depression. The authors address a relevant topic as the prevalence of LSS in the elderly is rising, impacting on health care utilization and increasing medical costs. The review protocol is clearly written and well-structured. In order to further improve the quality of the paper, the authors should address the following comments: Search strategy and selection of studies - Search strategy: the authors should specify the specific search string they intended to use. When they make the search, the authors should be able to give a specific date when the search was conducted, not a week as is currently stated (3.-10. July). - Related to the above mentioned point: do the authors have expertise to conduct a search or are the authors planning to involve an information specialist, e.g. an experienced librarian? In order to define and validate the search string, and to remove duplicates from the different electronic databases I would strongly advise you to consider this. - I would suggest to make the title and abstract screening in one stage, and then in a second stage for the full-texts. Potential moderators of between-group difference - The authors are planning to include depression as a potential moderator in their analyses. Specifically, authors are planning to include the “percentage of patients with depressive disorders” as a moderator variable. I see the potential problem that some studies will not report on depression, or will not explicitly state the percentage that you are looking for. Please state how you are going to include these publications. - Apart from depression, there may be other – even more important – parameters that may affect the effect sizes on health-related quality of life. Severity of LSS as measured with disability and pain scales, comorbidities, age will also play a role. Except for age, these confounders will not be accounted for through matching. The depression’s moderating effect that you aim to quantify may be actually confounded by a set of other variables. - There is no detailed description of the methods you are planning to use for estimating the moderating effect of depression, reported in patient and control groups. - In Table 3 you state that you are going to report mean and standard deviation for age. How will you include publications not reporting age, or reporting median and interquartile ranges or ranges instead? Outcomes - You describe that health related quality of life (QoL) will be either self-reported or interview based. Are you planning any subgroup analyses? - Why are you not considering the EQ-5D scale for measuring quality of life? - Please state how you are going to address the problem that multiple publications about the same cohort of LSS patients and controls may exist. Which one are you planning to include in the meta-analysis? Are you planning to include case series? - Will you exclude any study from the meta-analysis because of low methodological quality?

	Section: Meta-analytic procedure Summary measures - Please replace “random effect meta-analysis” by “random effects meta-analysis” throughout. - Please revise the sentence “For all the analyses, the p-value will be set at 0.05.” I would advise not to pre-specify a level of significance but rather quantify the evidence on a continuous scale. - You describe the categories of Cohen’s effect sizes, >= 0.8, 0.5-0.8, and <=0.2. What about the missing level from 0.2-0.5 – did you mean that this represents “low” effect size? - Will you need a minimum number of studies for pooling or for being able to quantify the moderating effect of depression? Publication bias - Please describe that the trim-and-fill method will result in a new summary estimate. Are you planning to report the new summary estimate (after trim-and-fill) together with the original estimate in every meta-analysis? Please also report on the number of studies “filled”. Moderator coding and analysis - You state that “if inconsistency between effect sizes is found” depression will be investigated as moderator. In fact, you should evaluate the moderating effect in all meta-analyses (if enough studies can be retrieved). Perhaps you want to describe pre-planned subgroup analyses in those studies with i) 100% patients with depression and ii) 0% patients with depression. How will you disentangle depression effects and LSS effects in studies where healthy controls did not show depressive disorders? - You are not defining any information on depression as inclusion criterion. Please describe how this will impact on the results. Please revise the conclusion of your abstract and discussion section accordingly. (Reviewer: Ulrike Held)

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Martin N. Stienen, Stanford University, CA (USA)

The rationale of the review and meta-analysis is given. The methods described in the article appear very sound; the analysis is registered and quality and risk of bias of each included study is assessed.

I congratulate the authors to their well-planned research and wish them luck that they will be able to include a satisfactory number of high-quality studies in order to obtain meaningful results. Response:

We warmly thank the Reviewer for this positive feedback and for the precious encouragement.

Please find some aspects below that the authors may find helpful:

- Why do the authors confine their methods to the QoL measures included in table 1? There are further, frequently used measures (e.g. SF-12, EQ5D, COMI, ...) that are used frequently for degenerative lumbar spine patients. Response: We thank the Reviewer for this comment. We have revised inclusion criteria regarding QoL measures as broad as possible to avoid missing any

measures available in the literature (given the large number of published QoL measures). Specifically, we have revised accordingly inclusion criterion "*d*" in the "Eligibility criteria" paragraph.

- Prior research indicated the increased use of "objective" functional tests as outcome measures, supplementing the use of PROMS (typically questionnaire-based) and adding a further dimension to the comprehensive outcome evaluation (for review, see Spine J. 2019 Jul;19(7):1276-1293. doi: 10.1016/j.spinee.2019.02.014). I would like to propose the authors consider objective outcome measures in their review, as well, as those were shown to be less biased by depressive comorbidity, compared to PROMs (compare e.g. Spine J. 2017 Jun;17(6):807-813. doi: 10.1016/j.spinee.2016.12.004). Response: We thank the Reviewer for providing us with this precious comment. Following this suggestion, we have revised our manuscript by adding measures of objective functional impairment in the objectives, eligibility criteria, and methods. Please, see the corresponding sections for the revisions.

Reviewer: 2

Ulrike Held, Dept. of Biostatistics, Epidemiology, Biostatlstics and Prevention Institute, University of Zurich, Switzerland

Ferretti and co-authors presented a systematic review protocol on the comparison of health-related quality of life between patients with lumbar spinal stenosis (LSS) and healthy controls, while accounting for depression. The authors address a relevant topic as the prevalence of LSS in the elderly is rising, impacting on health care utilization and increasing medical costs. The review protocol is clearly written and well-structured.

In order to further improve the quality of the paper, the authors should address the following comments:

Response: We thank the Reviewer for having read carefully our paper and for being interested in our work. We have revised the manuscript trying to follow as much as possible all these constructive comments.

Search strategy and selection of studies

- Search strategy: the authors should specify the specific search string they intended to use. When they make the search, the authors should be able to give a specific date when the search was conducted, not a week as is currently stated (3.-10. July). Response: As requested, a specific date when the search will be conducted is given. Please, see revised "Information sources and search procedure" paragraph.

- Related to the above mentioned point: do the authors have expertise to conduct a search or are the authors planning to involve an information specialist, e.g. an experienced librarian? In order to define and validate the search string, and to remove duplicates from the different electronic databases I would strongly advise you to consider this. Response: We thank the Reviewer for this advice. We have followed it and planned that an experienced librarian will conduct the search, particularly in validating the search string across different electronic databases. Please, see revised "Information sources and search procedure" paragraph.

- I would suggest to make the title and abstract screening in one stage, and then in a second stage for the full-texts. We have followed the Reviewer's suggestion. Please, see revised "Selection of studies" paragraph.

Potential moderators of between-group difference

- The authors are planning to include depression as a potential moderator in their analyses. Specifically, authors are planning to include the "percentage of patients with depressive disorders" as a moderator variable. I see the potential problem that some studies will not report on depression, or will not explicitly state the percentage that you are looking for. Please state how you are going to include these publications. Response, we thank the Reviewer for this precious comment. We have specified how we are going to include or handle these studies. Please, see revised "Subgroup and moderator analyses" paragraph as follows: "*If such data is not given in the study paper (i.e., the paper does not report on depression, or does not explicitly state the percentage of patients with comorbid depressive disorders), the corresponding author will be contacted to request this information. In this case, the study will be included in the analysis only if the corresponding author is available to provide the necessary data*.". In addition, this potential problem has been highlighted as a limitation in the Discussion section as follows: "*Another potential problem is that some studies will not report on the data necessary to code the moderators (e.g., they will not explicitly state the percentage of depressive disorders) or the authors are not available to provide them*.".

- Apart from depression, there may be other – even more important – parameters that may affect the effect sizes on health-related quality of life. Severity of LSS as measured with disability and pain scales, comorbidities, age will also play a role. Except for age, these confounders will not be accounted for through matching. The depression's moderating effect that you aim to quantify may be actually confounded by a set of other variables. Response: We thank the Reviewer for providing us with this important comment. As suggested, we have included additional moderators to be investigated. Please, see the added statements in "Rationale and objectives of the present protocol" and in "Subgroup and moderator analyses" paragraphs.

- There is no detailed description of the methods you are planning to use for estimating the moderating effect of depression, reported in patient and control groups. Response: We thank the Reviewer for this comment. We have added a statement which specifies the statistical methods planned to be used for estimating the moderating effects of all the moderators. Please, see revised "Subgroup and moderator analyses" paragraph as follows: "*The relationship between the effect sizes and all these moderators will be investigated by conducting weighted least squares meta-regression analyses*.".

- In Table 3 you state that you are going to report mean and standard deviation for age. How will you include publications not reporting age, or reporting median and interquartile ranges or ranges instead? Response: We thank the Reviewer for this precious comment. We have specified how mean age and standard deviation will be obtained if the paper does report them or report median and interquartile ranges. Please, see the changes added in Table 3 where coding of age is provided, as follows: "*If the study does not report these data, they will be requested from the corresponding author. If this is not the case, mean and standard deviation will be estimated from median and interquartile ranges through the formula proposed by Wan and colleagues [43]. Otherwise, the study will be excluded from the analyses involving data on age*.". In addition, the fact that some studies will not report the data for moderator coding or the authors are not available to provide them has been

highlighted as a potential limitation in the “Strengths and limitations of the study” bullet points and in the “Discussion” section.

Outcomes

- You describe that health related quality of life (QoL) will be either self-reported or interview based. Are you planning any subgroup analyses? Response: We thank the Reviewer for this precious comment. We have added a specific paragraph on subgroup analyses specifically on self-reported versus interview-based measures. Please, see the added statements in “Subgroup and moderator analyses” as follows: “*If significant inconsistency is found, subgroup analyses will be conducted for studies using (a) only clinician-administered interviews to measure health-related quality of life, (b) self-report questionnaires of health-related quality of life.”.*

- Why are you not considering the EQ-5D scale for measuring quality of life? Response: We thank the Reviewer for this comment. We recognize that there are other eligible measures of QoL. Similarly to our response to the first comment of Reviewer 1, we have made inclusion criteria regarding QoL measures as broad as possible in order to try to avoid missing any measures available in the literature (given the large number of QoL measures). Please, see the revised inclusion criterion “*d*” in the “Eligibility criteria” paragraph.

- Please state how you are going to address the problem that multiple publications about the same cohort of LSS patients and controls may exist. Which one are you planning to include in the meta-analysis? Are you planning to include case series? Response: We thank the Reviewer for pointing out this relevant point. We have added a statement which specifies the systematic methodology used to examine and handle potential duplications. Please, see the revision in “Selection of studies” paragraph as follows: “*During the whole selection process, potential duplicates will be handled and excluded by following the systematic detection heuristic proposed by Wood [42]*.”. We have decide to exclude case series because data in this type of publications are not suitable for our planned methods of meta-analysis. A statement regarding this point has been added in the “Eligibility criteria” paragraph.

- Will you exclude any study from the meta-analysis because of low methodological quality? Response: We have specified more thoroughly how the findings of the quality assessment will be used. In particular, since the NOS does not provide a cut-off score to allow us to exclude studies with poor quality and restrict the analysis on the high-quality ones, we have followed the Reviewer’s comment and we have chosen to investigate the association between study quality and effect sizes through a moderator analysis where the scores on the NOS are included as moderators. Please, see revised “Subgroup and moderator analyses” paragraph.

Section: Meta-analytic procedure

Summary measures

- Please replace “random effect meta-analysis” by “random effects meta-analysis” throughout. Response: We have corrected it.

- Please revise the sentence “For all the analyses, the p-value will be set at 0.05.” I would advise not to pre-specify a level of significance but rather quantify the evidence on a continuous scale. Response: We thank the Reviewer for this comment. As suggested, we have quantified the evidence on a continuous scale. Please, see the change in revised “Summary measures“ paragraph.

- You describe the categories of Cohen's effect sizes, >= 0.8, 0.5-0.8, and <=0.2. What about the missing level from 0.2-0.5 – did you mean that this represents "low" effect size? Response: We thank the Reviewer for this comment. We have corrected this typo. Please, see revised "Summary measures" paragraph.

- Will you need a minimum number of studies for pooling or for being able to quantify the moderating effect of depression? Response: We thank the Reviewer for this comment. We have added a statement specifying the minimum number of studies for pooling or quantifying the moderating effects. Please, see the revisions in "Subgroup and moderator analyses" paragraph, as follows: "*According to Valentine et al.'s recommendations [53], the minimum number of studies for pooling the data and performing effect size calculation will be 2. Following the guidelines for a continuous study level variable proposed by Fu et al. [54], at least 6 to 10 studies will be necessary to investigate the moderating effects through meta-regression*.".

Publication bias

- Please describe that the trim-and-fill method will result in a new summary estimate. Are you planning to report the new summary estimate (after trim-and-fill) together with the original estimate in every meta-analysis? Please also report on the number of studies "filled". Response: We thank the Reviewer for this relevant suggestion. We have added some statements which describe more thoroughly the procedure planned to perform the trim-and-fill method. Please, see revised "Publication bias" paragraph, as follows: "*The trim-and-fill method is aimed to assess the impact of adjustment for small study bias. It removes studies until symmetry in the funnel plot is achieved, recalculating the centre of the funnel before the removed studies are replaced together with their "missing" mirror-image counterparts [50]. This procedure will result in a revised summary estimate calculated using all of the original studies, together with the hypothetical "filled" studies. The new summary estimate (after trim-and-fill) will be reported together with the original estimate in every meta-analysis*.".

Moderator coding and analysis

- You state that "if inconsistency between effect sizes is found" depression will be investigated as moderator. In fact, you should evaluate the moderating effect in all meta-analyses (if enough studies can be retrieved). Perhaps you want to describe pre-planned subgroup analyses in those studies with i) 100% patients with depression and ii) 0% patients with depression. How will you disentangle depression effects and LSS effects in studies where healthy controls did not show depressive disorders? Response: We thank the Reviewer for this relevant comment. The reviewer points out an important problem and a potential limitation of our review. We expect that it is quite unlikely to find studies with 100% and 0% patients with depressive disorders (and also 100% controls with depressive disorders). With regard to the second aspect, we think that we might disentangle depression effects and LSS effects in studies with controls without depressive disorders by adding the percentage of controls with depressive disorders in the moderator analysis model. This strategy can allow us to include the studies with any percentages of depressive disorders in the control groups. If studies where healthy controls without depressive disorders are retrieved, we will include the percentage of healthy controls with depressive disorders (range = 0-100%) as moderator in the multivariate meta-regression model to examine whether the percentage of controls with depression moderate the effect sizes. We can

expect that in studies where the percentage of controls with comorbid depressive disorders is higher, the difference in the quality of life/functional impairment levels between ptients and control is lower. We think that this strategy may help us as much as possible to try to disentangle such effects. Please, see the revisions added in “Subgroup and moderator analyses” paragraph, as follows: “*If studies with controls without depressive disorders are retrieved, in order to disentangle depression effects and LSS effects in such studies, the percentage of controls with depressive disorders will be included as moderator. This strategy will aim to examine whether the percentage of controls with depression moderates the effect sizes. It can be expected that in studies where the percentage of controls with comorbid depressive disorders is higher, the difference in the quality of life/functional impairment levels between patients and control is lower*.”.

- You are not defining any information on depression as inclusion criterion. Please describe how this will impact on the results. Please revise the conclusion of your abstract and discussion section accordingly. Response: We thank the Reviewer for this precious comment. We have added more information about depression as inclusion criteria. Please, see the revisions in the Eligibility criteria paragraphs, as follows: “*If the study assessed the number of patients with comorbid depressive disorders, this comorbidity had to be evaluated by the criteria for a major depressive disorder according to an international standardized diagnostic system such as the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Classification of Diseases (ICD).”*. In addition, we have also discussed and revised the abstract and discussion sections accordingly. Please, see the revisions in the abstract and discussion sections, as follows: “*Higher percentages of LSS patients with depression may be expected to be related to poorer quality of life. Depressive comorbidity might impact negatively on quality of life because it is associated with dysfunctional coping, disability and psychophysiological symptoms*.”.

VERSION 2 – REVIEW

REVIEWER	Martin N. Stienen Stanford University Hospital, CA (USA)
REVIEW RETURNED	21-Sep-2019

GENERAL COMMENTS	The authors have responded very adequately to my previous comments. They also made several changes to the manuscript, reacting on the valuable comments of reviewer 2. Altogether I believe that their methodology is now well-described and I appreciate the authors pre-define their strategy before conducting the meta-analysis. I wish them good luck with their interesting project.

REVIEWER	Ulrike Held University of Zurich, Dept. of Biostatistics
REVIEW RETURNED	08-Oct-2019

GENERAL COMMENTS	The authors have fully addressed my comments, thank you for your replies. I have no further comments.