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)	Measurement properties of instruments assessing permanent functional impairment of the spine: a systematic review protocol
AUTHORS	Goes, Suelen; Trask, Catherine; Boden, Catherine; Bath, Brenna; Ribeiro, Daniel; Hendrick, Paul; Clay, Lynne; Zeng, Xiaoke; Milosavljevic, Stephan

VERSION 1 – REVIEW

REVIEWER	Dr. Fehlings
REVIEW RETURNED	04-Oct-2017

GENERAL COMMENTS	Article review Title: Measurement properties of instruments assessing permanent functional impairment of the spine: a systematic review protocol In the manuscript titled “Measurement properties of instruments assessing permanent functional impairment of the spine: a systematic review protocol” submitted for publication in BMJ Open, the authors propose to conduct a systematic literature review and potentially a meta-analysis to summarize the current level of evidence of measurement properties of instruments used for assessing permanent functional impairment (PFI) of the spine. The goal is to provide clinicians, researchers and health policy makers with a rigorous and up-to-date evaluation of the reliability and validity of instruments currently available for assessing PFI of the spine. This is a well-written manuscript and a very promising systematic review which addresses an important topic. Keeping in mind that this is a systematic review protocol, the recommendation is for this manuscript to be published with minor revisions. However, since it is a protocol and not a compete study, many of the Review Checklist questions were answered “NO” for the following reasons: Q1 and Q7, related to Statistics: the planned statistics seem appropriate and are described satisfactorily, but they have not been performed yet. Thus, it is impossible at this point to say whether or not the statistics used were appropriate and properly reported, and if this paper requires specialist statistical review. Q9 to Q12: Although the Methods presented align with the research question, the study has not been completed yet; there is no results section, the discussion does not relate to the results and the limitations of the study cannot be addressed.

	Besides, Q2: In the introduction, unless this manuscript is published in a specific section dedicated to publishing study protocols, it would be worthwhile specifying that this article presents the protocol of an ongoing systematic review. Specifically, state that the manuscript aims to synthetize and evaluate the current level of evidence of measurement properties of instruments used for assessing PFI. Minor points 1) In the Abstract, Strengths and limitations of this study, and Discussion: only an impact study can conclude what the findings of a specific study provided. For instance, in the Abstract, it would be recommended to state that “The findings may [or] are likely to provide a foundation and direction for future research priorities for assessing spinal PFI.” or that the authors anticipate that “The findings will provide a foundation and direction for future research priorities for assessing spinal PFI.” 2) There is redundancy throughout the paper; every piece of information should be addressed once in the appropriate section. For instance, only the first sentence of the third paragraph is relevant to the Introduction, the rest could be incorporated into the first paragraph of the Methods to clarify how the authors decided to defined the outcome, i.e. PFI. Similarly, the fifth paragraph of the introduction mostly pertains to defining PFI, which should be presented in the Methods. Hence, the third to fifth paragraph of the introduction could easily be condensed into one: Permanent functional impairment (PFI), or permanent impairment, is a rating systems used by compensation authorities, to establish an appropriate level of financial compensation5-9. Evaluation of PFI requires selecting appropriate outcome measures10. However, the metrics of PFI ratings are not uniformly specified nor universally adopted by workers’ compensation boards, varying in terms of specific PFI rating guides as well as the adoption of function-based criteria 6 9 11. The feasibility, reliability, validity, utility, and cost are important aspects that should be taken into consideration when selecting clinical measurement instruments for determination of PFI 12. For instance, impairment can contribute to limited function and ultimately may have consequences for physical functioning, yet a clinical examination finding of an impairment does not always correspond to a functional loss 15. Also, Screening, article selection and data extraction can be discussed under the same sub-heading since the overall reviewing technique is the same, such as: Citation screening based on title, abstract and full-text stages as well as data extraction will be performed by two reviewers independently. Any discrepancies will be solved by consensus and/or consultation with a third member of the research team prior to making a final decision. More specifically, a screening tool will be developed (…) specificity at these stages. Due to anticipated uncertainty (…) synthesis effort. To avoid analysing the same data (…) clarifications or missing data. We will extract information relating to (1) the measurement properties (i.e. reliability, validity, and responsiveness), which we will consider as distinct study dimension, of the measurement instruments for assessing PFI of the spine

	(e.g. assessment of RoM, muscle strength, coordination, endurance, and sensation, as well as ability to perform basic daily living and work activities, and duration of the impairment) and (2) the type of inferential test and (3) the test results for each article. The extraction process will be tracked using a standardized data extraction form. 3) The rationale for not limiting the search to any specific date range is valid. However, it might be worthwhile adding two other data extraction items pertaining to determining the number of citations and the date of last citation for each instrument measure identified. It would be useful to know how often an instrument was reportedly used, in what specific context and when it was last used for each specific context. 4) Table 2 mentions excluding articles that are not published in English, but this exclusion criterion is not described in the eligibility criteria. 5) The authors might want to further specify their definition of "working age" as to be between 18-65 years of age. Although it is not unusual nowadays to be working past 65 years old, it is unlikely that someone over 65 would be evaluated for PFI in order to obtain work-related compensation. 6) Also, it would improve the logical flow of the paper if the first sentence of the third paragraph of the eligibility criteria (line 178) followed the age criterion (line 165). 7) It seems appropriate to include all types of spinal conditions in individuals of working age evaluated for PFI. Since the aim of this study is to synthesize the literature concerning measurement properties of instruments used for assessing PFI of the spine, this should be done regardless of the etiology of the PFI, given that different people can have the same degree of PFI from different spinal conditions. 8) It would be appropriate to mention that systematic reviews and meta-analyses will be excluded but their referenc

REVIEWER	Arianne Verhagen Erasmus MC, The Netherlands
REVIEW RETURNED	10-Oct-2017

10-Oct-2017

GENERAL COMMENTS	This manuscript describes a the protocol of a systematic review of studies evaluating measurement instruments. This protocol has a lot of text in the method section that can be shortened and revised. General comments: This protocol has a lot of text in the method section that can be shortened and revised. I reads a bit like a novel instead of a research protocol. Specific comments: 1. Methods, design. Please use the subheadings as advised in the PRISMA statement, and the subheading 'design' should be followed with a short description of the design. Now it includes taxt on the description of what the authors consider a PFI (should be under inclusion criteria) and also a lot of text on the interpretation of PRISMA (can be deleted).

	2. Methods, eligibility criteria. Please shortly mention first all inclusion criteria, and next the exclusion criteria. Now there is no clear description of the inclusion criteria, but more a novel kind of text. Also the exclusion criteria are complementary to the inclusion criteria, this means they are redundant. Which studies that fit the inclusion criteria will be deleted? This should be mentioned as exclusion criteria. 3. Methods, study selection. Please mention that the procedure of study selection will be done in two steps. First two review authors independently screen all titles and abstracts (using a shorter form as described) and all references selected by one (or both) review authors should be retrieved for step 2: full paper selection. Step 2 will need consensus or third party adjudicion in case of disagreements. 4. Methods, data extraction. This paragraph is rather unclear. Please describe you developed a data extraction form that you will pilot test on 2 or 3 studies not included in the review. Also the procedure of data extraction should be explained (two review authors independently). 5. Methods, quality assessment. Please explain here the definitions of validity reliability etc. Also the procedure of quality assessment should be explained (two review authors independently). Why is the form not presented as a table? 6. Methods, analysis. The authors state they will perform a statistical summary of results using Revman. Please first describe how you will summarise all other info (study characteristics, kinds of tests etc). I do not think that a statistical summary is possible in studies evaluating measurement properties. In case you can, please explain how. Also I do not think you can use RevMan to do so. 7. Methods, please explain how you assess meta-bias and publication bias, you only explained how you are trying to assess reporting bias.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 - Dr. Fehlings

- This is a well-written manuscript and a very promising systematic review which addresses an important topic. Keeping in mind that this is a systematic review protocol, the recommendation is for this manuscript to be published with minor revisions. However, since it is a protocol and not a complete study, many of the Review Checklist questions were answered “NO” for the following reasons:

Q1 and Q7, related to Statistics: the planned statistics seem appropriate and are described satisfactorily, but they have not been performed yet. Thus, it is impossible at this point to say whether or not the statistics used were appropriate and properly reported, and if this paper requires specialist statistical review. / Q9 to Q12: Although the Methods presented align with the research question, the study has not been completed yet; there is no results section, the discussion does not relate to the results and the limitations of the study cannot be addressed.

 Authors' response: We would like to thank you for your thoughtful and thorough revision. It was a pleasure to read your comments because we truly believe that they substantially contribute to improve the quality of every section of the manuscript. We worked on the suggestion and corrections you pointed out. The changes in the manuscript are in red.

- Besides, Q2: In the introduction, unless this manuscript is published in a specific section dedicated to publishing study protocols, it would be worthwhile specifying that this article presents the protocol of an ongoing systematic review. Specifically, state that the manuscript aims to synthetize and evaluate the current level of evidence of measurement properties of instruments used for assessing PFI.

 Authors' response: We agree and have insert this information on Page 4, lines 120-123: "This manuscript presents the protocol of an ongoing systematic review with the objective of systematically review and synthesize the literature concerning measurement properties of the various and different instruments used for assessing PFI of the spine".

1) In the Abstract, Strengths and limitations of this study, and Discussion: only an impact study can conclude what the findings of a specific study provided. For instance, in the Abstract, it would be recommended to state that "The findings may [or] are likely to provide a foundation and direction for future research priorities for assessing spinal PFI." or that the authors anticipate that "The findings will provide a foundation and direction for future research priorities for assessing spinal PFI."

 Authors' response: We agree and have insert this information on Page 2, lines 64-65; 'Abstract' now reads: "Findings of this review may be applicable to clinicians, policy-makers, workers' compensation boards, other insurers, and health and safety organizations. The findings will likely provide a foundation and direction for future research priorities for assessing spinal PFI". Page 10, lines 279-286: 'Ethics and dissemination' section now reads: "This systematic review offers a feasible means for synthesizing the evidence specific to spinal PFI assessment; and our results will likely provide unique insights concerning the breadth and depth of literature in the area. Outcomes of this review will be applicable to clinicians, policy-makers, worker's compensation boards and health and safety organizations. In particular, findings will likely provide a foundation and direction in terms of research priorities for assessing PFI of the spine. Summarizing the nature and strength of the evidence regarding the reliability, validity and responsiveness of spinal PFI measures will also inform future research and policy in this field".

2) There is redundancy throughout the paper; every piece of information should be addressed once in the appropriate section. - For instance, only the first sentence of the third paragraph is relevant to the Introduction, the rest could be incorporated into the first paragraph of the Methods to clarify how the authors decided to defined the outcome, i.e. PFI. - Similarly, the fifth paragraph of the introduction mostly pertains to defining PFI, which should be presented in the Methods. Hence, the third to fifth paragraph of the introduction could easily be condensed into one: * Permanent functional impairment (PFI), or permanent impairment, is a rating systems used by compensation authorities, to establish an appropriate level of financial compensation5-9. Evaluation of PFI requires selecting appropriate outcome measures10. However, the metrics of PFI ratings are not uniformly specified nor universally adopted by workers' compensation boards, varying in terms of specific PFI rating guides as well as the adoption of function-based criteria (6 9 11). The feasibility, reliability, validity, utility, and cost are important aspects that should be taken into consideration when selecting clinical measurement instruments for determination of PFI (12). For instance, impairment can contribute to limited function and ultimately may have consequences for physical functioning, yet a clinical examination finding of an impairment does not always correspond to a functional loss (15).

 Authors' response: We agree and the paragraph was reorganized. Page 3, lines 97-106, which now reads: "Permanent functional impairment (PFI), or permanent impairment, is a rating systems used by compensation authorities, to establish an appropriate level of financial compensation 5-9. Evaluation of PFI requires selecting appropriate outcome measures 10. However, the metrics of PFI ratings are not uniformly specified nor universally adopted by workers' compensation boards, varying in terms of specific PFI rating guides as well as the adoption of function-based criteria 6 9 11. Feasibility, reliability, validity, utility, and cost are important aspects that should be taken into consideration when selecting clinical measurement instruments for determination of PFI 12. Although impairment can contribute to limited function and ultimately may have consequences for physical functioning, a clinical examination finding of impairment does not always correspond to a functional loss 13".

- Also, Screening, article selection and data extraction can be discussed under the same sub-heading since the overall reviewing technique is the same, such as: * Citation screening based on title, abstract and full-text stages as well as data extraction will be performed by two reviewers independently. Any discrepancies will be solved by consensus and/or consultation with a third member of the research team prior to making a final decision. More specifically, a screening tool will be developed (…) specificity at these stages* Due to anticipated uncertainty (…) synthesis effort. * To avoid analysing the same data (…) clarifications or missing data.* We will extract information relating to (1) the measurement properties (i.e. reliability, validity, and responsiveness), which we will consider as distinct study dimension, of the measurement instruments for assessing PFI of the spine (e.g. assessment of RoM, muscle strength, coordination, endurance, and sensation, as well as ability to perform basic daily living and work activities, and duration of the impairment) and (2) the type of inferential test and (3) the test results for each article. The extraction process will be tracked using a standardized data extraction form.

 Authors' response: We understand the reviewer's concern, but as per Reviwer 2, we are following the subheadings as advised in the PRISMA statement for reporting systematic reviews.

3) The rationale for not limiting the search to any specific date range is valid. However, it might be worthwhile adding two other data extraction items pertaining to determining the number of citations and the date of last citation for each instrument measure identified. It would be useful to know how often an instrument was reportedly used, in what specific context and when it was last used for each specific context.

 Authors' response: We understand the reviewer's point of view, but it seems like this would be more relevant for measuring usage of measures in research contexts, rather than in clinical practice, which would be out of the scope of our review.

4) Table 2 mentions excluding articles that are not published in English, but this exclusion criterion is not described in the eligibility criteria.

 Authors' response: We understand the inclusion and exclusion criteria were not very clear. Furthermore, we have reworded the eligibility criteria on Page 5, lines 159-163, which now reads: "We will exclude letters to the editor, book reviews, and short communications. We will also exclude clinical protocols, case reports and series, systematic reviews, meta-analysis, articles not published in English-language, studies intended for screening, diagnosis and prognosis of spinal pathologies as well as studies with specimen-, cadaveric-, cellular-, artificial-, and computer-based models." We have also updated Table 2 with more specific questions for full-text screening stage in order to answer our research question.

5) The authors might want to further specify their definition of “working age” as to be between 18-65 years of age. Although it is not unusual nowadays to be working past 65 years old, it is unlikely that someone over 65 would be evaluated for PFI in order to obtain work-related compensation.

 Authors’ response: We agree and are not using the expression “working age”. The sentence was corrected. Page 5, lines 157-158: “We will concentrate on studies of adults with spinal disorders between 18 and 65 years of age”.

6) Also, it would improve the logical flow of the paper if the first sentence of the third paragraph of the eligibility criteria (line 178) followed the age criterion (line 165).

 Authors’ response: We agree and have reworded the eligibility criteria, stating the inclusion and exclusion criteria in two separate paragraphs. Pages 5-6, lines 152-169.

7) It seems appropriate to include all types of spinal conditions in individuals of working age evaluated for PFI. Since the aim of this study is to synthesize the literature concerning measurement properties of instruments used for assessing PFI of the spine, this should be done regardless of the etiology of the PFI, given that different people can have the same degree of PFI from different spinal conditions.

 Authors’ response: We understand the reviewer’s point of view; however, workers compensation boards can provide different financial compensation for each specific spinal disease or disorder. Furthermore, specific rating systems might be applied to establish an appropriate level of financial compensation, in which each case could be adjudicated based on very particular characteristics of a spinal disease. For that reason, we will exclude studies of individuals with spinal conditions caused by congenital and developmental abnormalities, neoplasm, infection disorders, and systemic inflammatory disorders.

8) It would be appropriate to mention that systematic reviews and meta-analyses will be excluded but their referenc

 Authors’ response: We have added that information in Page 5, lines 159-160: “We will also exclude clinical protocols, case reports and series, systematic reviews, meta-analysis, […]”

Reviewer: 2 - Arianne Verhagen
This manuscript describes the protocol of a systematic review of studies evaluating measurement instruments. This protocol has a lot of text in the method section that can be shortened and revised.

General comments: This protocol has a lot of text in the method section that can be shortened and revised. It reads a bit like a novel instead of a research protocol.

 Authors’ response: We would like to thank you for your thoughtful and thorough revision. It was a pleasure to read your comments because we truly believe that they substantially contribute to improve the quality of every section of this manuscript. We worked on the suggestion and corrections you pointed out. In addition, as per your suggestion, we have concise the information in order to be more objective. The changes in the manuscript are in red.

1. Methods, design. Please use the subheadings as advised in the PRISMA statement, and the subheading ‘design’ should be followed with a short description of the design. Now it includes text on the description of what the authors consider a PFI (should be under inclusion criteria) and also a lot of text on the interpretation of PRISMA (can be deleted).

 Authors' response: We have reorganized the methods section following the subheadings as advised in the PRISMA statement for reporting systematic reviews. We also included our PFI definition in the eligibility criteria. The changes can be seen from Pages 4-9, lines 128-271.

2. Methods, eligibility criteria. Please shortly mention first all inclusion criteria, and next the exclusion criteria. Now there is no clear description of the inclusion criteria, but more a novel kind of text. Also the exclusion criteria are complementary to the inclusion criteria, this means they are redundant. Which studies that fit the inclusion criteria will be deleted? This should be mentioned as exclusion criteria.

 Authors' response: We have reworded the eligibility criteria, stating the inclusion and exclusion criteria in two separated paragraphs. Pages 5-6, lines 152-169: "We have elected not to limit publication range in order to be thorough in collecting documented evaluations of measurement properties of instruments assessing PFI of the spine and thus we will include peer-reviewed, full-text articles over the fully available date range. We will focus on cross-sectional and cohort studies investigating measurement instruments for assessing all components described as appraising PFI, as well as assessing the measurement properties of these assessment instruments in terms of validity, reliability, and responsiveness. We will concentrate on studies of adults with spinal disorders between 18 and 65 years of age.
We will exclude letters to the editor, book reviews, and short communications. We will also exclude clinical protocols, case reports and series, systematic reviews, meta-analysis, articles not published in English-language, studies intended for screening, diagnosis and prognosis of spinal pathologies as well as studies with specimen-, cadaveric-, cellular-, artificial-, and computer-based models. In addition, given the broad scope of spinal disorders and considering that workers compensation boards do not uniformly specify or universally provide financial compensation for all specific diseases and non-specific musculoskeletal disorders, we will exclude studies of individuals with spinal conditions caused by congenital and developmental abnormalities, neoplasm, infection disorders, and systemic inflammatory disorders 25. Exclusion of studies will take place at the screening stage instead of in the literature search phase to avoid the risk of excluding relevant articles, except for the non-articles published in journals."

3. Methods, study selection. Please mention that the procedure of study selection will be done in two steps. First two review authors independently screen all titles and abstracts (using a shorter form as described) and all references selected by one (or both) review authors should be retrieved for step 2: full paper selection. Step 2 will need consensus or third party adjudicion in case of disagreements.

 Authors' response: We inserted this information on Page 7, lines 194-196, which now reads: "Study selection will be undertaken in two steps. First, two reviewers will independently screen all titles and abstracts, and all citations selected by both reviewers will be retrieved for step two: full-text screening."

4. Methods, data extraction. This paragraph is rather unclear. Please describe you developed a data extraction form that you will pilot test on 2 or 3 studies not included in the review. Also the procedure of data extraction should be explained (two review authors independently).

 Authors' response: We agree and have inserted this information the first paragraph on Page 7, lines 206-208, which now reads: "We will develop data extraction forms, as outlined in table 3, which will be pilot tested on three studies not included in the review. Two reviewers will extract data independently. Any disagreement will be resolved through a third party adjudication."

5. Methods, quality assessment. Please explain here the definitions of validity reliability etc. Also the procedure of quality assessment should be explained (two review authors independently). Why is the form not presented as a table?

 Authors' response: We have reorganized the information presented in 'Quality assessment of individual studies' section as suggested by the reviewer. The definitions of the measurement properties are presented in the new table (table 4). Page 8-9, lines 233-247: "Included studies in each sub-group will be appraised independently by two reviewers as to their methodological quality using the consensus-based standards for the selection of health measurement instruments (COSMIN) criteria 27 28. Any disagreement will be resolved by consultation with a third member of the research team. The COSMIN checklist is a consensus-based tool designed to evaluate the methodological quality of studies investigating measurement properties. The instrument shows appropriate levels of agreement 27 and, based on its content validity, is a recommended tool for assessing the methodological quality of studies evaluating measurement properties of outcome measures within a systematic review 28. The tool evaluates the following measurement constructs: reliability; measurement error; content validity; structural validity; hypotheses testing; cross-cultural validity; and criterion validity. Responsiveness and interpretability with five to18 items concerning methodological standards for how each measurement property should be assessed (see Table 4 for definitions). The methodological quality of a study will be considered adequate if all items in a measurement property are considered adequate. Each item is scored on a four-point rating scale (i.e., ''poor'', ''fair'', ''good'', or ''excellent'')".

6. Methods, analysis. The authors state they will perform a statistical summary of results using Revman. Please first describe how you will summarise all other info (study characteristics, kinds of tests etc). I do not think that a statistical summary is possible in studies evaluating measurement properties. In case you can, please explain how. Also I do not think you can use RevMan to do so.

 Authors' response: We understand the reviewer's concern and we reworded the 'Planned methods of analysis' section. Page 9, lines 250- 266: "A narrative synthesis will be presented in text and table formats, with the intent of summarizing and discussing the sample and methodological aspects, as well as the findings regarding measurement properties of the included studies assessing PFI in individuals with spinal conditions. Tables will provide general information of the studies (i.e. authors, country, and population parameters, such as age, gender distribution, setting etc.), and will summarize reliability, validity and responsiveness data with associated study quality indicators (COSMIN checklist).
In order to determine the best available method for measuring PFI of the spine, each identified instrument will be ranked using a range of measurement performance metrics identified in the COSMIN checklist. The findings will be presented and possible hypotheses for the results will be generated and discussed. In addition, gaps in the literature will be identified and discussed.
Where appropriate, results will be statistically summarized (i.e. meta-analysis) into forest plots with estimates of heterogeneity; and sensitivity analysis will be pursued by comparing results from studies with high and low-quality ranking. Where possible, we will weight a meta-analysis using both the study's sample size and their quality assessment as determined by the COSMIN checklist 28. However, we predict some heterogeneity will be identified in the various assessments of spinal PFI measures, which will likely make a meta-analysis difficult to apply".

7. Methods, please explain how you assess meta-bias and publication bias, you only explained how you are trying to assess reporting bias.

 Authors' response: Meta-bias is the general evidence for biased selection of research findings. We won't be able to assess publication bias, so we apologize for the misleading information. We only kept the information on reporting-bias in the "Risk of bias across studies' section. Page 9-10, lines 268-

271: “Since protocol studies that describe forthcoming evaluation of measurement instruments (e.g. validity, reliability, and responsiveness) are scarce, we will verify whether the pre-specified primary outcomes have been reported, and contact authors when we lack important data or information from the included study”.

VERSION 2 – REVIEW

REVIEWER	Michael Fehlings Vice Chair Research, Department of Surgery Professor of Neurosurgery McLaughlin Scholar in Molecular Medicine University of Toronto Gerry and Tootsie Halbert Chair in Neural Repair and Regeneration Head, Spinal Program Toronto Western Hospital, University Health Network Canada
REVIEW RETURNED	07-Dec-2017

GENERAL COMMENTS	The authors have adequately addressed the previous comments and this manuscript is now suitable for publication in BMJ Open.

REVIEWER	Arianneianne Verhagen Erasmus University Medical Centre The Netherlands
REVIEW RETURNED	21-Nov-2017

GENERAL COMMENTS	This manuscript is revised, several points are addressed properly, and some are not addressed and, to be honest, I think that the authors can do much better. This manuscript still has a lot of text, especially in the method section, and can be shortened with at least 30%. See specific comments below. Specific comments: 1. Methods, design. I advised that the subheading ‘design’ should be followed with a short description of the design only. This the authors did not do. It now contains description of the PICO (still including text on the description of what the authors consider a PFI) and should be placed under inclusion criteria under the subheading of eligibility criteria. 2. Methods, eligibility criteria. Here still is no clear description of the inclusion criteria. The subheading ‘eligibility criteria’ is over one page long with largely irrelevant information. Please just describe here the inclusion and exclusion criteria. 3. Methods, study selection. Looking at table 2 I get the feeling that selection based on title and abstract will take way too long. I hope this table will be revised majorly after the pilot phase. 4. Methods, quality assessment. Please explain here the definitions of validity reliability etc. Also the procedure of quality assessment should be explained (two review authors independently). Why is the form not presented as a table?

	5. Methods, analysis. The authors state they will perform a statistical summary of results (pooling), but they need to describe how they will do that. In the analysis section the proposed analysis should be stated in a way that someone else is able to repeat it. Now the text is too vague. If meta-analysis is not possible, please describe how you will analyse the results than, how will you come to your conclusion? 6. Methods, risk of bias across studies. This section can be deleted. Risk of bias is dealt with in the quality assessment section.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1 - Michael Fehlings
General comment:
The authors have adequately addressed the previous comments and this manuscript is now suitable for publication in BMJ Open.
Authors' response: We would like to thank you for your thoughtful revision.

Reviewer: 2 - Arianne Verhagen
General comment:
This manuscript is revised, several points are addressed properly, and some are not addressed and, to be honest, I think that the authors can do much better. This manuscript still has a lot of text, especially in the method section, and can be shortened with at least 30%. See specific comments below.
Authors' response: We would like to thank you for your thoughtful revision. We worked on the suggestions and corrections you pointed out and the changes are written in red.

Specific comments:
1. Methods, design. I advised that the subheading 'design' should be followed with a short description of the design only. This the authors did not do. It now contains description of the PICO (still including text on the description of what the authors consider a PFI) and should be placed under inclusion criteria under the subheading of eligibility criteria.
Authors' response: We have reduced the information in the subheading 'design'. Page 4, lines 129-132. "This systematic review protocol has been registered with International Prospective Register of Systematic Reviews (PROSPERO) (registration number CRD42017060390). This protocol will be reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) statement and checklist 23".

2. Methods, eligibility criteria. Here still is no clear description of the inclusion criteria. The subheading 'eligibility criteria' is over one page long with largely irrelevant information. Please just describe here the inclusion and exclusion criteria.
Authors' response: We have reduced the information in the subheading 'eligibility criteria'. Page 5, lines 137-153. "Inclusion criteria: We will include peer-reviewed articles, with no date limitation, investigating measurement properties of instruments that assess PFI of the spine in individuals with spinal disorders aged from 18 to 65 years.

When evaluating PFI, acceptable studies will include measures of impairment (i.e. RoM, muscle strength, coordination, endurance, and sensation), functional limitation (i.e. self-report instruments of physical function and functional performance measures) and permanency of the impairment (i.e. duration of the impairment and the likelihood of improvement) 3 24. Exclusion criteria: We will exclude letters to the editor, book reviews, and short communications. We will also exclude clinical protocols,

case reports and series, systematic reviews, meta-analysis, articles not published in English-language, studies intended for screening, diagnosis and prognosis of spinal pathologies as well as studies with specimen-, cadaveric-, cellular-, artificial-, and computer-based models. We will also exclude studies of spinal conditions caused by congenital and developmental abnormalities, neoplasm, infection, and systemic inflammatory disorders 25. Exclusion will take place at the screening stage, following the literature search, in order to avoid the risk of excluding relevant articles."

3. Methods, study selection. Looking at table 2 I get the feeling that selection based on title and abstract will take way too long. I hope this table will be revised majorly after the pilot phase.
Authors' response: We understand the reviewer's concern. However, for title and abstract screening, we will use the first seven questions, and any of them having the answer 'No' will exclude the article.

4. Methods, quality assessment. Please explain here the definitions of validity reliability etc. Also the procedure of quality assessment should be explained (two review authors independently). Why is the form not presented as a table?
Authors' response: We have explain the definition of validity, reliability and responsiveness on page 8, lines 232-243: "In the context of this review, validity, in general, defines how well the instrument under evaluation measures the construct it purports to measure. Criterion validity is the degree to which measurements are an adequate reflection of a previously used 'gold standard'. Content validity is an adequate reflection of the construct to be measured; construct validity is based on an assumption that the instrument truly measures what it is meant to; and structural validity implies the scores of an instrument is an adequate reflection of the dimensionality of the construct to be assessed. Reliability refers to the extent to which scores for individuals who have not changed are the same for repeated measurement under several conditions. These include using different sets of items from the same instrument (internal consistency); over time (test-retest); by different assessors on the same occasion (inter-rater); or by the same assessors (i.e. raters or responders) on different occasions (intra-rater). Responsiveness is the ability of the measurement instrument to detect change over time in the construct to be measured 27." In addition, the procedure of quality assessment (two review authors independently) was explained on page 8, lines 217-220: "Included studies will be appraised independently by two reviewers as to their methodological quality using the consensus-based standards for the selection of health measurement instruments (COSMIN) criteria 27 28. Any disagreement will be resolved by consultation with a third member of the research team." Finally, the form is presented as a table, please see table 4 on page 21.

5. Methods, analysis. The authors state they will perform a statistical summary of results (pooling), but they need to describe how they will do that. In the analysis section the proposed analysis should be stated in a way that someone else is able to repeat it. Now the text is too vague. If meta-analysis is not possible, please describe how you will analyse the results than, how will you come to your conclusion?
Authors' response: We have provided more information on how we will analyze the results on page 9, lines 247-263: "Where it is possible, mean values of statistical analysis (e.g. Cronbach's alpha, intra-correlation coefficient, standard error of measurement, smallest detectable change, effect sizes, etc.) will be calculated from pooled data from methodologically similar studies and the results statistically summarized via meta-analysis into forest plots with estimates of heterogeneity. In addition, sensitivity analysis will be pursued by comparing results from studies with high and low-quality ranking.

We will then weight such meta-analyses using both the study's sample size and their quality assessment as determined by the COSMIN checklist 28. However, we predict some heterogeneity will be identified in the various assessments of spinal PFI measures, which will likely make a meta-analysis difficult to apply.

In the event meta-analysis is not possible, descriptive and narrative syntheses will be presented in text and table formats, with the intent of summarizing and discussing the sample and methodological aspects, as well as the findings regarding measurement properties of the included studies assessing PFI in individuals with spinal conditions. Tables will provide general and comparative information of these heterogeneous and disparate studies (i.e. authors, country, and population parameters, such as age, gender distribution, setting etc.), and will summarize reliability, validity and responsiveness data with associated study quality indicators (COSMIN checklist)."

6. Methods, risk of bias across studies. This section can be deleted. Risk of bias is dealt with in the quality assessment section.
Authors' response: This section was deleted.