

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)	Outcome Measures Evaluating Physical Functioning and their Measurement Properties in Adolescent Idiopathic Scoliosis: Protocol for a Systematic Review
AUTHORS	Alamrani, Samia; Rushton, Alison; Gardner, Adrian; Falla, Deborah; Heneghan, Nicola

VERSION 1 – REVIEW

REVIEWER	Jiao Jiao Li University of Sydney, Australia
REVIEW RETURNED	08-Oct-2019

GENERAL COMMENTS	<p>Introduction, line 85-86 – This sentence does not make sense, please rephrase. Also please define the abbreviations as they are first used in the main text.</p> <p>Introduction – It would be useful to add an additional sentence or two at the end of the introduction to describe the potential impact/significance of conducting the proposed systematic review on the field of AIS research, or any potential influences on clinical practice (highlights from the 'implications of the study' section that is included later).</p> <p>Objective – Please rephrase this, currently it is quite convoluted with repeated words (measurement/measures, evaluate/evaluating) in the same sentence which makes it difficult to read and comprehend.</p> <p>Study methods – The inclusion criteria is very broad, with inclusion of all study and publication types and no defined time frames. If a very large number of eligible studies are retrieved this could make data management and synthesis very difficult. A preliminary search may be useful to better define the scope of this study and/or to help propose appropriate limits on the search (e.g. time frames, type of literature, language).</p> <p>Discussion – This section is currently absent but would be good to include to give some additional details and discussions about the study. For example, the major research question that is answered by the study if completed, and how results of the study can potentially be utilised in the field of AIS research and more broadly in orthopaedic research. Limitations of the study design should be discussed in more detail and potential strategies to address or at least minimise the effect of these limitations should be discussed.</p>
-------------------------	---

REVIEWER	Eric Parent University of Alberta
-----------------	--------------------------------------

GENERAL COMMENTS

This is a very important research project. The protocol is nicely presented but I suggest additional details could be reported and the rationale could be outlined more clearly.

Keywords:

the following are MESH terms and could replace some of your free text keywords.:

Scoliosis

Outcome Assessment (Health Care)

reproducibility of results

Physical Functional Performance

Please add the Prospero registration info to this protocol paper.

ABSTRACT:

Define physical functioning rather than give generalities about quality of life.

Rather than impairment should you use the positive corresponding term from the ICF: Body structure and function measure.

Specify the search timeline for each databases.

Describe the two stage nature of the search and the general structure of how you build the search. Which concepts are combined with AND.

Specify whether search limitations or screening criteria will be used for age of the patients, treatment types, curve severity, or language.

Specify how Grade will be used. What will the conclusion be about. Will you aim to make recommendations within each category of PROM, PBOM and IM or will you summarize the evidence about measurement properties for each tool under each measurement type.

Could you end the abstract by describing the planned impact of the proposed research?

Will you develop a research agenda or make recommendations for a core set of outcomes?

Highlight the value of completing this review as a conclusion.

Keywords listed after the abstract are spelled different than those on the cover page of the PDF. See my key word comment above.

The strength and limitation section is a mix of that and key points.

IF you are to focus on key point and it is allowable by the journal please consider the following:

Rather than emphasize being the first please emphasise the value of systematically reviewing this information. State what the results to be presented might be about and why it may be impactful.

Re the third point: Rather than emphasize what the study is not about maybe define your operational definition of Outcome measure of physical functioning. Readers would like to know what type of measurements are included in your review.

	<p>With regards to limitations then please consider language limitation and age limitation in the review.</p> <p>L82 If there is already a core set then justify why do a systematic review. Add a justification for you're your review still needed. In the discussion you refer to the core set being for general spine deformity and maybe there is a need for a more specific review.</p> <p>L85> Define all abbreviations in the main text on first use.</p> <p>L85 "a commonly reported outcome measure in AIS. [24] " is unclear English. Redundant with IMs already mentioned. Is this provided as an example?</p> <p>L90 and throughout the paper: Use patient first language. Always say individuals with AIS rather than AIS individuals to be in accordance with ICF recommendations. Similarly rephrase AIS population.</p> <p>L102. I recommend adequate instead of establishment. It is not enough to determine what the properties are they must be good to avoid bias.</p> <p>L106. You cannot say absence of existing evidence. It would be illogical to plan a review if there was no evidence. Specify why a review may be valuable. It could be to outline a research agenda or make recommendations on best measurements.</p> <p>L110 The first objective is not stated clearly. Do you mean to inventory the available physical functioning outcome measures?</p> <p>L112 Could you present some hypotheses?</p> <p>L114 State in which field there was expertise on your team L115 Specify what kind of methodological expertise (review, research, measurement?)</p> <p>L116 Despite some overlap in methodology, this handbook (for review on the effect of interventions) has low relevance to a SR on outcome measurements. The following would be much more relevant: https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf</p> <p>L:126 This may be a problem. Often Risser stage is not reported in measurement studies. Also all adults with AIS could have Risser 5. It may be best to describe an upper limit of age? What will you do in cases where the info is missing?</p> <p>L126 State whether you will have restrictions related to treatments, severity, Evaluation settings or countries / languages. IF not specify that you do not.</p> <p>L134 Could you explain what you will do with mixed tools that may assess physical functioning as well as other qualities. The SRS-22r would be such an example by assessing mental health in addition to function. Will you only extract properties for the function scale?</p> <p>L139 Please use the Positive labels of the ICF categories. body</p>
--	---

	<p>structure and function. (rather than impairment)</p> <p>L140 Since impairments do not always lead to limitation in function it would be best to state "Which may limit" rather than which limit</p> <p>L145 Could you also specify if you will aim to track studies on interpretability including studies providing norms.</p> <p>L150 only validation. I recommend Metrological studies or studies on measurement properties which is more inclusive of all the properties you are interested in.</p> <p>L153 Your appendix describes 2 searches. Here it should be clear that you will perform to types of searches: One for the inventory and one of the measurement properties. Should you describe how the first search would be used to identify tools for the second search?</p> <p>L154 Plan on using the TERWEE filter for clinimetrics. You refer to it elsewhere but don't mention it here.</p> <p>L155 for each database: Refer to the date range searched.</p> <p>L157 Should other language databases and search terms also be used? Specify languages targeted.</p> <p>L157 Why search pubmed in addition to MEDLINE. Justify.</p> <p>L158 Scoliosis is the same journal as Scoliosis and Spinal Disorders which has had a name change. Only identify the last name: Scoliosis and Spinal Disorders.</p> <p>L160 State a database or prepare a list of targeted conferences for which you will review the proceedings and specify over which range of dates.</p> <p>L161 define EthOS or cite.</p> <p>L163. There is not really much value in running library database searches indepently. Specify which searches would be done independently.</p> <p>L168 Here your refer to the 2 stages of searching It was not sufficiently clear earlier.</p> <p>L171Detail the duplicate finding strategy. (which tool and will you only excluded exact or also close duplicates.</p> <p>L174 Should there be two screening processes. Is covidence used? Report whether you will do two stages of screening (titltes and abstract VS full-text). Explain which criteria may be applied only at a secondary stage. For example reliability information may appear in RCT and not be mentioned in the abstract. Having broader criteria for title and abstract may allow capturing such info.</p> <p>L180 Please plan on reporting Kappa also.</p> <p>L184 Describe the process of contacting author and timeline allowed for responses. Number of attempts?</p> <p>L190. the study sample size might differ from the sample size used in the Measurement property assessment.</p>
--	--

	<p>It would be best to extract sample used in each of the analyses and report missing data.</p> <p>I recommend you should extract Curve severity, Diagnoses (what will you do if not purely AIS) Could extract curve types.</p> <p>This list does not extract information about the name of the measurement tools, nor about instructions on how to administer the measurement and the scoring.</p> <p>IN this proposal, ideally you would list for each COSMIN category of reliability validity and responsiveness what should be extracted.</p> <p>EG. for validity there could be the subtype, and the tools used as gold standard or tools used in assessing convergent validity... The statistic computed the estimate and the confidence interval</p> <p>For all measurement studies it would be important to report the score descriptive statistics as this would allow examining the effect of score distribution on measurement properties.</p> <p>should details needed to appraise study quality also be extracted here?</p> <p>L192 I find it debatable that there would be no gold standard since you are measuring not only self-reported function but also observed performance and impairments.</p> <p>L199 Please describe how this rating is done (worst score on the item indicate risk of bias or some average or...). Report the training of your reviewers in using this tool.</p> <p>L204 Please report the Kappa,</p> <p>L204 Describe if you will report scores only or will report each item result. I would recommend the latter as this would be more informative in setting a future research agenda.</p> <p>L205 Describe if you plan to do comparisons of the measurement properties between tools and what you would use to draw conclusions that one is better than another. Will you compare properties among subgroups. For example ceiling effect and responsiveness may differ quite a bit between surgical or severely affected patients and those with mild curves and treated with observation or conservative treatments. Since research is conducted on specific samples such comparison may inform research in conservative vs surgical care...</p> <p>L208 outline a priori criteria to undertake or abandon meta-analysis. Outlining what would constitute too much heterogeneity in terms of patients, setting and measurement properties assessed would be useful. Will you do the summary by categories PROM, PBOM and IM or by tools. For example I foresee you will have enough data on the SRS22 to metanalyse results for this tool.</p> <p>L218 similarly Specify if the GRADE approach is applied for conclusion about PROM, PBOM and IM categories or for each tool.</p>
--	--

	<p>L228 In patient oriented research there are models governing involvement and reporting. Please identify the stakeholder involved (describe them), Describe the purpose of the engagement, Describe their roles and describe the influence they had on the protocol. Below is an example of patient engagement model https://www.google.com/url?sa=i&source=images&cd=&ved=2ahUK Ewj5-t6VlbjIAhWJu54KHeT7AisQjRx6BAgBEAQ&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FSpectrum-of-patient-engagement-in-health-research_fig1_312408187&psig=AOvVaw1-Zdx1O_2k3c0hvfTq2_IK&ust=1572118805405560</p> <p>L236 I would not agree with lifting objects being a common functional limitation in scoliosis.</p> <p>L241 I disagree. I recommend an addition to be true to why a Review is needed.</p> <p>L243 reduce emphasis on the first. There has been prior such review on the SRS22.</p> <p>L244 Good. consider also developing a research agenda by highlighting where gaps may be .</p> <p>L249 I would recommend developing a publication / presentation plan. There will likely be too much material for a single publication. Could you announce a priori your plans on how you may split the reporting into different publications.</p> <p>L278 Many references appear incomplete (only one author et al). Is this consistent with reporting style of this journal?</p> <p>PRISMA appendix: It would be enough to refer to where this table is presented and not necessary to include it in the present paper.</p> <p>INTERPRETABILITY : There is not enough in the methods on how you will find and extract interpretability information. I am not sure this Taxonomy table needs to be reproduced with the paper. A reference to where it is available may be enough.</p> <p>Please use titles and number the appendices.</p> <p>CRITERIA table: The hypotheses for validity and responsiveness should be defined in the context of your review. Will you use the hypotheses as presented by different studies or approach the review with your own set of hypotheses. IF the latter they should be presented in this protocol. Otherwise specify.</p> <p>Search appendix. Line 3 is redundant with Line 1.</p> <p>Lines 5 6. Would children or child* be relevant? Your lower limit is 10years old.</p> <p>Why not just search for function? Disability Dysfunction</p> <p>Isn't there full overlap between 35 and 36.</p>
--	--

	<p>Why not use all index terms as free text terms as well?</p> <p>Rather than the stage 2 search. Why not describe the much more complete TERWEE filter? Illustrate how this filter will be used and set criteria for when.</p> <p>You risk having many tools and therefore very many stage 2 searches.</p> <p>Could you combine the searches for different tools (search all at once or combine the results and an eliminate duplicates in your screening. Describe your strategy in the protocol. At the moment it appears you will conduct many separate stage 2 searches and it is not clear how you will proceed to combine those results.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 comments	
line 85-86 – This sentence does not make sense, please rephrase. Also, please define the abbreviations as they are first used in the main text.	Thank you for your suggestion. The sentence has been rephrased and the abbreviations are defined when they first used in the text. Page 4 line 109-110
Introduction – It would be useful to add an additional sentence or two at the end of the introduction to describe the potential impact/significance of conducting the proposed systematic review on the field of AIS research, or any potential influences on clinical practice (highlights from the ‘implications of the study’ section that is included later).	Thank you for your suggestion. Three sentences have been added to the end of Introduction describing the potential influence of conducting this systematic review. Page 5 line 130-134
Objective – Please rephrase this, currently it is quite convoluted with repeated words (measurement/measures, evaluate/evaluating) in the same sentence, which makes it difficult to read and comprehend.	Thank you for your suggestion. The objectives have been rephrased. Page 5 line 136-137.
Study methods – The inclusion criteria is very broad, with inclusion of all study and publication types and no defined time frames. If a very large number of eligible studies are retrieved this could make data management and synthesis very difficult. A preliminary search may be useful to better define the scope of this study and/or to help propose appropriate limits on the search (e.g. time frames, type of literature, language).	Thank you for your comments. Our search strategy comprises 2 searches listed below Search 1) will include all type of studies that assess physical functioning without any limitations (to avoid missing any outcome measures that has been used among individuals with AIS for assessment of physical functioning). This will generate a list of physical functioning outcome measure. Search 2) will be limited to the studies that assess measurement properties of physical functioning tools from first search. The search strategy and criteria for each search is now in details

	in the text. Page 6-8
Discussion – This section is currently absent but would be good to include to give some additional details and discussions about the study. For example, the major research question that is answered by the study if completed, and how results of the study can potentially be utilised in the field of AIS research and more broadly in orthopaedic research. Limitations of the study design should be discussed in more detail and potential strategies to address or at least minimise the effect of these limitations should be discussed.	A discussion paragraph has been added. Page 12 line 305- 320
Reviewer 2 comments	
This is a very important research project. The protocol is nicely presented but I suggest additional details could be reported and the rationale could be outlined more clearly	Thank you for your positive comments. Please see in the text the required additional details.
Keywords: the following are MESH terms and could replace some of your free text keywords.: Scoliosis, Outcome, Assessment (Health Care), Reproducibility of results, Physical Functional Performance	Thank you for your suggestions. We followed the recommendation of BMJ Journal to add keywords that are relevant to the content of the manuscript. The keywords has been updated page 2 line 54-55
Please add the Prospero registration info to this protocol paper.	The Prospero registration number has been added. Page 1 line 21
ABSTRACT: Define physical functioning rather than give generalities about quality of life.	Thank you for your comment. The definition of physical functioning has been added to the abstract. Page 2 line 30
Rather than impairment should you use the positive corresponding term from the ICF: Body structure and function measure.	Thank you for your suggestion. We replaced the term impairment measure with the Body structure and function measure. Page 2 line 32-33
Specify the search timeline for each database.	Thank you. The timeline for each database has been added. Page 2 line 41. Due to the limited number of words in the abstract the specific timeline for each database is in the text. Page 7 line 202-204

Describe the two stage nature of the search and the general structure of how you build the search. Which concepts are combined with AND.	Thank you for your comment. The two search nature and the structure of search is described in detail now in the text and the concepts were added. Page 2 line 42-44
Specify whether search limitations or screening criteria will be used for age of the patients, treatment types, curve severity, or language.	Thank you for suggestion. A general statement of li search limitation is added to abstract. Page 2 line 41-44 . The text has more search details as we are limited in the abstract with word limit.
Specify how Grade will be used. What will the conclusion be about. Will you aim to make recommendations within each category of PROM, PBOM and IM or will you summarize the evidence about measurement properties for each tool under each measurement type.	The details of GRADE have been added also description about the conclusion has been added. Page 2 line 47-50
Could you end the abstract by describing the planned impact of the proposed research? Will you develop a research agenda or make recommendations for a core set of outcomes? Highlight the value of completing this review as a conclusion.	Thank you for your comment. Concluding sentences has been added to the abstract. Page 2 line 51-52
Keywords listed after the abstract are spelled different than those on the cover page of the PDF. See my key word comment above.	Thank you for your comment. The keywords have been updated on the cover page of the PDF. Page 2 line 54-55
The strength and limitation section is a mix of that and key points. IF you are to focus on key point and it is allowable by the journal please consider the following: Rather than emphasize being the first please emphasise the value of systematically reviewing this information. State what the results to be presented might be about and why it may be impactful. Re the third point: Rather than emphasize what the study is not about maybe define your operational definition of Outcome measure of physical functioning. Readers would like to know what type of measurements are included in your review.	Thank you for your suggestion. The Journal recommended authors that the strength and limitations section relate specifically to the methods. We have rewritten this and emphasized the importance of conducting this review and the type of outcome measure included. Page 3 line 62-64. We have added a strength sentence about the review regarding the search strategy. Page 3 line 65-66
With regards to limitations then please consider language limitation and age limitation in the review.	The limitation section has been changed accordingly. Page 3 line 69-71
If there is already a core set then justify why do a systematic review. Add a justification for your review still needed. In the discussion you refer to the core set being for general spine deformity and maybe there is a need for a more specific review.	A justification sentence has been added to the introduction paragraph. Page 4 line 104-105
Define all abbreviations in the main text on first use.	All abbreviations are included in full first. Page 4 line 109-110
L85 "a commonly reported outcome measure in AIS. [24] " is unclear English. Redundant with IMs already mentionned. Is this provided as an example?	Thank you for your comment. The sentence has now been rephrased. Page 4 line 110-111

L90 and throughout the paper: Use patient first language. Always say individuals with AIS rather than AIS individuals to be in accordances with ICF recommendations. Similarly rephrase AIS population.	The manuscript has been revised throughout for the use of correct terminology.
L102. I recommend adequate instead of establishment. It is not enough to determine what the properties are they must be good to avoid bias.	The word has been replaced. Page 5 line 126
L106. You cannot say absence of existing evidence. It would be illogical to plan a review if there was no evidence. Specify why a review may be valuable. It could be to outline a research agenda or make recommendations on best measurements.	Thank you for your suggestion. The rationale sentence has been rephrased for accuracy. Page 5 line 130-134
L110 The first objective is not stated clearly. Do you mean to inventory the available physical functioning outcome measures?	The first objective is to identify outcome measures used to assess physical functioning in individuals with AIS. We have rewritten the first objective in the text. Page 5 line 136
L112 Could you present some hypotheses?	Thank you for your comment. Including hypotheses in systematic review is not standard practice. However, we will assess the hypothesis tested when we assess the construct validity for each tool. Alternatively, we will formulate our hypotheses for each tool during the review process.
L114 State in which field there was expertise on your team	The required information has been added to the relevant paragraph. Page 5 line 139
L115 Specify what kind of methodological expertise (review, research, measurement?)	The methodological expertise has been added to Page 5 line 140-141
L116 Despite some overlap in methodology, this handbook (for review on the effect of interventions) has low relevance to a SR on outcome measurements. The following would be much more relevant: https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf	Thank you for your recommendation. We have revised this accordingly. Page 5 line 143
L:126 This may be a problem. Often Risser stage is not reported in measurement studies. Also all adults with AIS could have Risser 5. It may be best to describe an upper limit of age? What will you do in cases where the info is missing?	The upper limit of age has now been added. Page 6 line 157 & Page 7 line 185. Authors will be contacted, in case of missing information.
L126 State whether you will have restrictions related to treatments, severity, Evaluation settings or countries / languages. IF not specify that you do not.	A statement of limitation is now added to the text. Page 6 line 155, page 7 181-183
134 Could you explain what you will do with mixed tools that may assess physical functioning as well as other qualities. The SRS-22r would be such an example by assessing mental health in addition to function. Will you only extract properties for the function scale?	When a tool assesses mixed qualities, the measurement properties of sub-domain will be evaluated. We will also check the psychometric studies of the function/physical functioning sub domain. Page 6 line 166-168

L139 Please use the Positive labels of the ICF categories. body structure and function. (rather than impairment)	The labels have been replaced. Page 6 line 171
L140 Since impairments do not always lead to limitation in function it would be best to state "Which may limit" rather than which limit	Thank you for your comment. The phrase has been changed. Page 6 line 172
L145 Could you also specify if you will aim to track studies on interpretability including studies providing norms.	Thank you for your comment. We aim to assess the three main measurement properties reliability, validity, and responsiveness. We will not track studies that include normative data. Page 7 line 191-193 & page 7, line 196-197
L150 only validation. I recommend Metrological studies or studies on measurement properties which is more inclusive of all the properties you are interested in.	The studies on measurement properties have been added. Page 7 line 180-181
L153 Your appendix describes 2 searches. Here it should be clear that you will perform to types of searches: One for the inventory and one of the measurement properties. Should you describe how the first search would be used to identify tools for the second search?	We have added a detailed description of the two searches to the text. Page 6-8
L154 Plan on using the TERWEE filter for clinimetrics. You refer to it elsewhere but don't mention it here.	The TERWEE filter will be adapted for search 2, it has been mentioned in the search strategy and it has now been added to the information sources. Page 7 line 201, Page 8 line 216
L155 for each database: Refer to the date range searched.	The date range has been added for each database. Page 7 line 202-204
L157 Should other language databases and search terms also be used? Specify languages targeted.	No language limitations will be applied in the search 1, however, the second search will be limited to full text article in English. Page 7 line 205
Why search pubmed in addition to MEDLINE. Justify.	We will search Pubmed in addition to Medline because Pubmed allows us to search through more content than Ovid Medline. We will have access to PubMed Central papers, articles that are "in process" that is, prior to being indexed with MeSH terms, and articles submitted by publishers "ahead of print."
L158 Scoliosis is the same journal as Scoliosis and Spinal Disorders which has had a name change. Only identify the last name: Scoliosis and Spinal Disorders.	Thank you for your comments. Only Scoliosis and Spinal Disorders journal is now in text. . Page 8 line 206

L160 State a database or prepare a list of targeted conferences for which you will review the proceedings and specify over which range of dates.	The Web of Science database and the dates have been added. Page 8 line 209-210
L161 define EthOS or cite.	The EthOS website is defined. Page 8 line 209
L163. There is not really much value in running library database searches independently. Specify which searches would be done independently.	The searches will be done by one researcher although informed in collaboration with specialists including subject and methodological experts. All other steps of SR will be completed by 2 researchers. Page 8 line 212-213
L168 Here you refer to the 2 stages of searching It was not sufficiently clear earlier.	We have described in detail the two searches. We hope this is now clearer for the reader.
L171 Detail the duplicate finding strategy. (which tool and will you only exclude exact or also close duplicates.	Thank you for your comment. The details were added. Page 8 line 221-222
L174 Should there be two screening processes. Is covidence used? Report whether you will do two stages of screening (titles and abstract VS full-text). Explain which criteria may be applied only at a secondary stage. For example reliability information may appear in RCT and not be mentioned in the abstract. Having broader criteria for title and abstract may allow capturing such info.	The screening process has been expanded on in response to this, so thank you. Page 8 line 224-227
L180 Please plan on reporting Kappa also.	The kappa reporting has been added. Page 8 line 231-232
L184 Describe the process of contacting author and timeline allowed for responses. Number of attempts?	Thank you for your comment. The author communication process is now described. Page 9 line 236-238
L190. the study sample size might differ from the sample size used in the Measurement property assessment. It would be best to extract sample used in each of the analyses and report missing data.	Thank you for your suggestion. The sample size information has been added to Table 1, page 9.
I recommend you should extract Curve severity, Diagnoses (what will you do if not purely AIS) Could extract curve types.	Thank you for your recommendation. The curve severity and the curve type are included in the data extraction table. In studies which included a mixed cohort, we will extract information about the number participants with AIS.
This list does not extract information about the name of the measurement tools, nor about instructions on how to administer the measurement and the scoring. IN this proposal, ideally you would list for each COSMIN category of reliability validity and responsiveness what should be extracted. EG. for validity there could be the subtype, and the tools used as	Thank you for your suggestions. We will follow the COSMIN guidelines for data extraction. We will extract data on the characteristic of the measure, sample, results of measurement properties, and Information about interpretability and feasibility of the score of the

<p>gold standard or tools used in assessing convergent validity... The statistic computed the estimate and the confidence interval</p> <p>For all measurement studies it would be important to report the score descriptive statistics as this would allow examining the effect of score distribution on measurement properties.</p> <p>should details needed to appraise study quality also be extracted here?</p>	<p>measure. It is now listed in Table 1 page 9.</p>
<p>L192 I find it debatable that there would be no gold standard since you are measuring not only self-reported function but also observed performance and impairments.</p>	<p>Thank you for your comment. The most frequently used outcome measure is radiographs using cobb method, and it is considered as gold standard for diagnosis and evaluation for the spinal curvature for individuals with AIS. However, no primary PROM or PBOM of physical functioning for individuals with AIS can be identified for this review. This review will therefore identify these tools and, in turn their measurement properties in AIS. Page 10 line 243-245</p>
<p>L199 Please describe how this rating is done (worst score on the item indicate risk of bias or some average or...). Report the training of your reviewers in using this tool.</p>	<p>Thank you for your suggestion. The rating will be based on worst score principle as recommended by COSMIN group. The overall methodological quality of study will be rated as inadequate if one item in measure property is rated inadequate. Page 10 line 252-254 The reviewers will practice using the tool on three studies and then reviewed with the supervisors who are experienced in using this tool.</p>
<p>L204 Please report the Kappa</p>	<p>The kappa is added to the text. Page 10 line 267-268</p>
<p>L204 Describe if you will report scores only or will report each item result. I would recommend the latter as this would be more informative in setting a future research agenda.</p>	<p>Thank you for your recommendation. The results of scoring for each item will be reported as you suggest. Page 10 line 260</p>
<p>L205 Describe if you plan to do comparisons of the measurement properties between tools and what you would use to draw conclusions that one is better than another.</p>	<p>Comparison between tools and the criteria to conclude the best tool has now been added. Page 11, line 286-289</p>

<p>Will you compare properties among subgroups? For example ceiling effect and responsiveness may differ quite a bit between surgical or severely affected patients and those with mild curves and treated with observation or conservative treatments. Since research is conducted on specific samples such comparison may inform research in conservative vs surgical care...</p>	<p>The comparison between subgroups will be undertaken if consistent results are found between homogenous subgroups. Page 11 line 277-279</p>
<p>L208 outline a priori criterion to undertake or abandon meta-analysis. Outlining what would constitute too much heterogeneity in terms of patients, setting and measurement properties assessed would be useful. Will you do the summary by categories PROM, PBOM and IM or by tools. For example I foresee you will have enough data on the SRS22 to metaanalyse results for this tool.</p>	<p>The criteria to undertake or abandon meta-analysis has now been added to the paragraph. Page 11 line 274-277</p>
<p>L218 similarly Specify if the GRADE approach is applied for conclusion about PROM, PBOM and IM categories or for each tool.</p>	<p>The GRADE approach is now described in the text. Page 11 line 291-292</p>
<p>L228 In patient oriented research there are models governing involvement and reporting. Please identify the stakeholder involved (describe them), Describe the purpose of the engagement, Describe their roles and describe the influence they had on the protocol. Below is an example of patient engagement model https://www.google.com/url?sa=i&source=images&cd=&ved=2ahUKEwj5-t6VlbjIAhWJu54KHeT7AisQjRx6BAgBEAQ&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FSpectrum-of-patient-engagement-in-health-research_fig1_312408187&psig=AOvVaw1-Zdx1O_2k3c0hvfTq2_IK&ust=1572118805405560Thank</p>	<p>Thank you for your recommendation. We followed the BMJ guidance in reporting the PPI. Further description of PPI is in detail in the text. Page 12 line 325- 326</p>
<p>L236 I would not agree with lifting objects being a common functional limitation in scoliosis.</p>	<p>We agree that lifting object is not common functional limitation, however it has been reported by individuals with AIS prior to treatment (Du, Yu et al. 2016). It now is removed from sentence. Page 12 line 332</p>
<p>L241 I disagree. I recommend an addition to be true to why a Review is needed.</p>	<p>Thank you for your recommendation. The sentence has now been rephrased. Page 12, 13 line 336-338</p>
<p>L243 reduce emphasis on the first. There has been prior such review on the SRS22.</p>	<p>Thank you for highlighting this. We rewrite the point and emphasize the importance of conducting this review. Page 13 line 336-337</p>
<p>L244 Good. consider also developing a research agenda by highlighting where gaps may be.</p>	<p>Thank you for your comment. The research agenda has been added to the sentence. Page 13 line 338-340</p>
<p>L249 I would recommend developing a publication / presentation plan. There will likely be too much material for a single publication. Could you announce a priori your plans on how you may split the reporting into different publications.</p>	<p>Thank you for your recommendation. We will report the results of this systematic review in peer review journal as well as international and national conferences.</p>

	In case there is a large of volume of data, the publications will split into different publications according to each category of outcome measures. We aim to participate with the results of this review in national and international conferences. Page 13 line 344-347
L278 Many references appear incomplete (only one author et al). Is this consistent with reporting style of this journal?	Thank you for your comments. The references list are updated and in line with BMJ reporting style.
PRISMA appendix: It would be enough to refer to where this table is presented and not necessary to include it in the present paper.	Thank you for your suggestion. The reference to the PRISMA will be included without appendix.
Interpretability There is not enough in the methods on how you will find and extract interpretability information.	The Interpretability is now defined in detail in the data extraction (Table 1 page 9)
I am not sure this Taxonomy table needs to be reproduced with the paper. A reference to where it is available may be enough.	Thank you for your suggestion. The reference to the table will be included without appendix.
Please use titles and number the appendices.	Appendices are now numbered.
CRITERIA table: The hypotheses for validity and responsiveness should be defined in the context of your review. Will you use the hypotheses as presented by different studies or approach the review with your own set of hypotheses. IF the latter they should be presented in this protocol. Otherwise specify.	Thank you for your suggestion. We will use the hypothesis as presented by each study.
Search appendix. Line 3 is redundant with Line 1.	The idiopathic scoliosis search term has now been removed from the search.
Lines 5 6. Would children or child* be relevant? Your lower limit is 10years old.	Thank you for your comment. Although AIS is mostly retrieved with MESH subject heading indexed by (adolescent or Adolescen\$.mp.) we do not wish to overlook potentially eligible studies given 10 is our lower age limit and would fit within the criteria of child
Why not just search for function? Disability Dysfunction	Our search terms were developed according to the common terminology used to capture the meaning of physical functioning. We used synonyms and related terms to physical functioning to avoid missing of any relevant studies.
Isn't there full overlap between 35 and 36. Why not use all index terms as free text terms as well?	Thank you for your comment. Walking and walking speed has been used interchangeably in the literature to describe walking ability. In order not to lose any of them, we used both terminology.

Rather than the stage 2 search. Why not describe the much more complete TERWEE filter? Illustrate how this filter will be used and set criteria for when.	Thank you for your comment. The TERWEE filter will be used in the second search and it will be adapted as it illustrated in the appendix.
<p>You risk having many tools and therefore very many stage 2 searches.</p> <p>Could you combine the searches for different tools (search all at once or combine the results and an eliminate duplicates in your screening. Describe your strategy in the protocol.</p> <p>At the moment it appears you will conduct many separate stage 2 searches and it is not clear how you will proceed to combine those results.</p>	Thank you for your comment. The search results from the search 1 will be screened to generate a list of physical functioning outcome measure and it will be categorized into groups according to its type (PROMs, PBOMs or body structure and function measure). Search 2 will be conducted for each group separately. For example, we will search all databases for PROMs identified in the search 1. This search will combine all tools found if possible, and then remove duplicate at the screening process. If it is not possible, we will conduct the search for measurement properties for each tool separately.

VERSION 2 – REVIEW

REVIEWER	Jiao Jiao Li University of Sydney, Australia
REVIEW RETURNED	15-Jan-2020

GENERAL COMMENTS	The authors have addressed my previous comments. I have no more comments to add.
-------------------------	--

REVIEWER	Eric Parent University of Alberta
REVIEW RETURNED	25-Jan-2020

GENERAL COMMENTS	<p>My notes refer to your track change manuscript.</p> <p>Abstract.</p> <p>L41 I recommend avoiding suggesting absence of evidence in your objective statement.</p> <p>L47 Why search MEDLINE if it is included in PUBMED?</p> <p>L63 Please keep the PROSPERO number in the abstract.</p> <p>L78 Spell out grade here.</p> <p>L77Point 2 seem incomplete. Rephrase to explain what each stage of the search will do.</p> <p>L142. What does it refer to in “It can give...”</p> <p>L145 Replace “and variants of such” by “and its variants”.</p> <p>L151 Rephrase the following to clarify then what? Although relevant, PROMs are influenced by patients’ perception of their abilities to perform activities and lack sensitivity to change. [21]</p> <p>L173. This sentence is unclear about which study it refers to: This study included all forms of spinal deformities, the heterogeneity limits applicability to individuals with AIS as a discrete population.</p>
-------------------------	---

	<p>L220. This is a redundant reference to language which was already mentioned on L210.</p> <p>L276 – 279. Specify the search interfaces. Why delete the references to OVID?</p> <p>L273 You should refer to the search strategies for each of the stages separately. Clarify here how the results of search one with inform the design of search 2. Include the search in this paper and insert here when you make it available.</p> <p>L285. Earlier you argued tha abstract could not provide enough evidence for extraction in stage 2. Why bother searching conferences proceedings then.?</p> <p>L276. I still don't understand why you search both medline and Pubmed. In your response to reviewer comments you suggest Pubmed will find more. Then why search Medline separately since it will already have been searched by PUBMED then?</p> <p>In supplementary file 1. Specify for which database the examples were prepared.</p> <p>The stage 1 and stage 2 examples provided should show clearly which parts of the searches correspond to filters found in the literature or in the manual. It does not seem that the current examples make use of the filters.</p> <p>Table 1. Will you be interested in data on content validation. If yes may need to add what would be extracted in this context.</p> <p>Under responsiveness and validity should the hypotheses tested be extracted? The anchor should be clearly identified in the context of anchor-based responsiveness studies.</p> <p>L412 the first part of the following is too long and somewhat unclear and the second part is missing a verb. Please rephrase. "The overall quality and strength of evidence will be assessed for pooled or summarized result for each measurement property per outcome measure per category by two reviewers, independently. Using a modified (GRADE) approach. [36]"</p> <p>L438 In the discussion you state the following but please add a description of how it benefits these groups: This review will provide a comprehensive assessment of current evidence which benefits practitioners, patients as well as policymakers.</p>
--	--

VERSION 2 – AUTHOR RESPONSE

Reviewer 2 comments:	
<p>Abstract.</p> <p>L41 I recommend avoiding suggesting absence of evidence in your objective statement.</p>	<p>Thank you, for your suggestion. Please see revised sentence in the manuscript. Page 2 line 35</p>
<p>L47 Why search MEDLINE if it is included in PUBMED?</p>	<p>Thank you for your comment. Although Medline is included in PUBMED, Medline allow a more focused and specific search than PUBMED. It gives slightly different results by searching each database separately. However, any duplicates can be removed in the screening stage.</p>

L63 Please keep the PROSPERO number in the abstract.	The PROSPERO number is in the abstract now. Page 2 line 52
L78 Spell out grade here.	Please see GRADE in full as suggested. Page 3 line 69-70
L77 Point 2 seem incomplete. Rephrase to explain what each stage of the search will do.	Point 2 is rephrased and two searches explained more. Page 3 line 64-67
L142. What does it refer to in “It can give...?”	Thank you for your comment. We rephrase the sentence to be clearer. Page 4,5 line 119-121
L145 Replace “and variants of such” by “and its variants”.	Thank you for your suggestion. The sentence has been replaced. Page 4 line 112-113
L151 Rephrase the following to clarify then what? Although relevant, PROMs are influenced by patients’ perception of their abilities to perform activities and lack sensitivity to change. [21]	The sentence has been rephrased. See in text page 4 line 114-116
L173. This sentence is unclear about which study it refers to: This study included all forms of spinal deformities, the heterogeneity limits applicability to individuals with AIS as a discrete population.	Thank you for your suggestion. The sentence was referring to the COS study for spinal deformities. We have added the COS study instead of this study. Page 5 line 125
L220. This is a redundant reference to language which was already mentioned on L210.	We removed limitation sentence from participant paragraph. Page 6 line 160
L276 – 279. Specify the search interfaces. Why delete the references to OVID?	Thank you for your comment. We added the search interfaces for databases. Page 8 line 215-218
L273 You should refer to the search strategies for each of the stages separately. Clarify here how the results of search one with inform the design of search 2. Include the search in this paper and insert here when you make it available.	Thank you for your suggestion. We include a separate description for each search strategy. Page7, line 177-185. Page 8, line 208-213. Further a diagram in figure 1 describing the search strategy is now included for clarity
L285. Earlier you argued that abstract could not provide enough evidence for extraction in stage 2. Why bother searching conferences proceedings then?	Conference proceeding will be searched in stage one to identify studies used physical functioning outcome measure among individuals with AIS. However, for the second stage we will not search the conference proceedings as we aim to assess full text article only. Page 8 line 219-220
L276. I still don’t understand why you search both Medline and PubMed. In your response to reviewer comments you suggest PubMed will find more. Then why search Medline separately since it will already have been searched by PUBMED then? In supplementary file 1. Specify for which database the examples were prepared.	Thank you for your comment. Please see response on this question in (author responses number 2) Thank you for your comment. We added the name of the database in the supplementary file.

The stage 1 and stage 2 examples provided should show clearly which parts of the searches correspond to filters found in the literature or in the manual. It does not seem that the current examples make use of the filters.	Thank you for your comment. We have revised this in the search strategy (P8, Line 241) to make it clear that recommended search filters will be used.
Table 1. Will you be interested in data on content validation? If yes may need to add what would be extracted in this context. Under responsiveness and validity should the hypotheses tested be extracted? The anchor should be clearly identified in the context of anchor-based responsiveness studies.	Thank you for your comment. We will follow COSMIN guidelines to extract data on measurement properties including reliability, validity and responsiveness. As well as the information about feasibility and interpretability. The Quality of Content validity will be assessed using COSMIN risk of bias tool but information about content validity will not be extracted. Hypotheses will be extracted for the responsiveness and validation studies. We will also draw on anchor/s for anchor-based method. See revisions in text Table 1
L412 the first part of the following is too long and somewhat unclear, and the second part is missing a verb. Please rephrase. "The overall quality and strength of evidence will be assessed for pooled or summarized result for each measurement property per outcome measure per category by two reviewers, independently. Using a modified (GRADE) approach. [36]"	Thank you for your comment. The sentence has been rephrased. Page 12 Line 299-301
L438 In the discussion you state the following but please add a description of how it benefits these groups: This review will provide a comprehensive assessment of current evidence which benefits practitioners, patients as well as policymakers	Thank you for your comment. Please see the description in text. Page 12, line 324-331

VERSION 3 – REVIEW

REVIEWER	Eric Parent University of Alberta
REVIEW RETURNED	29-Feb-2020
GENERAL COMMENTS	Thank you for addressing all my comments. Good luck with the research. This is important work and I look forward to the results. I have corrected minor typos and noted the following.

	<p>Line 40 the spelling of CINAHL should be corrected.</p> <p>L135. I could avoid referring to the SRS22 as a gold standard. There is no gold standard in QOL research. Reference standard may be best.</p> <p>L337 The is missing reference possibly identified by a ? See minor edits attached.</p>
--	---

VERSION 3 – AUTHOR RESPONSE

Reviewer 2 comments:	Authors responses
<p>Thank you for addressing all my comments. Good luck with the research. This is important work and I look forward to the results.</p>	<p>Thank you for your positive comments, and your valuable efforts in revising the manuscript.</p>
<p>Line 40 the spelling of CINAHL should be corrected.</p>	<p>Thank you for your comments. The CINAHL is now corrected. Page 2 Line 39 and page 8, line 217</p>
<p>L135. I could avoid referring to the SRS22 as a gold standard. There is no gold standard in QOL research. Reference standard may be best.</p>	<p>Thank you for your comment. We have replaced the gold standard with the reference standard. Page 5 Line 122.</p>
<p>L337 The is missing reference possibly identified by a ?</p>	<p>Thank you for your comment. We added “a” instead of “the” in the sentence. Page 13 line 337</p>