PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Measurement properties of the craniocervical flexion test: a systematic review protocol
AUTHORS	Araujo, Francisco; Ferreira, Giovanni; Scholl Schell, Maurício; Castro, Marcelo; Silva, Marcelo; Ribeiro, Daniel

VERSION 1 – REVIEW

REVIEWER	Gwendolen Jull The University of Queensland Australia Was instrumental in the development of the craniocervical flexion test, but no competing interests to this systematic review per se.
REVIEW RETURNED	11-Sep-2017

GENERAL COMMENTS	The proposed methodology for the systematic review is appropriate and well described. However the subject of the systematic review is confusing. The authors state that the aim of the review is to examine the measurement properties for the pressure biofeedback unit for the assessment of the deep cervical flexor muscles. However it would seem that they may be reviewing the measurement properties of the craniocervical flexion test, which would make more sense. The purpose is unclear. The pressure biofeedback, as the name implies, is a feedback tool to guide the subject's performance of the craniocervical flexion test. The measurement is the subject's performance in the test and, as the authors report, that is assessed by the clinician judging the quality of the muscle performance and the pattern of movement that the subjects use to flex the craniocervical region to five progressive increments of increasing craniocervical flexion range as guided by feedback from the pressure biofeedback unit. Thus the authors needs to clarify exactly what measurement they are evaluating and then use consistent language throughout the text. As mentioned, I think they are probably evaluating the measurement properties of the craniocervical flexion test, not the pressure biofeedback device itself as stated. Minor point: please use the last report of the Global Burden of Disease collaborators, neck pain has shifted to number one (along with low back pain) in the metric of years lived with a disability Global Burden Disease 2015 Disease and Injury Incidence and Prevalence Collaborators.

	Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet 2016;388:1545-602.

REVIEWER	Zaheen Ahmed Iqbal Rehabilitation research chair, King Saud Uiversity Riyadh, KSA
REVIEW RETURNED	14-Sep-2017

GENERAL COMMENTS	This paper illustrates how a systematic review shall be conducted to examine the measurement properties of pressure biofeedback unit for the assessment of deep cervical flexor muscles. 1. I understand that this is the first part of a research currently under process or to be done in future. If so, it should be clearly mentioned in the abstract and introduction section. 2. Article needs advice on English grammar. Future tense used is not appropriate. 3. Following references should be cited: a. 'Effect of Deep Cervical Flexor Muscles Training Using Pressure Biofeedback on Pain and Disability in School Teachers with Neck Pain' Journal of Physical Therapy Science 2013; 25(6) 657-661 b. 'EFFICACY OF PRESSURE-BIOFEEDBACK GUIDED DEEP CERVICAL FLEXOR TRAINING ON NECK PAIN AND MUSCLE PERFORMANCE IN VISUAL DISPLAY TERMINAL OPERATORS' Journal of Musculoskeletal Research, Vol. 16, No. 3 (2013) c. 'Reliability of an assessment of deep neck flexor muscle endurance test: A cross-sectional study' International Journal of Therapy and Rehabilitation, May 2014, Vol 21, No 5 d. 'Measurement properties of the pressure biofeedback unit in the evaluation of transversus abdominis muscle activity: a systematic review. Physiotherapy. 2011 Jun;97(2):100-6' 4. Authors should take care of any plagiarism. 5. Authors should also double check for journal requirements, especially for references All the best!

REVIEWER	Lidwine Mokkink VU University Medical Center, Amsterdam, the Netherlands I am one of the developers of the COSMIN checklist as well as of the updated COSMIN methodology and checklist
REVIEW RETURNED	07-Nov-2017

GENERAL COMMENTS	The manuscript bmjopen-2017-019486, entitled "Measurement properties of pressure biofeedback unit for the assessment of deep cervical flexor muscles: a systematic review protocol." is in my opinion a comprehensive protocol for performing such a review. I have a few issues which I hope the authors found constructive to use. P5 2nd paragraph: I fully agree to the things the authors write in this paragraph. However, are there any other methods or tools to measure craniocervical flexion? And if so, why are these not included in the review?

	If the goal is to select the best instrument (as it seems to be) the review should aim at including all tools for the specific construct. It would very much broaden the applicability of the review if all methods were included in the review. The aim is stated as ‘to assess the measurement properties of the [method]’. It now seems to me that the authors will perform such studies on the measurement properties themselves. I would suggest to change it into ‘our aim is to critically appraise and summarize the quality of the [method] in patients with neck pain’. An alternative could be 'our aim is to critically appraise, compare and summarize the quality of outcome measures of craniocervical flexion in patients with neck pain'. The authors will include studies of any design that have investigated and reported at least one measurement property of the tool under review (abstract and page 6). Do they also want to include all indirect evidence, so articles which did not had the aim to assess a measurement property of the pressure biofeedback unit but used it, and subsequently give indirect evidence, like a correlation with another tool. If not, the eligibility criteria could be changed into e.g. ‘articles that have the aim to investigate at least one measurement property or report on the development of the test procedure’. I would add to the 3rd ‘strength of the review’ that the strength is to take the methodological quality of the included papers into account when drawing conclusions on the quality of the tool by using the COSMIN methodology. P4 2nd paragraph: “Compared to asymptomatic….proprioception”. this sentence could grammatically be improved. P7 search strategy: the so-called ‘excluding search filter’ could additionally be used to delete case reports, comments, editorials, interviews, etc. and to lower the number of hits found in PubMed. P7 data extraction: Can the measurement procedure applied in different studies deviate from each other, or from the intended use? If so, I would also extract data about the specific methods/test procedures applied in a study. P8 Risk of Bias: The COSMIN checklist is developed for use in studies on PROMs. Should the COSMIN checklist be adopted for use in review on a performance-based method, and how should this be adapted? E.g. what about stability of the raters when assessing reliability and measurement error. Is internal consistency and structural validity relevant, or is the tool a single item tool or based on a formative model? Moreover, an update of the COSMIN checklist will be available soon (see website www.COSMIN.nl). As you are not yet started, I would suggest to use the new methodology.

REVIEWER	Chih-Hsiu Cheng School of Physical Therapy and Graduate Institute of Rehabilitation Science College of Medicine, Chang Gung University, Taoyuan, Taiwan, R.O.C.
REVIEW RETURNED	14-Nov-2017

GENERAL COMMENTS	The proposed protocol to review the measurement properties of the craniocervical flexion test tools is sound. The results of this review will provide valuable information for the researchers targeting neck pain related issues. Looking forward to seeing the complete review work soon.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Gwendolen Jull

The proposed methodology for the systematic review is appropriate and well described.

However the subject of the systematic review is confusing. The authors state that the aim of the review is to examine the measurement properties for the pressure biofeedback unit for the assessment of the deep cervical flexor muscles. However it would seem that they may be reviewing the measurement properties of the craniocervical flexion test, which would make more sense. The purpose is unclear.

The pressure biofeedback, as the name implies, is a feedback tool to guide the subject's performance of the craniocervical flexion test. The measurement is the subject's performance in the test and, as the authors report, that is assessed by the clinician judging the quality of the muscle performance and the pattern of movement that the subjects use to flex the craniocervical region to five progressive increments of increasing craniocervical flexion range as guided by feedback from the pressure biofeedback unit.

Thus the authors needs to clarify exactly what measurement they are evaluating and then use consistent language throughout the text. As mentioned, I think they are probably evaluating the measurement properties of the craniocervical flexion test, not the pressure biofeedback device itself as stated.

Response: thank you for this important comment. Indeed, our aim is to assess the craniocervical flexion text. The text was revised accordingly.

Minor point: please use the last report of the Global Burden of Disease collaborators, neck pain has shifted to number one (along with low back pain) in the metric of years lived with a disability
Global Burden Disease 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet 2016;388:1545-602.

Response: revised accordingly.

Reviewer: 2
Reviewer Name: Zaheen Ahmed Iqbal

This paper illustrates how a systematic review shall be conducted to examine the measurement properties of pressure biofeedback unit for the assessment of deep cervical flexor muscles.

1. I understand that this is the first part of a research currently under process or to be done in future. If so, it should be clearly mentioned in the abstract and introduction section.

Response: This manuscript is a protocol for a systematic review, as described in the title and in the methods section. To further clarify this issue, we have added the following sentence “This is a protocol for a systematic review” in the abstract and in the methods section (subheading 2.1 Protocol and Registration).

2. Article needs advice on English grammar. Future tense used is not appropriate.

Response: We have written all the manuscript using future tense to make it clear that this research is yet to be conducted.

3. Following references should be cited:
a. 'Effect of Deep Cervical Flexor Muscles Training Using Pressure Biofeedback on Pain and Disability in School Teachers with Neck Pain' Journal of Physical Therapy Science 2013; 25(6) 657-661
b. 'EFFICACY OF PRESSURE-BIOFEEDBACK GUIDED DEEP CERVICAL FLEXOR TRAINING ON NECK PAIN AND MUSCLE PERFORMANCE IN VISUAL DISPLAY TERMINAL OPERATORS' Journal of Musculoskeletal Research, Vol. 16, No. 3 (2013)
c. 'Reliability of an assessment of deep neck flexor muscle endurance test: A cross-sectional study' International Journal of Therapy and Rehabilitation, May 2014, Vol 21, No 5
d. 'Measurement properties of the pressure biofeedback unit in the evaluation of transversus abdominis muscle activity: a systematic review. Physiotherapy. 2011 Jun;97(2):100-6'

Response: thank you for your reference suggestions. The aim of this systematic review protocol is to evaluate the measurement properties of the craniocervical flexion test. We believe the references we have used are appropriate to present the rationale for this systematic review.
It is unclear to us why we should cite these papers. For instance, articles “a” and “b” evaluated clinical outcomes of an exercise program focusing on deep cervical flexor recruitment; article “c” tested the reliability of an endurance test that did not use the CCFT as originally described. These articles might be included in in our review if, after the electronic search and screening, they meet all pre-defined inclusion criteria. Article “d” is a systematic review that evaluated the measurement properties of the pressure biofeedback unit for the assessment of transversus abdominis muscle activity. Based on that, we deemed appropriate not to add these references in the protocol.

4. Authors should take care of any plagiarism.

Response: Thank you for your comment. We have revised all document to ensure there is no risk of plagiarism.

5. Authors should also double check for journal requirements, especially for references

Response: thank you for your comment. All references were carefully formatted and revised.

Reviewer: 3
Reviewer Name: Lidwine Mokkink

The manuscript bmjopen-2017-019486, entitled "Measurement properties of pressure biofeedback unit for the assessment of deep cervical flexor muscles: a systematic review protocol." is in my opinion a comprehensive protocol for performing such a review. I have a few issues which I hope the authors found constructive to use.

P5 2nd paragraph: I fully agree to the things the authors write in this paragraph.

However, are there any other methods or tools to measure craniocervical flexion? And if so, why are these not included in the review? If the goal is to select the best instrument (as it seems to be) the review should aim at including all tools for the specific construct. It would very much broaden the applicability of the review if all methods were included in the review.

Response: thank you for your comment. Indeed, there are other methods and tests described in the literature using the “craniocervical flexion” movement. However, these tests are designed to evaluate other muscle properties, such as the endurance of the neck flexors (Grimmer, K., 1994; Harris et al., 2005). These other tests are different from the craniocervical flexion test (CCFT), which is the test whose properties we are proposing to review, and are conducted without the biofeedback pressure unit. When designing the protocol, the review team came to the understanding that this would make the review excessively broad.

The aim is stated as ‘to assess the measurement properties of the [method]’. It now seems to me that the authors will perform such studies on the measurement properties themselves. I would suggest to change it into ‘our aim is to critically appraise and summarize the quality of the [method] in patients with neck pain’. An alternative could be 'our aim is to critically appraise, compare and summarize the quality of outcome measures of craniocervical flexion in patients with neck pain'.

Response: thank you for raising this point. We have agreed with the reviewer’s suggestion and revised the sentence.

The authors will include studies of any design that have investigated and reported at least one measurement property of the tool under review (abstract and page 6). Do they also want to include all indirect evidence, so articles which did not had the aim to assess a measurement property of the pressure biofeedback unit but used it, and subsequently give indirect evidence, like a correlation with another tool. If not, the eligibility criteria could be changed into e.g. ‘articles that have the aim to investigate at least one measurement property or report on the development of the test procedure’.

Response: in the proposed systematic review, we will accept indirect references. We believe it will allow us to thoroughly assess the craniocervical flexion test. We designed the inclusion criteria for it to be comprehensive.

I would add to the 3rd ‘strength of the review’ that the strength is to take the methodological quality of the included papers into account when drawing conclusions on the quality of the tool by using the COSMIN methodology.

Response: revised accordingly.

P4 2nd paragraph: “Compared to asymptomatic….proprioception”. this sentence could grammatically be improved.

Response: revised accordingly. It now reads “Compared to asymptomatic individuals, patients with neck pain exhibit increased activity of superficial neck flexors and reduced activity of the deep neck flexors;8 poor muscle endurance;9,10 altered kinematics of the cervical spine;11 delayed feedforward activity;12 and impaired proprioception.13–15”

P7 search strategy: the so-called ‘excluding search filter’ could additionally be used to delete case reports, comments, editorials, interviews, etc. and to lower the number of hits found in PubMed.

Response: Thank you for your suggestion. We designed our electronic search after consulting with an experienced health sciences librarian.

At this stage, we have started the electronic searches without this filter. We will consider this suggestion for future reviews. If you deem this is a mandatory revision to our protocol, we will re-run the electronic search using this filter. Otherwise, we will keep this in mind for future systematic reviews.

P7 data extraction: Can the measurement procedure applied in different studies deviate from each other, or from the intended use? If so, I would also extract data about the specific methods/test procedures applied in a study.

Response: thank you for your suggestion. We followed the deep cervical flexion test as described by Jull et al. (2008). We expect studies will follow the original description of the test. But your point is a valid one. Based on your comment, we have added this information on the data extraction section.

P8 Risk of Bias: The COSMIN checklist is developed for use in studies on PROMs. Should the COSMIN checklist be adopted for use in review on a performance-based method, and how should this be adapted? E.g. what about stability of the raters when assessing reliability and measurement error. Is internal consistency and structural validity relevant, or is the tool a single item tool or based on a formative model?

Response: thank you for this valid comment. You are correct. We are focusing on performance-based method. As you are aware, the COSMIN guidelines suggest this checklist can be used for performance-based methods. We will not focus on stability of raters, and will address this in the "limitation" section of the review.
Moreover, an update of the COSMIN checklist will be available soon (see website www.COSMIN.nl). As you are not yet started, I would suggest to use the new methodology.

Response: thank you for your suggestion. If possible we will wait for this update.

Reviewer: 4
Reviewer Name: Chih-Hsiu Cheng

The proposed protocol to review the measurement properties of the craniocervical flexion test tools is sound. The results of this review will provide valuable information for the researchers targeting neck pain related issues. Looking forward to seeing the complete review work soon.

Response: thank you for your interest in our paper.

References:
Grimmer, K. Measuring the endurance capacity of the cervical short flexor muscle group. Aust. J. Physiother. 40, 251–254 (1994).

Harris, K. D. et al. Reliability of a measurement of neck flexor muscle endurance. Phys. Ther. 85, 1349–1355 (2005).

Jull, G. A., O'Leary, S. P. & Falla, D. L. Clinical assessment of the deep cervical flexor muscles: the craniocervical flexion test. J. Manipulative Physiol. Ther. 31, 525–533 (2008).

VERSION 2 – REVIEW

REVIEWER	Gwendolen Jull School of Health and Rehabilitation Sciences the University of Queensland Australia

REVIEW RETURNED	07-Dec-2017

GENERAL COMMENTS	The authors have undertaken appropriate revisions and now made it clear that they are conducting a review of the measurement properties of the craniocervical flexion test (CCFT). Methods for the review are clearly described and all relevant materials are presented. A point for thought for the authors. The CCFT is a clinical test. The judgement of performance in the CCFT is reliant on the clinician's analysis of the patient's performance and the pressure biofeedback plays an assistive role. It provides feedback to the patient to guide them to the progressive stages of the test. It is an aid to the clinician in that it can provide a 'number', ie. the pressure level at which the clinician judged the patient's performance as satisfactory. Errata: Introduction, line 1 – better expression, neck pain is the 'leading' cause of…..

REVIEWER	Lidwine Mokkink VU University Medical Center, Amsterdam, the Netherlands I'm one of the developers of the COSMIN checklist, and the new COSMN methodology (manuscript accepted for publication)
REVIEW RETURNED	12-Dec-2017

GENERAL COMMENTS	The first eligibility criteria is still unclear to me. An RCT can be included (eligibility criteria 1), but an article on the effectiveness of an intervention will be excluded (page 6). This seems to be contradictory. So, an RCT can be included, if in addition, in the article also a study on the measurement property is reported? In that case, the first eligibility criteria seems not relevant and can be excluded. At page 7, the authors write: 'Reliability is defined as the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions;'. However, this is the definition of the measurement property reliability, and not of the domain, which should be given here (to be in line with the definitions of validity and responsiveness, and as the authors introduce the measurement properties afterwards). Please change it into 'the degree to which a measurement is free from measurement error.' The new COSMIN checklist has been accepted for publication. I would suggest to use the new methodology.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Gwendolen Jull

The authors have undertaken appropriate revisions and now made it clear that they are conducting a review of the measurement properties of the craniocervical flexion test (CCFT). Methods for the review are clearly described and all relevant materials are presented.

A point for thought for the authors. The CCFT is a clinical test. The judgement of performance in the CCFT is reliant on the clinician's analysis of the patient's performance and the pressure biofeedback plays an assistive role. It provides feedback to the patient to guide them to the progressive stages of the test. It is an aid to the clinician in that it can provide a 'number', ie. the pressure level at which the clinician judged the patient's performance as satisfactory.

Response: thank you for your valid comment. We will take this into account when interpreting and analyzing findings from this systematic review.

Errata:
Introduction, line 1 – better expression, neck pain is the 'leading' cause of…..

Response: revised accordingly.

Reviewer: 3
Reviewer Name: Lidwine Mokkink

The first eligibility criteria is still unclear to me. An RCT can be included (eligibility criteria 1), but an article on the effectiveness of an intervention will be excluded (page 6). This seems to be contradictory. So, an RCT can be included, if in addition, in the article also a study on the measurement property is reported? In that case, the first eligibility criteria seems not relevant and can be excluded.

Response: thank you for your suggestion. In fact, we will include an RCT only if the study reports findings regarding measurement properties of the test. Based on your suggestion, we have excluded the first eligibility criteria.

At page 7, the authors write: 'Reliability is defined as the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions;'. However, this is the definition of the measurement property reliability, and not of the domain, which should be given here (to be in line with the definitions of validity and responsiveness, and as the authors introduce the measurement properties afterwards). Please change it into 'the degree to which a measurement is free from measurement error.'

Response: revised accordingly.

The new COSMIN checklist has been accepted for publication. I would suggest to use the new methodology.
Response: thank you for your suggestion. We have now stated that we will use the new COSMIN checklist, and removed the section that described the previous version.

"The checklist consists of nine domains concerning measurement properties. The number of items for each domain varies from 5 to 18. Each item deals with design characteristics and statistical methods used and reported by authors. Each item will be scored based on a four-point rating scale as "excellent", "good", "fair", or "poor". The lowest rating score of a domain will be used for attributing the quality score for that specific domain."