

BMJ Open Measures of trunk muscle strength and their measurement properties: a protocol for a systematic review and narrative synthesis of clinical measures

Shouq Althobaiti,¹ Alison Rushton ,^{1,2} Deborah Falla ,¹ Nicola R Heneghan ¹

To cite: Althobaiti S, Rushton A, Falla D, *et al*. Measures of trunk muscle strength and their measurement properties: a protocol for a systematic review and narrative synthesis of clinical measures. *BMJ Open* 2021;**11**:e041499. doi:10.1136/bmjopen-2020-041499

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-041499>).

Received 10 June 2020
Revised 26 November 2020
Accepted 18 December 2020



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, UK
²School of Physical Therapy, Elbourn College, Western University, London, Ontario, Canada

Correspondence to

Dr Nicola R Heneghan;
n.heneghan@bham.ac.uk

ABSTRACT

Introduction Spinal musculoskeletal conditions, including low back and neck pain, are leading causes of disability globally. The trunk muscles, which comprise muscles in the thoracic and lumbar regions, are essential for functional activities, necessitating mobility, motor control and strength. To investigate the effectiveness of interventions directed at improving trunk muscle strength, it is essential to have valid, reliable and responsive performance-based outcome measures (PBOM). While isokinetic dynamometry is considered the gold-standard PBOM, the associated costs, size/weight and operational complexity of this equipment preclude its use in a clinical setting. There is, therefore, a need to evaluate the measurement properties of alternative accessible measures of trunk strength. This systematic review therefore aims to investigate the measurement properties of PBOM of trunk muscle strength measures appropriate for use in a clinical setting.

Methods and analysis This protocol has been designed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols. CINAHL, Web of Science, Pedro, MEDLINE, EMBASE, SPORTDiscuss will be searched systematically from the database start date up to 16 April 2020, along with reference checking and the grey literature searching. Eligibility criteria include studies investigating measurement properties of PROM for trunk muscle strength for use in a clinical setting in adults with and without spinal musculoskeletal complaints. Two independent reviewers will determine the eligibility of the studies through screening process of titles, abstract and the full text. Both reviewers will assess the risk of bias using COnsensus-based Standards for the selection of health status Measurement Instruments risk-of-bias tool and then extract the data from included studies. The overall quality of the included studies will be evaluated using the Grading of Recommendations Assessment, Development and Evaluation approach. A narrative synthesis will be carried out if meta-analysis is not applicable. Findings from this systematic review will aid clinicians and practitioners working in the field for example, sport, in using the most appropriate PBOM to measure trunk muscle strength.

Ethics and dissemination No research ethics application is needed as there are no patient data in this study. The results of this study will be submitted to a peer reviewed journal and presented at conferences.

Strengths and limitations of this study

- This is the first systematic review that will evaluate measurement properties of performance-based outcome measures, which will inform the selection of most reliable, valid and responsive tool to assess trunk muscle strength in a clinical setting.
- This study will use the consensus-based standards for the selection of health measurement instrument checklist, which therefore will improve the selection of trunk strength outcome measures in research and clinical practice.
- The term trunk is an umbrella term to reflect all muscles in the thoracic and lumbar spine, which may limit applicability of findings to specific clinical complaints in a specific spinal region.

PROSPERO registration number CRD42020167464.

INTRODUCTION

Musculoskeletal (MSK) conditions are common, not only among elderly but among children, teenagers and adults.¹ Back pain, neck pain, osteoarthritis and rheumatoid arthritis are considered among the most disabling MSK conditions that affect both physical and psychological capacities of individuals.² MSK disorders are considered the second highest contributor to years lived with disability.³ Low back pain (LBP) is the single-leading cause of disability worldwide³ and neck pain ranked as the fourth highest cause of years lived with disability.⁴ Therefore, spinal pain remains the main contributor of global disability.¹ Recent data suggest that the lifetime prevalence of spinal pain is 20%–70% for neck pain,⁴ 3.7%–77% for thoracic spine pain⁵ and 58%–84% for LBP,⁶ which place great economic burden on health services globally. In the UK, for instance, the estimated cost of managing chronic LBP alone is around £1.8–£2.3 billion.⁷

Muscles of the trunk which includes those with attachments to both the thoracic and lumbar spines, are central to providing mobility and stability of the spine during functional activities, including, gait and daily life activities.⁸ Research has identified that weak trunk muscles are associated with exaggerated spinal curves,^{9 10} and are a risk factor for spinal disorders^{11 12} and risk of falling.¹³ Several studies have suggested that patients with spinal MSK conditions may benefit from trunk strength training as essential part of the rehabilitation programmes.¹⁴ As a result, performance-based outcome measures (PBOM) of trunk muscle strength are important to evaluate patient clinical progression¹⁵ and to determine the effectiveness of therapeutic rehabilitation programmes.¹⁶

To evaluate trunk muscle strength, manual muscle testing,¹⁷ hand-held dynamometer,¹⁸ strain gauge tests,¹⁹ isostation²⁰ and isokinetic test²¹ have been described as available methods. However, the PROM need to obtain good level of measurement properties to be clinically and scientifically useful to help guide clinical decision making and treatment monitoring.^{22 23} A measurement property is the quality aspect of an instrument and due to variations in the terminology and definitions of these measurement properties, the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) initiative developed a consensus-based taxonomy of measurement properties.²⁴ The aforementioned taxonomy covers the three main domain (reliability, validity and responsiveness).²⁵

Several literature reviews have been published summarising and critically appraising the trunk muscle strength PBOM.²⁶⁻³¹ However, few have assessed their measurement properties. Two reviews have evaluated the measurement properties of trunk muscle strength using isomachines (isokinetic and isostation).^{28 29} Acceptable levels of reliability were reported for flexion and extension up to 120°/s, with limited and conflicting evidence regarding the reliability of trunk lateral bending strength.²⁹ However, conflicting evidence also exists regarding the validity of Iso-machines across both reviews.^{28 29}

Other reviews have reported some measurement properties of trunk muscle strength measures in neurological conditions.^{30 31} However, the reliability of the PBOM with one group does not necessarily generalise to another.³² and the value of these measures for use in a spinal MSK population is questionable.³³ Establishing measurement properties of any PBOM within a defined population is important to eliminate any potential bias and to have confidence in findings.²⁵ To the best of the author's knowledge, no systematic review has been published targeting the psychometric properties of the trunk muscle strength PBOM for use in a clinical setting. Therefore, a systematic review is needed to comprehensively evaluate the psychometric properties of the different clinical trunk strength outcome measures in healthy participants and patients with spinal MSK complaints. This systematic review aims to assess the measurement properties (validity, reliability

and responsiveness) of trunk muscle strength PBOM for use in a clinical setting.

Aim

To evaluate the measurement properties (validity, reliability and responsiveness) of the trunk muscle strength outcome measures appropriate for use in routine clinical practice.

METHODS

The systematic review protocol is designed using The Cochrane Handbook for Diagnostic Test Accuracy studies and the Centre for Reviews and Dissemination³⁴ and reported in line with The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Protocols guidelines checklist,³⁵ (see online supplemental file 1). The COSMIN group developed guidelines and practical tools for conducting studies on measurement properties or selecting the high-quality studies for systematic reviews on measurement properties.^{36 37} Hence, COSMIN risk-of-bias checklist will be used to improve the selection of the available trunk strength outcome measures available.

ELIGIBILITY CRITERIA

Inclusion criteria

Eligibility criteria will include adult participants aged ≥ 18 years who are either healthy or experiencing any spinal MSK condition. For the purpose of this review, spinal MSK conditions will be defined based on the International Classification of Diseases as any condition that affects the spine bones, joints, muscles and associated tissues such as ligaments and tendons.³⁸ (eg, arthritis, neck pain, thoracic pain, osteoporosis, scoliosis, LBP, etc). Including healthy participants will help to identify the cut-off points in trunk strength measures between healthy individuals and those with spinal MSK complaints. We will include studies which have investigated any PBOM of trunk muscle strength performed in a clinical or field-based setting, including manual, functional and mechanical methods. Eligible studies must have evaluated one or more aspect of the main three domains of the COSMIN Taxonomy (validity, reliability and responsiveness) of measurement properties will be included.²⁵ More details on the three domains of the COSMIN Taxonomy for measurement properties can be found in online supplemental file 2. Studies must report on the evaluation of the measurement properties of PBOM of trunk muscle strength which can be done in a clinical or field-based setting.

Exclusion criteria

Any study written in a language other than English will be excluded. Studies reporting measurement properties of PBOM which involve expensive, technical equipment or computerised instruments namely; isokinetic and isostation machines²⁷ and are not practical,^{15 39} that is, relevant

for clinicians working in home-based, community-based or field-based settings who need portable, easy to use devices⁴⁰ will be excluded.

Information sources

A comprehensive search strategy will be developed using both medical subject headings and free text, relevant keywords identified during the scanning search. Following the Cochrane collaboration recommendations, multiple electronic databases and a subject-specific database will be searched systematically, to cover the broadest available literature. These include CINAHL and SPORTDiscuss via (EBSCO interface), MEDLINE and EMBASE through (Ovid interface), Web of Science and Pedro. The search was from the database start date up to April 16, 2020, with no restrictions to the publication time frame although only articles published in English was included in this review. Handsearching through checking reference lists and grey literature searching through the main sources such as British National bibliography for report literature and open Grey will be conducted as well.

Search strategy

The search strategy was developed in discussion with the supervisory team (NRH, AR and DF) and a specialist librarian. Initially, the search strategy will be developed in the MEDLINE database, and then it will be adapted for each database. Specific key terms will include terms related to psychometric properties e.g. reliability, validity and responsiveness as well as terms describing the population of interest will be used. Additional search filters designed by COSMIN for retrieving studies on measurement properties will be applied where appropriate.⁴¹ See example of search strategy in online supplemental file 3. Relevant studies will be identified and selected by two independent reviewers SA and AA with specialist training and knowledge in MSK rehabilitation and research methods.

Study records

Data management

Comprehensive search on the after mentioned databases will be carried out by the main author SA. All search results will be exported and stored on EndNote V. X9 (Clarivate analytics) software programme. This will allow any duplicates to be identified and subsequently removed as well as storing of abstract and full texts.

Selection process

Secondary to searching process, two reviewers SA and AA will independently screen titles and abstract based on preidentified eligibility criteria and will subcategorise the identified studies into include/exclude/unsure.⁴² The second step comprises retrieving and reading the full text of potentially relevant articles which will then be independently examined by each reviewer against the eligibility criteria. If further information needed, authors will be contacted via email. Agreement between both reviewers is required for the study to be included in the review.

Table 1 Summary of data to be extracted from included studies

Content	Data items
Bibliographic data	Authors, year of publication
Study characteristics	Study design, sample size
Setting	Country, setting of measurement
Participants characteristics	Age, gender, healthy or with spinal musculoskeletal condition.
Outcome measures	Trunk muscle strength measures. Type of muscle contraction measured. Measurement procedure: warm-up, participants' position, fixation, examiner position (if any), line of force (resistance), cool down.
Measurement properties	Measurement properties (reliability, validity and responsiveness), statistical methods used and results.

Therefore, agreement will be assessed using Cohen's kappa (κ) statistic disagreement will be resolved by consensus or by the decision of a third reviewer NRH.⁴³ Information regarding the excluded studies and the reason for exclusion will be reported using PRISMA flow chart.

Data collection process

For each included study, a standard form will be used to extract the data. To ensure all the relevant information is captured, piloting the data collection form will be conducted. Both reviewers SA and AA will independently extract the data using the standardised form, the authors will be contacted for more clarification or if there is any missing data. In case of disagreement about extracted data between reviewers, discussion and/or involving a third reviewer (NRH) will be carried out until consensus reached.

Data items

Table 1 summarise the relevant data to be extracted from included studies.

Risk of bias in individual studies

The COSMIN Risk of Bias checklist for systematic reviews will be used to evaluate the risk of bias of included studies in this review.⁴⁴ The COSMIN Risk of Bias Checklist considered to have adequate reliability as it is developed from the original COSMIN tool which show high percentage inter-rater agreement.⁴⁵ The COSMIN Risk of Bias checklist includes standards for both design and the preferred statistical methods for each measurement property. The checklist covers nine different dimensions of reliability, validity and responsiveness. The COSMIN checklist was originally designed to evaluate patient-reported outcome measures (PROM). However, the COSMIN group have recommended adaptation of the tool for use with other types of measures such as clinician-reported outcome measures or PBOM.⁴¹ As with the



study selection process, two raters SA and AA will independently score each outcome measure as either 'very good', 'adequate', 'doubtful' or 'inadequate' quality.⁴⁴ Any disagreement between the reviewers will be resolved through discussion, if no consensus can be reached, a third reviewer will be consulted.

Data synthesis

Depending on the heterogeneity of included studies, either a meta-analysis or narrative synthesis will be conducted to synthesise the results which will follow the COSMIN guidelines for systematic reviews.⁴⁴ The results from different studies on single measurement property will be quantitatively pooled in meta-analysis if sufficient number of studies share the same reference standard, designs, population and measure the same movement. To find the estimate of test–retest reliability, standard generic inverse variance random effects model will be implemented to calculate the weighted mean intraclass correlation coefficients and 95% CIs. For construct validity, all correlations of PBOM with other PBOM that measure the same construct will be pooled.⁴⁴ Following the assessment of scoping searches of the currently available literature, pooling of data might not be possible due to an anticipated lack of homogeneity. Hence, a narrative synthesis will be conducted in line with the narrative synthesis in systematic reviews recommendation.⁴⁶ Synthesis will bring together evidence of measures of trunk strength, summary table will be generated to illustrates the pooled results per each measurement property per outcome measure per movement and rated against the updated criteria for good measurement properties as; sufficient (+), insufficient (-), inconsistent (\pm) or indeterminate (?).⁴⁴ Further analysis of the results will be presented in the discussion section in line with the quality of evidence.

Meta-bias

To eliminate any chance of publication bias, grey literature and conference papers will be searched.

Confidence in cumulative evidence

The overall quality of evidence regarding the measurement properties will then be assessed using a modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach. As recommended by COSMIN guidelines for systematic review, a modified GRADE approach will be used to assess how the pooled results from included studies are trustworthy.³⁶ The quality of evidence will be determined using four factors from the GRADE approach: risk of bias, inconsistency, imprecision and indirectness. Following the COSMIN recommendation, the fifth factor, that is, publication bias will not be used due to lack of registry data on measurement properties studies.

Patient and public involvement

The study question and systematic review protocol were informed following many years of working with patients and clinical experiences of managing patients with spinal

dysfunction. The question was informed following discussions with patient and public involvement meeting at the Centre of Precision Rehabilitation for Spinal Pain and specifically in light our earlier research investigating measurement properties of PBOM for spinal mobility. The group consist individuals with different MSK and spinal complaints. Since no patient data is needed, patients will not be involved in data collection or analysis. However, the results of the study will be shared with this group and other public engagement events.

Clinical implications of this study

By accurately measuring the trunk muscles strength, diagnosis of dysfunction as well as improvement can be monitored. Also, with the current assessment methods available to measure trunk strength, finding a valid, reliable and responsive tool as well as cost-effective for clinical use is a priority. Noteworthy, with the vast range of different test procedures and positions used, this review will summarise data regarding the measurement properties of different assessment methods and highlight the method which is superior in terms of psychometric properties, cost-effectiveness and time-saving. In doing so, healthcare professionals will be aware of the valid method to use within the clinical setting to assess the effectiveness of interventions directed to improve muscular function. Using accurate and objective muscle strength measures will facilitate the monitoring of rehabilitation programme efficacy and effectiveness of targeted interventions to improve trunk muscle strength.

ETHICS AND DISSEMINATION

No patient data will be collected, hence no ethical approval is needed for this systematic review. The results of this review will help to inform current healthcare practice and research on the most valid, reliable and responsive tool for measuring trunk muscle strength. Results of this review will be submitted to be published in a peer-review journal and presented at relevant conferences.

Twitter Alison Rushton @abrushton, Deborah Falla @Deb_Falla and Nicola R Heneghan @HeneghanNicola

Contributors SA is a PhD student at the University of Birmingham. NRH, AR and DF are supervisors. NRH and AR have expertise in the field of outcome measurement methods. SA, NRH, ABR and DF contributed to the systematic review topic. SA drafted the protocol with guidance and feedback from NRH, AR and DF. NRH, AR and DF reviewed the manuscript and commented on the protocol. AA is a PhD student who will act as a second reviewer. All authors have approved and contributed to the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content

includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Alison Rushton <http://orcid.org/0000-0001-8114-7669>

Deborah Falla <http://orcid.org/0000-0003-1689-6190>

Nicola R Heneghan <http://orcid.org/0000-0001-7599-3674>

REFERENCES

- Briggs AM, Woolf AD, Dreinhöfer K, *et al*. Reducing the global burden of musculoskeletal conditions. *Bull World Health Organ* 2018;96:366.
- Ingram M, Symmons DPM. The burden of musculoskeletal conditions. *Medicine* 2018;46:152–5.
- GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;392:1789–858.
- Blanpied PR, Gross AR, Elliott JM, *et al*. Neck pain: revision 2017. *J Orthop Sports Phys Ther* 2017;47:A1–83.
- Heneghan NR, Rushton A. Understanding why the thoracic region is the 'Cinderella' region of the spine. *Man Ther* 2016;21:274–6.
- Bernstein IA, Malik Q, Carville S, *et al*. Low back pain and sciatica: summary of NICE guidance. *BMJ* 2017;356:i6748.
- Hong J, Reed C, Novick D, *et al*. Costs associated with treatment of chronic low back pain: an analysis of the UK general practice research database. *Spine* 2013;38:75–82.
- Kong YS, Cho YH, Park JW. Changes in the activities of the trunk muscles in different kinds of bridging exercises. *J Phys Ther Sci* 2013;25:1609–12.
- Katzman WB, Wanek L, Shepherd JA, *et al*. Age-related hyperkyphosis: its causes, consequences, and management. *J Orthop Sports Phys Ther* 2010;40:352–60.
- Barczyk-Pawełec K, Piechura JR, Dziubek W, *et al*. Evaluation of isokinetic trunk muscle strength in adolescents with normal and abnormal postures. *J Manipulative Physiol Ther* 2015;38:484–92.
- Biering-Sorensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. *Spine* 1984;9:106–19.
- Steele J, Bruce-Low S, Smith D. A reappraisal of the deconditioning hypothesis in low back pain: review of evidence from a triumvirate of research methods on specific lumbar extensor deconditioning. *Curr Med Res Opin* 2014;30:865–911.
- Kato S, Murakami H, Demura S, *et al*. Abdominal trunk muscle weakness and its association with chronic low back pain and risk of falling in older women. *BMC Musculoskelet Disord* 2019;20:273.
- Gordon R, Bloxham S. A systematic review of the effects of exercise and physical activity on non-specific chronic low back pain. *Healthcare* 2016;4. doi:10.3390/healthcare4020022. [Epub ahead of print: 25 Apr 2016].
- El Mhandi L, Bethoux F. Isokinetic testing in patients with neuromuscular diseases: a focused review. *Am J Phys Med Rehabil* 2013;92:163–78.
- Demoulin C, Grosdent S, Smeets R. *Muscular performance assessment of trunk extensors: a critical appraisal of the literature: low back pain*. IntechOpen, 2012.
- Petty NJ, Examination N. *Neuromusculoskeletal examination and assessment, a Handbook for therapists with PAGEDURST access, 4: neuromusculoskeletal examination and assessment*. Elsevier Health Sciences, 2011.
- Durmus D, Akyol Y, Alayli G, *et al*. Effects of electrical stimulation program on trunk muscle strength, functional capacity, quality of life, and depression in the patients with low back pain: a randomized controlled trial. *Rheumatol Int* 2009;29:947–54.
- da Silva RA, Arsenault AB, Gravel D, *et al*. Back muscle strength and fatigue in healthy and chronic low back pain subjects: a comparative study of 3 assessment protocols. *Arch Phys Med Rehabil* 2005;86:722–9.
- Deutsch S. *B-200 back evaluation system version 3.0*. Pawtucket, RI: Occupational Orthopedic Center, 1989.
- Newton M, Thow M, Somerville D, *et al*. Trunk strength testing with iso-machines. Part 2: experimental evaluation of the Cybex II back testing system in normal subjects and patients with chronic low back pain. *Spine* 1993;18:812–24.
- Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil* 2005;84:719–23.
- Roach KE. Measurement of health outcomes: reliability, validity and responsiveness. *J Prosthet Orthot* 2006;18:P8–12.
- De Vet HC, Terwee CB, Mokkink LB. *Measurement in medicine: a practical guide*. Cambridge University Press, 2011.
- Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- Beimborn DS, Morrissey MC. A review of the literature related to trunk muscle performance. *Spine* 1988;13:655–60.
- Malliou P, Gioftsidou A, Beneka A, *et al*. Measurements and evaluations in low back pain patients. *Scand J Med Sci Sports* 2006;16:219–30.
- Mueller S, Stoll J, Mueller J, *et al*. Validity of isokinetic trunk measurements with respect to healthy adults, athletes and low back pain patients. *Isokinet Exerc Sci* 2012;20:255–66.
- Newton M, Waddell G. Trunk strength testing with iso-machines. Part 1: review of a decade of scientific evidence. *Spine* 1993;18:801–11.
- Jorgensen M, Dalgas U, Wens I, *et al*. Muscle strength and power in persons with multiple sclerosis - A systematic review and meta-analysis. *J Neurol Sci* 2017;376:225–41.
- Rabelo M, Nunes GS, da Costa Amante NM, *et al*. Reliability of muscle strength assessment in chronic post-stroke hemiparesis: a systematic review and meta-analysis. *Top Stroke Rehabil* 2016;23:26–35.
- Dijkers MPJM, Kropp GC, Esper RM, *et al*. Reporting on reliability and validity of outcome measures in medical rehabilitation research. *Disabil Rehabil* 2002;24:819–27.
- Portney LG, Mary P. *Foundations of clinical research: applications to practice*. 2 edn, 2002: 598–98.
- Deeks J, Wisniewski S, Davenport C. Guide to the contents of a Cochrane Diagnostic Test Accuracy Protocol. In: Deeks J, Bossuyt P, Gatsonis C, eds. *Cochrane Handbook for systematic reviews of diagnostic test accuracy version 100*. The Cochrane Collaboration, 2013.
- Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
- Mokkink LB, Prinsen CA, Patrick DL. *COSMIN methodology for systematic reviews of Patient-Reported outcome measures (PROMs) user manual*. 78. Netherlands, 2018.
- Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
- Campbell LC. Musculoskeletal disorders: addressing disparities in prevalence, severity, and treatment. *N C Med J* 2017;78:315–7.
- Stark T, Walker B, Phillips JK, *et al*. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review. *Pm R* 2011;3:472–9.
- Trudelle-Jackson E, Jackson AW, Frankowski CM, *et al*. Interdevice reliability and validity assessment of the Nicholas hand-held dynamometer. *J Orthop Sports Phys Ther* 1994;20:302–6.
- Prinsen CAC, Mokkink LB, Bouter LM, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
- McKenzie JE BS, Ryan RE, Thomson HJ. Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT TJ, Chandler J, Cumpston M, eds. *Cochrane Handbook for systematic reviews of interventions version 60*, 2019.
- Catalin Tufanaru ZM, Aromataris E, Campbell J. Systematic reviews of effectiveness. In: Aromataris EMZ, ed. *JBIR reviewer's manual*, 2017.
- Mokkink LB, de Vet HCW, Prinsen CAC, *et al*. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171–9.
- Mokkink LB, Terwee CB, Gibbons E, *et al*. Inter-rater agreement and reliability of the COSMIN (consensus-based standards for the selection of health status measurement instruments) checklist. *BMC Med Res Methodol* 2010;10:82.
- Popay J, Roberts H, Sowden A. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme 2006;1:b92.

Supplementary file 1:**PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol***

Section and topic	Item No	Checklist item	Page Number
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	1
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	N/A
Support:			
Sources	5a	Indicate sources of financial or other support for the review	11
Sponsor	5b	Provide name for the review funder and/or sponsor	
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	4,5
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	5
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	6
Information	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other	6

sources		grey literature sources) with planned dates of coverage	
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	7
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	7
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	7
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	7,8
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	8
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	N/A
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	8
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2 , Kendall's τ)	9
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	9
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	9

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

Supplementary file 2: COSMIN Definitions

COSMIN Taxonomy for measurement property (definitions)¹

Term			Definition
Domain	Measurement property	Aspects of measurement property	
Reliability			The degree to which the measurement is free from measurement error.
Reliability (extended definition)			The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same PROM (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater).
	Internal consistency		The degree of the interrelatedness among the items.
	Reliability		The proportion of the total variance in the measurements which is due to 'true'† differences between patients.
	Measurement error		The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured.
Validity			The degree to which a PROM measures the construct(s) it purports to measure
	Content validity		The degree to which the content of a PROM is an adequate reflection of the construct to be measured.
		Face validity	The degree to which (the items of) a PROM indeed looks as though they are an adequate reflection of the construct to be measured
	Construct validity		The degree to which the scores of a PROM are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the PROM validly measures the construct to be measured.
		Structural validity	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured.

		Hypotheses testing	Idem construct validity
		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version of the PROM.
	Criterion validity		The degree to which the scores of a PROM are an adequate reflection of a 'gold standard'.
Responsiveness			The ability of a PROM to detect change over time in the construct to be measured.
	Responsiveness		Idem responsiveness
Interpretability			Interpretability is the degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations - to a PROM's quantitative scores or change in scores.

† The word 'true' must be seen in the context of the CTT, which states that any observation is composed of two components – a true score and error associated with the observation. 'True' is the average score that would be obtained if the scale were given an infinite number of times. It refers only to the consistency of the score, and not to its accuracy (22) * Interpretability is not considered a measurement property, but an important characteristic of a measurement instrument

References:

1. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology* 2010;63(7):737-45.

Supplementary file 3:**Example of MEDLINE search strategy 1948– March 2020**

MEDLINE (Ovid interface) search strategy
1. Trunk musc* strength.ti,ab.
2. Trunk musc* power.ti,ab.
3. Trunk musc* torque.ti,ab.
4. Torso strength.ti,ab.
5. Torso power.ti,ab.
6. Isometric trunk strength.ti,ab.
7. Isotonic trunk strength.ti,ab.
8. Static trunk strength.ti,ab.
9. Back musc* strength.ti,ab.
10. Back exten* strength.ti,ab.
11. Lumbar exten* strength.ti,ab.
12. Lumbar exten* torque.ti,ab.
13. Abdomin\$ musc* strength.ti,ab.
14. Trunk Max* contract*.ti,ab.
15. Trunk MVC.ti,ab.
16. Trunk flex* strength.ti,ab.
17. Trunk flex* power.ti,ab.
18. Trunk flex* torque.ti,ab.
19. Trunk forward bend* strength.ti,ab.
20. Trunk forward bend* power.ti,ab.
21. Trunk forward bend* torque.ti,ab.
22. Trunk extens\$ strength.ti,ab.
23. Trunk extens\$ power.ti,ab.
24. Trunk extens\$ torque.ti,ab.
25. (Trunk rota* strength or Trunk rota* power or Trunk rota* torque).ti,ab.
26. (Trunk lateral\$ flex* strength or Trunk lateral\$ flex* power or Trunk lateral\$ flex* torque).ti,ab.
27. (Trunk lateral\$ bend* strength* or Trunk lateral\$ bend* power or Trunk lateral\$ bend* torque).ti,ab.
28. (core strength or core power or core torque).ti,ab.
29. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28
30. (Spinal musculoskeletal pain or musculoskeletal disease or musculoskeletal dysfunction* or spin* MUSCU* pain).ti,ab.
31. (Low* back pain or LBP or Chronic Low* back pain or CLBP).ti,ab.
32. (lumbago or dorsalgia).ti,ab.
33. back disorder\$.ti,ab.
34. (lumbar adj pain).ti,ab.
35. (slipped adj disc).ti,ab.
36. (slipped adj disk).ti,ab.
37. (prolap* adj disc).ti,ab.
38. (prolap* adj disk).ti,ab.
39. Spin*osteoarthritis.ti,ab.
40. Spine osteoarthritis.ti,ab.
41. spine spondylitis.ti,ab.

42. spine spondylosis.ti,ab.
43. (spine degenerative adj joint adj disease).ti,ab.
44. (Neck pain or Cervical pain or Chronic neck pain or CNP or cervicogenic).ti,ab.
45. (Thoracic spine pain or Mid back pain).ti,ab.
46. (Healthy adult* or Normal adult* or A symptomatic adult* or Physically active adult* or Athlete*).ti,ab.
47. 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46
48. (Performance-based tool* or performance-based test* or Clinical-based tool*).ti,ab.
49. (field- based test* or Assessment).ti,ab.
50. (Quantitative assessment or measurement* or evaluate\$).ti,ab.
51. (instrument\$ or tool* or test\$).ti,ab.
52. (Manual* musc* test\$ or MMT).ti,ab.
53. (mechanic* or Hand-held dynamometer* or HHD or Strain-gauge test*).ti,ab.
54. 48 or 49 or 50 or 51 or 52 or 53
55. 29 and 47 and 54
56. (Validation Studies or Comparative Study).pt.
57. exp Psychometrics/
58. psychometr*.ti,ab.
59. (clinimetr* or clinometr*).tw.
60. outcome assessment.ti,ab. or outcome measure*.tw. or exp Observer Variation/ or observer variation.ti,ab.
61. exp Health Status Indicators/
62. exp Reproducibility of Results/
63. reproducib*.ti,ab.
64. exp Discriminant Analysis/
65. (reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or internal consistency).ti,ab.
66. (cronbach* and (alpha or alphas)).ti,ab.
67. (item and (correlation* or selection* or reduction*)).ti,ab.
68. (agreement or precision or imprecision or precise values or test-retest).ti,ab.
69. (test and retest).ti,ab.
70. (reliab* and (test or retest)).ti,ab.
71. (stabil* or interrater or intrarater or intrarater or intertester or intratester or interobserver or interobserver or intraobserver or intertechnician or intratechnician or interexaminer or intraexaminer or interassay or intraassay or interindividual or intraindividual or interparticipant or intraparticipant or kappa or kappas or repeatab*).ti,ab.
72. ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab.
73. (generaliza* or generalisa* or concordance).ti,ab.
74. (intraclass and correlation*).ti,ab.
75. (item discriminant or interscale correlation* or error or errors).mp. or individual variability.ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
76. variability.mp. and (analysis or values).ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
77. (uncertainty and (measurement or measuring)).ti,ab.

78. (standard error of measurement or sensitiv* or responsive*).ti,ab.
79. ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab.
80. (small* and (real or detectable) and (change or difference)).ti,ab.
81. 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80
82. 55 and 81
83. limit 82 to English language