

BMJ Open Measurement properties of instruments assessing permanent functional impairment of the spine: a systematic review protocol

Suelen Meira Goes,¹ Catherine M Trask,² Catherine Boden,³ Brenna Bath,¹ Daniel Cury Ribeiro,⁴ Paul Hendrick,⁵ Lynne Clay,⁴ Xiaoke Zeng,² Stephan Milosavljevic¹

To cite: Goes SM, Trask CM, Boden C, *et al.* Measurement properties of instruments assessing permanent functional impairment of the spine: a systematic review protocol. *BMJ Open* 2018;**8**:e019276. doi:10.1136/bmjopen-2017-019276

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-019276>).

Received 22 August 2017
Revised 11 December 2017
Accepted 19 December 2017



¹School of Physical Therapy, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

²Canadian Centre for Health and Safety in Agriculture (CCHSA), University of Saskatchewan, Saskatoon, Saskatchewan, Canada

³Leslie and Irene Dubé Health Sciences Library, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

⁴School of Physiotherapy, University of Otago, Dunedin, New Zealand

⁵School of Health Sciences, University of Nottingham, Nottingham, UK

Correspondence to

Dr Suelen Meira Goes;
suelen.goes@usask.ca

ABSTRACT

Introduction Permanent functional impairment (PFI) of the spine is a rating system used by compensation authorities, such as workers compensation boards, to establish an appropriate level of financial compensation for persistent loss of function. Determination of PFI of the spine is commonly based on the assessment of spinal movement combined with other measures of physical and functional impairments; however, the reliability and validity of the measurement instruments used for these evaluations have yet to be established. The aim of this study is to systematically review and synthesise the literature concerning measurement properties of the various and different instruments used for assessing PFI of the spine.

Methods Three conceptual groups of terms (1) PFI, (2) spinal disorder and (3) measurement properties will be combined to search Medline, EMBASE, CINAHL, Web of Science, Scopus, PEDro, OTSeeker and Health and Safety Science Abstracts. We will examine peer-reviewed, full-text articles over the full available date range. Two reviewers will independently screen citations (title, abstract and full text) and perform data extraction. Included studies will be appraised as to their methodological quality using the COnsensus-based Standards for the selection of health Measurement INstruments criteria. Findings will be summarised and presented descriptively, with meta-analysis pursued as appropriate.

Ethics and dissemination This review will summarise the current level of evidence of measurement properties of instruments used for assessing PFI of the spine. Findings of this review may be applicable to clinicians, policy-makers, workers' compensation boards, other insurers and health and safety organisations. The findings will likely provide a foundation and direction for future research priorities for assessing spinal PFI.

PROSPERO registration number CRD42017060390.

INTRODUCTION

Spinal disorders are the leading cause of work disability, representing around 65% of all musculoskeletal injury/illness claims associated with lost work time, posing a

Strengths and limitations of this study

- This systematic review will include a broad range of instruments used to assess permanent functional impairment in individuals with spinal conditions in peer-reviewed articles.
- This systematic review protocol is presented in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses Protocols.
- A strength of this review is the use of the internationally recognised, validated COnsensus-based Standards for the selection of health Measurement INstruments guidelines to assess the methodological quality of the included studies.
- A limitation of this review is language bias, since only studies in English will be included. However, studies published in English that describe cross-cultural validation of instruments from English into other languages will be included.

considerable economic, personal and societal burden.¹ Internationally, occupational low back pain is the single most prevalent health problem related to absenteeism among workers, responsible for approximately 25% of costs from work-related conditions.²

Spinal disorders such as low back pain are often complex and impact a wide range of multidimensional aspects of health and physical function, such as disability.¹ Disability is an umbrella term for impairments, activity limitations and participation restrictions, referring to difficulties encountered in any or all of these three areas of functioning.³ Such complex and interactive processes can be challenging when trying to assess patients' disability using a dichotomous approach. For example, insurance providers might require thresholds to be set for impairment severity, activity limitations or participation restriction in order to determine eligibility for financial compensation or other benefits.^{3,4}

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

Instruments used to assess spinal PFI have been critically evaluated, since their measurements are usually based only on impairment.^{14–18} Poor validity and reliability of non-invasive clinical measurement instruments and the presence of significant measurement error^{14–17} are criticisms that have led to a number of alternative impairment-based spinal assessment systems such as: American Medical Association Guides' Diagnosis-Related Estimate,¹¹ Diagnosis-Based Impairment,⁵ McBride's method of spinal impairment evaluation¹⁹ or Physical Impairment Index.^{18,20,21} However, these alternative systems for rating PFI present poor descriptions of standardisation and normative values.^{10,22}

The outcome of spinal PFI assessment by compensation bodies can have considerable social, economic and health impacts, yet the reliability and validity of instruments currently available have not been systematically evaluated. To achieve a fair and accurate outcome, clinicians, researchers and government bodies must have access to the most accurate level of evidence regarding methods assessing spinal PFI. To our knowledge, no previous review has assessed the reliability and validity of available instruments for determining PFI of the spine. This manuscript presents the protocol of an ongoing systematic review with the objective to review and synthesise the literature concerning measurement properties of the instruments used for assessing PFI of the spine. Due to the diversity of instruments that might be used to assess PFI of the spine, we anticipate the identification of distinct metrics for measuring PFI will allow specific subgroup analyses for review and ultimate discussion on the strength of support for each instrument.

METHODS

Design

This systematic review protocol has been registered with International Prospective Register of Systematic Reviews (registration number CRD42017060390). This protocol will be reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols statement and checklist.²³

Eligibility criteria

Inclusion criteria

We will include peer-reviewed articles, with no date limitation, investigating measurement properties of instruments that assess PFI of the spine in individuals with spinal disorders aged from 18 to 65 years. When evaluating PFI, acceptable studies will include measures of impairment (ie, RoM, muscle strength, coordination, endurance and sensation), functional limitation (ie, self-report instruments of physical function and functional performance measures) and permanency of the impairment (ie, duration of the impairment and the likelihood of improvement).^{3,24}

Exclusion criteria

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

Information sources

Studies will be identified by searching electronic databases, scanning reference lists of articles and consultation with experts in the area. A preliminary search plan was developed and pretested by a professional health sciences librarian (CB). The following databases were searched: Medline—OVID (1946 to 20 April 2017); EMBASE—OVID (1947 to 21 April 2017); CINAHL—Cumulative Index to Nursing and Allied Health Literature (1937 to 25 April 2017); Web of Science (1900 to 21 April 2017); Scopus (1995 to 25 April 2017); PEDro (Inception to 22 April 2017); OTSeeker (Inception to 22 April 2017) and Health and Safety Science Abstracts—ProQuest (1988 to 22 April 2017).

Search strategy

To capture all measurement instruments used to evaluate PFI of the spine, our search strategy will focus on combining three conceptual groups of terms: (1) PFI, (2) spinal disorder and (3) measurement properties. The Medline search strategy is presented in [table 1](#).

The terms for the concept of 'measurement properties' were adapted from a search filter validated and published by Terwee *et al.*²⁶ and optimised for Medline. A similar search will be employed for other databases, optimised for database-specific search interfaces. A filter to exclude certain publication types (eg, addresses, editorials, letters and newspaper articles) will be applied, as recommended by Terwee *et al.*²⁶ References will be imported

Table 1 Preliminary search strategy in OVID Medline (1950 to present)

'Spinal disorder' search terms	'PFI' search terms	'Measurement properties' search terms
1. spine/or cervical vertebrae/or exp axis, cervical vertebra/or cervical atlas/or coccyx/ or intervertebral disc/or lumbar vertebrae/ or sacrum/or spinal canal/or epidural space/ or thoracic vertebrae/	33. exp 'Range of Motion, Articular'/	59. validation studies/
2. (spine or spinal or coccyx or 'intervertebral disc' or lumbar vertebrae or sacrum or 'spinal canal' or 'thoracic vertebrae' or 'cervical vertebrae').ab,ti.	34. range of motion.ab,ti.	60. Comparative Study/
3. low back pain/	35. Goniomet*.ab,ti.	61. Psychometrics/
4. back Pain/or neck pain/	36. Joint motion measurement*.ab,ti.	62. psychometr*.ab,ti.
5. back/or neck/or lumbosacral region/ or sacrococcygeal region/	37. Electrogoniomet*.ab,ti.	63. clinimetr*.tw.
6. (back or neck or 'lumbosacral region' or 'sacrococcygeal region').ab,ti.	38. (Movement adj2 evaluation*).ab,ti.	64. clinimetr*.tw.
7. Sciatica/	39. Inclinet*.ab,ti.	65. 'Outcome Assessment (Health Care)'/
8. Neck Injury/or Whiplash Injuries/	40. Joint flexibility.ab,ti.	66. 'outcome assessment'.ab,ti.
9. whiplash.ab,ti.	41. Arthromet*.ab,ti.	67. ('outcome measure' or 'outcome measures').tw.
10. Dorsalgia.ti,ab.	42. Joint mobility.ab,ti.	68. Observer Variation/
11. coccydynia.ti,ab.	43. Osteokinematic*.ab,ti.	69. 'observer variation'.ab,ti.
12. (('cervical vertebrae' or 'cervical spine' or craniovertebral or sacroiliac or verteb* or thoracic) adj2 (symptom* or injur* or disorder* or pain or dysfunction* or problem* or strain* or spain*)).ab,ti.	44. Flexion.ab,ti.	70. Health Status Indicators/
13. discitis.ti,ab.	45. Extension.ab,ti.	71. 'Reproducibility of Results'/
14. (disc adj degeneration).ti,ab.	46. Rotation/	72. reproducib*.ti,ab.
15. (disc adj prolapse).ti,ab.	47. side bend*.ab,ti.	73. Discriminant Analysis/
16. (disc adj herniation).ti,ab.	48. Work Capacity Evaluation/	74. reliab*.ab,ti.
17. (facet adj joints).ti,ab.	49. Disability Evaluation/and (Occupational Diseases/or Work/or Return to Work/)	75. unreliab*.ab,ti.
18. Intervertebral Disc/or annulus fibrosus/ or nucleus pulposus/	50. Occupational Diseases/or Work/or Return to Work/	76. valid*.ab,ti.
19. arachnoiditis.ti,ab.	51. (impair* or disabilit* or abilit* or handicap* or 'functional severity' or 'restriction of function' or 'capacity to work' or 'functional capacity' or 'disability rating' or 'impairment rating' or 'work fitness').ab,ti.	77. coefficient.ab,ti.
20. Spinal Fusion/	52. 50 and 51	78. homogeneity.ab,ti.
21. postlaminectomy.ti,ab.	53. ((work* or occupat*) adj4 (capacity or impair* or disabilit* or abilit* or handicap* or 'functional severity' or 'restriction of function')).ab,ti.	79. homogeneous.ab,ti.
22. Backache*.ti,ab.	54. (evaluation* or assessment* or instrument* or measurement* or tool* or scale* or questionnaire* or test* or determination*).ab,ti.	80. 'internal consistency'.ab,ti.
23. back injuries/or spinal injuries/or spinal fractures/	55. 53 and 54	81. cronbach*.ab,ti.
24. Spondylitis, Ankylosing/	56. 'permanent functional impairment'.ab,ti.	82. (alpha or alphas).ab,ti.
25. Spondylitis/	57. ('work performance evaluation' or 'work performance evaluations').ab,ti.	83. 81 and 82
26. spondylosis/or spondylolysis/or spondylolisthesis/	58. 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 or 47 or 48 or 49 or 52 or 55 or 56 or 57	84. item.ab,ti.

Continued

Table 1 Continued

'Spinal disorder' search terms	'PFI' search terms	'Measurement properties' search terms
27. Spinal Cord Compression/		85. (correlation* or selection* or reduction*).ab,ti.
28. (Musculoskeletal adj (symptom* or injur* or disorder* or pain or dysfunction* or problem*)).ab,ti.		86. 84 and 85
29. ((orthopedic or orthopaedic) adj (injur* or problem* or disorder* or dysfunction*)).ab,ti.		87. agreement.ab,ti.
30. Pelvis/or lesser Pelvis/or pelvic floor/		88. precision.ab,ti.
31. (pelvi* adj3 (symptom* or injur* or disorder* or pain or dysfunction* or problem*)).ab,ti.		89. imprecision.ab,ti.
32. 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 or 29 or 30 or 31		90. 'precise values'.ab,ti.
		91. test-retest.ab,ti.
		92. test.ab,ti.
		93. retest.ab,ti.
		94. 92 and 93
		95. reliab*.ab,ti.
		96. (test or retest).ab,ti.
		97. 95 and 96
		98. stability.ab,ti.
		99. interrater.ab,ti.
		100. inter-rater.ab,ti.
		101. intrarater.ab,ti.
		102. intra-rater.ab,ti.
		103. intertester.ab,ti.
		104. inter-tester.ab,ti.
		105. intratester.ab,ti.
		106. intra-tester.ab,ti.
		107. interobserver.ab,ti.
		108. inter-observer.ab,ti.
		109. intraobserver.ab,ti.
		110. intra-observer.ab,ti.
		111. intertechnician.ab,ti.
		112. intertechnician.ab,ti.
		113. intratechnician.ab,ti.
		114. intra-technician.ab,ti.
		115. interexaminer.ab,ti.
		116. inter-examiner.ab,ti.
		117. intraexaminer.ab,ti.
		118. intra-examiner.ab,ti.
		119. interassay.ab,ti.
		120. inter-assay.ab,ti.
		121. intraassay.ab,ti.
		122. intra-assay.ab,ti.
		123. inter-individual.ab,ti.
		124. interindividual.ab,ti.
		125. intraindividual.ab,ti.
		126. intra-individual.ab,ti.

Continued

Table 1 Continued

'Spinal disorder' search terms	'PFI' search terms	'Measurement properties' search terms
		127. interparticipant.ab,ti.
		128. inter-participant.ab,ti.
		129. intraparticipant.ab,ti.
		130. intra-participant.ab,ti.
		131. kappa*.ab,ti.
		132. repeatab*.ab,ti.
		133. (replicab* or repeated).ab,ti.
		134. (measure* or finding* or result* or test*).ab,ti.
		135. 133 and 134
		136. generaliza*.ab,ti.
		137. generalisa*.ab,ti.
		138. concordance.ab,ti.
		139. (intraclass or intra-class).ab,ti.
		140. correlation*.ab,ti.
		141. 139 and 140
		142. discriminative.ab,ti.
		143. 'known group'.ab,ti.
		144. 'factor analysis'.ab,ti.
		145. factor analyses.ab,ti.
		146. dimension*.ab,ti.
		147. subscale*.ab,ti.
		148. (multitrait and scaling and (analysis or analyses)).ab,ti.
		149. item discriminant.ab,ti.
		150. inter-scale correlation*.ab,ti.
		151. interscale correlation*.ab,ti.
		152. error.ab,ti.
		153. errors.ab,ti.
		154. 'individual variability'.ab,ti.
		155. (variability and (analysis or values)).ab,ti.
		156. (uncertainty and (measurement or measuring)).ab,ti.
		157. 'standard error of measurement'.ab,ti.
		158. sensitiv*.ab,ti.
		159. responsive*.ab,ti.
		160. ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ab,ti.
		161. (small* and (real or detectable) and (change or difference)).ab,ti.
		162. 'meaningful change'.ab,ti.
		163. 'ceiling effect'.ab,ti.
		164. 'floor effect'.ab,ti.
		165. 'item response model'.ab,ti.
		166. IRT.ab,ti.
		167. Rasch.ab,ti.
		168. 'differential item functioning'.ab,ti.
		169. DIF.ab,ti.

Continued

Table 1 Continued

'Spinal disorder' search terms	'PFI' search terms	'Measurement properties' search terms
		170. 'computer adaptive testing'.ab,ti.
		171. 'item bank'.ab,ti.
		172. 'cross-cultural equivalence'.ab,ti.
		173. ('gold standard' or 'criterion standard' or 'reference standard').ab,ti.
		174. 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 or 83 or 86 or 87 or 88 or 89 or 90 or 91 or 94 or 98 or 99 or 100 or 101 or 102 or 103 or 104 or 105 or 106 or 107 or 108 or 109 or 110 or 111 or 112 or 113 or 114 or 115 or 116 or 117 or 118 or 119 or 120 or 121 or 122 or 123 or 124 or 125 or 126 or 127 or 128 or 129 or 130 or 131 or 132 or 135 or 136 or 137 or 138 or 141 or 142 or 143 or 144 or 145 or 146 or 147 or 148 or 149 or 150 or 151 or 152 or 153 or 154 or 155 or 156 or 157 or 158 or 159 or 160 or 161 or 162 or 163 or 164 or 165 or 166 or 167 or 168 or 169 or 170 or 172 or 173
175. 32 and 58 and 174		
176. limit 175 to (addresses or autobiography or bibliography or biography or comment or congresses or directory or editorial or festschrift or interactive tutorial or interview or lectures or legal cases or legislation or letter or news or newspaper article or patient education handout or practice guideline or video-audio media or webcasts)		
177. 175 not 176		

PFI, permanent functional impairment.

into DistillerSR software (Distiller SR Evidence Partners, Ottawa, Canada). All duplicate articles will be removed prior to screening. Reference lists of included studies will be scrutinised during the process of data extraction for additional articles meeting our criteria. Identified citations will be downloaded and subjected to the same de-duplication and screening process.

Study selection

Study selection will be undertaken in two steps. First, two reviewers will independently screen all titles and abstracts, and all citations selected by both reviewers will be retrieved for step two: full-text screening. In cases where there is disagreement between the two reviewers, any discrepancies will be resolved by consensus and/or consultation with a third member of the research team prior to making a final decision.

A screening tool will be developed and piloted for the title, abstract and full-text screening stages (see [box](#)). Inclusion/exclusion will be determined using the same tool throughout the process, although the definitions may be refined as conflicts arise and are discussed. The same questions will be used for both title and abstract as well as full-text screening, with a focus on sensitivity rather than specificity at these stages.

Data extraction process

We will develop data extraction forms, as outlined in [table 2](#), which will be pilot tested on three studies not included in the review. Two reviewers will extract data independently. Any disagreement will be resolved through a third party adjudication. The nature of performance measures employed by included studies will drive the nature of the data extraction, as we anticipate that there may be some heterogeneity in the evaluation of measurement instruments.

To avoid analysing the same data from multiple articles, we will compare author names, sample sizes and outcomes. In the case of companion papers (eg, reports) of a primary study, all available data will be simultaneously evaluated, to maximise the yield of information for the primary study. Study authors may be contacted for clarifications. If contact with authors is unsuccessful, we will use the most complete and up-to-date data source available, thus avoiding the risk of double counting.

Data items

We will extract information relating to: (1) the measurement properties (ie, reliability, validity and responsiveness) of the measurement instruments for assessing PFI of the spine (eg, assessment of RoM, muscle strength,

Box Draft screening tool for use at the title, abstract and full-text review screening stages

Questions for all stages: title, abstract and the full-text (go from step 1 to 7)

1. Is the study written in English?
 - a. No—exclude
 - b. Yes or uncertain—go to step 2
2. Does the study deal with humans?
 - a. No—exclude
 - b. Yes or uncertain—go to step 3
3. Does the study deal with adults (between 18 and 65 years of age)?
 - a. No—exclude
 - b. Yes or uncertain—go to step 4
4. Does the article represent primary study (ie, no letters to the editor, book reviews, published study designs or trial protocols)?
 - a. No—exclude
 - b. Yes or uncertain—go to step 5
5. Does the study assess the spine (cervical, thoracic, lumbar spine, including neck and low back)?
 - a. No—exclude
 - b. Yes or uncertain—go to step 6
6. Does the study measure permanent functional impairment (PFI; permanent impairment, physical impairment, functional impairment or disability)?
 - a. No—exclude
 - b. Yes or uncertain—go to step 7
7. Is the study designed to evaluate measurement properties of measurement instruments/tools (eg, validity, reliability, responsiveness)?
 - a. No—exclude
 - b. Yes or uncertain—choose one of the following options:
 - i. Title and abstract screening stage—include
 - ii. Full-text screening stage—go to step 8

Additional question for Full-text stage only

8. Does this study investigate individuals with spinal conditions not caused by congenital and developmental abnormalities, neoplasm, infection disorders and systemic inflammatory disorders?
 - a. No—exclude
 - b. Yes or uncertain—go to step 9
9. Does this study have clinical measures of impairment (eg, range of motion, muscle performance) or objective measures of function (eg, functional impairment evaluation)?
 - a. No—exclude
 - b. Yes or uncertain—go to step 10
10. Does this study have a PFI definition that includes measures of chronic or permanent symptoms' duration as well as measures of functional limitation?
 - a. No—exclude
 - b. Yes or uncertain—go to step 11
11. Does this study use at least one statistical test between the measures listed on question number '9' and on question number '10'?
 - a. No—exclude
 - b. Yes or uncertain—Include

coordination, endurance and sensation, as well as ability to perform basic daily living and work activities, and duration of the impairment); (2) the type of inferential test and

(3) the test results for each article. The extraction process will be tracked using a standardised data extraction form.

Articles will be separated into subgroups based on the instruments used to assess PFI of the spine. The different subgroups will be analysed separately based on impairment-based measures (ie, RoM, muscle strength, coordination, endurance and sensation) and functional limitation assessment (ie, self-report instruments of physical function and functional performance measures). For clarity, we are likely to isolate articles in subgroups to better differentiate the instruments used to assess PFI of the spine, especially when considering that RoM has predominantly been used for assessing impairment of the spine.^{5 6 9 11}

Quality assessment of individual studies

Included studies will be appraised independently by two reviewers as to their methodological quality using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) criteria.^{27 28} Any disagreement will be resolved by consultation with a third member of the research team. The COSMIN checklist is a consensus-based tool designed to evaluate the methodological quality of studies investigating measurement properties. The instrument shows appropriate levels of agreement²⁷ and, based on its content validity, is a recommended tool for assessing the methodological quality of studies evaluating measurement properties of outcome measures within a systematic review.²⁸

The tool will evaluate the following measurement constructs: reliability; measurement error; content validity; structural validity; hypotheses testing; cross-cultural validity and criterion validity, responsiveness, and interpretability with 5–18 items concerning methodological standards for how each measurement property should be assessed (see [table 3](#)). The methodological quality of a study will be considered adequate if all items in a measurement property are considered adequate. Each item is scored on a four-point rating scale (ie, 'poor', 'fair', 'good' or 'excellent').

In the context of this review, validity, in general, defines how well the instrument under evaluation measures the construct it purports to measure. Criterion validity is the degree to which measurements are an adequate reflection of a previously used 'gold standard'. Content validity is an adequate reflection of the construct to be measured; construct validity is based on an assumption that the instrument truly measures what it is meant to and structural validity implies the scores of an instrument is an adequate reflection of the dimensionality of the construct to be assessed. Reliability refers to the extent to which scores for individuals who have not changed are the same for repeated measurement under several conditions. These include using different sets of items from the same instrument (internal consistency); over time (test-retest); by different assessors on the same occasion (inter-rater) or by the same assessors (ie, raters or responders) on different occasions (intra-rater). Responsiveness is the

Table 2 Draft extraction tool

Item	Definitions, decision rules, guidance and example data
Reference ID	Author, date
Instrument name	The name as described in the article, or as referred to in other articles (eg, 'Back-EST' or 'the Johnson method')
Instrument description	Will collect sub variables, such as: Equipment required Duration Clinicians performing test (eg, physiotherapist, occupational therapist, physician) Dimensions assessed (eg, range of motion, lifting capacity, etc)
Validity—criterion	Will collect subvariables, such as: Did the article assess this dimension? (yes/no) What was the inferential test used? (eg, intraclass correlation coefficient, kappa) What was the result? (eg, kappa score of 0.7)
Validity—content	Will collect sub variables, such as: Did the article assess this dimension? (yes/no) How was content validity assessed? What were the findings?
Validity—construct	Will collect subvariables, such as: Did the article assess this dimension? (yes/no) How was construct validity assessed? What were the findings?
Reliability—inter-rater	Will collect sub variables, such as: Did the article assess this dimension? (yes/no) What was the inferential test used? (eg, intraclass correlation coefficient, kappa) What was the result? (eg, kappa score of 0.7)
Reliability—intrarater	Will collect subvariables, such as: Did the article assess this dimension? (yes/no) What was the inferential test used? (eg, intraclass correlation coefficient, kappa) What was the result? (eg, kappa score of 0.7)
Responsiveness	Will collect sub variables, such as: Did the article assess this dimension? (yes/no) How was responsiveness assessed? What were the findings?
COSMIN Study Quality Metrics	Will adhere to the COSMIN tool, do not anticipate altering the definitions

These items may evolve as review progresses and studies are identified. This draft data dictionary describes the selected data extraction variables of the systematic review. The data dictionary would be used to extract data from articles identified as relevant during the screening process. The data extraction database itself would have these items listed in columns with the individual references representing rows such that these data are extracted for every article to form a summary database for later synthesis. Only key variables are presented here. It is anticipated that this list will expand based on the type of data presented in the articles; the definitions guide will also become more specific as a result of independent data extraction and discussion at team meetings.

COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments.

ability of the measurement instrument to detect change over time in the construct to be measured.²⁷

Planned methods of analysis

Where it is possible, mean values of statistical analysis (eg, Cronbach's alpha, intraclass correlation coefficient, SE of measurement, smallest detectable change, effect sizes, etc) will be calculated from pooled data from methodologically similar studies and the results statistically summarised via meta-analysis into forest plots with estimates of heterogeneity. In addition, sensitivity analysis will be pursued by comparing results from studies with high-quality and low-quality ranking. We will then weight such meta-analyses using both the study's sample

size and their quality assessment as determined by the COSMIN checklist.²⁸ However, we predict some heterogeneity will be identified in the various assessments of spinal PFI measures, which will likely make a meta-analysis difficult to apply.

In the event meta-analysis is not possible, descriptive and narrative syntheses will be presented in text and table formats, with the intent of summarising and discussing the sample and methodological aspects, as well as the findings regarding measurement properties of the included studies assessing PFI in individuals with spinal conditions. Tables will provide general and comparative information of these heterogeneous and disparate studies (ie, authors, country and population

Table 3 COSMIN definitions of domains, measurement properties and aspects of measurement properties

Domain	Measurement property	Aspect of a measurement property	Definition
Reliability			The extent to which scores for individuals who have not changed are the same for repeated measurement under several conditions.
	Internal consistency		Different sets of items from the same instrument.
	Test-retest		Changes are repeated measurement over time.
	Inter-rater		The degree of agreement between raters investigating the same property on the same patient.
	Intrarater		The degree of agreement between repeated measurements of a property on the same patient by the same rater.
Validity	Measurement error		The systemic and random error of a patient's score that is not attributed to true changes in the construct to be measured.
	Content validity		The degree to which the score of an instrument is adequate reflection of the construct to be measured.
		Face validity	The degree to which (the item of) an instrument is an adequate reflection of the construct to be measured.
	Construct validity		The degree to which the scores of an instrument are consistent with hypotheses (for instance, with regard to internal relationships, relationships to scores of other instruments or differences between relevant groups) based on an assumption that the instrument truly measures what it is meant to.
		Structural validity	The degree to which the score of an instrument is an adequate reflection of the dimensionality of the construct to be assessed.
		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument.
		Criterion validity	The degree to which measurements are an adequate reflection of a previously used 'gold standard'.
Responsiveness			The ability of the measurement instrument to detect change over time in the construct to be measured.
Interpretability			The degree to which one can assign qualitative meaning, that is, clinical or commonly understood connotations— to an instrument's quantitative scores or change in scores.

COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments.

parameters, such as age, gender distribution, setting etc), and will summarise reliability, validity and responsiveness data with associated study quality indicators (COSMIN checklist).

In order to determine the best available method for measuring PFI of the spine, each identified instrument will be ranked using a range of measurement performance metrics identified in the COSMIN checklist. The findings will be presented and possible hypotheses for the results will be generated and discussed. In addition, gaps in the literature will be identified and discussed.

ETHICS AND DISSEMINATION

A substantial number and variety of spine impairment ratings are calculated internationally and used for

financial award decisions, in both legal and compensation contexts.^{29–34} However, a systematic review in this area to support the use of instruments currently being used to determine PFI is lacking. The application of reliable and valid measurement instruments for assessing PFI of the spine is considered crucial.

This systematic review offers a feasible means for synthesising the evidence specific to spinal PFI assessment; and our results will likely provide unique insights concerning the breadth and depth of literature in the area. Outcomes of this review will be applicable to clinicians, policy-makers, worker's compensation boards and health and safety organisations. In particular, findings will likely provide a foundation and direction in terms of research priorities for assessing PFI of the spine. Summarising the nature and strength of the

evidence regarding the reliability, validity and responsiveness of spinal PFI measures will also inform future research and policy in this field.

Contributors SMG and SM conceptualised the initial review protocol and led the development of this manuscript. The entire team (SMG, SM, CB, CMT, BB, DCR, PH, LC and XZ) collaboratively drafted the manuscript, followed by numerous iterations and substantial input and appraisal from all authors. All authors have approved the final version of this manuscript.

Funding The Workers' Compensation Board of British Columbia (WorkSafeBC grant number RS2015-SR01).

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

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 and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Driscoll T, Jacklyn G, Orchard J, *et al.* The global burden of occupationally related low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014;73:975–81.
2. Helfenstein Junior M, Goldenfum MA, Siena C. Occupational low back pain. *Rev Assoc Med Bras* 2010;56:583–9.
3. WHO. *A practical manual for using the International Classification of Functioning, Disability and Health (ICF)*. Geneva: World Health Organization, 2013.
4. WHO. *World report on disability*. Geneva: WHO, 2011.
5. Rondinelli RD, Genovese E, Brigham CR, *et al.* *Guides to the evaluation of permanent impairment*. 6th ed. Chicago, IL: American Medical Association, 2008.
6. Comcare. *Guide to the assessment of the degree of permanent impairment*. 2.1 edn. Australia: Australian Government: Comcare, 2014.
7. WorkCover. *Guidelines for the assessment of permanent impairment*. 3rd edn. Tasmania, 2011.
8. SafeWork. *Permanent functional impairment rating schedule*. Newfoundland and Labrador - Canada: Workplace Health, Safety and Compensation Commission, 2009.
9. WorkSafeBC. *Permanent Disability Evaluation Schedule ("PDES") 2016 Review*. British Columbia - Canada: WorkSafeBC, 2016.
10. Kim YB, Lee SG, Park CW, *et al.* Korean guideline development for the evaluation of permanent impairment of the spine: proposal by the Korean Academy of Medical Sciences Committee. *J Korean Med Sci* 2009;24(Suppl 2):S307–13.
11. Cocchiarella L, Andersson GBJ. *Guide to the evaluation of permanent impairment*. 5th ed. Chicago, Ill: edAmerican Medical Association, 2000.
12. van der Wurff P, Hagmeijer RH, Meyne W. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 1: Reliability. *Man Ther* 2000;5:30–6.
13. Reiman MP, Manske RC. The assessment of function: how is it measured? A clinical perspective. *J Man Manip Ther* 2011;19:91–9.
14. Hirsch BP, Webb ML, Bohl DD, *et al.* Improving visual estimates of cervical spine range of motion. *Am J Orthop* 2014;43:E261–5.
15. Whitcroft KL, Massouh L, Amirfeyz R, *et al.* Comparison of methods of measuring active cervical range of motion. *Spine* 2010;35:E976–80.
16. Nitschke JE, Nattrass CL, Disler PB, *et al.* Reliability of the American Medical Association guides' model for measuring spinal range of motion. Its implication for whole-person impairment rating. *Spine* 1999;24:262–8.
17. Lowery WD, Horn TJ, Boden SD, *et al.* Impairment evaluation based on spinal range of motion in normal subjects. *J Spinal Disord* 1992;5:398–402.
18. Fritz JM, Piva SR. Physical impairment index: reliability, validity, and responsiveness in patients with acute low back pain. *Spine* 2003;28:1189–94.
19. McBride ED. Disability evaluation: principles of treatment of compensable injuries. 1936. *Clin Orthop Relat Res* 2007;457:11–15.
20. Trudelle-Jackson E, Sarvaiya-Shah SA, Wang SS. Interrater reliability of a movement impairment-based classification system for lumbar spine syndromes in patients with chronic low back pain. *J Orthop Sports Phys Ther* 2008;38:371–6.
21. Waddell G, Somerville D, Henderson I, *et al.* Objective clinical evaluation of physical impairment in chronic low back pain. *Spine* 1992;17:617–28.
22. Rondinelli RD. Changes for the new AMA Guides to impairment ratings, 6th Edition: implications and applications for physician disability evaluations. *Pm R* 2009;1:643–56.
23. 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.
24. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical perspective. *J Man Manip Ther* 2011;19:91–9.
25. Haldeman S, Kopansky-Giles D, Hurwitz EL, *et al.* Advancements in the management of spine disorders. *Best Pract Res Clin Rheumatol* 2012;26:263–80.
26. Terwee CB, JansmaEP, Riphagen II, *et al.* Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.
27. Mookink LB, Prinsen CA, Bouter LM, *et al.* The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther* 2016;20:105–13.
28. Terwee CB, Mookink LB, Knol DL, *et al.* Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
29. Davis MA, Onega T, Weeks WB, *et al.* Where the United States spends its spine dollars: expenditures on different ambulatory services for the management of back and neck conditions. *Spine* 2012;37:1693–701.
30. Martin BI, Deyo RA, Mirza SK, *et al.* Expenditures and health status among adults with back and neck problems. *JAMA* 2008;299:656–64.
31. Martin BI, Turner JA, Mirza SK, *et al.* Trends in health care expenditures, utilization, and health status among US adults with spine problems, 1997–2006. *Spine* 2009;34:2077–84.
32. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J* 2008;8:8–20.
33. McGee R, Bevan S, Quadrello T. *Fit for work? musculoskeletal disorders and the canadian labour market*. London, ON: The Work Foundation, 2009.
34. Dunning KK, Davis KG, Cook C, *et al.* Costs by industry and diagnosis among musculoskeletal claims in a state workers compensation system: 1999–2004. *Am J Ind Med* 2010;53:276–84.