BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019276
Article Type:	Protocol
Date Submitted by the Author:	22-Aug-2017
Complete List of Authors:	Goes, Suelen; University of Saskatchewan, School of Physical Therapy Trask, Catherine; University of Saskatchewan, Canadian Centre for Health and Safety in Agriculture Boden, Catherine; University of Saskatchewan, Leslie and Irene Dubé Health Sciences Library Bath, Brenna; University of Saskatchewan, School of Physical Therapy Ribeiro, Daniel; School of Physiotherapy Hendrick, Paul ; University of Nottingham, ; University of Nottingham, Clay, Lynne; University of Otago, School of Physiotherapy Zeng, Xiaoke; University of Saskatchewan, Canadian Centre for Health and Safety in Agriculture Milosavljevic, Stephan; University of Saskatchewan, College of Medicine
Primary Subject Heading:	Occupational and environmental medicine
Secondary Subject Heading:	Health policy, Rehabilitation medicine
Keywords:	OCCUPATIONAL & INDUSTRIAL MEDICINE, Back pain < ORTHOPAEDIC & TRAUMA SURGERY, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, Spine < ORTHOPAEDIC & TRAUMA SURGERY

SCHOLARONE™
Manuscripts

1 **TITLE PAGE**

2

3 **Title:**

4

5 Measurement properties of instruments assessing permanent functional impairment of the spine:

6 a systematic review protocol

7

8 **Authors, affiliations:**

9

10 **Suelen Meira Goes**: suelen.goes@usask.ca; School of Physical Therapy, University of

11 Saskatchewan - Health Sciences Building, E-Wing; Suite 3400 104 Clinic Place - Saskatoon,

12 Saskatchewan, S7N 2Z4 – Canada. *** Corresponding Author**

13

14 **Catherine Trask**: catherine.trask@usask.ca; Canadian Centre for Health and Safety in

15 Agriculture (CCHSA), University of Saskatchewan - 103 Hospital Drive, Saskatoon,

16 Saskatchewan, S7N 0W8 – Canada.

17

18 **Catherine Boden**: catherine.boden@usask.ca; Leslie and Irene Dubé Health Sciences Library,

19 Academic Health Sciences Building - 104 Clinic Place, Saskatoon, Saskatchewan, S7N 2Z4 –

20 Canada.

21

22 **Brenna Bath**: brenna.bath@usask.ca; School of Physical Therapy, University of Saskatchewan -

23 Health Sciences Building, E-Wing; Suite 3400 104 Clinic Place - Saskatoon, Saskatchewan, S7N

24 2Z4 – Canada.

25

26 **Daniel Cury Ribeiro**: daniel.ribeiro@otago.ac.nz; School of Physiotherapy, University of Otago

27 - 325 Great King Street, Dunedin, 9016 – New Zealand.

28

29 **Paul Hendrick**: paul.hendrick@nottingham.ac.uk; School of Health Sciences, University of

30 Nottingham – Clinical Sciences Building - Nottingham City Hospital - Hucknall Rd Nottingham,

31 NG5 1PB – United Kingdom.

32

33 **Lynne Clay**: lynne.clay@otago.ac.nz; School of Physiotherapy, University of Otago - 325 Great

34 King Street, Dunedin, 9016 – New Zealand.

35

36 **Xiaoke Zeng**: xiaoke.zeng@usask.ca; Canadian Centre for Health and Safety in Agriculture

37 (CCHSA), University of Saskatchewan - 103 Hospital Drive, PO Box 120, Saskatoon,

38 Saskatchewan, S7N 0W8 – Canada.

39

40 **Stephan Milosavljevic**: stephan.milosavljevic@usask.ca; School of Physical Therapy,

41 University of Saskatchewan - Health Sciences Building, E-Wing; Suite 3400 104 Clinic Place -

42 Saskatoon, Saskatchewan, S7N 2Z4- Canada.

43

44 **Word count:** 257 (abstract) + 2719 (protocol)

3

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 synthesize 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 (COSMIN) criteria. Findings will be summarized descriptively, with meta-analysis pursued as appropriate.

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

Keywords: disability, back, range of movement, psychometrics, validity, reliability, responsiveness

Systematic review registration: PROSPERO CRD42017060390.

Strengths and limitations of this study

- Feasible means for synthesizing the evidence specific to measurement properties of instruments used for assessing spinal PFI.

- Insights concerning the foundation and direction in terms of research priorities for assessing PFI of the spine.

- This systematic review will inform future research and policy in this field.

- Outcomes of this review will be applicable to clinicians, policy-makers, and worker's compensation boards.

81 **INTRODUCTION**

82 Spinal disorders are the leading cause of work disability, representing around 65% of all

83 musculoskeletal injury/illness claims associated with lost work time, posing a considerable

84 economic, personal, and societal burden [1]. Internationally, occupational low back pain is the

85 singular most prevalent health problem related to absenteeism amongst workers, responsible for

86 approximately 25% of costs from work-related conditions [2].

87 Spinal disorders such as low back pain are often complex and impact a wide range of

88 multidimensional aspects of health and physical function, such as disability [1]. Disability is an

89 umbrella term for impairments, activity limitations and participation restrictions, referring to

90 difficulties encountered in any or all of these three areas of functioning [3]. Such complex and

91 interactive processes can be challenging when trying to assess patients' disability using a

92 dichotomous approach. For example, insurance providers might require thresholds to be set for

93 impairment severity, activity limitations, or participation restriction in order to determine

94 eligibility for financial compensation or other benefits [3 4].

95 Permanent functional impairment (PFI), or permanent impairment, is a rating systems used by

96 compensation authorities, to establish an appropriate level of financial compensation [5-9].

97 Following review of current World Health Organisation ICF classifications [3] and American

98 Medical Association guides [5] for evaluation of permanent impairment, we have defined PFI as

99 any long-lasting disturbance of physiological structure and function, that leads to an inability to

100 perform a normal range of basic daily living and/or work-related activities. Furthermore, the

101 assessment of PFI needs to incorporate all of these components.

102 Evaluation of PFI requires selecting appropriate outcome measures [10]. However, the metrics of

103 PFI ratings are not uniformly specified nor universally adopted by workers' compensation

104 boards, varying in terms of specific PFI rating guides as well as the adoption of function-based

105 criteria [6 9 11]. The feasibility, reliability, validity, utility, and cost are important aspects that

106 should be taken into consideration when selecting clinical measurement instruments for

107 determination of PFI [12].

108 In view of the definition of PFI, spinal impairment assessment should take into account

109 impairment-based measures (i.e. range of motion, muscle strength, coordination, endurance, and

110 sensation), self-report instruments, and physical (functional) performance measures [5][13].

111 However, PFI assessment of individuals with work-related spinal disorders historically and

112 predominantly relies on a single impairment-based measure: spinal range of motion (RoM) [5][6][9]

113 [11]. Reduced spinal RoM, described as a diminished motion arc of the spine measured by values

114 of flexion-extension, rotation, and lateral bending [14], has been considered a key tenet to assist in

115 categorizing spinal PFI severity [5][6][9][11]. Impairment can contribute to limited function and

116 ultimately may have consequences for physical functioning, yet a clinical examination finding of

117 an impairment does not always correspond to a functional loss [15].

118 Instruments used to assess spinal PFI have been critically evaluated, since their measurements

119 are usually based only on impairment [16-20]. Poor validity and reliability of non-invasive clinical

120 measurement instruments and the presence of significant measurement error [16-19] are criticisms

121 that have led to a number of alternative impairment-based spinal assessment systems such as:

122 AMA Guides' Diagnosis-Related Estimate [11], Diagnosis-Based Impairment [5], McBride's method

123 of spinal impairment evaluation [21], or Physical Impairment Index [20][22][23]. However, these

124 alternative systems for rating PFI present poor descriptions of standardization and normative

125 values [10][13].

126 The outcome of spinal PFI assessment by compensation bodies can have considerable social,

127 economic and health impacts, yet the reliability and validity of instruments currently available

128 have not been systematically evaluated. To achieve a fair and accurate outcome, clinicians,

129 researchers and government bodies must have access to the most accurate level of evidence

130 regarding methods assessing spinal PFI. To our knowledge, no previous review has assessed the

131 reliability and validity of available instruments for determining PFI of the spine. The objective of

132 this study is to systematically review and synthesize the literature concerning measurement

133 properties of the various and different instruments used for assessing PFI of the spine. Due to the

134 diversity of instruments that might be used to assess PFI of the spine, we anticipate the

135 identification of distinct metrics for measuring PFI will allow specific sub-group analyses for

136 review and ultimate discussion on the strength of support for each instrument.

137   **METHODS**

138   This study will be a systematic review of the measurement properties (validity, reliability, and

139   responsiveness) of the different instruments identified for assessing spinal PFI, where PFI is

140   defined as any permanent disturbance or abnormality of physiological structure leading to loss of

141   function. Impairment is any problem in body function and structure; functional impairment is

142   described as restrictions in performance of the body systems [3], and permanent impairment is

143   determined when the impairment becomes well-stabilized with or without treatment, and further

144   significant medical improvement is unlikely [5].

145   While evaluating PFI, studies should include measures of impairment (i.e. RoM, muscle

146   strength, coordination, endurance, and sensation), functional limitation (i.e. self-report

147   instruments of physical function and functional performance measures) and permanency of the

148   impairment (i.e. duration of the impairment and the likelihood of improvement) [3 24].

149   Systematic review reporting will adhere to the Preferred Reporting Items for Systematic reviews

150   and Meta-Analyses (PRISMA) guidelines using the PICOS methodology [25]. This method

151   includes: **P**opulation (individuals aged between 18 to 65 years (i.e. typical working age

152   population) with spinal conditions not caused by congenital and developmental abnormalities,

153   neoplasm, infection disorders, and systemic inflammatory disorders); **I**ntervention/exposure

154   (measurement instruments used to assess PFI of the spine); **C**omparator (not applicable for this

155   review); **O**utcome (reliability, validity and responsiveness); and **S**tudy design (cross-sectional

156   and cohort studies reporting on measurement properties). To capture all measurement

157   instruments used to evaluate PFI of the spine, our search strategy will focus on combining three

158   conceptual groups of terms: (1) PFI, (2) spinal disorder, and (3) measurement properties.

159   **Eligibility Criteria**

160   We will examine peer-reviewed, full-text articles over the fully available date range. We

161   anticipate that older literature may have decreased relevance given changes in research

162   methodology. Nonetheless, we have elected not to limit publication range in order to be thorough

163   in collecting documented evaluations of measurement properties of instruments assessing PFI of

164   the spine. In order to capture typical 'working' age populations, we will focus on studies of

165 adults between 18 and 65 years of age. In addition, given the broad scope of spinal disorders and

166 considering that workers compensation boards do not uniformly specify or universally provide

167 financial compensation for all specific diseases and non-specific musculoskeletal disorders, we

168 will include studies of individuals with spinal conditions not caused by congenital and

169 developmental abnormalities, neoplasm, infection disorders, and systemic inflammatory

170 disorders [26].

171 Included articles will investigate measurement instruments for assessing all components

172 described as contributing to PFI and will assess the psychometric properties of these assessment

173 instruments in terms of validity, reliability, and responsiveness. We will include cross-sectional

174 and cohort studies reporting on measurement properties of at least one measure of PFI of the

175 spine in adults. This search strategy will likely detect two distinct categories of evaluations, one

176 intended for screening, diagnosis and prognosis of spinal pathologies, and the other on the

177 measurement properties of measurement instruments: our focus will be on the latter.

178 Exclusion of studies using a population sample under 18 and/or over 65 years will take place at

179 the screening stage instead of in the literature search phase to avoid the risk of excluding relevant

180 articles. Letters to the editor, book reviews, and other non-articles in journals will also be

181 excluded, as will published study designs and clinical protocols. Only studies in English-

182 language and on humans will be included.

183 **Search strategy**

184 A preliminary search plan was developed and pre-tested with the help of a professional health

185 sciences librarian (CB). The following databases will be searched:

186 1. Medline (OVID) –1946 to current

187 2. EMBASE (OVID) – 1947 to current

188 3. CINAHL (Cumulative Index to Nursing and Allied Health Literature) –1937 to current

189 4. Web of Science –1900 to current

190 5. Scopus – 1995 to current

191 6. PEDro – Inception to current

192 7. OTSeeker – Inception to current

193   8.   Health and Safety Science Abstracts (ProQuest) – 1988 to current

194   The Medline search strategy is presented in Table 1. The terms for the concept of "measurement

195   properties" were adapted from a search filter validated and published by Terwee et al [27] and

196   optimized for Medline. A similar search will be employed for other databases, optimized for

197   database-specific search interfaces. A filter to exclude certain publication types (e.g., addresses,

198   editorials, letters, and newspaper articles) will be applied, as recommended by Terwee et al [27].

199   References will be imported into DistillerSR software (Distiller SR Evidence Partners Limited,

200   Ottawa, Canada). All duplicate articles will be removed prior to screening. Reference lists of

201   included studies will be scrutinized during the process of data extraction for additional articles

202   meeting our criteria. Identified citations will be downloaded and subjected to the same de-

203   duplication and screening process.

204   Table 1 about here

Screening and selecting articles

206   Two reviewers will screen citations independently at the title, abstract, and full-text stages.

207   Citations that both reviewers agree should be included will then be sent to the next stage of

208   screening and citations that both reviewers agree should be excluded will be eliminated from the

209   study. In cases where there is disagreement between the two reviewers, any discrepancies will be

210   solved by consensus and/or consultation with a third member of the research team prior to

211   making a final decision.

212   A screening tool will be developed and piloted for the title, abstract and full-text screening stages

213   (see table 2). Inclusion/exclusion will be determined using the same tool throughout the process,

214   although the definitions may be refined as conflicts arise and are discussed. The same questions

215   will be used for both title and abstract as well as full-text screening, with a focus on sensitivity

216   rather than specificity at these stages.

217   Table 2 about here

9

218 **Data extraction**

219 Following full-text screening and identification of a final list of included studies, we will begin

220 data extraction. Due to anticipated uncertainty regarding the nature and extent of information to

221 be extracted, we will collectively develop and pilot preliminary data extraction forms as outlined

222 in table 3. The data extraction instruments will be developed in parallel with the screening phase

223 at team meetings. After reviewing the first five articles, the team will re-convene to discuss

224 whether updates to the instruments or process are required. The criteria in the final extraction

225 tool will drive the synthesis efforts.

226 | Table 3 about here |

227 Two reviewers will extract data independently. Any disagreement will be resolved through

228 discussion with research team members. The nature of performance measures employed by

229 included studies will drive the nature of the data extraction, as we anticipate that there may be

230 some heterogeneity in the evaluation of measurement instruments.

231 To avoid analysing the same data from multiple articles, we will compare author names, sample

232 sizes and outcomes. In the case of companion papers (e.g. protocols, reports) of a primary study,

233 all available data will be simultaneously evaluated, to maximize the yield of information. Study

234 authors may be contacted for clarifications or missing data.

235 We will extract information relating to the measurement properties (i.e. reliability, validity, and

236 responsiveness) of the measurement instruments for assessing PFI of the spine (e.g. assessment

237 of RoM, muscle strength, coordination, endurance, and sensation, as well as ability to perform

238 basic daily living and work activities, and duration of the impairment) for each article. This

239 process will involve tracking the dimension reported by published studies (i.e. reliability,

240 validity, responsiveness), as well as the type of inferential test and the test results. The extraction

241 process will be tracked using a standardized data extraction form.

242 Articles will be separated into sub-groups based on the instruments used to assess PFI of the

243 spine. The different sub-groups will be analyzed separately based on impairment-based measures

244 (i.e. RoM, muscle strength, coordination, endurance, and sensation) and functional limitation

245 assessment (i.e. self-report instruments of physical function and functional performance

246 measures). For clarity, we are likely to isolate articles in sub-groups to better differentiate the

247 instruments used to assess PFI of the spine, especially when considering that RoM has

248 predominantly been used for assessing impairment of the spine [5][6][9][11].

249 In the context of this review, validity, in general, defines how well the instrument under

250 evaluation measures the construct it purports to measure. Criterion validity is the degree to which

251 measurements are an adequate reflection of a previously used 'gold standard'. Content validity is

252 an adequate reflection of the construct to be measured; construct validity is based on an

253 assumption that the instrument truly measures what it is meant to; and structural validity implies

254 the scores of an instrument is an adequate reflection of the dimensionality of the construct to be

255 assessed. Reliability refers to the extent to which scores for individuals who have not changed

256 are the same for repeated measurement under several conditions. These include using different

257 sets of items from the same instrument (internal consistency); over time (test-retest); by different

258 assessors on the same occasion (inter-rater); or by the same assessors (i.e. raters or responders)

259 on different occasions (intra-rater). Responsiveness is the ability of the measurement instrument

260 to detect change over time in the construct to be measured [28].

261 **Quality assessment**

262 Included studies in each sub-group will be appraised as to their methodological quality using the

263 consensus-based standards for the selection of health measurement instruments (COSMIN)

264 criteria [28][29]. The COSMIN checklist is a consensus-based tool designed to evaluate the

265 methodological quality of studies investigating measurement properties. The instrument shows

266 appropriate levels of agreement [28] and, based on its content validity, is a recommended tool for

267 assessing the methodological quality of studies evaluating measurement properties of outcome

268 measures within a systematic review [29]. The tool evaluates the following measurement

269 constructs: reliability; measurement error; content validity; structural validity; hypotheses

270 testing; cross-cultural validity; and criterion validity. Responsiveness and interpretability with

271 five to18 items concerning methodological standards for how each measurement property should

272 be assessed. The methodological quality of a study will be considered adequate if all items in a

273 measurement property are considered adequate. Each item is scored on a four-point rating scale

274 (i.e., ''poor'', ''fair'', ''good'', or ''excellent'').

Data synthesis and analysis

276 Statistical summary of results (i.e. meta-analysis when applicable) will be pursued (Review

277 Manager Software – Cochrane Collaboration - RevMan 5.1.6). When appropriate, results will be

278 summarized into forest plots with estimates of heterogeneity; and sensitivity analysis will be

279 performed by comparing results from studies with high and low-quality ranking. We anticipate

280 some heterogeneity in how PFI of the spine measures are assessed in each article. For example,

281 agreement between repeated tests, between assessors, or between two instruments can be

282 reported using kappa statistics, intraclass correlation coefficient, limits of agreement, or

283 correlation coefficients that would complicate statistical summary via meta-analysis. In cases

284 where the data required for meta-analysis calculations are not available or there is too much

285 diversity in exposure metrics, findings will be summarized qualitatively. Where possible, we will

286 weight a meta-analysis using both the study's sample size and their quality assessment as

287 determined by the COSMIN checklist [29]. Literature gaps, which limit interpretation of research

288 questions, will be identified and discussed.

289 Evidence for biased selection of research findings (meta-bias; publication bias; reporting bias)

290 will also be considered. Since protocol studies that describe forthcoming evaluation of

291 measurement instruments (e.g. validity, reliability, and responsiveness) are scarce, we will verify

292 whether the pre-specified primary outcomes have been reported, and contact authors when we

293 lack important data or information from the included study.

294

DISCUSSION

296 A significant amount of spine impairment ratings are calculated annually across the world and

297 used for financial award decisions, in both legal and compensation contexts [30-35]. However, a

298 systematic review in this area to support the use of instruments currently being used to determine

299   PFI is lacking. The application of reliable and valid measurement instruments for assessing PFI

300   of the spine is considered crucial.

301   This systematic review offers a feasible means for synthesizing the evidence specific to spinal

302   PFI assessment; results will provide unique insights concerning the breadth and depth of

303   literature in the area. Outcomes of this review will be applicable to clinicians, policy-makers,

304   worker's compensation boards and health and safety organizations. In particular, findings will

305   provide a foundation and direction in terms of research priorities for assessing PFI of the spine.

306   Summarizing the nature and strength of the evidence regarding the reliability, validity and

307   responsiveness of spinal PFI measures will inform future research and policy in this field.

308

309   **LIST OF ABBREVIATIONS**

310   AMA: American Medical Association

311   CINAHL: Cumulative Index to Nursing and Allied Health Literature

312   COSMIN: COnsensus-based Standards for the selection of health Measurement Instruments

313   Embase: Excerpta Medica Database

314   HR-PRO: health related patient-reported outcomes.

315   ICF: International Classification of Functioning, Disability and Health

316   OTseeker: Occupational Therapy Systematic Evaluation of Evidence

317   PEDro: Physiotherapy Evidence Database

318   PFI: Permanent Functional Impairment

319   PICOS: Population; Interventions or exposure; Comparator; Outcomes or endpoint; Study design

320   PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

321   RoM: Range of Motion

322   WHO: World Health Organization

323

AUTHORS' CONTRIBUTIONS

325   SMG and SM conceptualized the initial review protocol and led the development of this

326   manuscript. The entire team (SMG, SM, CB, CT, BB, DCR, PH, LC, and XZ) collaboratively

327   drafted the manuscript, followed by numerous iterations and substantial input and appraisal from

328   all authors. All authors have approved the final version of this manuscript.

FUNDING

330   This research is funded by The Workers' Compensation Board of British Columbia

331   (*WorkSafeBC grant number RS2015-SR01*).

COMPETING INTERESTS

333   The views, findings, opinions and conclusion expressed herein do not necessarily represent the

334   views of the funder.

REFERENCES

336   1. Driscoll T, Jacklyn G, Orchard J, et al. The global burden of occupationally related low back
337   pain: estimates from the Global Burden of Disease 2010 study. *Annals of the rheumatic*
338   *diseases* 2014;73(6):975-81. doi: 10.1136/annrheumdis-2013-204631 [published Online
339   First: 2014/03/26]
340   2. Helfenstein M, Goldenfum MA, Siena C. OCCUPATIONAL LOW BACK PAIN. *Revista Da*
341   *Associacao Medica Brasileira* 2010;56(5):583-89.
342   3. WHO. A Practical Manual for using the International Classifcation of Functioning, Disability
343   and Health (ICF). Geneva: World Health Organization, 2013.
344   4. WHO. World report on disability. Geneva, 2011.
345   5. Rondinelli RD, Genovese E, Brigham CR. American Medical Association. Guides to the
346   Evaluation of Permanent Impairment. 6th ed. Chicago, IL: American Medical
347   Association 2008.
348   6. Comcare. Guide to the assessment of the degree of permanet impairment. 2.1 ed. Australia,
349   2014.

350 7. WorkCover. Guidelines for the assessment of permanent impairment. 3rd ed. Tasmania, 2011.
351 8. SafeWork. Permanent Functional Impairment Rating Schedule. Newfoundland and Labrador -
352   Canada: Workplace Health, Safety and Compensation Commission, 2009.
353 9. WorkSafeBC. Permanent Disability Evaluation Schedule ("PDES") 2016 Review. British
354   Columbia - Canada, 2016.
355 10. Kim YB, Lee SG, Park CW, et al. Korean guideline development for the evaluation of
356   permanent impairment of the spine: proposal by the Korean Academy of Medical
357   Sciences Committee. *J Korean Med Sci* 2009;24 Suppl 2:S307-13. doi:
358   10.3346/jkms.2009.24.S2.S307 [published Online First: 2009/06/13]
359 11. Cocchiarella L, Andersson GBJ. Guide to the Evaluation of Permanent Impairment. 5th ed.
360   ed. Chicago, Ill: Americal Medical Association 2000.
361 12. van der Wurff P, Hagmeijer RH, Meyne W. Clinical tests of the sacroiliac joint. A systematic
362   methodological review. Part 1: Reliability. *Manual therapy* 2000;5(1):30-6. doi:
363   10.1054/math.1999.0228 [published Online First: 2000/02/26]
364 13. Rondinelli RD. Changes for the new AMA Guides to impairment ratings, 6th Edition:
365   implications and applications for physician disability evaluations. *Pm r* 2009;1(7):643-
366   56. doi: 10.1016/j.pmrj.2009.04.012 [published Online First: 2009/07/25]
367 14. Norkin CC, White DJ. Measurement of joint motion: a guide to goniometry: FA Davis 2016.
368 15. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical
369   perspective. *Journal of Manual and Manipulative Therapy* 2011;19(2):91-99. doi:
370   http://dx.doi.org/10.1179/106698111X12973307659546
371 16. Hirsch BP, Webb ML, Bohl DD, et al. Improving visual estimates of cervical spine range of
372   motion. *Am J Orthop (Belle Mead NJ)* 2014;43(11):E261-5. [published Online First:
373   2014/11/08]
374 17. Whitcroft KL, Massouh L, Amirfeyz R, et al. Comparison of methods of measuring active
375   cervical range of motion. *Spine* 2010;35(19):E976-80. doi:
376   10.1097/BRS.0b013e3181cd6176 [published Online First: 2010/08/12]
377 18. Nitschke JE, Nattrass CL, Disler PB, et al. Reliability of the American Medical Association
378   guides' model for measuring spinal range of motion. Its implication for whole-person
379   impairment rating. *Spine (Phila Pa 1976)* 1999;24(3):262-8. [published Online First:
380   1999/02/20]
381 19. Lowery WD, Jr., Horn TJ, Boden SD, et al. Impairment evaluation based on spinal range of
382   motion in normal subjects. *J Spinal Disord* 1992;5(4):398-402. [published Online First:
383   1992/12/11]
384 20. Fritz JM, Piva SR. Physical impairment index: reliability, validity, and responsiveness in
385   patients with acute low back pain. *Spine (Phila Pa 1976)* 2003;28(11):1189-94. doi:
386   10.1097/01.brs.0000067270.50897.db [published Online First: 2003/06/05]
387 21. McBride ED. Disability evaluation: principles of treatment of compensable injuries. 1936.
388   *Clinical orthopaedics and related research* 2007;457:11-5. doi:
389   10.1097/BLO.0b013e31803dcab1 [published Online First: 2007/04/07]
390 22. Trudelle-Jackson E, Sarvaiya-Shah SA, Wang SS. Interrater reliability of a movement
391   impairment-based classification system for lumbar spine syndromes in patients with
392   chronic low back pain. *J Orthop Sports Phys Ther* 2008;38(6):371-6. doi:
393   10.2519/jospt.2008.2760 [published Online First: 2008/06/03]

394   23. Waddell G, Somerville D, Henderson I, et al. Objective clinical evaluation of physical
395   impairment in chronic low back pain. *Spine* 1992;17(6):617-28. [published Online First:
396   1992/06/01]
397   24. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical
398   perspective. *The Journal of manual & manipulative therapy* 2011;19(2):91-9. doi:
399   10.1179/106698111x12973307659546 [published Online First: 2012/05/02]
400   25. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic
401   reviews and meta-analyses of studies that evaluate health care interventions: explanation
402   and elaboration. *J Clin Epidemiol* 2009;62(10):e1-34. doi: 10.1016/j.jclinepi.2009.06.006
403   [published Online First: 2009/07/28]
404   26. Haldeman S, Kopansky-Giles D, Hurwitz EL, et al. Advancements in the Management of
405   Spine Disorders. *Best Practice & Research Clinical Rheumatology* 2012;26(2):263-80.
406   doi: 10.1016/j.berh.2012.03.006
407   27. Terwee CB, Jansma EP, Riphagen, II, et al. Development of a methodological PubMed
408   search filter for finding studies on measurement properties of measurement instruments.
409   *Qual Life Res* 2009;18(8):1115-23. doi: 10.1007/s11136-009-9528-5 [published Online
410   First: 2009/08/28]
411   28. Mokkink LB, Prinsen CA, Bouter LM, et al. The COnsensus-based Standards for the
412   selection of health Measurement INstruments (COSMIN) and how to select an outcome
413   measurement instrument. *Braz J Phys Ther* 2016;20(2):105-13. doi: 10.1590/bjpt-
414   rbf.2014.0143 [published Online First: 2016/01/21]
415   29. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic
416   reviews of studies on measurement properties: a scoring system for the COSMIN
417   checklist. *Qual Life Res* 2012;21(4):651-7. doi: 10.1007/s11136-011-9960-1 [published
418   Online First: 2011/07/07]
419   30. Davis MA, Onega T, Weeks WB, et al. Where the United States spends its spine dollars:
420   expenditures on different ambulatory services for the management of back and neck
421   conditions. *Spine* 2012;37(19):1693-701. doi: 10.1097/BRS.0b013e3182541f45
422   [published Online First: 2012/03/22]
423   31. Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back
424   and neck problems. *Jama* 2008;299(6):656-64. doi: 10.1001/jama.299.6.656 [published
425   Online First: 2008/02/14]
426   32. Martin BI, Turner JA, Mirza SK, et al. Trends in health care expenditures, utilization, and
427   health status among US adults with spine problems, 1997-2006. *Spine* 2009;34(19):2077-
428   84. doi: 10.1097/BRS.0b013e3181b1fad1 [published Online First: 2009/08/14]
429   33. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies
430   in the United States and internationally. *Spine J* 2008;8(1):8-20. doi:
431   10.1016/j.spinee.2007.10.005 [published Online First: 2008/01/01]
432   34. McGee R, Bevan S, Quadrello T. Fit For Work? Musculoskeletal Disorders and the Canadian
433   Labour Market. London, ON: The Work Foundation, 2009.
434   35. Dunning KK, Davis KG, Cook C, et al. Costs by industry and diagnosis among
435   musculoskeletal claims in a state workers compensation system: 1999-2004. *Am J Ind
436   Med* 2010;53(3):276-84. doi: 10.1002/ajim.20774 [published Online First: 2009/11/26]

437

438 **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. 35. Goniomet*.ab,ti. 36. Joint motion measurement*.ab,ti. 37. Electrogoniomet*.ab,ti. 38. (Movement adj2 evaluation*).ab,ti. 39. Inclinomet*.ab,ti.	60. Comparative Study/ 61. Psychometrics/ 62. psychometr*.ab,ti. 63. clinimetr*.tw. 64. clinometr*.tw. 65. "Outcome Assessment (Health Care)"/
3. low back pain/	40. Joint flexibility.ab,ti.	66. "outcome assessment".ab,ti.
4. back pain/ or neck pain/	41. Arthromet*.ab,ti.	67. ("outcome measure" or "outcome measures").tw.
5. back/ or neck/ or lumbosacral region/ or sacrococcygeal region/	42. Joint mobility.ab,ti. 43. Osteokinematic*.ab,ti.	68. Observer Variation/ 69. "observer variation".ab,ti.
6. (back or neck or "lumbosacral region" or "sacrococcygeal region").ab,ti.	44. Flexion.ab,ti. 45. Extension.ab,ti.	70. Health Status Indicators/ 71. "Reproducibility of Results"/
7. Sciatica/	46. Rotation/	72. reproducib*.ti,ab.
8. Neck Injury/ or Whiplash Injuries/	47. side bend*.ab,ti.	73. Discriminant Analysis/
9. whiplash.ab,ti.	48. Work Capacity Evaluation/	74. reliab*.ab,ti.
10. Dorsalgia.ti,ab.	49. Disability Evaluation/ and (Occupational Diseases/ or Work/ or Return to Work/)	75. unreliab*.ab,ti. 76. valid*.ab,ti.
11. coccydynia.ti,ab.	50. Occupational Diseases/ or Work/ or Return to Work/	77. coefficient.ab,ti. 78. homogeneity.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.	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.	79. homogeneous.ab,ti. 80. "internal consistency".ab,ti. 81. cronbach*.ab,ti. 82. (alpha or alphas).ab,ti. 83. 81 and 82
13. discitis.ti,ab.	52. 50 and 51	84. item.ab,ti.
14. (disc adj degeneration).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.	85. (correlation* or selection* or reduction*).ab,ti.
15. (disc adj prolapse).ti,ab.		86. 84 and 85
16. (disc adj herniation).ti,ab.		
17. (facet adj joints).ti,ab.	54. (evaluation* or assessment* or instrument* or measurement* or tool* or scale* or questionnaire* or test* or determination*).ab,ti.	87. agreement.ab,ti. 88. precision.ab,ti. 89. imprecision.ab,ti.
18. Intervertebral Disc/ or annulus fibrosus/ or nucleus pulposus/		90. "precise values".ab,ti.
19. arachnoiditis.ti,ab.	55. 53 and 54	91. test-retest.ab,ti.
20. Spinal Fusion/	56. "permanent functional impairment".ab,ti.	92. test.ab,ti.
21. postlaminectomy.ti,ab.		93. retest.ab,ti.
22. Backache*.ti,ab.	57. ("work performance evaluation" or "work performance evaluations").ab,ti.	94. 92 and 93
23. back injuries/ or spinal injuries/ or spinal fractures/	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	95. reliab*.ab,ti.
24. Spondylitis, Ankylosing/		96. (test or retest).ab,ti.
25. Spondylitis/		
26. spondylosis/ or spondylolysis/ or		

spondylolisthesis/

27. Spinal Cord Compression/

28. (Musculoskeletal adj (symptom* or injur* or disorder* or pain or dysfunction* or problem*)).ab,ti.

29. ((orthopedic or orthopaedic) adj (injur* or problem* or disorder* or dysfunction*)).ab,ti.

30. pelvis/ or lesser pelvis/ or pelvic floor/

31. (pelvi* adj3 (symptom* or injur* or disorder* or pain or dysfunction* or problem*)).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

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. inter-technician.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.

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.

18

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.

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

439   PFI: Permanent Functional Impairment

440

441

442 **Table 2 - Draft screening tool for use at the title, abstract, and full-text review screening**
443 **stages.**

Title, Abstract and Full-text screening questions

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 two

2) Does the study deal with humans?
    a) No – exclude
    b) Yes or uncertain – go to step three

3) Does the study deal with adults (between 18 and 65 years of age)?
    a) No – exclude
    b) Yes or uncertain – go to step four

4) Does the article represent primary study (i.e. no letters to the editor, book reviews, published study designs or trial protocols)
    a) No – exclude
    b) Yes or uncertain – go to step five

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 six

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 seven

7) Is the study designed to evaluate measurement properties of measurement instruments/tools (e.g. validity, reliability, responsiveness)?
    a) No – exclude
    b) Yes or uncertain – choose one of the following options:
        (1) Title and abstract screening stage – Include
        (2) Full-text screening stage – go to step eight

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 – Include

---

444 **Note:** Exclusion occurs if the answer to any of the questions is "no".

445

446

447 **Table 3 - Draft extraction tool**

Item	Description	Definitions, decision rules, guidance, and example data
Reference ID	Identifies the article	Author, date
Instrument name	Identifies the PFI assessment instrument	The name as described in the article, or as referred to in other articles (e.g. 'Back-EST' or 'the Johnson method')
Instrument description	Describes the PFI assessment instrument	Will collect sub variables, such as: * Equipment required Duration Clinicians performing test (e.g. physiotherapist, occupational therapist, physician) Dimensions assessed (e.g. Range of motion, lifting capacity, etc.)
Validity-criterion	The degree to which the scores of an Health Related-Patient Reported Outcomes instrument are an **adequate reflection of a 'gold standard'** (as per COSMIN tool)	Will collect sub variables, such as: * Did the article assess this dimension? (yes/no) What was the inferential test used? (e.g. intraclass correlation coefficient, kappa) What was the result? (e.g. kappa score of 0.7)
Validity-content	The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured (as per COSMIN tool)	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	The degree to which the scores of an HR-PRO 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 the assumption that the HR-PRO instrument **validly measures the construct to be measured** (as per COSMIN tool)	Will collect sub variables, such as: * Did the article assess this dimension? (yes/no) How was construct validity assessed? What were the findings?
Reliability – inter-rater	Degree of agreement between raters investigating the same property on the same patient	Will collect sub variables, such as: * Did the article assess this dimension? (yes/no) What was the inferential test used? (e.g. intraclass correlation coefficient, kappa) What was the result? (e.g. kappa score of 0.7)
Reliability – intra-rater	Degree **of agreement between repeated measurements of a property on the same patient by the same rater**	Will collect sub variables, such as: * Did the article assess this dimension? (yes/no) What was the inferential test used? (e.g. intraclass correlation coefficient, kappa) What was the result? (e.g. kappa score of 0.7)
Responsiveness	The ability of an HR-PRO instrument to **detect change over time in the construct to be measured** (as per COSMIN tool)	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	There are multiple binary variables over several domains: **hypothesis testing, clarity of instruments reporting, and multiple study design elements** (as per COSMIN tool)	Will adhere to the COSMIN tool, do not anticipate altering the definitions

448   PFI: Permanent Functional Impairment; COSMIN: COnsensus-based Standards for the selection
449   of health Measurement Instruments; HR-PRO: health related patient-reported outcomes.
450   **Note:** These Items may evolve as review progresses and studies are identified. This draft data
451   dictionary describes the selected data extraction variables of the systematic review. The data
452   dictionary would be used to extract data from articles identified as relevant during the screening
453   process. The data extraction database itself would have these items listed in columns with the
454   individual references representing rows such that these data are extracted for every article to
455   form a summary database for later synthesis. Only key variables are presented here. It is
456   anticipated that this list will expand based on the type of data presented in the articles; the
457   definitions guide will also become more specific as a result of independent data extraction and
458   discussion at team meetings.

459

1

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

Manuscript Title: Measurement properties of instruments assessing permanent functional impairment of the spine: a systematic review protocol

Section and topic	Item No	Checklist item	Information reported		Page, Line number(s)
			Yes	N/A	
ADMINISTRATIVE INFORMATION					
Title					
Identification	1a	Identify the report as a protocol of a systematic review	☒	☐	Page 1, Line 6
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	☐	☒	N/A
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	☒	☐	Page 3, Line 70
Authors					
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	☒	☐	Pages 2-3, Lines 10-42
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	☒	☐	Page 13, Lines 325-328
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	☒	☐	Page 15, Lines 322-323
Sponsor	5b	Provide name for the review funder and/or sponsor	☒	☐	Page 13, Lines 329-331
Role of	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the	☒	☐	Page 13, Lines

2

Section and topic	Item No	Checklist item	Information reported		Page, Line number(s)
			Yes	N/A	
sponsor/funder		protocol			332-334
INTRODUCTION					
Rationale	6	Describe the rationale for the review in the context of what is already known	☒	☐	Pages 4-5, Lines 81-131
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	☒	☐	Page 5, Lines 131-136
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	☒	☐	Pages 6-7, Lines 159-182
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	☒	☐	Pages 7-8, Lines 184-203
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	☒	☐	Pages 16-19, "Table 1"
Study Records					
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	☒	☐	Page 8 , Lines 199-200
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)	☒	☐	Page 9, Lines 205-216
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	☒	☐	Pages 9-10, Lines 218-260
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	☒	☐	Pages 21-22, "Table 3"

3

Section and topic	Item No	Checklist item	Information reported		Page, Line number(s)
			Yes	N/A	
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	☒	☐	Page 6, Lines 145-148
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	☒	☐	Pages 10-11, Lines 261-274
Data Synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	☒	☐	Page 11, Lines 275-288
	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)	☒	☐	Page 11, Lines 276-285
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	☒	☐	Page 11, Lines 276-288
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	☒	☐	Page 11, Lines 283-285
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	☒	☐	Page 11, Lines 289-293
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	☒	☐	Page 11, Lines 285-287

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.

BMJ Open

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

SCHOLARONE™
Manuscripts

1 **TITLE PAGE**

2

3 **Title:**

4

5 Measurement properties of instruments assessing permanent functional impairment of the spine:
6 a systematic review protocol

7

8 **Authors, affiliations:**

9

10 **Suelen Meira Goes**: suelen.goes@usask.ca; School of Physical Therapy, University of
11 Saskatchewan - Health Sciences Building, E-Wing; Suite 3400 104 Clinic Place - Saskatoon,
12 Saskatchewan, S7N 2Z4 – Canada. *** Corresponding Author**

13

14 **Catherine Trask**: catherine.trask@usask.ca; Canadian Centre for Health and Safety in
15 Agriculture (CCHSA), University of Saskatchewan - 103 Hospital Drive, Saskatoon,
16 Saskatchewan, S7N 0W8 – Canada.

17

18 **Catherine Boden**: catherine.boden@usask.ca; Leslie and Irene Dubé Health Sciences Library,
19 Academic Health Sciences Building - 104 Clinic Place, Saskatoon, Saskatchewan, S7N 2Z4 –
20 Canada.

21

22 **Brenna Bath**: brenna.bath@usask.ca; School of Physical Therapy, University of Saskatchewan -
23 Health Sciences Building, E-Wing; Suite 3400 104 Clinic Place - Saskatoon, Saskatchewan, S7N
24 2Z4 – Canada.

25

26 **Daniel Cury Ribeiro**: daniel.ribeiro@otago.ac.nz; School of Physiotherapy, University of Otago
27 - 325 Great King Street, Dunedin, 9016 – New Zealand.

28

29 **Paul Hendrick**: paul.hendrick@nottingham.ac.uk; School of Health Sciences, University of
30 Nottingham – Clinical Sciences Building - Nottingham City Hospital - Hucknall Rd Nottingham,
31 NG5 1PB – United Kingdom.

32

33 **Lynne Clay**: lynne.clay@otago.ac.nz; School of Physiotherapy, University of Otago - 325 Great
34 King Street, Dunedin, 9016 – New Zealand.

35

36 **Xiaoke Zeng**: xiaoke.zeng@usask.ca; Canadian Centre for Health and Safety in Agriculture
37 (CCHSA), University of Saskatchewan - 103 Hospital Drive, PO Box 120, Saskatoon,
38 Saskatchewan, S7N 0W8 – Canada.

39

40 **Stephan Milosavljevic**: stephan.milosavljevic@usask.ca; School of Physical Therapy,
41 University of Saskatchewan - Health Sciences Building, E-Wing; Suite 3400 104 Clinic Place -
42 Saskatoon, Saskatchewan, S7N 2Z4- Canada.

43

44 **Word count:** 260 (abstract) + 2402 (protocol)

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 synthesize 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 (COSMIN) criteria. Findings will be summarized descriptively, with meta-analysis pursued as appropriate.

Ethics and Dissemination: This review will summarize 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 organizations. The findings will likely provide a foundation and direction for future research priorities for assessing spinal PFI.

Keywords: disability, back, range of movement, psychometrics, validity, reliability, responsiveness

Systematic review registration: PROSPERO CRD42017060390.

Strengths and limitations of this study

- This systematic review will include a broad range of instruments used to assess PFI in individuals with spinal conditions in peer-reviewed articles.

- This systematic review protocol is presented in accordance with the Preferred Reporting Items for Systematic review and Meta-Analyses Protocols (PRISMA-P).

- A strength of this review is the use of the internationally recognised, validated COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) 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.

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 considerable economic, personal, and societal burden [1]. Internationally, occupational low back pain is the singular most prevalent health problem related to absenteeism amongst 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:

111 AMA Guides' Diagnosis-Related Estimate [11], Diagnosis-Based Impairment [5], McBride's method

112 of spinal impairment evaluation [19], or Physical Impairment Index [18 20 21]. However, these

113 alternative systems for rating PFI present poor descriptions of standardization and normative

114 values [10 22].

115 The outcome of spinal PFI assessment by compensation bodies can have considerable social,

116 economic and health impacts, yet the reliability and validity of instruments currently available

117 have not been systematically evaluated. To achieve a fair and accurate outcome, clinicians,

118 researchers and government bodies must have access to the most accurate level of evidence

119 regarding methods assessing spinal PFI. To our knowledge, no previous review has assessed the

120 reliability and validity of available instruments for determining PFI of the spine. This manuscript

121 presents the protocol of an ongoing systematic review with the objective to review and

122 synthesize the literature concerning measurement properties of the instruments used for assessing

123 PFI of the spine. Due to the diversity of instruments that might be used to assess PFI of the spine,

124 we anticipate the identification of distinct metrics for measuring PFI will allow specific sub-

125 group analyses for review and ultimate discussion on the strength of support for each instrument.

126

127 **METHODS**

128 **Design**

129 A systematic review protocol of the measurement properties (validity, reliability, and

130 responsiveness) of the different instruments identified for assessing spinal PFI. Reporting will

131 adhere to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA)

132 guidelines using the PICOS methodology [23]. This includes: **P**opulation (individuals aged

133 between 18 to 65 years with spinal conditions not caused by congenital and developmental

134 abnormalities, neoplasm, infection disorders, and systemic inflammatory disorders);

135 **I**ntervention/exposure (measurement instruments used to assess PFI of the spine); **C**omparator

136 (not applicable for this review); **O**utcome (reliability, validity and responsiveness); and **S**tudy

137 design (cross-sectional and cohort studies reporting on measurement properties).

Eligibility Criteria

139 Following review of current World Health Organisation ICF classifications [3] and the American

140 Medical Association guides [5] for evaluation of permanent impairment, we have defined PFI as

141 any long-lasting disturbance of physiological structure and function, that leads to an inability to

142 perform a normal range of basic daily living and/or work-related activities. Furthermore, the

143 assessment of PFI needs to incorporate measures of impairment, functional impairment, and

144 permanent impairment. Impairment is any problem in body function and structure; functional

145 impairment is described as restrictions in performance of the body systems [3], and permanent

146 impairment is determined when the impairment becomes well-stabilized with or without

147 treatment, and further significant medical improvement is unlikely [5]. While evaluating PFI,

148 studies should include measures of impairment (i.e. RoM, muscle strength, coordination,

149 endurance, and sensation), functional limitation (i.e. self-report instruments of physical function

150 and functional performance measures) and permanency of the impairment (i.e. duration of the

151 impairment and the likelihood of improvement) [3 24].

152 We have elected not to limit publication range in order to be thorough in collecting documented

153 evaluations of measurement properties of instruments assessing PFI of the spine and thus we will

154 include peer-reviewed, full-text articles over the fully available date range. We will focus on

155 cross-sectional and cohort studies investigating measurement instruments for assessing all

156 components described as appraising PFI, as well as assessing the measurement properties of

157 these assessment instruments in terms of validity, reliability, and responsiveness. We will

158 concentrate on studies of adults with spinal disorders between 18 and 65 years of age.

159 We will exclude letters to the editor, book reviews, and short communications. We will also

160 exclude clinical protocols, case reports and series, systematic reviews, meta-analysis, articles not

161 published in English-language, studies intended for screening, diagnosis and prognosis of spinal

162 pathologies as well as studies with specimen-, cadaveric-, cellular-, artificial-, and computer-

163 based models. In addition, given the broad scope of spinal disorders and considering that workers

164 compensation boards do not uniformly specify or universally provide financial compensation for

165 all specific diseases and non-specific musculoskeletal disorders, we will exclude studies of

166 individuals with spinal conditions caused by congenital and developmental abnormalities,

6

167 neoplasm, infection disorders, and systemic inflammatory disorders [25]. Exclusion of studies will

168 take place at the screening stage instead of in the literature search phase to avoid the risk of

169 excluding relevant articles, except for the non-articles published in journals.

Information Sources

171 Studies will be identified by searching electronic databases, scanning reference lists of articles

172 and consultation with experts in the area. A preliminary search plan was developed and pre-

173 tested by a professional health sciences librarian (CB). The following databases were searched:

174 Medline – OVID (1946 to April 20, 2017); EMBASE – OVID (1947 to April 21, 2017);

175 CINAHL – Cumulative Index to Nursing and Allied Health Literature (1937 to April 25, 2017);

176 Web of Science (1900 to April 21, 2017); Scopus (1995 to April 25, 2017); PEDro (Inception to

177 April 22, 2017); OTSeeker (Inception to April 22, 2017); and Health and Safety Science

178 Abstracts – ProQuest (1988 to April 22, 2017).

Search strategy

180 To capture all measurement instruments used to evaluate PFI of the spine, our search strategy

181 will focus on combining three conceptual groups of terms: (1) PFI, (2) spinal disorder, and (3)

182 measurement properties. The Medline search strategy is presented in Table 1.

183 The terms for the concept of "measurement properties" were adapted from a search filter

184 validated and published by Terwee et al [26] and optimized for Medline. A similar search will be

185 employed for other databases, optimized for database-specific search interfaces. A filter to

186 exclude certain publication types (e.g., addresses, editorials, letters, and newspaper articles) will

187 be applied, as recommended by Terwee et al [26]. References will be imported into DistillerSR

188 software (Distiller SR Evidence Partners Limited, Ottawa, Canada). All duplicate articles will be

189 removed prior to screening. Reference lists of included studies will be scrutinized during the

190 process of data extraction for additional articles meeting our criteria. Identified citations will be

191 downloaded and subjected to the same de-duplication and screening process.

192 Table 1 about here

193 **Study Selection**

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

199 A screening tool will be developed and piloted for the title, abstract and full-text screening stages
200 (see table 2). Inclusion/exclusion will be determined using the same tool throughout the process,
201 although the definitions may be refined as conflicts arise and are discussed. The same questions
202 will be used for both title and abstract as well as full-text screening, with a focus on sensitivity
203 rather than specificity at these stages.

204 | Table 2 about here |

205 **Data extraction process**

206 We will develop data extraction forms, as outlined in table 3, which will be pilot tested on three
207 studies not included in the review. Two reviewers will extract data independently. Any
208 disagreement will be resolved through a third party adjudication. The nature of performance
209 measures employed by included studies will drive the nature of the data extraction, as we
210 anticipate that there may be some heterogeneity in the evaluation of measurement instruments.

211 | Table 3 about here |

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

218 **Data Items**

219 We will extract information relating to (1) the measurement properties (i.e. reliability, validity,

220 and responsiveness) of the measurement instruments for assessing PFI of the spine (e.g.

221 assessment of RoM, muscle strength, coordination, endurance, and sensation, as well as ability to

222 perform basic daily living and work activities, and duration of the impairment); (2) the type of

223 inferential test; and (3) the test results for each article. The extraction process will be tracked

224 using a standardized data extraction form.

225 Articles will be separated into sub-groups based on the instruments used to assess PFI of the

226 spine. The different sub-groups will be analyzed separately based on impairment-based measures

227 (i.e. RoM, muscle strength, coordination, endurance, and sensation) and functional limitation

228 assessment (i.e. self-report instruments of physical function and functional performance

229 measures). For clarity, we are likely to isolate articles in sub-groups to better differentiate the

230 instruments used to assess PFI of the spine, especially when considering that RoM has

231 predominantly been used for assessing impairment of the spine [5 6 9 11].

232 **Quality assessment of individual studies**

233 Included studies in each sub-group will be appraised independently by two reviewers as to their

234 methodological quality using the consensus-based standards for the selection of health

235 measurement instruments (COSMIN) criteria [27 28]. Any disagreement will be resolved by

236 consultation with a third member of the research team. The COSMIN checklist is a consensus-

237 based tool designed to evaluate the methodological quality of studies investigating measurement

238 properties. The instrument shows appropriate levels of agreement [27] and, based on its content

239 validity, is a recommended tool for assessing the methodological quality of studies evaluating

240 measurement properties of outcome measures within a systematic review [28]. The tool evaluates

241 the following measurement constructs: reliability; measurement error; content validity; structural

242 validity; hypotheses testing; cross-cultural validity; and criterion validity. Responsiveness and

243 interpretability with five to18 items concerning methodological standards for how each

244 measurement property should be assessed (see Table 4 for definitions). The methodological

245 quality of a study will be considered adequate if all items in a measurement property are

246 considered adequate. Each item is scored on a four-point rating scale (i.e., ''poor'', ''fair'',

247 ''good'', or ''excellent'').

248 | Table 4 about here

Planned methods of analysis

250 A narrative synthesis will be presented in text and table formats, with the intent of summarizing

251 and discussing the sample and methodological aspects, as well as the findings regarding

252 measurement properties of the included studies assessing PFI in individuals with spinal

253 conditions. Tables will provide general information of the studies (i.e. authors, country, and

254 population parameters, such as age, gender distribution, setting etc.), and will summarize

255 reliability, validity and responsiveness data with associated study quality indicators (COSMIN

256 checklist).

257 In order to determine the best available method for measuring PFI of the spine, each identified

258 instrument will be ranked using a range of measurement performance metrics identified in the

259 COSMIN checklist. The findings will be presented and possible hypotheses for the results will

260 be generated and discussed. In addition, gaps in the literature will be identified and discussed.

261 Where appropriate, results will be statistically summarized (i.e. meta-analysis) into forest plots

262 with estimates of heterogeneity; and sensitivity analysis will be pursued by comparing results

263 from studies with high and low-quality ranking. Where possible, we will weight a meta-analysis

264 using both the study's sample size and their quality assessment as determined by the COSMIN

265 checklist [28]. However, we predict some heterogeneity will be identified in the various

266 assessments of spinal PFI measures, which will likely make a meta-analysis difficult to apply.

Risk of bias across studies

268 Since protocol studies that describe forthcoming evaluation of measurement instruments (e.g.

269 validity, reliability, and responsiveness) are scarce, we will verify whether the pre-specified

270 primary outcomes have been reported, and contact authors when we lack important data or

271 information from the included study.

ETHICS AND DISSEMINATION

273 A substantial number and variety of spine impairment ratings are calculated internationally and
274 used for financial award decisions, in both legal and compensation contexts [29-34]. However, a
275 systematic review in this area to support the use of instruments currently being used to determine
276 PFI is lacking. The application of reliable and valid measurement instruments for assessing PFI
277 of the spine is considered crucial.

278 This systematic review offers a feasible means for synthesizing the evidence specific to spinal
279 PFI assessment; and our results will likely provide unique insights concerning the breadth and
280 depth of literature in the area. Outcomes of this review will be applicable to clinicians, policy-
281 makers, worker's compensation boards and health and safety organizations. In particular,
282 findings will likely provide a foundation and direction in terms of research priorities for
283 assessing PFI of the spine. Summarizing the nature and strength of the evidence regarding the
284 reliability, validity and responsiveness of spinal PFI measures will also inform future research
285 and policy in this field.

286

LIST OF ABBREVIATIONS

288 AMA: American Medical Association

289 CINAHL: Cumulative Index to Nursing and Allied Health Literature

290 COSMIN: COnsensus-based Standards for the selection of health Measurement Instruments

291 Embase: Excerpta Medica Database

292 HR-PRO: health related patient-reported outcomes.

293 ICF: International Classification of Functioning, Disability and Health

294 OTseeker: Occupational Therapy Systematic Evaluation of Evidence

295   PEDro: Physiotherapy Evidence Database

296   PFI: Permanent Functional Impairment

297   PICOS: Population; Interventions or exposure; Comparator; Outcomes or endpoint; Study design

298   PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

299   RoM: Range of Motion

300   WHO: World Health Organization

301

302   **AUTHORS' CONTRIBUTIONS**

303   SMG and SM conceptualized the initial review protocol and led the development of this

304   manuscript. The entire team (SMG, SM, CB, CT, BB, DCR, PH, LC, and XZ) collaboratively

305   drafted the manuscript, followed by numerous iterations and substantial input and appraisal from

306   all authors. All authors have approved the final version of this manuscript.

307

308   **FUNDING**

309   This research is funded by The Workers' Compensation Board of British Columbia

310   (*WorkSafeBC grant number RS2015-SR01*).

311

312   **COMPETING INTERESTS**

313   The views, findings, opinions and conclusion expressed herein do not necessarily represent the

314   views of the funder.

315 **REFERENCES**

316 1. Driscoll T, Jacklyn G, Orchard J, et al. The global burden of occupationally related low back
317   pain: estimates from the Global Burden of Disease 2010 study. *Annals of the rheumatic*
318   *diseases* 2014;73(6):975-81. doi: 10.1136/annrheumdis-2013-204631 [published Online
319   First: 2014/03/26]
320 2. Helfenstein M, Goldenfum MA, Siena C. OCCUPATIONAL LOW BACK PAIN. *Revista Da*
321   *Associacao Medica Brasileira* 2010;56(5):583-89.
322 3. WHO. A Practical Manual for using the International Classifcation of Functioning, Disability
323   and Health (ICF). Geneva: World Health Organization, 2013.
324 4. WHO. World report on disability. Geneva, 2011.
325 5. Rondinelli RD, Genovese E, Brigham CR. American Medical Association. Guides to the
326   Evaluation of Permanent Impairment. 6th ed. Chicago, IL: American Medical
327   Association 2008.
328 6. Comcare. Guide to the assessment of the degree of permanet impairment. 2.1 ed. Australia,
329   2014.
330 7. WorkCover. Guidelines for the assessment of permanent impairment. 3rd ed. Tasmania, 2011.
331 8. SafeWork. Permanent Functional Impairment Rating Schedule. Newfoundland and Labrador -
332   Canada: Workplace Health, Safety and Compensation Commission, 2009.
333 9. WorkSafeBC. Permanent Disability Evaluation Schedule ("PDES") 2016 Review. British
334   Columbia - Canada, 2016.
335 10. Kim YB, Lee SG, Park CW, et al. Korean guideline development for the evaluation of
336   permanent impairment of the spine: proposal by the Korean Academy of Medical
337   Sciences Committee. *J Korean Med Sci* 2009;24 Suppl 2:S307-13. doi:
338   10.3346/jkms.2009.24.S2.S307 [published Online First: 2009/06/13]
339 11. Cocchiarella L, Andersson GBJ. Guide to the Evaluation of Permanent Impairment. 5th ed.
340   ed. Chicago, Ill: Americal Medical Association 2000.
341 12. van der Wurff P, Hagmeijer RH, Meyne W. Clinical tests of the sacroiliac joint. A systematic
342   methodological review. Part 1: Reliability. *Manual therapy* 2000;5(1):30-6. doi:
343   10.1054/math.1999.0228 [published Online First: 2000/02/26]
344 13. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical
345   perspective. *Journal of Manual and Manipulative Therapy* 2011;19(2):91-99. doi:
346   http://dx.doi.org/10.1179/106698111X12973307659546
347 14. Hirsch BP, Webb ML, Bohl DD, et al. Improving visual estimates of cervical spine range of
348   motion. *Am J Orthop (Belle Mead NJ)* 2014;43(11):E261-5. [published Online First:
349   2014/11/08]
350 15. Whitcroft KL, Massouh L, Amirfeyz R, et al. Comparison of methods of measuring active
351   cervical range of motion. *Spine* 2010;35(19):E976-80. doi:
352   10.1097/BRS.0b013e3181cd6176 [published Online First: 2010/08/12]
353 16. Nitschke JE, Nattrass CL, Disler PB, et al. Reliability of the American Medical Association
354   guides' model for measuring spinal range of motion. Its implication for whole-person
355   impairment rating. *Spine (Phila Pa 1976)* 1999;24(3):262-8. [published Online First:
356   1999/02/20]
357 17. Lowery WD, Jr., Horn TJ, Boden SD, et al. Impairment evaluation based on spinal range of
358   motion in normal subjects. *J Spinal Disord* 1992;5(4):398-402. [published Online First:
359   1992/12/11]

360   18. Fritz JM, Piva SR. Physical impairment index: reliability, validity, and responsiveness in
361   patients with acute low back pain. *Spine (Phila Pa 1976)* 2003;28(11):1189-94. doi:
362   10.1097/01.brs.0000067270.50897.db [published Online First: 2003/06/05]
363   19. McBride ED. Disability evaluation: principles of treatment of compensable injuries. 1936.
364   *Clinical orthopaedics and related research* 2007;457:11-5. doi:
365   10.1097/BLO.0b013e31803dcab1 [published Online First: 2007/04/07]
366   20. Trudelle-Jackson E, Sarvaiya-Shah SA, Wang SS. Interrater reliability of a movement
367   impairment-based classification system for lumbar spine syndromes in patients with
368   chronic low back pain. *J Orthop Sports Phys Ther* 2008;38(6):371-6. doi:
369   10.2519/jospt.2008.2760 [published Online First: 2008/06/03]
370   21. Waddell G, Somerville D, Henderson I, et al. Objective clinical evaluation of physical
371   impairment in chronic low back pain. *Spine* 1992;17(6):617-28. [published Online First:
372   1992/06/01]
373   22. Rondinelli RD. Changes for the new AMA Guides to impairment ratings, 6th Edition:
374   implications and applications for physician disability evaluations. *Pm r* 2009;1(7):643-
375   56. doi: 10.1016/j.pmrj.2009.04.012 [published Online First: 2009/07/25]
376   23. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic
377   reviews and meta-analyses of studies that evaluate health care interventions: explanation
378   and elaboration. *J Clin Epidemiol* 2009;62(10):e1-34. doi: 10.1016/j.jclinepi.2009.06.006
379   [published Online First: 2009/07/28]
380   24. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical
381   perspective. *The Journal of manual & manipulative therapy* 2011;19(2):91-9. doi:
382   10.1179/106698111x12973307659546 [published Online First: 2012/05/02]
383   25. Haldeman S, Kopansky-Giles D, Hurwitz EL, et al. Advancements in the Management of
384   Spine Disorders. *Best Practice & Research Clinical Rheumatology* 2012;26(2):263-80.
385   doi: 10.1016/j.berh.2012.03.006
386   26. Terwee CB, Jansma EP, Riphagen, II, et al. Development of a methodological PubMed
387   search filter for finding studies on measurement properties of measurement instruments.
388   *Qual Life Res* 2009;18(8):1115-23. doi: 10.1007/s11136-009-9528-5 [published Online
389   First: 2009/08/28]
390   27. Mokkink LB, Prinsen CA, Bouter LM, et al. The COnsensus-based Standards for the
391   selection of health Measurement INstruments (COSMIN) and how to select an outcome
392   measurement instrument. *Braz J Phys Ther* 2016;20(2):105-13. doi: 10.1590/bjpt-
393   rbf.2014.0143 [published Online First: 2016/01/21]
394   28. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic
395   reviews of studies on measurement properties: a scoring system for the COSMIN
396   checklist. *Qual Life Res* 2012;21(4):651-7. doi: 10.1007/s11136-011-9960-1 [published
397   Online First: 2011/07/07]
398   29. Davis MA, Onega T, Weeks WB, et al. Where the United States spends its spine dollars:
399   expenditures on different ambulatory services for the management of back and neck
400   conditions. *Spine* 2012;37(19):1693-701. doi: 10.1097/BRS.0b013e3182541f45
401   [published Online First: 2012/03/22]
402   30. Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back
403   and neck problems. *Jama* 2008;299(6):656-64. doi: 10.1001/jama.299.6.656 [published
404   Online First: 2008/02/14]

405 31. Martin BI, Turner JA, Mirza SK, et al. Trends in health care expenditures, utilization, and
406   health status among US adults with spine problems, 1997-2006. *Spine* 2009;34(19):2077-
407   84. doi: 10.1097/BRS.0b013e3181b1fad1 [published Online First: 2009/08/14]
408 32. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies
409   in the United States and internationally. *Spine J* 2008;8(1):8-20. doi:
410   10.1016/j.spinee.2007.10.005 [published Online First: 2008/01/01]
411 33. McGee R, Bevan S, Quadrello T. Fit For Work? Musculoskeletal Disorders and the Canadian
412   Labour Market. London, ON: The Work Foundation, 2009.
413 34. Dunning KK, Davis KG, Cook C, et al. Costs by industry and diagnosis among
414   musculoskeletal claims in a state workers compensation system: 1999-2004. *Am J Ind*
415   *Med* 2010;53(3):276-84. doi: 10.1002/ajim.20774 [published Online First: 2009/11/26]
416

417 **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/
	34. range of motion.ab,ti.	60. Comparative Study/
	35. Goniomet*.ab,ti.	61. Psychometrics/
	36. Joint motion measurement*.ab,ti.	62. psychometr*.ab,ti.
	37. Electrogoniomet*.ab,ti.	63. clinimetr*.tw.
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.	38. (Movement adj2 evaluation*).ab,ti.	64. clinometr*.tw.
	39. Inclinomet*.ab,ti.	65. "Outcome Assessment (Health Care)"/
	40. Joint flexibility.ab,ti.	66. "outcome assessment".ab,ti.
	41. Arthromet*.ab,ti.	67. ("outcome measure" or "outcome measures").tw.
3. low back pain/	42. Joint mobility.ab,ti.	68. Observer Variation/
4. back pain/ or neck pain/	43. Osteokinematic*.ab,ti.	69. "observer variation".ab,ti.
5. back/ or neck/ or lumbosacral region/ or sacrococcygeal region/	44. Flexion.ab,ti.	70. Health Status Indicators/
	45. Extension.ab,ti.	71. "Reproducibility of Results"/
6. (back or neck or "lumbosacral region" or "sacrococcygeal region").ab,ti.	46. Rotation/	72. reproducib*.ti,ab.
	47. side bend*.ab,ti.	73. Discriminant Analysis/
7. Sciatica/	48. Work Capacity Evaluation/	74. reliab*.ab,ti.
8. Neck Injury/ or Whiplash Injuries/	49. Disability Evaluation/ and (Occupational Diseases/ or Work/ or Return to Work/)	75. unreliab*.ab,ti.
9. whiplash.ab,ti.		76. valid*.ab,ti.
10. Dorsalgia.ti,ab.	50. Occupational Diseases/ or Work/ or Return to Work/	77. coefficient.ab,ti.
11. coccydynia.ti,ab.		78. homogeneity.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.	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.	79. homogeneous.ab,ti.
		80. "internal consistency".ab,ti.
		81. cronbach*.ab,ti.
		82. (alpha or alphas).ab,ti.
	52. 50 and 51	83. 81 and 82
13. discitis.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.	84. item.ab,ti.
14. (disc adj degeneration).ti,ab.		85. (correlation* or selection* or reduction*).ab,ti.
15. (disc adj prolapse).ti,ab.		
16. (disc adj herniation).ti,ab.		86. 84 and 85
17. (facet adj joints).ti,ab.	54. (evaluation* or assessment* or instrument* or measurement* or tool* or scale* or questionnaire* or test* or determination*).ab,ti.	87. agreement.ab,ti.
18. Intervertebral Disc/ or annulus fibrosus/ or nucleus pulposus/		88. precision.ab,ti.
		89. imprecision.ab,ti.
19. arachnoiditis.ti,ab.		90. "precise values".ab,ti.
20. Spinal Fusion/	55. 53 and 54	91. test-retest.ab,ti.
21. postlaminectomy.ti,ab.	56. "permanent functional impairment".ab,ti.	92. test.ab,ti.
22. Backache*.ti,ab.		93. retest.ab,ti.
23. back injuries/ or spinal injuries/ or spinal fractures/	57. ("work performance evaluation" or "work performance evaluations").ab,ti.	94. 92 and 93
24. Spondylitis, Ankylosing/	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	95. reliab*.ab,ti.
25. Spondylitis/		96. (test or retest).ab,ti.
26. spondylosis/ or spondylolysis/ or		

spondylolisthesis/

27. Spinal Cord Compression/

28. (Musculoskeletal adj (symptom* or injur* or disorder* or pain or dysfunction* or problem*)).ab,ti.

29. ((orthopedic or orthopaedic) adj (injur* or problem* or disorder* or dysfunction*)).ab,ti.

30. pelvis/ or lesser pelvis/ or pelvic floor/

31. (pelvi* adj3 (symptom* or injur* or disorder* or pain or dysfunction* or problem*)).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

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. inter-technician.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.

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.

170. "computer adaptive testing".ab,ti.

171. "item bank".ab,ti.

172. "cross-cultural equivalence".ab,ti.

173. ("gold standard" or "criterion

18

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

418   PFI: Permanent Functional Impairment

419

420

421 **Table 2 - Draft screening tool for use at the title, abstract, and full-text review screening**
422 **stages**
423

Title, Abstract and Full-text screening questions
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 two

2) Does the study deal with humans?
   a) No – exclude
   b) Yes or uncertain – go to step three

3) Does the study deal with adults (between 18 and 65 years of age)?
   a) No – exclude
   b) Yes or uncertain – go to step four

4) Does the article represent primary study (i.e. no letters to the editor, book reviews, published study designs or trial protocols)
   a) No – exclude
   b) Yes or uncertain – go to step five

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 six

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 seven

7) Is the study designed to evaluate measurement properties of measurement instruments/tools (e.g. 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 eight

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 nine

9) Does this study have clinical measures of impairment (e.g. range of motion, muscle performance) or objective measures of function (e.g. functional impairment evaluation)?
   a) No – exclude
   b) Yes or uncertain – go to step ten

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 eleven

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

424 **Note:** Exclusion occurs if the answer to any of the questions is "no".
425

426 **Table 3 - 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 (e.g. 'Back-EST' or 'the Johnson method')
Instrument description	Will collect sub variables, such as: * Equipment required Duration Clinicians performing test (e.g. physiotherapist, occupational therapist, physician) Dimensions assessed (e.g. Range of motion, lifting capacity, etc.)
Validity- criterion	Will collect sub variables, such as: * Did the article assess this dimension? (yes/no) What was the inferential test used? (e.g. intraclass correlation coefficient, kappa) What was the result? (e.g. 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 sub variables, 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? (e.g. intraclass correlation coefficient, kappa) What was the result? (e.g. kappa score of 0.7)
Reliability – intra-rater	Will collect sub variables, such as: * Did the article assess this dimension? (yes/no) What was the inferential test used? (e.g. intraclass correlation coefficient, kappa) What was the result? (e.g. 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

427 PFI: Permanent Functional Impairment; COSMIN: COnsensus-based Standards for the selection of health
428 Measurement Instruments; **Note:** These Items may evolve as review progresses and studies are identified. This draft
429 data dictionary describes the selected data extraction variables of the systematic review. The data dictionary would
430 be used to extract data from articles identified as relevant during the screening process. The data extraction database
431 itself would have these items listed in columns with the individual references representing rows such that these data
432 are extracted for every article to form a summary database for later synthesis. Only key variables are presented here.
433 It is anticipated that this list will expand based on the type of data presented in the articles; the definitions guide will
434 also become more specific as a result of independent data extraction and discussion at team meetings.

435 **Table 4** – COSMIN definitions of domains, measurement properties, and aspects of
436 measurement properties
437

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
	Intra-rater		The degree of agreement between repeated measurements of a property on the same patient by the same rater
	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
Validity			How well the instrument under evaluation measures the construct it purports to measure
	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

438 **COSMIN:** COnsensus-based Standards for the selection of health Measurement Instruments.

1

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

Manuscript Title: Measurement properties of instruments assessing permanent functional impairment of the spine: a systematic review protocol

Section and topic	Item No	Checklist item	Information reported		Page, Line number(s)
			Yes	N/A	
ADMINISTRATIVE INFORMATION					
Title					
Identification	1a	Identify the report as a protocol of a systematic review	☒	☐	Page 1, Line 6
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	☐	☒	N/A
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	☒	☐	Page 2, Line 70
Authors					
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	☒	☐	Page 1, Lines 10-42
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	☒	☐	Page 11, Lines 303-306
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	☒	☐	Page 11, Lines 309-310
Sponsor	5b	Provide name for the review funder and/or sponsor	☒	☐	Page 11, Lines 309-310
Role of	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the	☒	☐	Page 11, Lines

Section and topic	Item No	Checklist item	Information reported Yes	Information reported N/A	Page, Line number(s)
sponsor/funder		protocol			313-314
INTRODUCTION					
Rationale	6	Describe the rationale for the review in the context of what is already known	☒	☐	Pages 3-4, Lines 83-125
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	☒	☐	Page 4, Lines 120-123
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	☒	☐	Pages 5-6, Lines 138-169
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	☒	☐	Pages 6, Lines 170-178
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	☒	☐	Pages 6-7, Lines 179-192 + "Table 1"
Study Records					
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	☒	☐	Page 7 , Lines 193-204
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)	☒	☐	Page 7, Lines 205-217
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	☒	☐	Pages 9, Lines 249-266
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding	☒	☐	Pages 8, Lines

3

Section and topic	Item No	Checklist item	Information reported		Page, Line number(s)
			Yes	N/A	
		sources), any pre-planned data assumptions and simplifications			218-231 + "Table 3"
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	☒	☐	Page 8, Lines 219-224
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	☒	☐	Pages 8-9, Lines 232-248 + "Table 4"
Data Synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	☒	☐	Page 9, Lines 249-266
	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)	☒	☐	Page 9, Lines 249-266
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	☒	☐	Page 9, Lines 261-266
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	☒	☐	Page 9, Lines 250-256
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	☒	☐	Page 9, Lines 267-271
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	☒	☐	Pages 8-9, Lines 232-248

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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

BMJ Open

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

SCHOLARONE™
Manuscripts

1 **TITLE PAGE**

2

3 **Title:**

4

5 Measurement properties of instruments assessing permanent functional impairment of the spine:
6 a systematic review protocol

7

8 **Authors, affiliations:**

9

10 **Suelen Meira Goes**: suelen.goes@usask.ca; School of Physical Therapy, University of
11 Saskatchewan - Health Sciences Building, E-Wing; Suite 3400 104 Clinic Place - Saskatoon,
12 Saskatchewan, S7N 2Z4 – Canada. *** Corresponding Author**

13

14 **Catherine Trask**: catherine.trask@usask.ca; Canadian Centre for Health and Safety in
15 Agriculture (CCHSA), University of Saskatchewan - 103 Hospital Drive, Saskatoon,
16 Saskatchewan, S7N 0W8 – Canada.

17

18 **Catherine Boden**: catherine.boden@usask.ca; Leslie and Irene Dubé Health Sciences Library,
19 Academic Health Sciences Building - 104 Clinic Place, Saskatoon, Saskatchewan, S7N 2Z4 –
20 Canada.

21

22 **Brenna Bath**: brenna.bath@usask.ca; School of Physical Therapy, University of Saskatchewan -
23 Health Sciences Building, E-Wing; Suite 3400 104 Clinic Place - Saskatoon, Saskatchewan, S7N
24 2Z4 – Canada.

25

26 **Daniel Cury Ribeiro**: daniel.ribeiro@otago.ac.nz; School of Physiotherapy, University of Otago
27 - 325 Great King Street, Dunedin, 9016 – New Zealand.

28

29 **Paul Hendrick**: paul.hendrick@nottingham.ac.uk; School of Health Sciences, University of
30 Nottingham – Clinical Sciences Building - Nottingham City Hospital - Hucknall Rd Nottingham,
31 NG5 1PB – United Kingdom.

32

33 **Lynne Clay**: lynne.clay@otago.ac.nz; School of Physiotherapy, University of Otago - 325 Great
34 King Street, Dunedin, 9016 – New Zealand.

35

36 **Xiaoke Zeng**: xiaoke.zeng@usask.ca; Canadian Centre for Health and Safety in Agriculture
37 (CCHSA), University of Saskatchewan - 103 Hospital Drive, PO Box 120, Saskatoon,
38 Saskatchewan, S7N 0W8 – Canada.

39

40 **Stephan Milosavljevic**: stephan.milosavljevic@usask.ca; School of Physical Therapy,
41 University of Saskatchewan - Health Sciences Building, E-Wing; Suite 3400 104 Clinic Place -
42 Saskatoon, Saskatchewan, S7N 2Z4- Canada.

43

44 **Word count:** 262 (abstract) + 2291 (protocol)

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 synthesize 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 (COSMIN) criteria. Findings will be summarized and presented descriptively, with meta-analysis pursued as appropriate.

Ethics and Dissemination: This review will summarize 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 organizations. The findings will likely provide a foundation and direction for future research priorities for assessing spinal PFI.

Keywords: disability, back, range of movement, psychometrics, validity, reliability, responsiveness

Systematic review registration: PROSPERO CRD42017060390.

Strengths and limitations of this study

- This systematic review will include a broad range of instruments used to assess PFI in individuals with spinal conditions in peer-reviewed articles.

- This systematic review protocol is presented in accordance with the Preferred Reporting Items for Systematic review and Meta-Analyses Protocols (PRISMA-P).

- A strength of this review is the use of the internationally recognised, validated COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) 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.

83 **INTRODUCTION**

84 Spinal disorders are the leading cause of work disability, representing around 65% of all
85 musculoskeletal injury/illness claims associated with lost work time, posing a considerable
86 economic, personal, and societal burden [1]. Internationally, occupational low back pain is the
87 singular most prevalent health problem related to absenteeism amongst workers, responsible for
88 approximately 25% of costs from work-related conditions [2].

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

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

107 Instruments used to assess spinal PFI have been critically evaluated, since their measurements
108 are usually based only on impairment [14-18]. Poor validity and reliability of non-invasive clinical
109 measurement instruments and the presence of significant measurement error [14-17] are criticisms
110 that have led to a number of alternative impairment-based spinal assessment systems such as:

4

111 AMA Guides' Diagnosis-Related Estimate [11], Diagnosis-Based Impairment [5], McBride's method

112 of spinal impairment evaluation [19], or Physical Impairment Index [18 20 21]. However, these

113 alternative systems for rating PFI present poor descriptions of standardization and normative

114 values [10 22].

115 The outcome of spinal PFI assessment by compensation bodies can have considerable social,

116 economic and health impacts, yet the reliability and validity of instruments currently available

117 have not been systematically evaluated. To achieve a fair and accurate outcome, clinicians,

118 researchers and government bodies must have access to the most accurate level of evidence

119 regarding methods assessing spinal PFI. To our knowledge, no previous review has assessed the

120 reliability and validity of available instruments for determining PFI of the spine. This manuscript

121 presents the protocol of an ongoing systematic review with the objective to review and

122 synthesize the literature concerning measurement properties of the instruments used for assessing

123 PFI of the spine. Due to the diversity of instruments that might be used to assess PFI of the spine,

124 we anticipate the identification of distinct metrics for measuring PFI will allow specific sub-

125 group analyses for review and ultimate discussion on the strength of support for each instrument.

126

METHODS

127

Design

128

129 This systematic review protocol has been registered with International Prospective Register of

130 Systematic Reviews (PROSPERO) (registration number CRD42017060390). This protocol will

131 be reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses

132 Protocols (PRISMA-P) statement and checklist [23].

133

134

135

136 **Eligibility Criteria**

137 Inclusion criteria

138 We will include peer-reviewed articles, with no date limitation, investigating measurement

139 properties of instruments that assess PFI of the spine in individuals with spinal disorders aged

140 from 18 to 65 years. When evaluating PFI, acceptable studies will include measures of

141 impairment (i.e. RoM, muscle strength, coordination, endurance, and sensation), functional

142 limitation (i.e. self-report instruments of physical function and functional performance measures)

143 and permanency of the impairment (i.e. duration of the impairment and the likelihood of

144 improvement) [3 24].

145 Exclusion criteria

146 We will exclude letters to the editor, book reviews, and short communications. We will also

147 exclude clinical protocols, case reports and series, systematic reviews, meta-analysis, articles not

148 published in English-language, studies intended for screening, diagnosis and prognosis of spinal

149 pathologies as well as studies with specimen-, cadaveric-, cellular-, artificial-, and computer-

150 based models. We will also exclude studies of spinal conditions caused by congenital and

151 developmental abnormalities, neoplasm, infection, and systemic inflammatory disorders [25].

152 Exclusion will take place at the screening stage, following the literature search, in order to avoid

153 the risk of excluding relevant articles.

154 **Information Sources**

155 Studies will be identified by searching electronic databases, scanning reference lists of articles

156 and consultation with experts in the area. A preliminary search plan was developed and pre-

157 tested by a professional health sciences librarian (CB). The following databases were searched:

158 Medline – OVID (1946 to April 20, 2017); EMBASE – OVID (1947 to April 21, 2017);

159 CINAHL – Cumulative Index to Nursing and Allied Health Literature (1937 to April 25, 2017);

160 Web of Science (1900 to April 21, 2017); Scopus (1995 to April 25, 2017); PEDro (Inception to

161 April 22, 2017); OTSeeker (Inception to April 22, 2017); and Health and Safety Science

162 Abstracts – ProQuest (1988 to April 22, 2017).

163 **Search strategy**

164 To capture all measurement instruments used to evaluate PFI of the spine, our search strategy

165 will focus on combining three conceptual groups of terms: (1) PFI, (2) spinal disorder, and (3)

166 measurement properties. The Medline search strategy is presented in Table 1.

167 The terms for the concept of "measurement properties" were adapted from a search filter

168 validated and published by Terwee et al [26] and optimized for Medline. A similar search will be

169 employed for other databases, optimized for database-specific search interfaces. A filter to

170 exclude certain publication types (e.g., addresses, editorials, letters, and newspaper articles) will

171 be applied, as recommended by Terwee et al [26]. References will be imported into DistillerSR

172 software (Distiller SR Evidence Partners Limited, Ottawa, Canada). All duplicate articles will be

173 removed prior to screening. Reference lists of included studies will be scrutinized during the

174 process of data extraction for additional articles meeting our criteria. Identified citations will be

175 downloaded and subjected to the same de-duplication and screening process.

176 | Table 1 about here |

177 **Study Selection**

178 Study selection will be undertaken in two steps. First, two reviewers will independently screen

179 all titles and abstracts, and all citations selected by both reviewers will be retrieved for step two:

180 full-text screening. In cases where there is disagreement between the two reviewers, any

181 discrepancies will be resolved by consensus and/or consultation with a third member of the

182 research team prior to making a final decision.

183 A screening tool will be developed and piloted for the title, abstract and full-text screening stages

184 (see table 2). Inclusion/exclusion will be determined using the same tool throughout the process,

185 although the definitions may be refined as conflicts arise and are discussed. The same questions

186 will be used for both title and abstract as well as full-text screening, with a focus on sensitivity

187 rather than specificity at these stages.

188 | Table 2 about here |

189 **Data extraction process**

190 We will develop data extraction forms, as outlined in table 3, which will be pilot tested on three

191 studies not included in the review. Two reviewers will extract data independently. Any

192 disagreement will be resolved through a third party adjudication. The nature of performance

193 measures employed by included studies will drive the nature of the data extraction, as we

194 anticipate that there may be some heterogeneity in the evaluation of measurement instruments.

195 | Table 3 about here

196 To avoid analysing the same data from multiple articles, we will compare author names, sample

197 sizes and outcomes. In the case of companion papers (e.g. reports) of a primary study, all

198 available data will be simultaneously evaluated, to maximize the yield of information for the

199 primary study. Study authors may be contacted for clarifications. If contact with authors is

200 unsuccessful, we will use the most complete and up-to-date data source available, thus avoiding

201 the risk of double counting.

202 **Data Items**

203 We will extract information relating to (1) the measurement properties (i.e. reliability, validity,

204 and responsiveness) of the measurement instruments for assessing PFI of the spine (e.g.

205 assessment of RoM, muscle strength, coordination, endurance, and sensation, as well as ability to

206 perform basic daily living and work activities, and duration of the impairment); (2) the type of

207 inferential test; and (3) the test results for each article. The extraction process will be tracked

208 using a standardized data extraction form.

209 Articles will be separated into sub-groups based on the instruments used to assess PFI of the

210 spine. The different sub-groups will be analyzed separately based on impairment-based measures

211 (i.e. RoM, muscle strength, coordination, endurance, and sensation) and functional limitation

212 assessment (i.e. self-report instruments of physical function and functional performance

213 measures). For clarity, we are likely to isolate articles in sub-groups to better differentiate the

214 instruments used to assess PFI of the spine, especially when considering that RoM has

215 predominantly been used for assessing impairment of the spine [5][6][9][11].

Quality assessment of individual studies

216

217 Included studies will be appraised independently by two reviewers as to their methodological

218 quality using the consensus-based standards for the selection of health measurement instruments

219 (COSMIN) criteria [27 28]. Any disagreement will be resolved by consultation with a third member

220 of the research team. The COSMIN checklist is a consensus-based tool designed to evaluate the

221 methodological quality of studies investigating measurement properties. The instrument shows

222 appropriate levels of agreement [27] and, based on its content validity, is a recommended tool for

223 assessing the methodological quality of studies evaluating measurement properties of outcome

224 measures within a systematic review [28].

225 The tool will evaluate the following measurement constructs: reliability; measurement error;

226 content validity; structural validity; hypotheses testing; cross-cultural validity; and criterion

227 validity, responsiveness, and interpretability with five to18 items concerning methodological

228 standards for how each measurement property should be assessed (see Table 4). The

229 methodological quality of a study will be considered adequate if all items in a measurement

230 property are considered adequate. Each item is scored on a four-point rating scale (i.e., ''poor'',

231 ''fair'', ''good'', or ''excellent'').

232 In the context of this review, validity, in general, defines how well the instrument under

233 evaluation measures the construct it purports to measure. Criterion validity is the degree to which

234 measurements are an adequate reflection of a previously used 'gold standard'. Content validity is

235 an adequate reflection of the construct to be measured; construct validity is based on an

236 assumption that the instrument truly measures what it is meant to; and structural validity implies

237 the scores of an instrument is an adequate reflection of the dimensionality of the construct to be

238 assessed. Reliability refers to the extent to which scores for individuals who have not changed

239 are the same for repeated measurement under several conditions. These include using different

240 sets of items from the same instrument (internal consistency); over time (test-retest); by different

241 assessors on the same occasion (inter-rater); or by the same assessors (i.e. raters or responders)

242 on different occasions (intra-rater). Responsiveness is the ability of the measurement instrument

243 to detect change over time in the construct to be measured [27].

244 | Table 4 about here

245

Planned methods of analysis

247 Where it is possible, mean values of statistical analysis (e.g. Cronbach's alpha, intra-correlation
248 coefficient, standard error of measurement, smallest detectable change, effect sizes, etc.) will be
249 calculated from pooled data from methodologically similar studies and the results statistically
250 summarized via meta-analysis into forest plots with estimates of heterogeneity. In addition,
251 sensitivity analysis will be pursued by comparing results from studies with high and low-quality
252 ranking. We will then weight such meta-analyses using both the study's sample size and their
253 quality assessment as determined by the COSMIN checklist [28]. However, we predict some
254 heterogeneity will be identified in the various assessments of spinal PFI measures, which will
255 likely make a meta-analysis difficult to apply.

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

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

268

ETHICS AND DISSEMINATION

10

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

271 used for financial award decisions, in both legal and compensation contexts [29-34]. However, a

272 systematic review in this area to support the use of instruments currently being used to determine

273 PFI is lacking. The application of reliable and valid measurement instruments for assessing PFI

274 of the spine is considered crucial.

275 This systematic review offers a feasible means for synthesizing the evidence specific to spinal

276 PFI assessment; and our results will likely provide unique insights concerning the breadth and

277 depth of literature in the area. Outcomes of this review will be applicable to clinicians, policy-

278 makers, worker's compensation boards and health and safety organizations. In particular,

279 findings will likely provide a foundation and direction in terms of research priorities for

280 assessing PFI of the spine. Summarizing the nature and strength of the evidence regarding the

281 reliability, validity and responsiveness of spinal PFI measures will also inform future research

282 and policy in this field.

283

284 **LIST OF ABBREVIATIONS**

285 AMA: American Medical Association

286 CINAHL: Cumulative Index to Nursing and Allied Health Literature

287 COSMIN: COnsensus-based Standards for the selection of health Measurement Instruments

288 Embase: Excerpta Medica Database

289 HR-PRO: health related patient-reported outcomes.

290 ICF: International Classification of Functioning, Disability and Health

291 OTseeker: Occupational Therapy Systematic Evaluation of Evidence

292 PEDro: Physiotherapy Evidence Database

293   PFI: Permanent Functional Impairment

294   PICOS: Population; Interventions or exposure; Comparator; Outcomes or endpoint; Study design

295   PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

296   RoM: Range of Motion

297   WHO: World Health Organization

298

299   **AUTHORS' CONTRIBUTIONS**

300   SMG and SM conceptualized the initial review protocol and led the development of this

301   manuscript. The entire team (SMG, SM, CB, CT, BB, DCR, PH, LC, and XZ) collaboratively

302   drafted the manuscript, followed by numerous iterations and substantial input and appraisal from

303   all authors. All authors have approved the final version of this manuscript.

304

305   **FUNDING**

306   This research is funded by The Workers' Compensation Board of British Columbia

307   (*WorkSafeBC grant number RS2015-SR01*).

308

309   **COMPETING INTERESTS**

310   The views, findings, opinions and conclusion expressed herein do not necessarily represent the

311   views of the funder.

12

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. *Annals of the rheumatic diseases* 2014;73(6):975-81. doi: 10.1136/annrheumdis-2013-204631 [published Online First: 2014/03/26]

2. Helfenstein M, Goldenfum MA, Siena C. OCCUPATIONAL LOW BACK PAIN. *Revista Da Associacao Medica Brasileira* 2010;56(5):583-89.

3. WHO. A Practical Manual for using the International Classifcation of Functioning, Disability and Health (ICF). Geneva: World Health Organization, 2013.

4. WHO. World report on disability. Geneva, 2011.

5. Rondinelli RD, Genovese E, Brigham CR. American Medical Association. 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 permanet impairment. 2.1 ed. Australia, 2014.

7. WorkCover. Guidelines for the assessment of permanent impairment. 3rd ed. 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, 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. doi: 10.3346/jkms.2009.24.S2.S307 [published Online First: 2009/06/13]

11. Cocchiarella L, Andersson GBJ. Guide to the Evaluation of Permanent Impairment. 5th ed. ed. Chicago, Ill: Americal 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. *Manual therapy* 2000;5(1):30-6. doi: 10.1054/math.1999.0228 [published Online First: 2000/02/26]

13. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical perspective. *Journal of Manual and Manipulative Therapy* 2011;19(2):91-99. doi: http://dx.doi.org/10.1179/106698111X12973307659546

14. Hirsch BP, Webb ML, Bohl DD, et al. Improving visual estimates of cervical spine range of motion. *Am J Orthop (Belle Mead NJ)* 2014;43(11):E261-5. [published Online First: 2014/11/08]

15. Whitcroft KL, Massouh L, Amirfeyz R, et al. Comparison of methods of measuring active cervical range of motion. *Spine* 2010;35(19):E976-80. doi: 10.1097/BRS.0b013e3181cd6176 [published Online First: 2010/08/12]

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 (Phila Pa 1976)* 1999;24(3):262-8. [published Online First: 1999/02/20]

356 17. Lowery WD, Jr., Horn TJ, Boden SD, et al. Impairment evaluation based on spinal range of
357   motion in normal subjects. *J Spinal Disord* 1992;5(4):398-402. [published Online First:
358   1992/12/11]
359 18. Fritz JM, Piva SR. Physical impairment index: reliability, validity, and responsiveness in
360   patients with acute low back pain. *Spine (Phila Pa 1976)* 2003;28(11):1189-94. doi:
361   10.1097/01.brs.0000067270.50897.db [published Online First: 2003/06/05]
362 19. McBride ED. Disability evaluation: principles of treatment of compensable injuries. 1936.
363   *Clinical orthopaedics and related research* 2007;457:11-5. doi:
364   10.1097/BLO.0b013e31803dcab1 [published Online First: 2007/04/07]
365 20. Trudelle-Jackson E, Sarvaiya-Shah SA, Wang SS. Interrater reliability of a movement
366   impairment-based classification system for lumbar spine syndromes in patients with
367   chronic low back pain. *J Orthop Sports Phys Ther* 2008;38(6):371-6. doi:
368   10.2519/jospt.2008.2760 [published Online First: 2008/06/03]
369 21. Waddell G, Somerville D, Henderson I, et al. Objective clinical evaluation of physical
370   impairment in chronic low back pain. *Spine* 1992;17(6):617-28. [published Online First:
371   1992/06/01]
372 22. Rondinelli RD. Changes for the new AMA Guides to impairment ratings, 6th Edition:
373   implications and applications for physician disability evaluations. *Pm r* 2009;1(7):643-
374   56. doi: 10.1016/j.pmrj.2009.04.012 [published Online First: 2009/07/25]
375 23. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and
376   meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews* 2015;4:1. doi:
377   10.1186/2046-4053-4-1 [published Online First: 2015/01/03]
378 24. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical
379   perspective. *The Journal of manual & manipulative therapy* 2011;19(2):91-9. doi:
380   10.1179/106698111x12973307659546 [published Online First: 2012/05/02]
381 25. Haldeman S, Kopansky-Giles D, Hurwitz EL, et al. Advancements in the Management of
382   Spine Disorders. *Best Practice & Research Clinical Rheumatology* 2012;26(2):263-80.
383   doi: 10.1016/j.berh.2012.03.006
384 26. Terwee CB, Jansma EP, Riphagen, II, et al. Development of a methodological PubMed
385   search filter for finding studies on measurement properties of measurement instruments.
386   *Qual Life Res* 2009;18(8):1115-23. doi: 10.1007/s11136-009-9528-5 [published Online
387   First: 2009/08/28]
388 27. Mokkink LB, Prinsen CA, Bouter LM, et al. The COnsensus-based Standards for the
389   selection of health Measurement INstruments (COSMIN) and how to select an outcome
390   measurement instrument. *Braz J Phys Ther* 2016;20(2):105-13. doi: 10.1590/bjpt-
391   rbf.2014.0143 [published Online First: 2016/01/21]
392 28. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic
393   reviews of studies on measurement properties: a scoring system for the COSMIN
394   checklist. *Qual Life Res* 2012;21(4):651-7. doi: 10.1007/s11136-011-9960-1 [published
395   Online First: 2011/07/07]
396 29. Davis MA, Onega T, Weeks WB, et al. Where the United States spends its spine dollars:
397   expenditures on different ambulatory services for the management of back and neck
398   conditions. *Spine* 2012;37(19):1693-701. doi: 10.1097/BRS.0b013e3182541f45
399   [published Online First: 2012/03/22]

14

400   30. Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back
401   and neck problems. *Jama* 2008;299(6):656-64. doi: 10.1001/jama.299.6.656 [published
402   Online First: 2008/02/14]
403   31. Martin BI, Turner JA, Mirza SK, et al. Trends in health care expenditures, utilization, and
404   health status among US adults with spine problems, 1997-2006. *Spine* 2009;34(19):2077-
405   84. doi: 10.1097/BRS.0b013e3181b1fad1 [published Online First: 2009/08/14]
406   32. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies
407   in the United States and internationally. *Spine J* 2008;8(1):8-20. doi:
408   10.1016/j.spinee.2007.10.005 [published Online First: 2008/01/01]
409   33. McGee R, Bevan S, Quadrello T. Fit For Work? Musculoskeletal Disorders and the Canadian
410   Labour Market. London, ON: The Work Foundation, 2009.
411   34. Dunning KK, Davis KG, Cook C, et al. Costs by industry and diagnosis among
412   musculoskeletal claims in a state workers compensation system: 1999-2004. *Am J Ind*
413   *Med* 2010;53(3):276-84. doi: 10.1002/ajim.20774 [published Online First: 2009/11/26]

414

415 **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/
	35. Goniomet*.ab,ti.	61. Psychometrics/
	36. Joint motion measurement*.ab,ti.	62. psychometr*.ab,ti.
	37. Electrogoniomet*.ab,ti.	63. clinimetr*.tw.
	38. (Movement adj2 evaluation*).ab,ti.	64. clinometr*.tw.
3. low back pain/	39. Inclinomet*.ab,ti.	65. "Outcome Assessment (Health Care)"/
4. back pain/ or neck pain/	40. Joint flexibility.ab,ti.	66. "outcome assessment".ab,ti.
5. back/ or neck/ or lumbosacral region/ or sacrococcygeal region/	41. Arthromet*.ab,ti.	67. ("outcome measure" or "outcome measures").tw.
6. (back or neck or "lumbosacral region" or "sacrococcygeal region").ab,ti.	42. Joint mobility.ab,ti.	68. Observer Variation/
	43. Osteokinematic*.ab,ti.	69. "observer variation".ab,ti.
7. Sciatica/	44. Flexion.ab,ti.	70. Health Status Indicators/
8. Neck Injury/ or Whiplash Injuries/	45. Extension.ab,ti.	71. "Reproducibility of Results"/
9. whiplash.ab,ti.	46. Rotation/	72. reproducib*.ti,ab.
10. Dorsalgia.ti,ab.	47. side bend*.ab,ti.	73. Discriminant Analysis/
11. coccydynia.ti,ab.	48. Work Capacity Evaluation/	74. reliab*.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.	49. Disability Evaluation/ and (Occupational Diseases/ or Work/ or Return to Work/)	75. unreliab*.ab,ti.
	50. Occupational Diseases/ or Work/ or Return to Work/	76. valid*.ab,ti.
	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.
13. discitis.ti,ab.		78. homogeneity.ab,ti.
14. (disc adj degeneration).ti,ab.	52. 50 and 51	79. homogeneous.ab,ti.
15. (disc adj prolapse).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.	80. "internal consistency".ab,ti.
16. (disc adj herniation).ti,ab.		81. cronbach*.ab,ti.
17. (facet adj joints).ti,ab.		82. (alpha or alphas).ab,ti.
18. Intervertebral Disc/ or annulus fibrosus/ or nucleus pulposus/	54. (evaluation* or assessment* or instrument* or measurement* or tool* or scale* or questionnaire* or test* or determination*).ab,ti.	83. 81 and 82
19. arachnoiditis.ti,ab.		84. item.ab,ti.
20. Spinal Fusion/		85. (correlation* or selection* or reduction*).ab,ti.
21. postlaminectomy.ti,ab.	55. 53 and 54	86. 84 and 85
22. Backache*.ti,ab.	56. "permanent functional impairment".ab,ti.	87. agreement.ab,ti.
23. back injuries/ or spinal injuries/ or spinal fractures/		88. precision.ab,ti.
	57. ("work performance evaluation" or "work performance evaluations").ab,ti.	89. imprecision.ab,ti.
24. Spondylitis, Ankylosing/		90. "precise values".ab,ti.
25. Spondylitis/	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	91. test-retest.ab,ti.
26. spondylosis/ or spondylolysis/ or		92. test.ab,ti.
		93. retest.ab,ti.
		94. 92 and 93
		95. reliab*.ab,ti.
		96. (test or retest).ab,ti.

16

spondylolisthesis/

27. Spinal Cord Compression/

28. (Musculoskeletal adj (symptom* or injur* or disorder* or pain or dysfunction* or problem*)).ab,ti.

29. ((orthopedic or orthopaedic) adj (injur* or problem* or disorder* or dysfunction*)).ab,ti.

30. pelvis/ or lesser pelvis/ or pelvic floor/

31. (pelvi* adj3 (symptom* or injur* or disorder* or pain or dysfunction* or problem*)).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

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. inter-technician.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.

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.

170. "computer adaptive testing".ab,ti.

171. "item bank".ab,ti.

172. "cross-cultural equivalence".ab,ti.

173. ("gold standard" or "criterion

18

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

416   PFI: Permanent Functional Impairment

417

418

419 **Table 2 - Draft screening tool for use at the title, abstract, and full-text review screening**
420 **stages**
421

Title, Abstract and Full-text screening questions
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 two

2) Does the study deal with humans?
 a) No – exclude
 b) Yes or uncertain – go to step three

3) Does the study deal with adults (between 18 and 65 years of age)?
 a) No – exclude
 b) Yes or uncertain – go to step four

4) Does the article represent primary study (i.e. no letters to the editor, book reviews, published study designs or trial protocols)
 a) No – exclude
 b) Yes or uncertain – go to step five

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 six

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 seven

7) Is the study designed to evaluate measurement properties of measurement instruments/tools (e.g. 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 eight

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 nine

9) Does this study have clinical measures of impairment (e.g. range of motion, muscle performance) or objective measures of function (e.g. functional impairment evaluation)?
 a) No – exclude
 b) Yes or uncertain – go to step ten

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 eleven

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

422 **Note:** Exclusion occurs if the answer to any of the questions is "no".
423

20

424 **Table 3 - 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 (e.g. 'Back-EST' or 'the Johnson method')
Instrument description	Will collect sub variables, such as: * Equipment required Duration Clinicians performing test (e.g. physiotherapist, occupational therapist, physician) Dimensions assessed (e.g. Range of motion, lifting capacity, etc.)
Validity- criterion	Will collect sub variables, such as: * Did the article assess this dimension? (yes/no) What was the inferential test used? (e.g. intraclass correlation coefficient, kappa) What was the result? (e.g. 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 sub variables, 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? (e.g. intraclass correlation coefficient, kappa) What was the result? (e.g. kappa score of 0.7)
Reliability – intra-rater	Will collect sub variables, such as: * Did the article assess this dimension? (yes/no) What was the inferential test used? (e.g. intraclass correlation coefficient, kappa) What was the result? (e.g. 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

425 PFI: Permanent Functional Impairment; COSMIN: COnsensus-based Standards for the selection of health
426 Measurement Instruments; **Note:** These Items may evolve as review progresses and studies are identified. This draft
427 data dictionary describes the selected data extraction variables of the systematic review. The data dictionary would
428 be used to extract data from articles identified as relevant during the screening process. The data extraction database
429 itself would have these items listed in columns with the individual references representing rows such that these data
430 are extracted for every article to form a summary database for later synthesis. Only key variables are presented here.
431 It is anticipated that this list will expand based on the type of data presented in the articles; the definitions guide will
432 also become more specific as a result of independent data extraction and discussion at team meetings.

433 **Table 4** – COSMIN definitions of domains, measurement properties, and aspects of
434 measurement properties
435

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
	Intra-rater		The degree of agreement between repeated measurements of a property on the same patient by the same rater
	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
Validity			How well the instrument under evaluation measures the construct it purports to measure
	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

436 **COSMIN:** COnsensus-based Standards for the selection of health Measurement Instruments.

1

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

Manuscript Title: Measurement properties of instruments assessing permanent functional impairment of the spine: a systematic review protocol

Section and topic	Item No	Checklist item	Information reported		Page, Line number(s)
			Yes	N/A	
ADMINISTRATIVE INFORMATION					
Title					
Identification	1a	Identify the report as a protocol of a systematic review	☒	☐	Page 1, Line 6
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	☐	☒	N/A
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	☒	☐	Page 2, Line 70
Authors					
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	☒	☐	Page 1, Lines 10-42
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	☒	☐	Page 11, Lines 303-306
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	☒	☐	Page 11, Lines 309-310
Sponsor	5b	Provide name for the review funder and/or sponsor	☒	☐	Page 11, Lines 309-310
Role of	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the	☒	☐	Page 11, Lines

Section and topic	Item No	Checklist item	Information reported		Page, Line number(s)
			Yes	**N/A**	
sponsor/funder		protocol			313-314
INTRODUCTION					
Rationale	6	Describe the rationale for the review in the context of what is already known	☒	☐	Pages 3-4, Lines 83-125
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	☒	☐	Page 4, Lines 120-123
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	☒	☐	Pages 5, Lines 138-154
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	☒	☐	Pages 6, Lines 156-163
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	☒	☐	Pages 6, Lines 165-172 + "Table 1"
Study Records					
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	☒	☐	Page 6, Lines 172-173
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)	☒	☐	Page 6, Lines 179-189
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	☒	☐	Pages 7, Lines 191-202
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding	☒	☐	Pages 7, Lines

3

Section and topic	Item No	Checklist item	Information reported		Page, Line number(s)
			Yes	**N/A**	
		sources), any pre-planned data assumptions and simplifications			204-216 + "Table 3"
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	☒	☐	Page 7, Lines 204-209
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	☒	☐	Pages 8, Lines 218-244 + "Table 4"
Data Synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	☒	☐	Page 9, Lines 249-266
	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)	☒	☐	Page 9, Lines 248-268
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	☒	☐	Page 9, Lines 248-256
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	☒	☐	Page 9, Lines 257-264
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	☐	☒	N/A
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	☒	☐	Pages 9, Lines 265-268

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.