BMJ Open

Physical functioning outcome measures in the lumbar spinal surgery population and measurement properties of the physical outcome measures: protocol for a systematic review

Katie L Kowalski ,[1] Michael J Lukacs ,[1] Jai Mistry,[1,2] Maren Goodman ,[3] Alison B Rushton [1]

To cite: Kowalski KL, Lukacs MJ, Mistry J, *et al*. Physical functioning outcome measures in the lumbar spinal surgery population and measurement properties of the physical outcome measures: protocol for a systematic review. *BMJ Open* 2022;**12**:e060950. doi:10.1136/bmjopen-2022-060950

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-060950).

Received 10 January 2022
Accepted 18 May 2022

[1]School of Physical Therapy, Faculty of Health Sciences, Western University, London, Ontario, Canada
[2]Physiotherapy, St George's Hospital, London, UK
[3]Western Libraries, Western University, London, Ontario, Canada

Correspondence to
Dr Katie L Kowalski; kkowals7@uwo.ca

ABSTRACT

Introduction Low back pain can lead to substantial decline in physical functioning. For disabling pain not responsive to conservative management, surgical intervention can enhance physical functioning. Measurements of physical functioning include patient-reported outcome measures and physical outcome measures using evaluations of impairments, performance on a standardised task or activity in a natural environment. Selecting outcome measures with adequate measurement properties is fundamental to evaluating effectiveness of interventions. The purpose of this systematic review is to identify outcome measures (patient reported and physical) used to evaluate physical functioning (stage 1) and assess the measurement properties of physical outcome measures of physical functioning (stage 2) in the lumbar spinal surgery population.

Methods and analysis This protocol aligns with the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines and Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols. Using a two-staged approach, searches will be performed in MEDLINE, EMBASE, Health and Psychosocial Instruments, CINAHL, Web of Science, Scopus, PEDro and the grey literature from inception until 15 December 2021. Stage 1 will identify studies evaluating physical functioning with patient-reported or physical outcome measures in the lumbar spinal surgery population. Stage 2 will search for studies evaluating measurement properties (validity, reliability, responsiveness) of the physical outcome measures identified in stage 1 in the lumbar spinal surgery population. Two independent reviewers will evaluate studies for inclusion, extract data, assess risk of bias (COSMIN risk of bias tool and checklist) and quality of evidence (modified Grading of Recommendations Assessment, Development and Evaluation approach). Results for each measurement property per physical outcome measure will be quantitatively pooled if there is adequate clinical and methodological homogeneity or qualitatively synthesised if there is high heterogeneity in studies.

Ethics and dissemination Ethics approval is not required. Results will be disseminated through peer-reviewed journal publication and conference presentation.

PROSPERO registration number CRD42021293880.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This protocol is designed according to the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines for conducting systematic reviews of physical outcome measurement instruments.
⇒ Reporting of the protocol aligns with Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols and the protocol is registered with PROSPERO.
⇒ Two stages of searches were developed in collaboration with a library information specialist and independently peer-reviewed using the Peer Review of Electronic Search Strategies checklist.
⇒ Two independent reviewers will perform study eligibility assessments, data extraction, risk of bias evaluation using COSMIN Risk of Bias assessment for physical outcome measurement instruments and determine quality of evidence using a modified Grading of Recommendations Assessment, Development and Evaluation approach.
⇒ Focusing on the lumbar spinal surgery population limits the generalisability of results to other interventions and regions of the spine.

INTRODUCTION

Low back pain is a leading cause of disability worldwide with substantial burden of disease from adolescence through older adulthood.[1] This burden has a significant impact at societal and individual levels, such as high healthcare resource use, reduced participation in work and declines in physical functioning.[2] Most episodes of low back pain resolve within weeks to months, however, recurrence is

common and a small percentage of people will develop persistent pain.[2–4] For severe and disabling low back pain that is not responsive to conservative management, surgical interventions can reduce pain and enhance physical functioning.[5–8]

The Core Outcome Measures in Effectiveness Trials Initiative defines physical functioning as the impact of a disease or condition on physical activities of daily living, such as walking, self-care, performance status and disability index.[9] Physical functioning is a multidimensional construct with several interacting facets, including body structure and function, performance of physical activities, and social and role functioning.[10–12] It is an important component of health-related quality of life[2 9 10] and recommended as a key domain within core outcome sets to measure the effectiveness of interventions for low back pain.[13 14] Physical functioning can be measured by self-report using patient-reported outcome measures (PROMs) or with physical outcome measures using evaluations of impairments (eg, range of motion), performance on a standardised task (eg, Timed Up and Go) or activity in a natural environment (eg, accelerometry).[10 15]

In the lumbar spinal surgery population, PROMs are the most common method to evaluate physical functioning.[16] The Oswestry Disability Index (ODI) questionnaire is used most frequently,[16 17] with disability referring to impairments in body structure and function, limitations in physical activities and restrictions in participation.[11] The ODI is a well-established legacy PROM as it has been used for over four decades with translations into almost 80 languages and cultural contexts.[18] Accordingly, most national and international spine registries use the ODI to evaluate physical functioning in lumbar spinal pain population.[19] However, recent advances in outcome measure development and research[13 20] warrant a wider consideration of PROMs other than the ODI to evaluate physical functioning. Yet there is no resource which comprehensively outlines PROMs that evaluate physical functioning, beyond the ODI, in the lumbar spinal surgery population.

While the ODI and other PROMs are easy to implement and measure physical functioning directly from the patient's perspective, they have limitations. Content validity is the degree to which the outcome measure comprehensively evaluates all relevant aspects of a construct in the population of interest.[21] It is the most important measurement property of an outcome measure.[21] However, given the multidimensional nature of physical functioning, content validity of the ODI and other physical functioning PROMs is lacking in the lumbar spinal pain population.[17 22] This is also suggested in the lumbar spinal surgery population as PROMs do not provide a comprehensive evaluation of physical functioning.[5 6 23] Further, a range of contextual factors can influence a patient's perspective of their physical functioning, such as pain and psychological functioning,[10 24 25] limiting the responsiveness and interpretability of PROMs.[26 27] These limitations suggest that only using PROMs is an insufficient means to accurately measure the multiple domains of physical functioning in the lumbar spinal surgery population.

Compared with PROMs, physical outcome measures evaluate distinct aspects of physical functioning, and when used in combination, provide a comprehensive evaluation of physical functioning.[5 6 10 15 23 24] International recommendations in other surgical disciplines (eg, hip and knee arthroplasty) are to measure both PROMs and physical outcomes in evaluations of physical functioning.[28 29] The use of physical outcome measures in lumbar spinal pain populations has recently increased[30] and evidence suggests both types of measures are necessary to comprehensively evaluate physical functioning. For example, after lumbar spinal surgery the trajectory of recovery differs between PROMs and physical outcome measures as there is a rapid improvement in ODI scores with slower improvement in Timed Up and Go performance and accelerometry measures of daily activity.[5 6] Further highlighting the importance of physical outcome measures, studies suggest good outcome after lumbar spinal surgery is associated with preoperative and postoperative physical outcome measures of physical functioning.[23 31–33]

Fundamental to accurately evaluating effectiveness of interventions and avoiding risk of bias is selecting outcome measures with adequate measurement properties.[34] One aim of COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) is to improve the selection of outcome measurement instruments through a systematic evaluation of measurement properties, including validity, reliability and responsiveness.[34 35] As physical outcome measures require involvement of patients, practitioners and sometimes equipment, they are complex and influenced by more sources of variation than PROMs. Accordingly, COSMIN has developed a tool to assess the risk of bias of studies on reliability and measurement error in physical outcome measures.[35] This tool enables systematic reviews of measurement properties of physical outcome measures.

Despite a recent increase in the use of physical outcome measures,[30] there is no systematic review evaluating measurement properties of physical outcome measures in the lumbar spinal surgery population. Recent advances in outcome measurement development and research[13 20] also warrant development of a resource outlining PROMs of physical functioning, beyond the ODI, in the lumbar spinal surgery population.

Objectives

1. To identify outcome measures (patient reported and physical) that are used to evaluate physical functioning in the lumbar spinal surgery population.
2. To assess the measurement properties of physical outcome measures of physical functioning in the lumbar spinal surgery population.

METHODS AND ANALYSIS

This protocol is designed according to the COSMIN guidelines for conducting systematic reviews of outcome

measurement instruments.[34 35] Reporting of the protocol aligns with Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols.[36] The protocol has been registered with PROSPERO, which will be updated with any protocol amendments. The protocol methodology consists of two stages of searching that align with the two review objectives. Stage 1 will identify studies evaluating physical functioning using either PROMs (excluding ODI) or physical outcome measures in the lumbar spinal surgery population. This will enable the generation of a comprehensive list of physical functioning outcome measures used in the lumbar spinal surgery population, beyond the ODI. Using the list of identified physical outcome measures, stage 2 will search for studies evaluating measurement properties of the physical outcome measures in the lumbar spinal surgery population. PROMs will not inform stage 2. Stage 1 search was executed on 15 December 2021, and study completion is anticipated for July 2022.

Stage 1: Identify physical functioning outcome measures (PROMs and physical)

Eligibility criteria

Participants

Inclusion criteria include adults aged 18 years or older who either are listed for or previously had lumbar spinal surgery for low back and/or low back-related leg pain. No restrictions will be placed on the history of lumbar surgery. Surgery due to trauma, fracture, space occupying mass (eg, tumour, cyst), inflammatory conditions, infection, osteoporosis, congenital scoliosis, cauda equina syndrome and extra-spinal causes of back and/or leg pain will be excluded.

Intervention

Lumbar spinal surgery at one or more levels will be eligible, including thoracolumbar and lumbosacral surgical interventions as a component of lumbar spinal surgery.

Outcome measures

Studies will be eligible for inclusion if within the title or abstract there is an indication of using an outcome measure, scale or subscale that evaluates physical functioning in the lumbar spinal surgery population. Physical functioning may be evaluated using PROMs or physical measures (ie, impairment-based, performance-based or activity in natural environment). The ODI will be excluded as it is already a well-established, legacy PROM of physical functioning in the lumbar spinal surgery population.[16 19] Outcome measures that are not practical within clinical, hospital or community Physical Therapy settings will be excluded. This may include outcomes such as imaging, electrophysiological measures (eg, EMG), and motion capture gait analysis (eg, force plates, 3D video analysis).

Study design

All study designs will be included.

Stage 2: Assess measurement properties of physical outcome measures of physical functioning

Eligibility criteria

Participants

Inclusion criteria include adults aged 18 years or older who either are listed for or previously had lumbar spinal surgery for low back and/or low back-related leg pain. No restrictions will be placed on the history of lumbar surgery. Surgery due to trauma, fracture, space occupying mass (eg, tumour, cyst), inflammatory conditions, infection, osteoporosis, congenital scoliosis, cauda equina syndrome and extra-spinal cause of back and/or leg pain will be excluded.

Intervention

Lumbar spinal surgery at one or more levels will be eligible, including thoracolumbar and lumbosacral surgical interventions as a component of lumbar spinal surgery.

Outcome measures

For stage 2 search, outcome measures of interest are measurement properties of the physical measures identified in the stage 1 search. Measurement property domains of validity, reliability and responsiveness will be included.

Study design

For stage 2, the COSMIN measurement properties sensitive search and exclusion filters[37] will be used to search for studies evaluating the measurement properties of physical outcome measures. The filter restricts retrieval of irrelevant records, such as case reports, editorials and animal studies. Studies will be excluded if there is no original study data (eg, systematic review), insufficient information to thoroughly assess measurement properties (eg, conference abstract) or normative data is presented only.

Search strategy for stage 1 and 2

Stage 1 and 2 searches have been developed in collaboration with a library information specialist (MG). Search strategies in stage 1 and 2 are informed by the National Institute for Health and Care Excellence guidelines for low back pain and sciatica in over 16s.[38] The stage 2 search incorporates the COSMIN sensitive search and exclusion filters for retrieving studies on measurement properties.[37] Prior to executing searches, stage 1 and 2 search strategies were independently peer-reviewed by a second library information specialist using the Peer Review of Electronic Search Strategies checklist.[39] No revisions to the search strategy were recommended for either stage.

Stage 1 will identify studies evaluating physical functioning with either PROMs or physical outcome measures in the lumbar spinal surgery population. To ensure a comprehensive search during stage 1, the type and name of physical functioning outcome measures will not be specified, and the measurement properties construct will not be searched. This will enable the generation of a comprehensive list of physical functioning outcome measures used in the lumbar spinal surgery population,

beyond the ODI. Measures will be classified as PROMs or physical outcome measures. Using the names of physical outcome measures (eg, Timed Up and Go), stage 2 will search for studies evaluating measurement properties of the identified physical outcome measures in the lumbar spinal surgery population. No language restrictions will be used for either stage of searching, though each will be restricted to human studies only. Online supplemental file 1 contains example stage 1 and 2 search strategies developed in MEDLINE (Ovid).

Information sources

For stage 1, a comprehensive search will be performed from inception to 15 December 2021 using key databases. Stage 2 search will be performed immediately after stage 1 data extraction. Searches have been developed in MEDLINE (Ovid) and a library information specialist will adapt searches for use in EMBASE (Ovid), Health and Psychosocial Instruments (Ovid), CINAHL (EBSCOhost), Web of Science Core Collection, Scopus and PEDro. No date or language limits will be applied. The grey literature will be searched using ProQuest Dissertations and Theses. No trial registries will be searched as there are none for studies of measurement properties and additional studies will not be sought through contacting experts, manufacturers or authors. For stage 2, references lists of articles that meet inclusion criteria will be screened to identify potential articles for inclusion.

Study records

Data management

Citations identified in stage 1 and 2 searches will be imported and stored in Covidence, a web-based software platform for systematic reviews. Duplicates will be automatically identified by the software and removed. After title and abstract screening, full texts will be uploaded and stored in Covidence. Eligibility screening at both the title/abstract and full-text stages will also occur in Covidence.

Selection process

Eligibility assessment will be performed for stage 1 and 2 searches independently by two reviewers. Title and abstracts will be screened against eligibility criteria with full texts obtained for studies in which both reviewers agree on inclusion or there is insufficient information in the title and abstract for determination of eligibility. Full texts will then be screened independently in duplicate for determination of inclusion. Articles will be included if both reviewers agree eligibility criteria have been met. Disagreements at both steps will be discussed and if consensus is not achieved, a third reviewer (ABR) will mediate. Agreement between reviewers will be evaluated with Cohen's kappa[40] using SPSS (V.27; IBM). The article selection process will be summarised for both stages of searching with reasons for exclusion documented in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram (figure 1).[41]

Data collection process

Two reviewers will independently extract data from eligible studies using standardised data collection forms for each stage of searching. Data collection forms will be piloted on five articles with modifications made as needed to ensure they are fit for purpose. Discrepancies in data extracted will be resolved through discussion between reviewers, with a third reviewer (ABR) to mediate if needed. If data are unclear or not presented in eligible studies, corresponding authors will be contacted with a request for additional information. In the event of multiple reports of a single study, corresponding authors will be contacted for further information to ensure data are not duplicated in the review. As needed, two follow-up reminder emails will be sent at 2-week intervals.

Data items

Data to be extracted from eligible studies during stage 1 and 2 are summarised in table 1.

Outcomes and prioritisation

Stage 1: Outcome measures evaluating physical functioning in the lumbar spinal surgery population will be classified as one of the following:

1. PROM including questionnaires, scales or subscales assessing one or more aspects of physical functioning.
2. Impairment-based physical outcome measure evaluating structure or function of a specific body part or system[15] (eg, range of motion).
3. Performance-based physical outcome measure evaluating performance on a defined task in a standardised environment[10 15] (eg, Timed Up and Go).
4. Physical outcome measure evaluating activity in a natural environment[10] (eg, accelerometry).

Stage 2: Measurement properties of physical outcome measures (classifications 2–4 above) in the lumbar spinal surgery population. Measurement property domains of validity, reliability and responsiveness will be included. Secondary outcomes will include feasibility and interpretability of the physical outcome measures.

Risk of bias in individual studies

Risk of bias of included studies will be evaluated using the COSMIN Risk of Bias checklist[42] and extended tool for reliability and measurement error of outcome measurement instruments.[35] Two reviewers will independently assess risk of bias for all included studies. Disagreements will be resolved through discussion and if consensus is not achieved, a third reviewer (ABR) will mediate. Agreement between reviewers will be evaluated with Cohen's kappa[40] and reported in results.

Risk of bias of each single study of a measurement property will be assessed separately.[42] As recommended by COSMIN, risk of bias related to reliability and measurement error will be assessed using the extended tool for outcome measurement instruments,[35] while all other measurement properties will be assessed against standards for PROMs.[34 42] For each study of a measurement property,

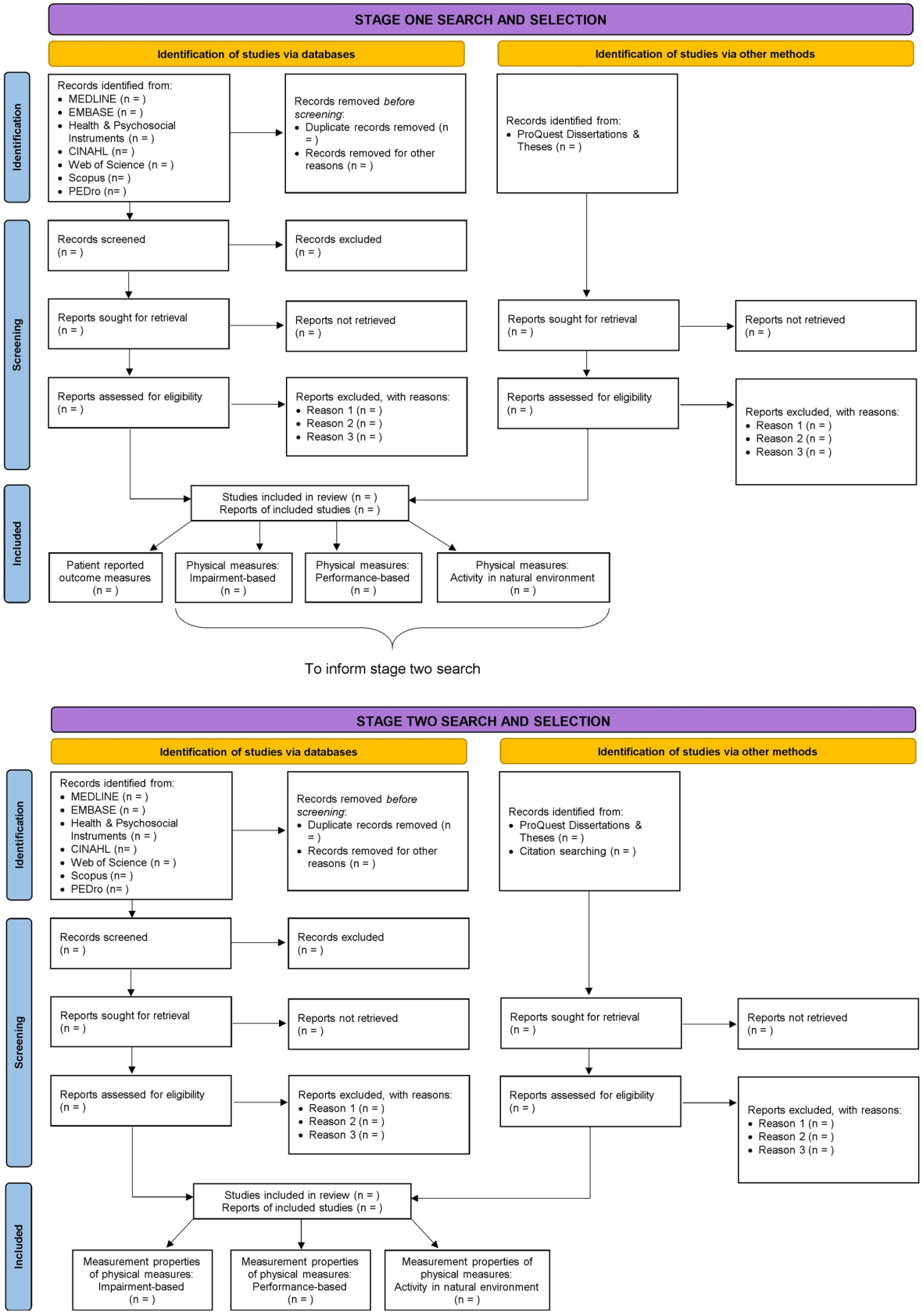

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 flow diagram of stage 1 and 2 search and study selection processes.[41]

Table 1 Summary of data to be extracted from eligible studies

Stage 1 and 2	
Study characteristics	Authors, year of publication, study design, country of study
Study participants	Age, gender, sample size, condition, surgery
Outcome measure	Name, version and/or subscale as needed, language, type (eg, performance-based), physical construct evaluated (eg, aerobic capacity), equipment required
Stage 2 only	
Outcome measure	Preparation required, data collection, processing and storage, assignment of scores, determination of score value, number of assessments, outcome measure descriptive statistics
Validity	Type of validity, method of evaluation including design, statistical methods and hypothesis as needed, comparator or predictor outcome, results
Reliability	Type of reliability, statistical methods and results
Responsiveness	Method of evaluation, statistical methods and hypothesis as needed, time interval, results
Interpretability	Distribution of scores, percentage of missing items and total scores, floor and ceiling effects, scores and change scores available for relevant (sub)groups, minimal important change or difference (MIC, MID), information on response shift
Feasibility	Patient and clinician comprehensibility, type and ease of administration, length of instrument, completion time, patient's required mental and physical ability level, ease of standardisation, ease of score calculation, copyright, cost of instrument, required equipment, availability in different settings, regulatory agency's requirement for approval

risk of bias of relevant standards for design requirements and preferred statistical methods will be rated as very good, adequate, doubtful or inadequate. Overall rating of risk of bias of each single study on a measurement property will follow the worst score counts principle, in which the lowest rating of any standard is applied to the whole study for each specific measurement property. Risk of bias reporting will include ratings for each standard and the overall risk of bias of each measurement property study.

Data synthesis

Results of each study on a measurement property will be rated against criteria for good measurement properties.[34 35] Results will be rated as insufficient (+), insufficient (−) or indeterminate (?) and reported. To summarise the evidence on a measurement property for each outcome measure, results will be quantitatively pooled or qualitatively summarised, as appropriate. If there is adequate clinical and methodological homogeneity, results of measurement properties for each outcome measure from different studies will be quantitatively pooled by calculating weighted means and 95% CIs.[34] Data will be quantitatively pooled when: (a) the lumbar spinal surgery population has similar characteristics in terms of symptom severity and surgical intervention; (b) similar baseline physical functioning score; (c) same time interval and (d) same statistical parameters.[43] Results will be rated as sufficient (+) or insufficient (−), based on criteria for good measurement properties.[34 35] If the results of measurement properties are inconsistent, explanations will be explored (eg, subgroups of surgical interventions). If satisfactory explanations are found for inconsistent results, overall outcome measure ratings will be given for relevant subgroups with consistent results. If no satisfactory explanation is found for inconsistent results, an inconsistent (±) rating will be given. If not enough information is available, the overall rating will be indeterminate (?). If there is insufficient clinical and methodological homogeneity, results will be qualitatively synthesised for each measurement property per physical outcome measure.

Meta-bias(es)

If protocols are identified within our searches, reporting bias will be evaluated through consistency of study protocols and published results.

Confidence in cumulative evidence

Grading of Recommendations Assessment, Development and Evaluation (GRADE) will be used to rate the overall quality of evidence for each measurement property per outcome measure.[34] This approach will assess the pooled or summarised results of criteria for good measurement properties and rate the evidence as high, moderate, low or very low. In accordance with COSMIN guidelines, the standard GRADE approach will be modified for evaluating measurement properties in a systematic review.[34] Four factors will contribute to the determination of quality of evidence, including: risk of bias, inconsistency, imprecision and indirectness. Two reviewers will independently assess quality of evidence, with disagreements resolved through discussion and if consensus is not achieved, a third reviewer will mediate.

Patient and public involvement

Results of this systematic review will be discussed with the spinal pain research Patient Partner Advisory Group in the School of Physical Therapy at Western University.

The purpose of the discussion will be to compare physical functioning outcome measure results with patient partner experiences of having physical functioning measured and how they prefer physical functioning be measured. This discussion will inform the development of future research projects.

DISCUSSION

Physical functioning is an important component of health-related quality of life[10 44–46] and a multidimensional construct.[2 10 12 24] In the lumbar spinal surgery population, PROMs are commonly used to measure physical functioning.[16 47] While this is an important mechanism to capture a patient's perspective of the impact of their symptoms, PROMs are insufficient to measure the multiple domains of physical functioning.[5 6 12 23 24] Complementing the use of PROMs with physical outcome measures, including evaluations of impairments, performance on a standardised task and activity in a natural environment, will provide a comprehensive evaluation of physical functioning in the lumbar spinal surgery population.[10 12 15]

Physical outcome measures of physical functioning can be also used to inform decision-making within the care pathway, such as decisions related to the need for rehabilitation before or after lumbar spinal surgery. For example, preoperative and early postoperative physical outcome measures of physical functioning predict good outcome following lumbar spinal surgery.[23 32 33] Early identification of deficits in physical outcome measures will enable targeted rehabilitation interventions to improve patient outcomes following lumbar spinal surgery. Use of physical outcome measures also aligns with recently updated physical therapy clinical practice guidelines for the management of acute and chronic low back pain.[48] These guidelines suggest physical outcome measures should be evaluated within physical therapy interventions for surgical and non-surgical lumbar spinal pain populations.[48]

Key to accurately measuring physical functioning and effectiveness of interventions is selecting outcome measures with adequate measurement properties. Given the importance and recent increase in the use of physical outcome measures in the lumbar spinal surgery populations,[5 6 23 30 33] a systematic review is necessary to synthesise the literature on the measurement properties of physical outcome measures. Results will inform clinicians and researchers in selecting the best available physical outcome measures of physical functioning in the lumbar spinal surgery population to evaluate effectiveness of interventions and inform decisions about the need for rehabilitation before or after lumbar spinal surgery. This review may also highlight important gaps in the literature, such as many physical outcome measures being used but they may not all be validated for use in the lumbar spinal surgery population or at different points within the care pathway (eg, preoperative, early postoperative, late postoperative). Review results will be limited to the lumbar spinal surgery population with limited generalisability to other interventions and regions of the spine.

Twitter Alison B Rushton @abrushton

Acknowledgements The authors thank Meagan Stanley for peer-reviewing the electronic search strategy.

Contributors KLK, MJL, JM, MG and ABR contributed to conception and design of the protocol. JM and KLK will screen articles for eligibility and extract data. All authors will contribute to data interpretation. KLK drafted and edited the manuscript. MJL, JM, MG, ABR critically reviewed and provided feedback on the manuscript. All authors have approved the final version of the manuscript to be published. KLK is the guarantor of the review.

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

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Katie L Kowalski http://orcid.org/0000-0003-2130-652X
Michael J Lukacs http://orcid.org/0000-0002-9113-4834
Maren Goodman http://orcid.org/0000-0002-0216-0994
Alison B Rushton http://orcid.org/0000-0001-8114-7669

REFERENCES

1 GBD. Diseases and injuries Collaborators. global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 2019;2020:1204–22.
2 Hartvigsen J, Hancock MJ, Kongsted A, *et al*. What low back pain is and why we need to pay attention. *Lancet* 2018;391:2356–67.
3 da C Menezes Costa L, Maher CG, Hancock MJ, *et al*. The prognosis of acute and persistent low-back pain: a meta-analysis. *CMAJ* 2012;184:E613–24.
4 da Silva T, Mills K, Brown BT, *et al*. Risk of recurrence of low back pain: a systematic review. *J Orthop Sports Phys Ther* 2017;47:305–13.
5 Coronado RA, Master H, White DK, *et al*. Early postoperative physical activity and function: a descriptive case series study of 53 patients after lumbar spine surgery. *BMC Musculoskelet Disord* 2020;21:1–10.
6 Smuck M, Muaremi A, Zheng P, *et al*. Objective measurement of function following lumbar spinal stenosis decompression reveals improved functional capacity with stagnant real-life physical activity. *Spine J* 2018;18:15–21.
7 Koenders N, Rushton A, Verra ML, *et al*. Pain and disability after first-time spinal fusion for lumbar degenerative disorders: a systematic review and meta-analysis. *Eur Spine J* 2019;28:696–709.

8 Rushton A, Heneghan NR, Heymans MW, *et al*. Clinical course of pain and disability following primary lumbar discectomy: systematic review and meta-analysis. *Eur Spine J* 2020;29:1660–70.
9 Dodd S, Clarke M, Becker L, *et al*. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *J Clin Epidemiol* 2018;96:84–92.
10 Taylor AM, Phillips K, Patel KV, *et al*. Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. *Pain* 2016;157:1836–50.
11 World Health Organization. Towards a common language for functioning, disability and health ICF; 2002.
12 Wright AA, Hegedus EJ, Baxter GD, *et al*. Measurement of function in hip osteoarthritis: developing a standardized approach for physical performance measures. *Physiother Theory Pract* 2011;27:253–62.
13 Chiarotto A, Boers M, Deyo RA, *et al*. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain* 2018;159:481–95.
14 Clement RC, Welander A, Stowell C, *et al*. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthop* 2015;86:523–33.
15 Reiman MP, Manske RC. The assessment of function: how is it measured? a clinical perspective. *J Man Manip Ther* 2011;19:91–9.
16 Stokes OM, Cole AA, Breakwell LM, *et al*. Do we have the right PROMs for measuring outcomes in lumbar spinal surgery? *Eur Spine J* 2017;26:816–24.
17 Wiitavaara B, Heiden M. Content and psychometric evaluations of questionnaires for assessing physical function in people with low back disorders. A systematic review of the literature. *Disabil Rehabil* 2020;42:163–72.
18 Oswestry Disability Index (ODI). Mapi Res. trust, 2021. Available: https://eprovide.mapi-trust.org/instruments/oswestry-disability-index#languages [Accessed 30 Dec 2021].
19 van Hooff ML, Jacobs WC, Willems PC. Evidence and practice in spine registries: a systematic review, and recommendations for future design of registries. *Acta Orthop* 2015;86:534–44.
20 Chiarotto A, Terwee CB, Ostelo RW. Choosing the right outcome measurement instruments for patients with low back pain. *Best Pract Res Clin Rheumatol* 2016;30:1003–20.
21 Terwee CB, Prinsen CAC, Chiarotto A, *et al*. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a delphi study. *Qual Life Res* 2018;27:1159–70.
22 Chiarotto A, Ostelo RW, Boers M, *et al*. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *J Clin Epidemiol* 2018;95:73–93.
23 Master H, Pennings JS, Coronado RA, *et al*. Physical performance tests provide distinct information in both predicting and assessing patient-reported outcomes following lumbar spine surgery. *Spine* 2020;45:E1556–63.
24 Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol* 2006;59:160–7.
25 Bean JF, Ölveczky DD, Kiely DK, *et al*. Performance-Based versus patient-reported physical function: what are the underlying predictors? *Phys Ther* 2011;91:1804–11.
26 Cook CE, Wright A, Wittstein J, *et al*. Five recommendations to address the limitations of patient-reported outcome measures. *J Orthop Sports Phys Ther* 2021;51:562–5.
27 Chiarotto A, Maxwell LJ, Terwee CB, *et al*. Roland-Morris disability questionnaire and Oswestry disability index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? systematic review and meta-analysis. *Phys Ther* 2016;96:1620–37.
28 Fitzgerald GK, Hinman RS, Zeni J, *et al*. OARSI clinical trials recommendations: design and conduct of clinical trials of rehabilitation interventions for osteoarthritis. *Osteoarthritis Cartilage* 2015;23:803–14.
29 Dobson F, Hinman RS, Roos EM, *et al*. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage* 2013;21:1042–52.
30 Stienen MN, Ho AL, Staartjes VE, *et al*. Objective measures of functional impairment for degenerative diseases of the lumbar spine: a systematic review of the literature. *Spine J* 2019;19:1276–93.
31 Janssen ER, Osong B, van Soest J, *et al*. Exploring associations of preoperative physical performance with postoperative outcomes after lumbar spinal fusion: a machine learning approach. *Arch Phys Med Rehabil* 2021;102:1324–30.
32 Jakobsson M, Brisby H, Gutke A, *et al*. Prediction of objectively measured physical activity and self-reported disability following lumbar fusion surgery. *World Neurosurg* 2019;121:e77–88.
33 Gilmore SJ, Hahne AJ, Davidson M, *et al*. Predictors of substantial improvement in physical function six months after lumbar surgery: is early post-operative walking important? A prospective cohort study. *BMC Musculoskelet Disord* 2019;20:418.
34 Prinsen CAC, Mokkink LB, Bouter LM, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
35 Mokkink LB, Boers M, van der Vleuten CPM, *et al*. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol* 2020;20:1–13.
36 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
37 Terwee CB, Jansma EP, Riphagen II, *et al*. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.
38 Nice guideline. low back pain and sciatica in over 16S: assessment and management NICE guideline, 2016. Available: www.nice.org.uk/guidance/ng59 [Accessed 19 Sep 2021].
39 McGowan J, Sampson M, Salzwedel DM, *et al*. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol* 2016;75:40–6.
40 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
41 Page MJ, Mckenzie JE, Bossuyt PM. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372.
42 Mokkink LB, de Vet HCW, Prinsen CAC, *et al*. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171–9.
43 Alamrani S, Rushton A, Gardner A, *et al*. Outcome measures evaluating physical functioning and their measurement properties in adolescent idiopathic scoliosis: a protocol for a systematic review. *BMJ Open* 2020;10:e034286.
44 WHOQOL Group. The world Health organization quality of life assessment (WHOQOL): position paper from the world Health organization. *Soc Sci Med* 1995;41:1403–9.
45 Stewart AL, Hays RD, Ware JE. The mos short-form general health survey. reliability and validity in a patient population. *Med Care* 1988;26:724–35.
46 Herdman M, Gudex C, Lloyd A, *et al*. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–36.
47 Finkelstein JA, Schwartz CE. Patient-Reported outcomes in spine surgery: past, current, and future directions. *J Neurosurg Spine* 2019;31:155–64.
48 George SZ, Fritz JM, Silfies SP. Clinical practice guidelines interventions for the management of acute and chronic low back pain: revision 2021. *J Orthop Sport Phys Ther* 2021;51:1–60.