Research Paper

PAIN

OPEN

Core outcome measurement instruments for clinical trials in nonspecific low back pain

Alessandro Chiarotto[a,b,*], Maarten Boers[a,c], Richard A. Deyo[d], Rachelle Buchbinder[e,f], Terry P. Corbin[g], Leonardo O.P. Costa[h], Nadine E. Foster[i], Margreth Grotle[j,k], Bart W. Koes[l], Francisco M. Kovacs[m], C.-W. Christine Lin[n], Chris G. Maher[n], Adam M. Pearson[o], Wilco C. Peul[p], Mark L. Schoene[q], Dennis C. Turk[r], Maurits W. van Tulder[b], Caroline B. Terwee[a], Raymond W. Ostelo[a,b]

Abstract

To standardize outcome reporting in clinical trials of patients with nonspecific low back pain, an international multidisciplinary panel recommended physical functioning, pain intensity, and health-related quality of life (HRQoL) as core outcome domains. Given the lack of a consensus on measurement instruments for these 3 domains in patients with low back pain, this study aimed to generate such consensus. The measurement properties of 17 patient-reported outcome measures for physical functioning, 3 for pain intensity, and 5 for HRQoL were appraised in 3 systematic reviews following the COSMIN methodology. Researchers, clinicians, and patients (n = 207) were invited in a 2-round Delphi survey to generate consensus (≥67% agreement among participants) on which instruments to endorse. Response rates were 44% and 41%, respectively. In round 1, consensus was achieved on the Oswestry Disability Index version 2.1a for physical functioning (78% agreement) and the Numeric Rating Scale (NRS) for pain intensity (75% agreement). No consensus was achieved on any HRQoL instrument, although the Short Form 12 (SF12) approached the consensus threshold (64% agreement). In round 2, a consensus was reached on an NRS version with a 1-week recall period (96% agreement). Various participants requested 1 free-to-use instrument per domain. Considering all issues together, recommendations on core instruments were formulated: Oswestry Disability Index version 2.1a or 24-item Roland-Morris Disability Questionnaire for physical functioning, NRS for pain intensity, and SF12 or 10-item PROMIS Global Health form for HRQoL. Further studies need to fill the evidence gaps on the measurement properties of these and other instruments.

Keywords: Core outcome set, Recommendations, Outcome measurement instruments, Low back pain, Clinical trials

1. Introduction

Low back pain (LBP) represents the leading cause of years lived with disability globally, ranking first in both developed and developing countries.[46] The mean lifetime prevalence of LBP is estimated to be 39%, with a mean point prevalence of 18%.[58] The costs of LBP constitute a major burden to health care systems and society.[32,76] Most commonly, a specific pathoanatomical cause cannot be identified for LBP, so its most prevalent form is nonspecific LBP (nsLBP).[79] The number of randomized controlled trials assessing the effectiveness of health interventions in nsLBP has substantially increased over the past 2 decades.[12]

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

[a] Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, the Netherlands, [b] Department of Health Sciences, Faculty of Science, Amsterdam Movement Sciences Research Institute, Vrije Universiteit, Amsterdam, the Netherlands, [c] Amsterdam Rheumatology and Immunology Center, VU University Medical Center, Amsterdam, the Netherlands, [d] Department of Family Medicine, Department of Internal Medicine, and Oregon Institute of Occupational Health Sciences, Oregon Health and Science University, Portland, OR, USA, [e] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia, [f] Monash Department of Clinical Epidemiology, Cabrini Institute, Malvern, Australia, [g] Cochrane Collaboration, Back and Neck Review Group, Maple Grove, MN, USA, [h] Masters and Doctoral Programs in Physical Therapy, Universidade Cidade de Sao Paulo, Sao Paulo, Brazil, [i] Arthritis Research UK Primary Care Centre, Research Institute for Primary Care and Health Sciences, Keele University, Keele, United Kingdom, [j] Oslo and Akershus University College, Faculty of Health Science, Oslo, Norway, [k] Communication and Research Unit for Musculoskeletal Disorders (FORMI), Oslo University Hospital & University of Oslo, Oslo, Norway, [l] Department of General Practice, Erasmus MC University Medical Center, Rotterdam, the Netherlands, [m] Spanish Back Pain Research Network, Hospital Universitario HLA-Moncloa, Madrid, Spain, [n] Sydney School of Public Health, Sydney Medical School, University of Sydney, Sydney, Australia, [o] Department of Orthopaedic Surgery, Dartmouth-Hitchcock Medical Center, Lebanon, PA, USA, [p] Department of Neurosurgery, Leiden University Medical Center, Leiden, the Netherlands, [q] Cochrane Collaboration, Back and Neck Review Group, Newbury, MA, USA, [r] Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, USA

**Corresponding author. Address: Department of Health Sciences, Faculty of Science, Amsterdam Public Health Research Institute, Amsterdam Movement Sciences Research Institute, Vrije Universiteit, de Boelelaan 1085, Room U-601, 1081HV Amsterdam, the Netherlands. E-mail address: a.chiarotto@vu.nl (A. Chiarotto).*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.painjournalonline.com).

PAIN 159 (2018) 481–495

http://dx.doi.org/10.1097/j.pain.0000000000001117

Heterogeneity in the choice of outcomes and measurement instruments assessed in clinical trials hampers comparisons between studies and systematic reviews summarizing them.[72,73] In several medical fields including nsLBP, this is a major issue.[53,70,77] It can be addressed by agreeing on a standardized set of outcomes that should be measured and reported in all clinical trials on a specific health condition: a core outcome set (COS).[7,19,113] A COS does not preclude the choice of primary or secondary outcomes that are not in the COS, but ensures that important outcomes are consistently assessed.[7,19,113] A COS specific to LBP was introduced 20 years ago by a group of experienced researchers and clinicians.[8,30]

Deyo et al.[30] and Bombardier[8] proposed 5 core outcome domains to be measured in LBP clinical research: back-specific function, pain symptoms, generic health status, work disability, and satisfaction with care; for each of these domains, 1 or 2 patient-reported outcome measures (PROMs) were also suggested. More recently, we initiated an international Steering Committee to build on this existing proposal, by consulting up-to-date methodology of Core Outcome Measures in Effectiveness Trials (COMETs) and Outcome Measures in Rheumatology (OMERACT) initiatives[6,7,92,104,111,112] to develop a COS applicable to clinical trials in patients with nsLBP.[22]

Developing a COS is a 2-step consensus process that involves, first, determining the core outcome domains ("core domain set"), and second, selecting the best outcome measurement instruments to measure these domains ("core outcome measurement set").[7,19,113] For nsLBP, a consensus was achieved on 4 core outcome domains: physical functioning, pain intensity, health-related quality of life (HRQoL), and number of deaths.[16] The domain number of deaths was included in line with OMERACT mandatory requirement to have at least 1 domain in the core area "Death"[7] and because it is good practice for any trial to report on this domain; it can be covered with a simple statement reporting how many deaths occurred in a trial.[16] However, there is no consensus on measurement instruments for the other 3 core outcome domains. The selection of core outcome measurement instruments comprises the following steps: (1) identifying potential core instruments, (2) evaluating their measurement properties and feasibility, and (3) reaching a consensus on those that should be recommended.[6,92] The objective of this study was to formulate recommendations on core outcome measurement instruments for clinical trials in patients with nsLBP.

2. Methods

An international Steering Committee, including 19 members, worked on the development of this COS: 17 researchers and/or clinicians (A.C., M.B., R.A.D., R.B., L.O.P.C., N.E.F., M.G., B.W. K., F.M.K., C.-W.C.L., C.G.M., A.M.P., W.C.P., D.C.T., M.W.v. T., C.B.T., and R.W.O.) and 2 patients' representatives (T.P.C. and M.L.S.). A 4-member project team comprising a subset of the Steering Committee (A.C., M.B., C.B.T., and R.W.O.) oversaw the initiative. The committee expertise included the following: anesthesiology, epidemiology, internal medicine, orthopaedics, physical therapy, neurosurgery, primary care, psychology, rehabilitation, and rheumatology.

The intent was to develop a COS applicable to the measurement of efficacy or effectiveness of health interventions assessed in all clinical trials for patients with nsLBP, defined as "LBP not attributable to a recognizable, known specific pathology (eg, infection, tumour, fracture, and axial spondyloarthritis)."[22] Therefore, this COS applies to all interventions, regardless of type, setting, frequency, or mode of administration. Following COMET and OMERACT definitions,[7,113] this COS does not prescribe primary outcomes. Rather, it recommends outcome domains and measurement instruments that should be included in each individual trial, alongside additional trial-specific outcomes. The selection of instruments for physical functioning, pain intensity, and HRQoL was guided by the OMERACT handbook,[6] and the consensus-based guidance of the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative in cooperation with COMET.[92]

In the Netherlands, this type of study does not fall within the score of the Dutch Medical Research in Human Subjects Act (WMO), therefore it was exempt from ethical approval of a University Ethics Committee.

2.1. Identification of potential core outcome measurement instruments

The Steering Committee selected a preliminary set of outcome measurement instruments for the core domains, choosing among those frequently used in clinical trials[15,44] and those recommended by other initiatives aimed at standardizing measurements for LBP[8,24,30,31] or chronic pain.[34] It was considered that these criteria (ie, already in frequent use and recommended by others) would facilitate implementation of this COS. The project team performed an initial screening to determine whether an instrument had good face validity to measure the domain and was feasible (eg, accessibility, cost prohibitive, and availability of translations) for inclusion in a COS.[6] A previous systematic review linking LBP-specific PROMs content to the International Classification of Functioning was consulted to support decisions on face validity.[49] Only PROMs were selected because they are feasible and the most frequently used and recommended tools in the LBP literature.[8,15,24,30,31,34,44]

2.2. Appraisal of measurement properties of outcome measurement instruments

The COSMIN initiative[83] previously identified 9 measurement properties relevant for PROMs: internal consistency, test-retest reliability, measurement error, construct validity, structural validity, criterion validity, cross-cultural validity, and responsiveness.[85] Three systematic reviews (for physical functioning, pain intensity and HRQoL) summarized and appraised the evidence on these measurement properties in patients with nsLBP (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review: Unpublished data; Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review: Unpublished data; and Ref. 18). These reviews were conducted according to the recently updated COSMIN methodology for this type of reviews (Prinsen et al., 2018. COSMIN guideline for systematic reviews of patient-reported outcome measures: Unpublished data); a more detailed description of their methodology is presented elsewhere (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review: Unpublished data; Chiarotto et al., 2018. Evidence on

the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review: Unpublished data; and Ref. 18).

2.3. Delphi study

A consensus procedure is recommended to find an agreement on core outcome measurement instruments.[7,92] An online modified Delphi survey was chosen as it is a widely used method to establish a consensus on various health- and research-related issues[47,63,74,85,105]; allows participation of a broad, international, and multistakeholder panel of 'experts'; enables reconsideration of participants' views based on responses from others; and preserves anonymity among respondents.[51,98] Authors of at least 2 publications comprising psychometric or clinimetric studies, randomized clinical trials, or systematic reviews of clinical trials in patients with nsLBP were selected to participate. This selection was performed among 280 people invited to participate in the Delphi study on core outcome domains for nsLBP (selected with a systematic approach, as explained elsewhere[16,22]), members of the Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT) executive, authors of the 2 most recent IMMPACT publications,[37,103] and 39 members of the OMERACT pain working group. To retrieve the publications, a PubMed search was performed on October 18, 2016, by 1 reviewer (A.C.) combining authors' names with MESH terms and key words referring to LBP. All eligible authors were invited for Delphi participation; all Steering Committee members were also invited.

Two Delphi rounds were run: the first between October 19 and November 9, 2016, the second between December 13, 2016, and January 17, 2017. Before invitation, the content of each round was pilot tested by at least 4 Steering Committee members. Selected participants were invited to participate in both rounds, unless they explicitly indicated that they did not wish to participate. During each round, 2 reminders were sent to people who had not responded. Participants were asked about sociodemographic (eg, nationality and sex) and professional characteristics (eg, current role and number of clinical trials in nsLBP). Given the high LBP point prevalence,[58] all participants were asked whether they currently had nsLBP, and those answering positively were specifically requested to also consider their patient perspective when responding to the Delphi survey. These professionals were also considered as part of the patient stakeholder group, together with patient representatives. Proposals were presented in the Delphi survey as closed questions in which participants could answer on a 5-point Likert scale ranging from "Strongly disagree/Absolutely no" to "Strongly agree/Absolutely yes" and give reasons for their answers. Because Delphi studies rely on reaching a consensus, no sample size calculation was required. A consensus was set a priori at 67% of total number of participants (dis)agreeing with a proposal (ie, "Strongly (dis)agree" and "(Dis)Agree" answers were pooled together). This criterion is in line with previous Delphi studies (Terwee et al., 2018. COSMIN standards and criteria for evaluating the content validity of patient-reported outcome measures: a Delphi study: Unpublished data; and Refs. 16, 87, 88, 90). Consistency of results was assessed by separately calculating proportions of each stakeholder group (ie, researchers, clinicians, and patients). The online software SurveyMonkey (SurveyMonkey, Palo Alto, CA) was used.

2.3.1. Delphi round 1

There is a consensus that the minimum requirement to include a PROM in a COS is that it has high quality evidence for sufficient content validity,[92] but in the systematic reviews this criterion was not met by any instrument (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review: Unpublished data; Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review: Unpublished data; and Ref. 18). Despite this, a proposal was made in the first round, before the actual consensus procedure commenced, for recommending core instruments based on the following reasoning: the absence of high quality evidence does not equate to insufficient content validity, not endorsing any instrument may hamper design and conduct of future trials, and there is a need to update the 20-year old recommendations.[8,30] Subsequently, participants were asked whether they agreed or disagreed with the endorsement of each potential core instrument for inclusion in the COS, taking into account the instrument itself, its measurement properties, and characteristics (synthesized in a table comparing multiple PROMs for the same domain). To facilitate the interpretation of the summary of evidence on measurement properties, colored smiley faces were used for each measurement property of each instrument (eg, a green happy smiley face indicated a high or moderate quality evidence of sufficient results). The order of PROM presentation was randomized across participants. Finally, 2 open questions were asked to participants for additional potential core instruments and for generic feedback on the Delphi and the COS development process. One reviewer (A.C.) read all comments and selected the most consistent and/or substantial ones for discussion together with quantitative results in face-to-face meetings with the other members of the project team.

2.3.2. Delphi round 2

In the second round, participants were presented with the results of Round 1, including their own ratings, those of the total Delphi panel and those of each stakeholder group; a selection of illustrative comments describing participants' reasoning was also displayed. The full feedback report with all comments was emailed to the participants. Patient-reported outcome measures for which there was a consensus for endorsement in the first round were rediscussed only to address some specific aspects (eg, feasibility and characteristics). Patient-reported outcome measures without a consensus were presented again for voting only if they had at least 50% of participants in favor of the endorsement or if any substantial remark favored their endorsement. If no consensus was found on any instrument for a domain, all potential core instruments for that domain were presented again for rating. The round concluded with an open question asking for suggestions for the research agenda.

2.4. Recommendations on core outcome measurement instruments

The Delphi results were discussed in a face-to-face meeting of the project team. A first proposal on recommendations for core

Table 1

Patient-reported outcome measures selected as potential core outcome measurement instruments to measure physical functioning, pain intensity and health-related quality of life in clinical trials in non-specific low back pain.

PROM	Name abbreviation	Reference(s) original development	Characteristics				Recommended by other initiatives aimed at fostering standardization for LBP or chronic pain
			Number of items	Response options	Total score range	Recall period	
				Physical functioning			
Oswestry Disability Index version 1.0	ODI 1.0	Fairbank 1980[38]	10	0-5 rating scale	0-100	Undefined	Original core set for LBP clinical research[30]
Oswestry Disability Index version 2.1a	ODI 2.1a	Fairbank 1980,[38] Meade 1986,[81] Baker 1989[5]	10	0-5 rating scale	0-100	Undefined	Original core set for LBP clinical research; ICHOM standard set for LBP[24, 30]
Chiropractic version Low Back Pain Disability Questionnaire	CLBPDQ	Fairbank 1980,[38] Hudson-Cook 1989[59]	10	0-5 rating scale	0-100	Undefined	
Modified version Low Back Pain Disability Questionnaire	MLBPDQ	Fairbank 1980,[38] Fritz 2001[42]	10	0-5 rating scale	0-100	Today	
24-item Roland Morris Disability Questionnaire	RMDQ-24	Roland 1983[94]	24	0-1 yes/no	0-24	Today	Original core set for LBP clinical research[30]
23-item Roland Morris Disability Questionnaire	RMDQ-23	Roland 1983,[94] Patrick 1995[89]	23	0-1 yes/no	0-23	Today	Original core set for LBP clinical research[30]
18-item Roland Morris Disability Questionnaire	RMDQ-18	Roland 1983,[94] Stratford 1997[102]	18	0-1 yes/no	0-18	Today	
Pain Interference subscale of Brief Pain Inventory	BPI-PI	Daut 1983,[28] Cleeland 1994,[23] Cleeland 2009[22]	7	0-10 numeric scale	0-10	Last 24 h	IMMPACT for chronic pain clinical trials[34]
Pain Interference items of Multidimensional Pain Inventory	MPI-PI	Kerns 1985[71]	9	0-6 rating scale	0-6	Undefined	IMMPACT for chronic pain clinical trials[34]
Physical Functioning subscale of 36-item Short Form Health Survey	SF36-PF	Stewart 1992,[101] Ware 1992[108]	10	1-3 rating scale	0-100	Now	
Disability Index of Low Back Pain Rating Scale	LBPRS-DI	Manniche 1994[80]	15	0-2 rating scale	0-30	Undefined	
Quebec Back Pain Disability Scale	QBPDS	Kopec 1996[75]	20	0-5 rating scale	0-80	Today	
4-item Patient-Reported Outcomes Measurement Information System Physical Function short form	PROMIS-PF-4	Cella 2007,[14] DeWalt 2007,[29] Bruce 2009,[10] Fries 2009,[40] Cella 2010,[13] Rose 2014,[95] PROMIS scientific standards[1]	4	1-5 rating scale	4-20	Undefined	NIH Task Force for research standards in chronic LBP[31]
6-item Patient-Reported Outcomes Measurement Information System Physical Function short form	PROMIS-PF-6	Cella 2007,[14] DeWalt 2007,[29] Bruce 2009,[10] Fries 2009,[40] Cella 2010,[13] Rose 2014,[95] PROMIS scientific standards[1]	6	1-5 rating scale	6-30	Undefined	
8-item Patient-Reported Outcomes Measurement Information System Physical Function short form	PROMIS-PF-8	Cella 2007,[14] DeWalt 2007,[29] Bruce 2009,[10] Fries 2009,[40] Cella 2010,[13] Rose 2014,[95] PROMIS scientific standards[1]	8	1-5 rating scale	8-40	Undefined	
10-item Patient-Reported Outcomes Measurement	PROMIS-PF-10	Cella 2007,[14] DeWalt 2007,[29] Bruce 2009,[10] Fries 2009,[40] Cella 2010,[13]	10	1-5 rating scale	10-50	Undefined	

(*continued on next page*)

Table 1 (*continued*)

PROM	Name abbreviation	Reference(s) original development	Characteristics				Recommended by other initiatives aimed at fostering standardization for LBP or chronic pain
			Number of items	Response options	Total score range	Recall period	
Information System Physical Function short form		Rose 2014,[92] PROMIS scientific standards[1]					
20-item Patient-Reported Outcomes Measurement Information System Physical Function short form	PROMIS-PF-20	Cella 2007,[14] DeWalt 2007,[29] Bruce 2009,[10] Fries 2009,[40] Cella 2010,[13] Rose 2014,[95] PROMIS scientific standards[1]	20	1-5 rating scale	20-99	Undefined	
Pain intensity							
Visual Analogue Scale	VAS	Huskisson 1974[64]	1	0-100 scale	0-100	Varying	
Numeric Rating Scale	NRS	Downie 1978[33]	1	0-10 numeric scale	0-10	Varying	Original core set for LBP clinical research; ICHOM standard set for LBP; NIH Task Force for research standards in chronic LBP; IMMPACT for chronic pain clinical trials[24, 30, 31, 34]
Pain Severity subscale of Brief Pain Inventory	BPI-PS	Daut 1983,[29] Cleeland 1994,[23] Cleeland 2009[22]	4	0-10 numeric scale	0-10	Varying	
Health-related quality of life							
36-item Short Form Health Survey	SF36	Ware 1992[108]	36	Varying	0-100*	Varying	Original core set for LBP clinical research[30]
12-item Short Form Health Survey	SF12	Ware 1996[107]	12	Varying	0-100†	Varying	Original core set for LBP clinical research[30]
EuroQol Five Dimensions questionnaire	EQ-5D	EuroQol Group 1990,[36] Brooks 1996[9]	5 (items); 1 (visual analogue scale)	1-3 rating scale (items); 0-100 visual analogue scale	0-1‡ (items); 0-100 (visual analogue scale)	Today	Original core set for LBP clinical research[30]
Nottingham Health Profile	NHP	Hunt 1981[62]	45	0-1 yes/no	0-100§	At the moment	
10-item Patient-Reported Outcomes Measurement Information System Global Health short form	PROMIS-GH-10	Cella 2007,[14] DeWalt 2007,[29] Hays 2009,[54] Cella 2010,[13] PROMIS scientific standards[1]	9 (items); 1 (numeric scale)	1-5 rating scale (items); 0-10 (numeric scale)	4-20†	Undefined	

* This is the total score range for each of the 8 subscales of SF36.
† This is the total score for physical component and mental component summary scores.
‡ This is a utility score.
§ This is the total score range for the 6 domains measured by NHP part 1 (38 items); the rating for each individual item is provided for part 2 (7 items).
ICHOM, International Consortium for Health Outcomes Measurement; LBP, low back pain; PROM, patient-reported outcome measure.

outcome measurement instruments for clinical trials in nsLBP was formulated and sent to all members of the Steering Committee for review. The committee feedback was considered in a second face-to-face meeting of the project team, after which a refined proposal was sent to the Steering Committee for further revision. Once approval was obtained from all committee members, the recommendations were considered ready for reporting.

3. Results

3.1. Potential core outcome measurement instruments

Seventeen PROMs were selected as potential core instruments for physical functioning, 3 for pain intensity, and 5 for HRQoL (**Table 1**).[1,5,9,10,13,14,22,23,28,29,33,36,38,40,42,54,59,62,64,71,75,80,81,89,94,95,101,102,107,108] There are multiple versions of both the Roland-Morris Disability Questionnaire (RMDQ) and Oswestry Disability Index (ODI), the most widely used physical functioning PROMs in LBP.[15,44] Several versions with sufficient face validity were included (**Table 1**). The Pain Interference subscale of the Brief Pain Inventory (BPI-PI) and the Pain Interference items of the Multidimensional Pain Inventory (MPI-PI) were included because they had been recommended as generic instruments to measure physical functioning in chronic pain.[34]

The NIH Task Force report for research standards for chronic LBP recommended the 4-item Patient-Reported Outcomes Measurement Information System Physical Function short form (PROMIS-PF-4) to measure physical functioning[31]; in this Delphi the standard 4-, 6-, 8-, 10-, and 20-item PROMIS-PF short forms[2,40,95] were included as potential core instruments. The 36-item Short Form Health Survey (SF36) is the most frequently used PROM to measure HRQoL in LBP[15] and its physical functioning subscale (SF36-PF) was also included as a standalone instrument for physical functioning (**Table 1**). The Sickness Impact Profile is one of the most frequently used tools to measure HRQoL in LBP,[15] but it was not selected because its length (ie, 136 items) was considered excessively burdensome for inclusion in a COS. The 10-item PROMIS Global Health short form (PROMIS-GH-10) is not broadly used, but it was included for HRQoL as its face validity was judged to be similar to that of the other selected PROMs and because recently it was recommended by another core set initiative[96] (**Table 1**).

3.2. Measurement properties of the potential core outcome measurement instruments

The systematic review on physical functioning PROMs revealed low or very low quality evidence underpinning the content validity of all the PROMs, with the exception of the 24-item RMDQ (RMDQ-24), which displayed high quality evidence of insufficient comprehensiveness and sufficient comprehensibility.[18] High quality evidence of insufficient unidimensionality was found for ODI 1.0, RMDQ-24, and RMDQ-18; unidimensionality of other PROMs was underpinned by moderate quality evidence, or no studies were found (Appendix 2, available online as supplemental digital content at http://links.lww.com/PAIN/A511).[18] The systematic review on pain intensity PROMs highlighted that content validity of visual analogue scale (VAS), Numeric Rating Scale (NRS), and pain severity subscale of the Brief Pain Inventory (BPI-PS) was underpinned by (very) low quality evidence (Appendix 2, available online as supplemental digital content at http://links.lww.com/PAIN/A511) (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review. Unpublished data). High quality evidence was found only for insufficient measurement error of the NRS. Moderate quality evidence was found for sufficient structural validity and internal consistency of BPI-PS, inconsistent construct validity of BPI-PS, and inconsistent responsiveness of NRS. There was lower quality evidence or no studies on the other measurement properties of these 3 instruments (Appendix 2, available online as supplemental digital content at http://links.lww.com/PAIN/A511) (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review. Unpublished data). In the systematic review on HRQoL PROMs, very low quality evidence was found on the content validity of each PROM (Appendix 2, available online as supplemental digital content at http://links.lww.com/PAIN/A511) (Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review. Unpublished data). High quality evidence was found only for insufficient construct validity of EuroQol 5D (EQ-5D) utility and VAS scores. Moderate quality evidence was

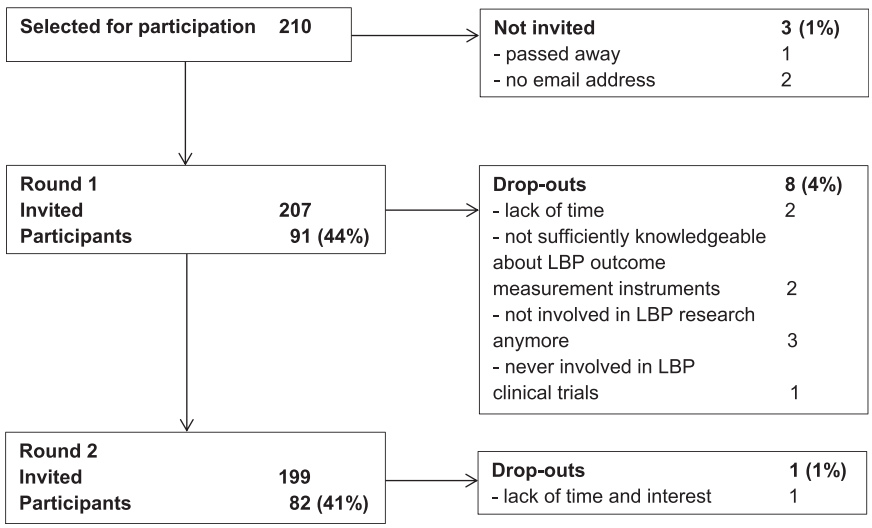

Figure 1. Flowchart of participants in the Delphi study on core outcome measurement instruments for clinical trials in nonspecific low back pain (LBP).

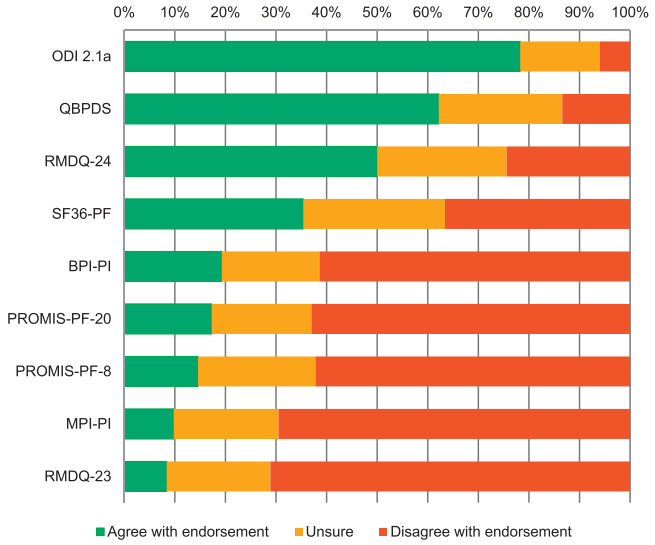

Figure 2. Delphi endorsement of 9 physical functioning tools as core outcome measurement instruments for clinical trials in nonspecific low back pain (round 1).

found for inconsistent construct validity of component summaries of the SF36 and for inconsistent responsiveness of the EQ-5D utility score. All other measurement properties were underpinned by lower quality evidence or not assessed (Appendix 2, available online as supplemental digital content at http://links.lww.com/PAIN/A511) (Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review. Unpublished data). A detailed presentation of results of these reviews is available elsewhere (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review. Unpublished data; Unpublished data; Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review. Unpublished data; and Ref. 18).

3.3. Delphi study

In total, 207 people were invited to participate in the Delphi study, and response rates in the 2 rounds were 44% and 41%, respectively (**Fig. 1**). Most participants were from the United States, the Netherlands, United Kingdom, and Australia; the most represented

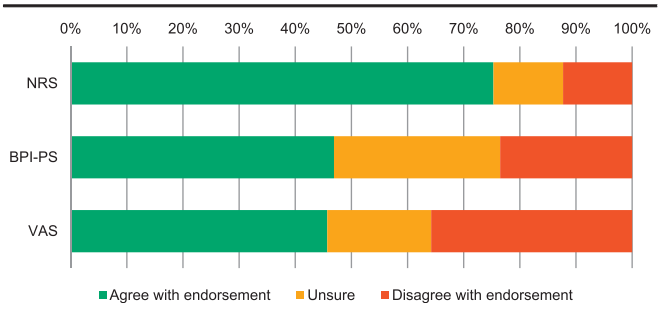

Figure 3. Delphi endorsement of 3 pain intensity tools as core outcome measurement instruments for clinical trials in nonspecific low back pain (round 1).

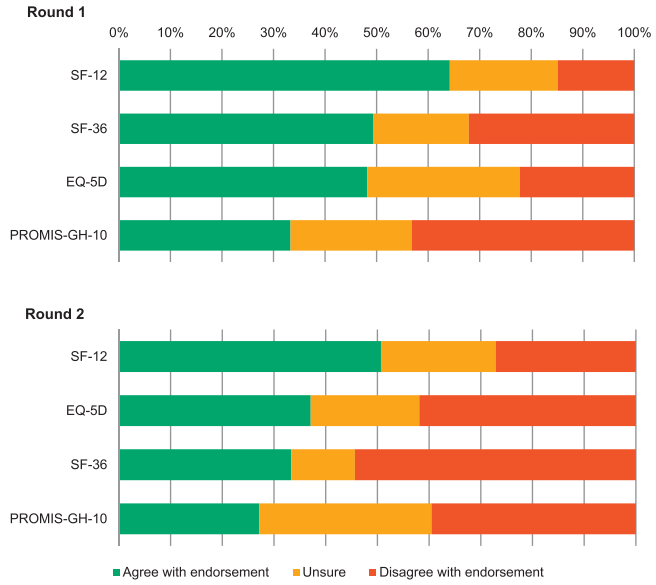

Figure 4. Delphi endorsement of 4 health-related tools as core outcome measurement instruments for clinical trials in nonspecific low back pain (round 1 and round 2).

disciplines were epidemiology, physical therapy, human movement sciences, psychology, and orthopedics (**Table 2**). In Round 1, 13 participants had LBP: 11 were male, mean (SD) age was 56 (8) years, 7 were classified as nsLBP by a health care professional, 11 with pain lasting for more than 1 year, 1 with pain spreading down the legs, and none having received a LBP operation or disability compensation. In round 2, 14 participants reported LBP with similar characteristics.

3.3.1. Delphi round 1

In the first round, there was a consensus (90%) to provisionally recommend core outcome measurement instruments, despite the absence of adequate evidence to support the PROMs' content validity. Several participants emphasized that core instruments should be recommended because COS development and/or PROM validity are moving fields in which results are always provisional, meaning that this should not refrain from providing recommendations on the best available instruments. There was also a consensus (90%) to reduce the list of potential core instruments for physical functioning because for 8 of them (ODI 1.0, Chiropractic Low Back Pain Disability Questionnaire [CLBPDQ], Modified Low Back Pain Disability Questionnaire [MLBPDQ], RMDQ-18, LBPRS, PROMIS-PF-4, PROMIS-PF-6, and PROMIS-PF-10), there were also convincing arguments for not being endorsed. Main reasons were that some of these PROMs were cross culturally adapted in very few languages (ie, ODI 1.0, CLBPDQ, MLBPDQ, and LBPRS-DI)[18] or they could be extracted from other instruments included in the list of potential core instruments (ie, RMDQ-18, PROMIS-PF-4, PROMIS-PF-6, and PROMIS-PF-10).

Regarding the remaining physical functioning PROMs, 78% of the panel agreed to endorse ODI 2.1a as a core outcome measurement instrument, whereas 71% and 70% agreed on not endorsing RMDQ-23 and MPI-PI, respectively (**Fig. 2**). No consensus was reached on the other 6 PROMs, with Quebec Back Pain Disability Scale (QBPDS) (62% in favor and 24% unsure) and RMDQ-24 (50% in favor and 26% unsure) being the

Table 2

Characteristics of participants in the Delphi study.

Characteristics	Round 1 (n = 91)	Round 2 (n = 82)
Complete answers on all questions, %	89	99
Nationality, %		
United States of America	26	28
Dutch	19	16
British	13	14
Australian	9	10
Italian	5	5
Spanish	3	2
Canadian	3	4
German	2	2
Brazilian	2	2
Danish	2	1
Finnish	2	2
Norwegian	2	4
Other	12	10
Work country, %		
United States of America	28	31
The Netherlands	19	15
United Kingdom	12	11
Australia	10	11
Canada	6	8
Brazil	3	4
Italy	3	4
Spain	3	2
Finland	2	2
Norway	2	4
Switzerland	2	2
Denmark	2	1
Germany	2	2
Other	6	3
Female, %	25	24
Current job activity, %		
Researcher	56	56
Researcher and clinician	36	36
Clinician	6	6
Patient representative	2	2
Current presence of LBP, %	14	17
Educational background, %*		
Physical therapy	35	32
Epidemiology	30	33
Human movement science	13	10
Psychology	12	9
Orthopaedics	11	12
Rheumatology	9	9
Physical medicine and rehabilitation	9	9
Internal medicine	8	7
Chiropractic	7	5
Occupational medicine	4	7
Statistics	4	1
General practice/family medicine	3	4
Osteopathy	2	1
Other	13	15
Current field of work, %*		
Epidemiology	40	43
Physical therapy	29	26
Physical medicine and rehabilitation	18	16
Orthopaedics	13	12
Psychology	11	7
Rheumatology	10	10
Internal medicine	7	6
Statistics	6	4
Human movement science	6	4
Chiropractic	3	2
Occupational medicine	3	4
Anaesthesiology	2	4
General practice/family medicine	2	1
Other	18	15
Experience with clinical trials in patients with nsLBP, %		
None	10	9
1 clinical trial	12	12
2 clinical trials	14	16
3 or more clinical trials	64	55
Missing information	0	9
Experience with systematic reviews in patients with nsLBP, %		
None	25	22
1 systematic review	17	15
2 systematic reviews	18	11
3 or more systematic reviews	40	45
Missing information	0	9
Experience in developing measurement instruments for patients with nsLBP, %		
None	46	45
1 measurement instrument	20	21
2 measurement instruments	10	6
3 or more measurement instruments	24	19
Missing information	0	9
Experience in assessing measurement properties of instruments in nsLBP, %		
None	26	23
1 study	19	18
2 studies	13	15
3 or more studies	41	35
Missing information	0	9
Experience in developing core outcome sets, %		
None	48	42
1 study	32	32
2 studies	11	9
3 or more studies	8	10
Missing information	0	9

* Each participant could indicate more than one discipline.
nsLBP, nonspecific low back pain.

second and third highest in endorsement (**Fig. 2**). These results were consistent across stakeholder groups.

For pain intensity, NRS was endorsed (75%), but the panel was split on BPI-PS (47% in favor and 24% against) and VAS (46% in favor and 36% against) (**Fig. 3**). For HRQoL, the panel was unsure for all included instruments, with the Short Form Health Survey 12 (SF12) being the closest to endorsement (64% in favor and 21% unsure) (**Fig. 4A**). Single participants suggested 11 additional potential instruments, whereas 2 participants suggested the PROMIS pain interference instrument. Two participants highlighted that the generic information supplied on the costs of using PROMs may not be correct and that more precise costs for each instrument should have been reported. Four participants expressed the concern that the instruments considered may be "dated" and 2 of these participants suggested that new instruments should be developed. Two other participants criticized our systematic reviews for pain intensity and HRQoL PROMs on the basis that they should have included studies in all pain conditions.

3.3.2. Delphi round 2

In the second round, the exact cost for the use of each instrument was presented together with information on characteristics and measurement properties. Given the inconsistency of suggestions for additional potential core instruments, none were added to this round. For physical functioning, because a consensus on endorsing ODI 2.1a was reached in Round 1, participants were asked whether they could see any major argument against its endorsement. Eleven participants responded that they were concerned with its fees (350€/study for funded academic research and 0€/study for nonfunded academic research),[3] arguing against any fee to use instruments for measuring core domains, expressing concerns that it could represent a barrier for funded academic research in low- and middle-income countries, and that fees might be increased once an instrument is recommended as core (Appendix 3, available online as supplemental digital content at http://links.lww.com/PAIN/A511). QBPDS and RMDQ-24 were presented again but no consensus was reached on their endorsement (ie, 54% in favor and 27% against for QBPDS, 52% in favor and 33% against for RMDQ-24).

For pain intensity, because a consensus on endorsing NRS was achieved in round 1, participants were asked whether they agreed on endorsing an NRS referring to "average LBP intensity over the last week" in the introductory statement (Appendix 1, available online as supplemental digital content at http://links.lww.com/PAIN/A511), similar to other recommendations for LBP.[24,31] A strong consensus (96%) was achieved on endorsing this NRS version. For HRQoL, results were similar to round 1, with the SF12 being the highest on endorsement (51% in favor and 22% unsure) (**Fig. 4B**). The main reasons against endorsing these instruments were overlap of their content with physical functioning and pain intensity instruments; scarce validity for measuring HRQoL for EQ-5D; unfamiliarity and lack of testing in nsLBP for PROMIS-GH-10; high costs for SF36 and SF12; and excessive length of SF36.

Various suggestions for the research agenda were made by the participants, with the most consistent being to investigate the measurement properties not fully assessed so far (9 participants), perform head-to-head comparison studies on measurement studies of recommended and not recommended PROMs (6), take PROMIS instruments more into account (4), develop a better outcome measurement instrument for LBP (3), develop a new instrument for HRQoL (2), develop an instrument for LBP that takes into account other constructs (eg, social participation) (2), use instruments that can be administered with computerized adaptive testing (CAT) (2), consider the recently developed Musculoskeletal Health Questionnaire[56] in future clinimetric studies (2), and assess the minimal important difference of the various instruments to explore whether it differs depending on patient characteristics and interventions (2).

3.4. Recommendations on core outcome measurement instruments

Considering the Delphi process results, the Steering Committee discussed and formulated a set of recommendations on measurement instruments to be used in nsLBP clinical trials (**Table 3**). This includes ODI 2.1a and NRS to measure physical functioning and pain intensity, respectively. Given the concerns of Delphi participants and some committee members on the ODI 2.1a fees, the instrument's distributor was contacted to ask iwhether it was possible to eliminate or reduce the ODI 2.1a fee for funded academic research. Because this was not possible, the Steering Committee decided to also recommend the RMDQ-24 for physical functioning because it achieved the highest level of consensus among the free-to-use instruments (**Fig. 2**), but also because its measurement properties resemble those of ODI 2.1a in head-to-head comparisons studies.[17] Despite a similar level of endorsement and measurement properties, the QBPDS was not recommended because of the same fee issue as the ODI 2.1a and also to limit the number of instruments for a single core domain.

The NRS with a 1-week recall period (Appendix 1, available online as supplemental digital content at http://links.lww.com/PAIN/A511) should be used to measure pain intensity in nsLBP trials. Because it is a free tool that obtained ample consensus in the Delphi, the Steering Committee does not recommend another instrument for pain intensity. However, researchers should note the limitations in its use for acute nsLBP trials when participants may have had pain for less than 1 week at baseline.[41,110] In these trials, the addition of an NRS with a 24-hour recall period is suggested.

Despite the lack of a consensus for measuring HRQoL, to reduce measurement variability for this domain, we recommend the use of the SF12 as it was closest to a consensus (**Fig. 4**), but because it is not free of charge, the PROMIS-GH-10 is also recommended (**Table 3**). Both PROMs provide a physical and a mental summary score (**Table 1**), which allows pooling of their results in meta-analysis. The SF36 is not recommended because of its length. The EQ-5D is not recommended because of its cost; it results in a utility index, which is not possible to pool with data from other instruments and its content is strongly redundant given the domains physical functioning and pain intensity. However, the Steering Committee suggests inclusion of the EQ-5D (preferably EQ-5D-5L version[55,65]) in nsLBP clinical trials if there is an economic evaluation.

No specific recommendations regarding time frames of outcome assessment and reporting of adverse events are made in line with the NIH Task Force Report for chronic LBP suggestion.[31] Time frames should match the specific goals and feasibility of each clinical trial. Potential adverse events should preferably be specified before the start of a clinical trial and measured prospectively. The Steering Committee suggests the use of previous consensus-based recommendations for reporting of outcome results[43] and for interpreting change scores on core instruments.[35,86]

4. Discussion

This study formulates recommendations on core outcome measurement instruments for use in nsLBP trials (**Table 3**). They comprise the ODI 2.1a or RMDQ-24 for physical functioning, NRS with a 1-week recall period for pain intensity, and SF12 or PROMIS-GH-10 for HRQoL. In addition, a simple statement reporting whether any death occurred in a clinical trial is recommended.[16] These recommendations update the previous LBP outcome recommendations of Deyo et al.[30] and Bombardier.[8] This COS applies to both acute and chronic nsLBP, and in the latter group, it complements the baseline research standards recommended by the NIH Task Force Report.[31]

4.1. Recommendations for future research

A recommended process that involved identification and review of measurement properties for candidate instruments and a consensus process for final selection was followed.[6,92] This core outcome measurement set is preliminary because high quality evidence is lacking for several measurement properties of various PROMs (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain:

Table 3

Core outcome measurement instruments for clinical trials in nonspecific low back pain.

Core outcome domain	Instrument	Free of charge?	Availability
Physical functioning	Oswestry Disability Index version 2.1a (ODI 2.1a)	Yes for not funded academic users; no for funded academic and commercial users	https://eprovide.mapi-trust.org/instruments/oswestry-disability-index
	24-item Roland Morris Disability Questionnaire (RMDQ-24)	Yes	http://www.rmdq.org/download.htm
Pain intensity	Numeric Rating Scale (NRS)*	Yes	Appendix 1
Health-related quality of life	Short Form Health Survey 12 (SF12)	No, costs are established on a per study basis	https://campaign.optum.com/optum-outcomes/what-we-do/health-surveys/sf-12v2-health-survey.html
	10-item PROMIS Global Health (PROMIS-GH-10)	Yes	http://www.healthmeasures.net/administrator/components/com_instruments/uploads/Global%20Health%20Scale%20v1.2%2008.22.2016.pdf
No. of deaths	A simple statement on the number of deaths occurring in the trial	Yes	

* A consensus was achieved on a Numeric Rating Scale version that refers to average low back pain intensity over the past week in the introductory statement, as presented in Appendix 1, available online as supplemental digital content at http://links.lww.com/PAIN/A511.

a systematic review. Unpublished data; Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review. Unpublished data; and Ref. 18). In particular, there is an urgent need to better assess and compare content validity, structural validity, reliability, and responsiveness of the recommended instruments with other instruments (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review. Unpublished data; Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review. Unpublished data; and Ref. 18). Developing a COS is an iterative process that should be updated if new evidence emerges on outcome domains or measurement instruments. Therefore, these recommendations are likely to evolve in the future.

Cross-cultural validity has not been investigated for the recommended instruments or other candidate PROMs (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review: Unpublished data; Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review: Unpublished data; and Ref. 18). This measurement property assesses whether the performance of the items on a translated or culturally adapted PROM is an adequate reflection of the performance of the items of the original version.[85] It can be evaluated using data from several countries to assess differential item functioning,[26,100] and it would give a clear indication on the appropriateness of pooling data on the same PROM from different countries.

4.1.1. Physical functioning

Roland-Morris Disability Questionnaire and ODI were included in earlier recommendations for physical functioning in LBP,[8,30] but this report gives more precise recommendations on which versions to use (**Table 3**). The International Consortium for Health Outcomes Measurement standard set for LBP also recommended ODI 2.1a to measure physical functioning because it "is the most heavily studied, providing superior interpretability" and "the most feasible to implement as it has been validated in 14 languages (…) and is relatively short."[24] One systematic review showed that from a measurement point of view, there are no strong reasons to prefer ODI 2.1a over RMDQ-24 in patients with nsLBP.[17] Moreover, the RMDQ-24 is available in some languages in which the ODI 2.1a is not.[18] There is high quality evidence suggesting that RMDQ-24 has limitations in key aspects of validity such as comprehensiveness and unidimensionality,[18] but its (content and structural) validity has never been directly compared with that of ODI 2.1a in the same group of patients with LBP.[17]

Direct head-to-head comparisons of instruments should be extended to include other recently suggested instruments to measure physical functioning in LBP (eg, QBPDS or PROMIS-PF short forms).[21,31,99] Comparing the content validity has the highest priority because this is the first measurement property that should be evaluated when selecting PROMs for a COS.[92] The measurement properties of PROMIS-PF instruments have been assessed in the generic population or in a heterogeneous spine or pain population,[10,25,27,40,60,61,88,95,97] but there is little evidence in patients with nsLBP. A recent study compared unidimensionality and item response theory performance of PROMIS-PF short forms with the RMDQ-24 in patients with chronic nsLBP, finding promising results in favor of PROMIS-PF short forms (Chiarotto et al., 2018. The 4-, 6-, 8- and 10-item PROMIS Physical Function short forms have better psychometric performance than the 24-item Roland Morris Disability Questionnaire: Unpublished data). It should be noted that there is a lively debate on the question whether generic instruments should be tested in each specific disease population or not.[78,113]

The PROMIS-PF item bank was also developed to administer computerized adaptive testing (CAT) forms (ie, PROMIS-PF-CAT[40]), however CAT instruments have not been considered for LBP outcome standardization because they are not yet feasible for use in every trial internationally. Nonetheless, researchers should also test CAT forms because CAT simulations were demonstrated to provide increased measurement efficiency and precision.[27,40] Some participants of this Delphi study suggested that new outcome measurement instruments should be developed for LBP, but we are hesitant

to suggest this as a high research priority because many PROMs to measure physical functioning are already available[48] and efforts may be better spent on generating evidence on the key measurement properties of these instruments.

4.1.2. Pain intensity

An NRS with a 1-week recall period has been repeatedly suggested as a key instrument for pain intensity in LBP,[21,24,31] and these previous suggestions strengthen our recommendation. Although the evidence base for this tool was of low quality in nsLBP (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review. Unpublished data). There is a larger body of evidence in other pain conditions suggesting that its measurement properties are satisfactory.[52,57,67] Nevertheless, pain-rating scales definitely present some shortcomings, such as capturing multiple dimensions of the pain experience, and not only its intensity.[39,93,109] For this reason, we decided to add the key word "intensity" in the recommended NRS, and more studies exploring the patients' perspective on these tools are needed. A few studies have directly compared the measurement performance of single-item NRS with that of multiitem instruments (eg, BPI-PS) and suggested that single-item instruments may be acceptable.[66,68,69]

4.1.3. Health-related quality of life

Reaching a consensus on a single instrument for HRQoL proved to be challenging. This highlights various issues with the domain and its instruments. Compared with physical functioning and pain intensity, HRQoL displayed a lower level of consensus for inclusion in this COS;[16] it has a broad definition, is multidimensional in nature, and has been less frequently assessed in LBP clinical trials.[46] Moreover, only the construct validity of commonly used PROMs has been adequately assessed in patients with nsLBP (Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review. Unpublished data). Low back pain is considered as a multidimensional biopsychosocial pain disorder,[90,106] and some authors have advocated the use of multidimensional instruments to fully capture the complexity of treatment response.[45,50] Health-related quality of life is a domain that meets the LBP multidimensional nature, and this may be a sufficient reason to make an effort to better define this domain for patients with nsLBP, taking into account all the aspects that impact and burden their life.[11,45] New back-specific or musculoskeletal-specific PROMs, such as instruments based on the International Classification of Functioning LBP core set[4] or the Musculoskeletal Health Questionnaire,[56] should be considered in future clinimetric studies for a direct comparison with the generic instruments recommended here.

4.2. Strengths and weaknesses

Overall, the main strengths of the current study are the thorough assessment of the measurement properties of candidate instruments (Chiarotto et al., 2018. Measurement properties of Numeric Rating Scale, Visual Analogue Scale and Pain Severity subscale of Brief Pain Inventory in patients with low back pain: a systematic review: Unpublished data; Chiarotto et al., 2018. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review: Unpublished data; and Ref. 18) and transparency in each stage of the study (eg, providing full feedback reports to Delphi participants). The systematic reviews were conducted according to the most recent COSMIN methodology (Prinsen et al., 2018. COSMIN guideline for systematic reviews of patient-reported outcome measures: Unpublished data; Terwee et al., 2018. COSMIN standards and criteria for evaluating the content validity of patient-reported outcome measures: a Delphi study: Unpublished data; and Refs. 82,85). and included a thorough assessment of the content validity of the instruments as well as information about their development phase. The Delphi participants were presented with summary information on the potential core instruments, including measurement properties and availability and, therefore, had the opportunity to make informed decisions, taking into account the instruments' content also. This is the first study to perform a consensus procedure on core outcome measurement instruments for nsLBP and the first one to use a Delphi survey to seek a consensus on instruments for any health condition. Another strength of this project is that the selected outcome domains and measurement instruments represent those for which there is a consensus across relevant stakeholders in the nsLBP field. Therefore, it is reasonable to suggest that these recommendations may also apply to observational studies or routine clinical practice.

A limitation of our study regards the Delphi panel selection. It included a selected sample of researchers, clinicians, and patient representatives that may not generalize to the whole LBP community. We attempted to be comprehensive in inviting participants and we have described the sample appropriately (**Table 2**), but our sample may not be fully representative. Another potential limitation is that "ordinary" patients were not involved in the consensus procedure. Nevertheless, it should be underlined that it remains unclear how patients can contribute to the selection of core instruments taking aspects like measurement properties into account, and methodological research in this field is lacking. In addition, all existing studies in which patients with nsLBP were asked about their perspective on the potential core instruments were included in the 3 systematic reviews and this became part of the content validity evidence synthesis presented in the Delphi survey. Another limitation may be that potential core instruments were selected among those most frequently used and recommended, potentially overlooking some more recent, less frequently used, and/or investigated tools; however, it should be also noted that PROMIS instruments were included in our consensus procedure to partly address this issue. Delphi open-ended questions were reviewed and categorized by only 1 reviewer with no double checking by a second one; this may also represent a potential limitation of this study.

5. Conclusions

In summary, this study has formulated a preliminary core outcome measurement set specifying instruments to be included in every clinical trial in patients with nsLBP (**Table 3**). These recommendations will be updated as further

evidence on the measurement properties of recommended and alternative instruments becomes available.

Conflict of interest statement

The authors have no conflict of interest to declare.

R. Buchbinder, C.-W.C. Lin, and C.G. Maher are supported by Australian National Health and Medical Research Council (NHMRC) Research Fellowships. N.E. Foster is supported by a UK National Institute for Health Research (NIHR) Research Professorship (NIHR-RP-011-015). These funding bodies did not have any role in designing the study, in collecting, analysing and interpreting the data, in writing this manuscript, and in deciding to submit it for publication.

Acknowledgements

The authors acknowledge the researchers, clinicians, and patients representatives who completed at least 1 round of the Delphi study, here listed in alphabetical order (members of the Steering Committee are excluded from this list): William A. Abdu, Gunnar Andersson, Adri T. Apeldoorn, Steven J. Atlas, Ralf Baron, Dorcas Beaton, Mark D. Bishop, Paul Bishop, David Borenstein, Alan Breen, Cristina Cabral, Christine Cedraschi, Roger Chou, Robin Christensen, Steven P. Cohen, Pierre Coté, Peter Croft, Ric Day, Rob de Bie, Anthony Delitto, Henrika C.W. de Vet, Clermont E. Dionne, Kate Dunn, Wendy T. Enthoven, John T. Farrar, Silvano Ferrari, Timothy W. Flynn, Julie Fritz, Robert Froud, Robert J. Gatchel, Andrew John Haig, Mark Hancock, Ian Harris, Jan Hartvigsen, Martijn W. Heymans, Jan Hildebrandt, Eric L. Hurwitz, Wilco C. Jacobs, Steven J. Kamper, Jaro Karppinen, Francis J. Keefe, Peter Kent, Robert D. Kerns, Jane Latimer, Charlotte Leboeuf-Yde, Martyn Lewis, Patrick Loisel, Pim A.J. Luijsterburg, Jon D. Lurie, Luciana Macedo, Anne Mannion, James McAuley, Alison McGregor, Luciola Menezes Costa, Stephan Milosavljevic, Marco Monticone, Peter O'Sullivan, Tamar Pincus, Serge Poiraudeau, James Rainville, Ana Royuela, Jesus Seco Calvo, Marcus Schiltenwolf, Gay Schoene, William S. Shaw, Karen J. Sherman, Shannon Smith, Matthew Smuck, Bart Staal, Simon Somerville, Kjersti Storheim, Liv Inger Strand, Simo Taimela, Peter Tugwell, Martin Underwood, Danielle van der Windt, Hans van Helvoirt, Willem van Mechelen, Arianne Verhagen, Steven Vogel, and Gustavo Zanoli. The authors also acknowledge the EUROSPINE Task Force Research for providing funding for this study (EUROSPINE TFR 5-2015).

Appendix A. Supplemental digital content

Supplemental digital content associated with this article can be found online at http://links.lww.com/PAIN/A511.

Article history:
Received 7 August 2017
Received in revised form 9 November 2017
Accepted 15 November 2017
Available online 24 November 2017

References

[1] PROMIS Instrument Development and Validation Scientific Standards version 2.0. 2013. p. 1–72. Available at: http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf. Accessed 5 August 2017.

[2] A brief guide to the PROMIS Physical Function instruments. 2015. Available at: https://assessmentcenter.net/documents/PROMIS%20Physical%20Function%20Scoring%20Manual.pdf. Accessed 5 August 2017.

[3] Oswestry Disability Index. 2017. Available at: https://eprovide.mapi-trust.org/instruments/oswestry-disability-index. Accessed 5 August 2017.

[4] Bagraith KS, Strong J, Meredith PJ, McPhail SM. Rasch analysis supported the construct validity of self-report measures of activity and participation derived from patient ratings of the ICF low back pain core set. J Clin Epidemiol 2017;84:161–72.

[5] Baker D, Pynsent P, Fairbank J. The Oswestry Disability Index revisited: its reliability, repeatability and validity, and a comparison with the St Thomas's Disability Index. In: Roland M, Jenner J, editors. Back pain: new approaches to rehabilitation and education. Manchester: Manchester University Press, 1989. p. 174–86.

[6] Boers M, Kirwan JR, Tugwell P, Beaton DE, Bingham CO III, Conaghan PG, D'Agostino MA, de Wit M, Gossec L, March L, Simon LS, Singh JA, Strand V, Wells GA. The OMERACT handbook. 2017. Available at: https://www.omeract.org/pdf/OMERACT_Handbook.pdf. Accessed 9 May 2017.

[7] Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, Conaghan PG, Bingham CO, Brooks P, Landewé R, March L, Simon LS, Singh JA, Strand V, Tugwell P. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol 2014;67:745–53.

[8] Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine (Phila Pa 1976) 2000;25:3100–3.

[9] Brooks R; EuroQol Group. EuroQol: the current state of play. Health policy 1996;37:53–72.

[10] Bruce B, Fries JF, Ambrosini D, Lingala B, Gandek B, Rose M, Ware JE. Better assessment of physical function: item improvement is neglected but essential. Arthritis Res Ther 2009;11:R191.

[11] Buchbinder R, Batterham R, Elsworth G, Dionne CE, Irvin E, Osborne RH. A validity-driven approach to the understanding of the personal and societal burden of low back pain: development of a conceptual and measurement model. Arthritis Res Ther 2011;13:R152.

[12] Castellini G, Gianola S, Banfi G, Bonovas S, Moja L. Mechanical low back pain: secular trend and intervention topics of randomized controlled trials. Physiother Can 2016;68:61–3.

[13] Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol 2010;63:1179–94.

[14] Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007;45(5 suppl 1):S3.

[15] Chapman JR, Norvell DC, Hermsmeyer JT, Bransford RJ, DeVine J, McGirt MJ, Lee MJ. Evaluating common outcomes for measuring treatment success for chronic low back pain. Spine (Phila Pa 1976) 2011;36:S54–68.

[16] Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, Costa LO, Foster NE, Grotle M, Koes BW, Kovacs FM, Lin CW, Maher CG, Pearson AM, Peul WC, Schoene ML, Turk DC, van Tulder MW, Ostelo RW. Core outcome domains for clinical trials in non-specific low back pain. Eur Spine J 2015;24:1127–42.

[17] Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P, Ostelo RW. 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.

[18] Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in low back pain. J Clin Epidemiol 2018;95:73–93.

[19] Chiarotto A, Ostelo RW, Turk DC, Buchbinder R, Boers M. Core outcome sets for research and clinical practice. Braz J Phys Ther 2017;21:77–84.

[20] Chiarotto A, Terwee CB, Deyo RA, Boers M, Lin CWC, Buchbinder R, Corbin TP, Costa LO, Foster NE, Grotle M, Koes BW, Kovacs FM, Maher CG, Pearson AM, Peul WC, Schoene ML, Turk DC, van Tulder MW, Ostelo RW. A core outcome set for clinical trials on non-specific low back pain: study protocol for the development of a core domain set. Trials 2014;15:511.

[21] Chiarotto A, Terwee CB, Ostelo RW. Choosing the right outcome measurement instruments for low back pain. Best Pract Res Clin Rheumatol 2016;30:1003–20.

[22] Cleeland CS. The Brief Pain Inventory. 2009. Available at: https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf. Accessed 9 May 2017.
[23] Cleeland CS, Ryan K. Pain assessment: global used of the Brief Pain Inventory. Ann Acad Med Singapore 1994;23:129–38.
[24] Clement RC, Welander A, Stowell C, Cha TD, Chen JL, Davies M, Fairbank JC, Foley KT, Gehrchen M, Hagg O, Jacobs WC, Kahler R, Khan SN, Lieberman IH, Morisson B, Ohnmeiss DD, Peul WC, Shonnard NH, Smuck MW, Solberg TK, Stromqvist BH, Hooff ML, Wasan AD, Willems PC, Yeo W, Fritzell P. A proposed set of metrics for standardized outcome reporting in the management of low back pain. Acta Orthop 2015;86:523–33.
[25] Cook KF, Jensen SE, Schalet BD, Beaumont JL, Amtmann D, Czajkowski S, Dewalt DA, Fries JF, Pilkonis PA, Reeve BB. PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. J Clin Epidemiol 2016;73:89–102.
[26] Crane PK, Gibbons LE, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques. DIFdetect and difwithpar. Med Care 2006;44(11 suppl 3):S115–123.
[27] Crins MHP, Terwee CB, Klausch T, Smits N, de Vet HCW, Westhovens R, Cella D, Cook KF, Revicki DA, van Leeuwen J, Boers M, Dekker J, Roorda LD. The Dutch-Flemish PROMIS Physical Function item bank exhibited strong psychometric properties in patients with chronic pain. J Clin Epidemiol 2017;87:47–58.
[28] Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. Pain 1983;17:197–210.
[29] DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. Med Care 2007;45(5 suppl 1):S12.
[30] Deyo RA, Battie M, Beurskens A, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcome measures for low back pain research: a proposal for standardized use. Spine (Phila Pa 1976) 1998;23:2003–13.
[31] Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, Carrino J, Chou R, Cook K, DeLitto A, Goertz C, Khalsa P, Loeser J, Mackey S, Panagis J, Rainville J, Tosteson T, Turk D, Von Korff M, Weiner DK. Report of the NIH Task Force on research standards for chronic low back pain. J Pain 2014;15:569–85.
[32] Dieleman JL, Baral R, Birger M, Bui AL, Bulchis A, Chapin A, Hamavid H, Horst C, Johnson EK, Joseph J. US spending on personal health care and public health, 1996–2013. JAMA 2016;316:2627–46.
[33] Downie W, Leatham P, Rhind V, Wright V, Branco J, Anderson J. Studies with pain rating scales. Ann Rheum Dis 1978;37:378–81.
[34] Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, Kerns RD, Stucki G, Allen RR, Bellamy N, Carr DB, Chandler J, Cowan P, Dionne R, Galer BS, Hertz S, Jadad AR, Kramer LD, Manning DC, Martin S, McCormick CG, McDermott MP, McGrath P, Quessy S, Rappaport BA, Robbins W, Robinson JP, Rothman M, Royal MA, Simon L, Stauffer JW, Stein W, Tollett J, Wernicke J, Witter J. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. PAIN 2005;113:9–19.
[35] Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC, McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105–21.
[36] EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199–208.
[37] Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, Hansson P, Haroutounian S, Arendt-Nielsen L, Attal N, Baron R, Brell J, Bujanover S, Burke LB, Carr D, Chappell AS, Cowan P, Etropolski M, Fillingim RB, Gewandter JS, Katz NP, Kopecky EA, Markman JD, Nomikos G, Porter L, Rappaport BA, Rice AS, Scavone JM, Scholz J, Simon LS, Smith SM, Tobias J, Tockarshewsky T, Veasley C, Versavel M, Wasan AD, Wen W, Yarnitsky D. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. PAIN 2016;157:1851–71.
[38] Fairbank J, Couper J, Davies J, O'brien J. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;66:271–3.
[39] Franchignoni F, Salaffi F, Tesio L. How should we use the visual analogue scale (VAS) in rehabilitation outcomes? I: How much of what? The seductive VAS numbers are not true measures. J Rehabil Med 2012;44:798–9.
[40] Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. J Rheumatol 2009;36:2061–6.
[41] Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. Spine (Phila Pa 1976) 2003;28:1363–71.
[42] Fritz JM, Irrgang JJ. A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec back pain disability scale. Phys Ther 2001;81:776.
[43] Froud R, Eldridge S, Kovacs F, Breen A, Bolton J, Dunn K, Fritz J, Keller A, Kent P, Lauridsen HH, Ostelo R, Pincus T, van Tulder M, Vogel S, Underwood M. Reporting outcomes of back pain trials: a modified Delphi study. Eur J Pain 2011;15:1068–74.
[44] Froud R, Patel S, Rajendran D, Bright P, Bjørkli T, Buchbinder R, Eldridge S, Underwood M. A systematic review of outcome measures use, analytical approaches, reporting methods, and publication volume by year in low back pain trials published between 1980 and 2012: respice, adspice, et prospice. PLoS One 2016;11:e0164573.
[45] Froud R, Patterson S, Eldridge S, Seale C, Pincus T, Rajendran D, Fossum C, Underwood M. A systematic review and meta-synthesis of the impact of low back pain on people's lives. BMC Musculoskelet Disord 2014;15:50.
[46] GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388:1545–602.
[47] Gianola S, Frigerio P, Agostini M, Bolotta R, Castellini G, Corbetta D, Gasparini M, Gozzer P, Guariento E, Li LC. Completeness of outcomes description reported in low back pain rehabilitation interventions: a survey of 185 randomized trials. Physiother Can 2016;68:267–74.
[48] Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing important health outcomes for comparative effectiveness research: an updated review and user survey. PLoS One 2016;11:e0146444.
[49] Grotle M, Brox JI, Vøllestad NK. Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. Spine (Phila Pa 1976) 2005;30:130–40.
[50] Hancock MJ, Hill JC. Are small effects for back pain interventions really surprising? J Orthop Sports Phys Ther 2016;46:317–19.
[51] Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000;32:1008–15.
[52] Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent And Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res 2011;63:S240–52.
[53] Hayden JA, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. Cochrane Database Syst Rev 2005:CD000335.
[54] Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009;18:873–80.
[55] Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20:1727–36.
[56] Hill JC, Kang S, Benedetto E, Myers H, Blackburn S, Smith S, Dunn KM, Hay E, Rees J, Beard D, Glyn-Jones S, Barker K, Ellis B, Fitzpatrick R, Price A. Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. BMJ Open 2016;6:e012331.
[57] Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, Fainsinger R, Aass N, Kaasa S; European Palliative Care Research Collaborative (EPCRC). Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manag 2011;41:1073–93.
[58] Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, Woolf A, Vos T, Buchbinder R. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012;64:2028–37.

[59] Hudson-Cook N, Tomes-Nicholson K, Breen A. The revised Oswestry low-back pain disability questionnaire. In: Roland M, Jenner J, editors. Back pain: new approaches to rehabilitation and Education. New York: Manchester University Press, 1989. p. 187–204.
[60] Hung M, Clegg DO, Greene T, Saltzman CL. Evaluation of the PROMIS physical function item bank in orthopaedic patients. J Orthop Res 2011; 29:947–53.
[61] Hung M, Hon SD, Franklin JD, Kendall RW, Lawrence BD, Neese A, Cheng C, Brodke DS. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. Spine (Phila Pa 1976) 2014;39:158–63.
[62] Hunt SM, McKenna S, McEwen J, Williams J, Papp E. The Nottingham Health Profile: subjective health status and medical consultations. Social Sci Med A 1981;15:221–9.
[63] Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E. Consolidated health economic evaluation reporting standards (CHEERS) statement. Cost Eff Resour Alloc 2013;11:6.
[64] Huskisson E. Measurement of pain. Lancet 1974;304:1127–31.
[65] Janssen M, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Swinburn P, Busschbach J. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res 2013;22:1717–27.
[66] Jensen MP, Hu X, Potts SL, Gould EM. Single vs composite measures of pain intensity: relative sensitivity for detecting treatment effects. PAIN 2013;154:534–8.
[67] Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R, editors. Handbook of pain assessment. New York: The Guilford Press, 1992. p. 19–41.
[68] Jensen MP, Tome-Pires C, Sole E, Racine M, Castarlenas E, de la Vega R, Miro J. Assessment of pain intensity in clinical trials: individual ratings vs composite scores. Pain Med 2015;16:141–8.
[69] Jensen MP, Wang W, Potts SL, Gould EM. Reliability and validity of individual and composite recall pain measures in patients with cancer. Pain Med 2012;13:1284–91.
[70] Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, Guzman J, van Tulder MW. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev 2014:CD000963.
[71] Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). PAIN 1985;23:345–56.
[72] Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ 2010;340:c365.
[73] Kirkham JJ, Gargon E, Clarke M, Williamson PR. Can a core outcome set improve the quality of systematic reviews?–a survey of the Co-ordinating Editors of Cochrane Review Groups. Trials 2013;14:21.
[74] Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Moher D, Schmitt J, Tugwell P. Core outcome set–standards for reporting: the COS-STAR statement. PLos Med 2016;13:e1002148.
[75] Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI. The Quebec back pain disability scale: conceptualization and development. J Clin Epidemiol 1996;49:151–61.
[76] Lambeek LC, van Tulder MW, Swinkels IC, Koppes LL, Anema JR, van Mechelen W. The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. Spine (Phila Pa 1976) 2011;36:1050–8.
[77] Maas ET, Ostelo RW, Niemisto L, Jousimaa J, Hurri H, Malmivaara A, van Tulder MW. Radiofrequency denervation for chronic low back pain. Cochrane Database Syst Rev 2015:CD008572.
[78] Magasi S, Ryan G, Revicki D, Lenderking W, Hays RD, Brod M, Snyder C, Boers M, Cella D. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. Qual Life Res 2012; 21:739–46.
[79] Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet 2017;389:736–47.
[80] Manniche C, Asmussen K, Lauritsen B, Vinterberg H, Kreiner S, Jordan A. Low Back Pain Rating scale: validation of a tool for assessment of low back pain. PAIN 1994;57:317–26.
[81] Meade T, Browne W, Mellows S, Townsend J, Webb J, North W, Frank A, Fyfe I, Williams K, Lowe L. Comparison of chiropractic and hospital outpatient management of low back pain: a feasibility study. J Epidemiol Commun Health 1986;40:12–17.
[82] Mokkink LB, de Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN checklist 2.0 for assessing the methodological quality of studies on the measurement properties of Paitent-Reported Outcome Measures. Qual Life Res 2017. doi:10.1007/s11136-017-1765-4 [Epub ahead of print].
[83] Mokkink LB, Prinsen CA, Bouter LM, de Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther 2016;20:105–13.
[84] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539–49.
[85] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–45.
[86] Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008;33:90–4.
[87] Page MJ, Huang H, Verhagen AP, Buchbinder R, Gagnier JJ. Identifying a core set of outcome domains to measure in clinical trials for shoulder disorders: a modified Delphi study. RMD Open 2016;2:e000380.
[88] Papuga MO, Mesfin A, Molinari R, Rubery PT. Correlation of PROMIS physical function and pain CAT instruments with Oswestry Disability Index and Neck Disability Index in spine patients. Spine (Phila Pa 1976) 2016;41:1153–9.
[89] Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. Spine (Phila Pa 1976) 1995;20:1899–908.
[90] Pincus T, Kent P, Bronfort G, Loisel P, Pransky G, Hartvigsen J. Twenty-five years with the biopsychosocial model of low back pain—is it time to celebrate? A report from the twelfth international forum for primary care research on low back pain. Spine (Phila Pa 1976) 2013;38:2118–23.
[91] Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, Williamson PR, Terwee CB. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set"—a practical guideline. Trials 2016;17:449.
[92] Robinson-Papp J, George MC, Dorfman D, Simpson DM. Barriers to chronic pain measurement: a qualitative study of patient perspectives. Pain Med 2015;16:1256–64.
[93] Roland M, Morris R. A study of the natural history of back pain: part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976) 1983;8:141–4.
[94] Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014;67:516–26.
[95] Salinas J, Sprinkhuizen SM, Ackerson T, Bernhardt J, Davie C, George MG, Gething S, Kelly AG, Lindsay P, Liu L, Martins SC, Morgan L, Norrving B, Ribbers GM, Silver FL, Smith EE, Williams LS, Schwamm LH. An international standard set of patient-centered outcome measures after stroke. Stroke 2016;47:180–6.
[96] Schalet BD, Hays RD, Jensen SE, Beaumont JL, Fries JF, Cella D. Validity of PROMIS physical function measured in diverse clinical samples. J Clin Epidemiol 2016;73:112–18.
[97] Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. PLos Med 2011;8:e1000393.
[98] Smeets R, Köke A, Lin CW, Ferreira M, Demoulin C. Measures of function in low back pain/disorders: low back pain rating scale (LBPRS), Oswestry disability index (ODI), progressive isoinertial lifting evaluation (PILE), Quebec back pain disability scale (QBPDS), and Roland-Morris disability questionnaire (RDQ). Arthritis Care Res 2011;63:S158–73.
[99] Stark S, Chernyshenko OS, Drasgow F. Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. J Appl Psychol 2006;91:1292–306.
[100] Stewart A, Kamberg C. Physical functioning measures. In: Stewart A, Ware JE Jr, editors. Measuring functioning and well-being: the Medical Outcomes Study approach. Durham: Duke University Press, 1992. p. 86–101.
[101] Stratford PW, Binkley JM. Measurement properties of the RM-18: a modified version of the Roland-Morris disability scale. Spine (Phila Pa 1976) 1997;22:2416–21.
[102] Taylor AM, Phillips K, Patel KV, Turk DC, Dworkin RH, Beaton D, Clauw DJ, Gignac MA, Markman JD, Williams DA, Bujanover S, Burke LB, Carr DB, Choy EH, Conaghan PG, Cowan P, Farrar JT, Freeman R, Gewandter J, Gilron I, Goli V, Gover TD, Haddox JD, Kerns RD, Kopecky EA, Lee DA, Malamut R, Mease P, Rappaport BA, Simon LS, Singh JA, Smith SM, Strand V, Tugwell P, Vanhove GF, Veasley C, Walco GA,

Wasan AD, Witter J. Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. PAIN 2016;157:1836–50.
[103] Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials 2007;8:38.
[104] Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, Knipschild PG. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998;51:1235–41.
[105] Waddell G. 1987 Volvo Award in Clinical Sciences: a new clinical model for the treatment of low-back pain. Spine (Phila Pa 1976) 1987;12: 632–44.
[106] Ware JE Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220–33.
[107] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473–83.
[108] Williams ACdC, Davies HTO, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. PAIN 2000;85:457–63.
[109] Williams CM, Maher CG, Latimer J, McLachlan AJ, Hancock MJ, Day RO, Lin CWC. Efficacy of paracetamol for acute low-back pain: a double-blind, randomised controlled trial. Lancet 2014;384: 1586–96.
[110] Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. J Health Serv Res Pol 2012;17:1–2.
[111] Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, Clarke M, Gargon E, Gorst S, Harman N, Kirkham JJ, McNair A, Prinsen CAC, Schmitt J, Terwee CB, Young B. The COMET handbook: version 1.0. Trials 2017;18(suppl 3):280.
[112] Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13:132.
[113] Witter JP. Introduction: PROMIS a first look across diseases. J Clin Epidemiol 2016;73:87.