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

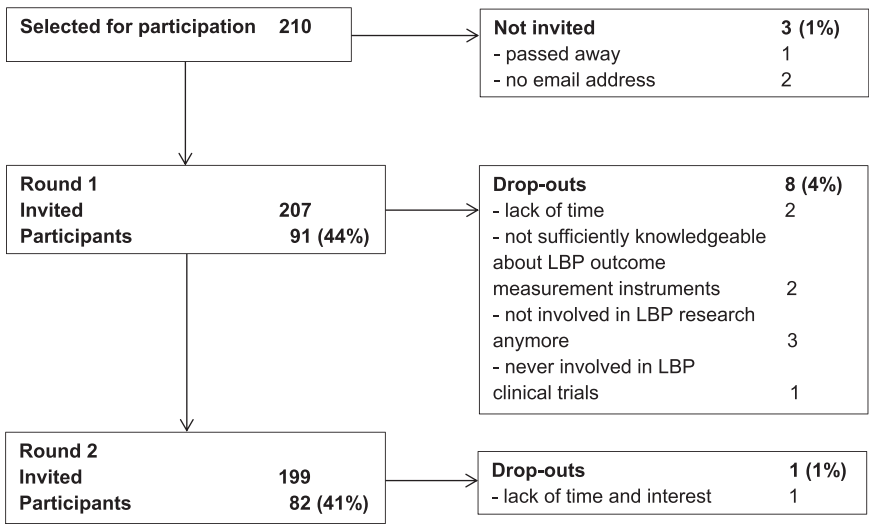

Figure 1. Flowchart of participants in the Delphi study on core outcome measurement instruments for clinical trials in nonspecific low back pain (LBP).