PAIN®

Mixed-methods development of a new patient-reported outcome instrument for chronic low back pain: part 2—The Patient Assessment for Low Back Pain–Impacts (PAL-I)

Donald M. Bushnell[a,*], Steven I. Blum[b], Hiltrud Liedgens[c], Mona L. Martin[a], Rainer Freynhagen[d], Mark Wallace[e], Charles Argoff[f], Mariëlle Eerdekens[c], Maurits Kok[c], Donald L. Patrick[g]

Abstract

We describe qualitative and quantitative development and preliminary validation of the Patient Assessment for Low Back Pain–Impacts (PAL-I), a patient-reported outcome measure for use in chronic low back pain (cLBP) clinical trials. Concept elicitation and cognitive interviews (qualitative methods) were used to identify and refine symptom concepts. Classical test theory and Rasch measurement theory (quantitative methods) were used to evaluate item-level and scale-level performance of the PAL-I using an iterative approach between qualitative and quantitative methods. Patients with cLBP participated in concept elicitation interviews (N = 43), cognitive interviews (N = 38), and assessment of paper-to-electronic format equivalence (N = 8). A web-based sample of self-reported patients with cLBP participated in quantitative studies to evaluate preliminary (N = 598) and revised (n = 401) drafts and patients with physician-diagnosed cLBP (N = 45) participated in preliminary validation of the PAL-I. The instrument contained 9 items describing cLBP impacts (walking, sitting, standing, lifting, sleep, social activities, travelling, climbing, and body movements). Item-level performance, scale structure, and scoring seemed to be appropriate. One-week test–retest reproducibility was acceptable (intraclass correlation coefficient 0.88 [95% confidence interval, 0.78-0.94]). Convergent validity was demonstrated with PAL-I total score and Roland-Morris Disability Questionnaire (Pearson correlation 0.82), MOS-36 Physical Functioning (−0.71), and MOS-36 Bodily Pain (−0.71). Individual item scores and total score discriminated between numeric rating scale tertile groups and painDETECT categories. Interpretation of paper and electronic administration modes was equivalent. The PAL-I demonstrated content validity and is potentially useful to assess treatment benefit in clinical trials of cLBP therapies.

Keywords: Low back pain, Patient-reported outcomes, Qualitative research

1. Introduction

Low back pain (LBP) is described as "pain, muscle tension, or stiffness localized below the costal margin and above the gluteal folds, with or without leg pain."[1,5,20] Low back pain is a common and painful condition that has been reported to have a global prevalence up to 84% depending on the case definition.[2] The prevalence of chronic LBP (cLBP), defined as LBP with duration >12 weeks,[21] is estimated to be ~23%.[2] Low back pain causes more global disability than any other condition[10] and is frequently work-related. Treatments for cLPB include nonsteroidal anti-inflammatory drugs, acetaminophen, muscle relaxants, opioid pain medications, and physical therapy, which are effective for most patients.[11] For moderate-to-severe pain, epidural steroid injections, facet injections, and radiofrequency ablations are used.[11] Several promising new therapies are in phase 2/3 clinical trials for the treatment of cLBP, including ethanol gel for use in chemonucleolysis protocols (a process of injecting proteolytic enzymes into the intervertebral disk), platelet-rich plasma, stem-cell therapy for disk regeneration, tanezumab (an anti–nerve growth factor antibody) for nociceptor modulation, artemin (a neurotrophic growth factor) for neuronal regrowth,[11] and cebranopadol, a nociceptin/orphanin FQ peptide and opioid peptide receptor agonist.[6]

Patient-reported outcome (PRO) measures are used to evaluate patient perspective in clinical trials. Patient-reported outcome measures capture the health experience of patients and can be used to document treatment benefit and support labeling claims.[17,19] Many PRO measures/instruments are currently used in clinical trials of cLBP, including the Neuropathic Pain Symptom Inventory (NPSI), painDETECT, Roland-Morris Disability Questionnaire (RMDQ), Pain

Quality Assessment Scale–Revised, Revised Short-Form McGill Pain Questionnaire, Low Back Pain Impact Questionnaire, Oswestry Disability Index, Pain Disability Index, Brief Pain Inventory and Brief Pain Inventory–Short Form, Musculoskeletal Outcomes Data Evaluation and Management System Spine Module, Orebro Musculoskeletal Pain Questionnaire, and the West Haven-Yale Multidimensional Pain Inventory Interference Scale. Most of the instruments measure pain and/or disability, hallmarks of cLBP.

The U.S. Food and Drug Administration (FDA) has published a guidance for drug sponsors outlining requirements for the development of PRO measures to ensure that they reliably measure the claimed concept in the patient population enrolled in the clinical trial.[19] We conducted a literature search to identify PRO instruments that could be used for a label claim for cLBP and found that most available measures were not developed in accordance with the U.S. FDA guideline,[18] particularly the requirement to demonstrate relevance to the intended patient population. This requires initial qualitative evidences, but these were not found in the published literature for the measures that were evaluated. Several measures lacked specificity to a cLBP population, and others had difficulties that were apparent in the clarity of item construction and the singularity of concepts presented. Therefore, we have developed and conducted preliminary validation studies on 2 PRO instruments for use in cLPB: the Patient Assessment for Low Back Pain–Symptoms (PAL-S)[14] and the PAL–Impacts (PAL-I). We now report the mixed-methods development of the PAL-I.

2. Methods

2.1. Study design and development steps for PAL-I questionnaire

The PAL-I was developed in parallel with the PAL-S, and the study design and development have been fully described.[14] Briefly, a mixed-methods approach, which includes both qualitative and quantitative methods, was used to finalize the questionnaires (**Fig. 1**). Qualitative methods were used to generate and refine the concepts included in the PRO, and quantitative methods were used to evaluate the concepts in the PRO. Because of the iterative nature (alternating use of qualitative and quantitative methods as they inform each other) of this approach, the methods and results are presented in chronological order.

A survey of the LBP literature was conducted to identify impact concepts and to review currently available instruments that could be relevant to patients with cLBP. Although many instruments were identified, none was sufficient for use in labeling claims[18] based on FDA guidance.[19] Qualitative methods used to develop the PAL-I included concept elicitation (CE) interviews and cognitive interviews, and the quantitative development was based on administration of the instrument to a large sample population with cLBP. The equivalence of paper and electronic platforms was also established using cognitive interviews.

This study was performed in accordance with Good Clinical Practice and applicable regulatory requirements. The study was approved by the Quorum Review Institutional Review Board for sites in the United States. For sites in the EU, market research facilities in Germany and the United Kingdom recruited patients from their databases and used their standard consent form process. No medical information or records were accessed or used with these groups, and therefore, no formal ethics review was required.

2.2. Patient population and recruitment

Patient recruitment included patients with cLBP across a spectrum of pain severity as well as demographic characteristics, and

the sample populations have been described.[14] The PAL-I is intended for use in global clinical trials. To capture global patient experiences, patients were recruited from 5 clinical sites in the United States, 1 site in Germany, and 1 site in the United Kingdom. Seven cohorts of patients were recruited for (1) CE to identify impact concepts (N = 43); (2) first round of cognitive interviews to evaluate the instrument (N = 30); (3) second round of cognitive interviews to evaluate modifications to the instrument (N = 8); (4) cognitive interviews to evaluate paper to electronic equivalence (N = 8); (5) first quantitative administration (N = 598); (6) second quantitative administration (a subset of the first quantitative administration; n = 401); and (7) preliminary psychometric validation (N = 45).

Patients with a current pain score ≥4 on a 0- to 10-point numeric rating scale (NRS; moderate-to-severe pain) were
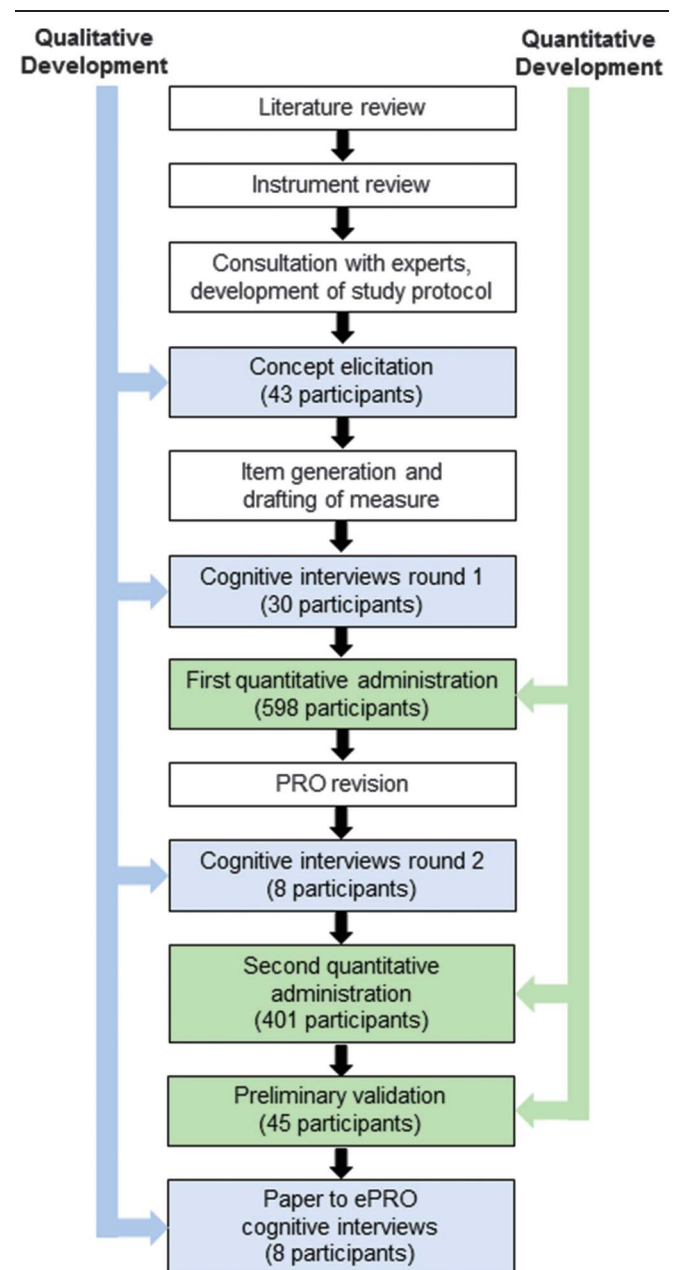
Figure 1. Steps in mixed-methods development of the PAL-I. Qualitative methods (blue) and quantitative methods (green) are shown for chronological steps in the development of the PAL-I. (Reproduced from Martin et al.[14])

eligible to participate in CE, cognitive interviews, and preliminary psychometric validation, and patients with pain NRS scores of 0 to 10 were eligible to participate in quantitative administrations. Pain likely to be neuropathic LBP was identified based on a painDETECT[8] score >19 during screening[9]; painDETECT scores were descriptive only and used to guide recruitment targets (at least 40% with neuropathic pain component in the quantitative population).

2.3. Concept elicitation interviews

Concept elicitation interviews were conducted using an interview guide that was designed to obtain both spontaneous and prompted input from patients about the impact of cLBP symptoms, as previously described for the PAL-S.[14] Concept elicitation interviews were audio recorded and transcribed, and the transcripts were coded and organized using Atlas.ti software[15] to develop a coding framework. The coding framework was used to organize and group information with similar content and was revised as needed to fit the concept information from the interview transcripts.

2.4. Quality indicators for qualitative data

Saturation of concept occurs when no new information is forthcoming from the CE interviews. Transcripts were chronologically ordered and grouped into quartiles of 10 and 11 transcripts. New concept codes for each subsequent transcript group were compared with codes from the preceding group until no new information was forthcoming from subsequent interviews, indicating saturation of concept.

Codes were identified and coded by 2 coders. The degree of consistency between coders was evaluated by dual coding of 10% of the transcript database and comparing each pair of coded transcripts for differences.

2.5. Item generation

Items to be included in the instrument were identified from expert input, a review of existing cLBP PRO instruments,[18] and CE interviews. Draft items were constructed to be at the appropriate reading level using patient-derived language and terminology.

2.6. Cognitive interviews (round 1) for instrument refinement

Cognitive interviews were conducted to evaluate the accuracy and consistency of patient comprehension of concepts presented in the draft items.[16] Patients were also asked about instrument instructions, response options, the recall period, and specific terminology.

2.7. First quantitative administration (pilot data collection)

The first quantitative administration was conducted to evaluate item performance and determine whether any items required revision. Classical test theory was used for item reduction, and included evaluations of missing data, ceiling and floor effects, item-to-item correlations, item-to-total correlations, factor analysis, and estimations of reliability. Rasch measurement theory (RMT) analyses were used to examine the measurement model and scoring of the PAL-I instrument, and to assess item-level performance.

The PAL-I instrument was administered using an existing web-based panel through Ipsos Observer (https://www.ipsos.com/en-us/online-research). Individuals who had previously participated in the Ipsos research panel and had reported cLPB were invited to complete a web-based survey to confirm their eligibility.

Eligibility criteria included as follows: confirmation of clinical diagnosis and current cLBP, duration of cLBP, pain intensity, absence of recent low back surgery or planned low back surgery in the next 30 days, and recent epidural injections or spinal cord stimulation therapy. Participants completed the PAL-I and a questionnaire that included demographic characteristics, clinical characteristics (severity and location of back pain, pain movement, and sciatica/neuropathic pain assessment), and items regarding treatment (medication type, current treatment or not, duration of treatment, satisfaction with current treatment, and other nonmedication treatment).

Analyses included descriptive statistics and floor/ceiling effects for individual item responses. An item-to-item correlation matrix based on Pearson's r was constructed for each item; coefficients >0.70 suggested a potential redundancy between items. Item-to-total score correlations based on bivariate Pearson's r were evaluated (excluding the item of interest from the total score); coefficients <0.40 suggested potential nonassociations with the remaining items in the hypothesized scale. Items were evaluated for appropriate psychometric scaling based on RMT analysis, and required that item response options were ordered and the items formed a unidimensional construct (root-mean-square error of approximation). Items that did not fit the RMT model were considered candidates for item reduction or revision. Category probability curves (item characteristic curves) were used to identify items that did not demonstrate monotonically increased responses. Simulation analyses were conducted for items that exhibited disordered thresholds (inconsistent responses) by collapsing response categories to assess potential improvements in item characteristics. Consistency of response was examined using a person–item distribution map, which displays persons and items on a logit scale with the most able persons and most difficult items on one side and the least able persons and easiest items on the other side. To avoid large gaps in measurement, the distance between items should be <0.30 logits.[3] Cronbach's alpha was used to assess internal consistency.

2.8. Cognitive interviews (round 2) to test modifications

After the first quantitative administration, the PAL-I was revised and evaluated in 2 waves of cognitive interviews.

2.9. Second quantitative administration and preliminary psychometric validation

After evaluation in the second round of cognitive interviews, the revised PAL-I instrument was tested using 2 separate U.S.-based cohorts of patients with cLBP. The first sample to complete the web-based survey of the revised instrument was a subset of the participants in the first quantitative administration (401 patients of the original 598 patients). The second sample was a clinic-based cohort of patients with physician-diagnosed cLBP who were recruited to conduct preliminary psychometric validation of the instrument. This cohort comprised 45 patients with cLBP who were identified by patient records. These patients completed (on paper) the PAL-I, painDETECT, MOS-36 (a multi-item scale that assesses 8 health concepts of which we used 2: limitations in physical activities because of health problems and bodily pain, scored on a 0-100-point scale), RMDQ (a 24-item health status checklist designed to assess physical disability due to LBP, scored by summing the number of items checked by the patient with a range of 0-24), and NPSI (a self-administered questionnaire specifically designed to evaluate different symptoms of neuropathic pain with descriptors reflecting spontaneous, ongoing, or

Table 1

Demographic and clinical characteristics of patients participating in concept elicitation and cognitive interviews (reproduced from Martin et al.[14]).

	Concept elicitation patients (N = 43)	Initial cognitive interview patients (N = 30)	Cognitive interviews to evaluate modifications (N = 8)	Paper to electronic interview patients (N = 8)
Age, mean years (range)	48.6 (21-73)	43.6 (20-77)	47.2 (30-78)	41.5 (29-60)
Sex, n female (%)	23 (53.5)	17 (56.7)	4 (50.0)	4 (50.0)
Marital status, n (%)				
Married or living as married	24 (55.8)	14 (46.7)	2 (25.0)	3 (37.5)
Living with partner	5 (11.6)	5 (16.7)	2 (25.0)	0
Widowed	3 (7.0)	0	1 (12.5)	0
Separated	1 (2.3)	1 (3.3)	1 (12.5)	2 (25.0)
Divorced	8 (18.6)	4 (13.3)	2 (25.0)	0
Never married	2 (4.7)	6 (20.0)	0	3 (37.5)
Highest education level, n (%)				
Less than high school	3 (7.0)	0	0	1 (12.5)
High school	13 (30.2)	6 (20.0)	3 (37.5)	1 (12.5)
Some college	13 (30.2)	13 (43.3)	4 (50.0)	4 (50.0)
Bachelor's degree	5 (11.6)	3 (10.0)	0	2 (25.0)
Graduate or professional school	4 (9.3)	7 (23.3)	1 (12.5)	0
Missing	5 (11.6)	1 (3.3)	0	0
Employment outside home, n (%)				
Not employed outside home	4 (9.3)	1 (4.0)	1 (12.5)	0
Full-time	21 (48.8)	15 (50.0)	4 (50.0)	4 (50.0)
Part-time	2 (4.7)	7 (23.3)	0	1 (12.5)
Retired	6 (14.0)	3 (10.0)	1 (12.5)	0
Not employed/student	10 (23.3)	4 (13.3)	2 (25.0)	3 (37.5)
Race/ethnicity, n (%)				
White/Caucasian (non-Hispanic)	32 (74.4)	20 (66.7)	4 (50.0)	4 (50.0)
Hispanic	2 (4.7)	1 (3.3)	0	2 (25.0)
Black/African American	4 (9.3)	5 (16.7)	1 (12.5)	2 (25.0)
Asian	2 (4.7)	3 (10.0)	0	0
American Indian (Hispanic)	1 (2.3)	0	0	0
Other/mixed race	2 (4.7)	1 (3.3)	3 (37.5)	0
Overall rating of health, n (%)				
Excellent	2 (4.7)	5 (16.7)	0	0
Very good	13 (30.2)	9 (30.0)	2 (25.0)	1 (12.5)
Good	18 (41.9)	10 (33.3)	5 (62.5)	6 (75.0)
Fair	9 (20.9)	5 (16.7)	1 (12.5)	1 (12.5)
Poor	1 (2.3)	0	0	0
Missing	0	1 (3.3)	0	0
Current medication for pain, n (%)				
NSAID	27 (62.8)	20 (66.7)	5 (62.5)	6 (75.0)
Muscle relaxants	19 (44.2)	9 (30.0)	4 (50.0)	3 (37.5)
Anticonvulsants	6 (14.0)	3 (10.0)	1 (12.5)	0
Antidepressants	3 (7.0)	1 (3.3)	0	0
Opioids	25 (58.1)	15 (50.0)	7 (87.5)	5 (62.5)
Other	1 (2.3)	2 (6.7)	3 (37.5)	0
None	2 (4.7)	3 (10.0)	0	1 (12.5)
Pain intensity, mean NRS score* (SD) [range]	6.7 (1.2) [4-10]	7.0 (1.4) [4-10]	7.0 (1.4) [5-9]	6.0 (1.5) [4-7]
painDETECT subgroup, n (%)				
Unlikely neuropathic (score 0-12)	14 (32.5)	9 (45.0)	3 (37.5)	6 (75.0)
Unclear (score 13-18)	15 (35.0)	2 (10.0)	1 (12.5)	4 (50.0)
Likely neuropathic (score ≥19)	14 (32.5)	9 (45.0)	4 (50.0)	2 (25.0)

* Scale of 0 to 10, with higher scores indicating greater pain intensity.
NRS, numeric rating scale; NSAID, nonsteroidal anti-inflammatory drug.

Table 2

Summary of impact concept code frequencies.

LBP impact subdomains and concepts	No. (%) of 2220 total symptom expressions	No. (%) of 42 transcripts contributing to concept expression
Physical activity limitations and restrictions	270 (12)	
Exercise	74 (3.3)	32 (76.2)
General limitations	16 (0.7)	13 (31.0)
Sports	38 (1.7)	18 (42.9)
Stairs	18 (0.8)	13 (31.0)
Walking: distance	44 (2.0)	22 (52.4)
Walking: general	56 (2.5)	27 (64.3)
Walking: incline	7 (0.3)	6 (14.3)
Walking: speed	11 (0.5)	8 (19.0)
Walking: surface	6 (0.3)	6 (14.3)
Restricted body movement	347 (16)	
Bending	63 (2.8)	30 (71.4)
Carrying/lifting	48 (2.2)	25 (59.5)
Rising	23 (1.0)	19 (45.2)
Shuffling/limping/wobbling	13 (0.6)	7 (16.7)
Sitting	101 (4.5)	37 (88.1)
Standing	49 (2.2)	23 (54.8)
Turning/twisting	10 (0.5)	8 (19.0)
Other restricted body movements	40 (1.8)	22 (52.4)
Difficulty getting around	110 (5)	
Driving	72 (3.2)	31 (73.8)
Getting around	19 (0.9)	12 (28.6)
Travel	19 (0.9)	16 (38.1)
Difficulty doing daily activities	393 (18)	
Childcare activities	10 (0.5)	6 (14.3)
General daily activity limitations	101 (4.5)	32 (76.2)
Household activities	127 (5.7)	34 (81.0)
Personal care activities	42 (1.9)	24 (57.1)
Work activities	113 (5.1)	33 (78.6)
Social/lifestyle limitations and restrictions	227 (10)	
Leisure activities	85 (3.8)	31 (73.8)
Relationships	63 (2.8)	27 (64.3)
Religious activities	7 (0.3)	6 (14.3)
Sexual activities	21 (0.9)	14 (33.3)
Social engagements	51 (2.3)	27 (64.3)
Emotional impacts	183 (8)	
Anger	18 (0.8)	12 (28.6)
Irritation	30 (1.4)	18 (42.9)
Anxiety	17 (0.8)	11 (26.2)
Depression	35 (1.6)	19 (45.2)
Frustration	46 (2.1)	21 (50.0)
Low self-esteem	7 (0.3)	6 (14.3)
Worry/fear	20 (0.9)	10 (23.8)
Other emotional impacts	10 (0.5)	7 (16.7)
Aspects of burden	66 (3)	
Financial burden	21 (0.9)	17 (40.5)
Needing assistance	45 (2.0)	19 (45.2)
Sleep difficulties	139 (6)	
Sleep position	28 (1.3)	18 (42.9)
Sleep quality	45 (2.0)	22 (52.4)
Waking from pain	66 (3.0)	23 (54.8)
Low-energy difficulties	59 (3)	
Low energy	44 (2.0)	24 (57.1)
Stamina	15 (0.7)	11 (26.2)
Coping behaviors	421 (19)	
Coping behaviors	421 (19.0)	42 (100)
Other impacts	5 (0)	
Other impacts	5 (0.2)	2 (4.8)

LBP, low back pain.

paroxysmal pain, evoked pain [ie, mechanical and thermal allodynia/hyperalgesia], and dysesthesia/paresthesia; with each item quantified on a 0-10 numerical scale). Intraclass correlation coefficients were used to evaluate test–retest reproducibility. Pearson correlations were used to assess convergent validity of the PAL-I with the painDETECT, MOS-36, RMDQ, and NPSI instruments. Known-groups validity was assessed by 2 methods: comparison with pain NRS tertiles and painDETECT groups.

2.10. Cognitive interviews to assess paper to electronic equivalence

Cognitive interviews were conducted per ISPOR recommendations (Ref. 7 page 423) to confirm that the intent and meaning of the items, response options, and instructions were unaffected by how the instrument was administered. The interview guide was specifically intended to capture the patient's comprehension of items and ability to complete the PAL-I instrument using a paper or electronic format. Questions in the interview process asked about: the comprehension and relevance of the individual items and how they may have differed between the paper or electronic versions; the fit of the response scales; the language used; and any lack of clarity of items, terminology, instructions, or sentence structure.

3. Results

3.1. Concept elicitation results

A total of 43 patients participated in CE interviews to identify cLPB impacts (**Table 1**). Saturation of concept was achieved by the end of the second transcript group. Interrater agreement ranged from 82.5% to 86.9% between the coders for the identification of concepts being expressed by patients, and ranged from 97.2% to 99.2% for the assignment of specific concept codes to concepts identified.

Based on the number of times a particular concept was expressed in the overall transcript data set, the predominant impact-related concepts were impacts on coping behaviors, household activities, and work activities (**Table 2**). The most common spontaneously offered impacts were those on walking (reported by 65.1% of participants), sitting (62.8%), exercise (58.1%), and leisure activities (58.1%). The most difficult impacts were on climbing stairs (mean difficulty rating 9.5 on NRS 0-10), coping behaviors (8.5), standing (8.4), and sports (8.4).

3.2. Item generation results

A total of 14 items were identified as appropriate for inclusion in the preliminary version of the PAL-I instrument (**Table 3**).

3.3. Cognitive interviews round 1 results

Thirty patients participated in the first round (4 waves) of cognitive interviews (**Table 1**). Patient input from these interviews led to changes in the instructions for the instrument (to better highlight the response procedure and recall period) and changes to the wording of some items.

3.4. First quantitative administration results

A total of 598 patients participated in the first quantitative administration of the instrument (**Table 4**). All PAL-I items were endorsed in the first quantitative administration. Item mean scores ranged from 2.15 (needed assistance with daily activities; SD 1.50) to 5.32 (affected your sexual activity; SD 2.23). High floor effects (participants selecting least impactful option) were seen with 4 items (affected your climbing up or down stairs, caused you to feel sad or down, caused you to feel worried or nervous, and needed assistance with daily activities). No item-to-item correlations were problematic. Item-to-total correlations ranged from 0.29 to 0.61. Cronbach's alpha for the 13 items was 0.81, with alphas remaining at 0.79 to 0.82 if any item was deleted.

The RMT analysis showed only 2 items with an ordered threshold (affected your lifting, affected your turning, twisting, and

---

Table 3

Low back pain concepts selected for PAL-I instrument.

Targeted symptom concept	Actions on concepts and final item in PAL-I
Personal care	Removed
Lifting	Item 4: limited your lifting
Walking	Item 1: limited your walking
Sitting	Item 2: limited your sitting
Standing	Item 3: limited your standing
Sleeping	Item 5: limited your sleep
Sex life	Removed
Social life	Item 6: limited your social activities (going out and seeing friends)
Travelling	Item 7: limited your travelling (daily driving and taking trips)
Climbing stairs	Item 8: affected your climbing up or down stairs (at home or at other locations)
Body movement	Item 9: limited your turning, twisting, or bending
Feeling depressed	Removed
Feeling anxious	Removed
Needing assistance to get around	Removed

bending); the remaining items were disordered. Disordering seemed to be based on how the response categories were presented: some items were not linear (eg, walking) and some items mixed impacts from the activity with impacts from the cLBP (eg, climbing stairs). These findings indicated that the response options in the preliminary version were not ideal, and the instrument was modified to expand the distributions for better scaling. Four items were removed (caused you to feel sad or down, caused you to feel worried or nervous, affected your sexual activity, and needed assistance with daily activities). Response options for the remaining 9 items were changed to a 4-point rating scale with a fifth option available for not doing the activity for reasons other than cLBP.

3.5. Cognitive interviews round 2 results

Modifications made to the PAL-I based on item-level analyses in the first quantitative administration were evaluated in a second round of cognitive interviews with 8 patients (**Table 1**). Additional changes were made to the instrument based on the results of these interviews. Instructions were changed from "Mark one box," to "Select one," to better facilitate migration from paper to electronic formatting. The order of items was adjusted so that the physical activity–related items were organized by increasing level of exertion necessary, which was followed by items assessing impacts in other areas of daily life. The walking item was altered to specify "during daily activities," to clarify the intent of the item and

---

Table 4

Demographic and clinical characteristics of patients participating in quantitative analyses (reproduced from Martin et al.[14]).

	First quantitative administration (N = 598)	Second quantitative administration (n = 401)	Preliminary psychometric validation (N = 45)
Age, mean years (range)	55.5 (21-80)	55.3 (20-80)	53.0 (29-77)
Sex, n female (%)	406 (67.9)	272 (67.8)	23 (52.3)
Marital status, n (%)			
Married or living as married	323 (54.0)	263 (65.6)	22 (50.0)
Widowed	47 (7.9)	35 (8.7)	4 (9.1)
Separated	11 (1.8)	7 (1.7)	3 (6.8)
Divorced	124 (20.7)	53 (13.2)	6 (13.6)
Never married	93 (15.6)	43 (10.7)	9 (20.5)
Education level, mean highest grade/year (range)	14.4 (6-20)	14.8 (4-20)	14.6 (8-20)
Employment outside home, n (%)			
Full-time	128 (21.4)	98 (24.4)	16 (36.4)
Part-time	50 (8.4)	33 (8.2)	1 (2.3)
Self-employed	43 (7.2)	29 (7.2)	2 (4.5)
Unemployed	27 (4.5)	21 (5.2)	6 (13.6)
Homemaker	59 (9.9)	40 (10.0)	1 (2.3)
Student	6 (1.0)	9 (2.2)	2 (4.5)
Retired	186 (31.1)	123 (30.7)	7 (15.9)
Unable to work	97 (16.2)	45 (11.2)	9 (20.5)
Other	2 (0.3)	0	0
Race/ethnicity, n (%)			
White	526 (88.0)	352 (87.8)	35 (79.5)
Black/African American	28 (4.7)	24 (6.0)	7 (15.9)
Hispanic	14 (2.3)	11 (2.7)	2 (4.5)
Asian	9 (1.5)	7 (1.7)	0
Other/missing	21 (3.5)	4 (1.0)	0
Overall rating of health, n (%)			
Excellent	9 (1.5)	18 (4.5)	0
Very good	105 (17.6)	96 (23.9)	9 (20.5)
Good	232 (38.8)	165 (41.1)	26 (59.1)
Fair	204 (34.1)	101 (25.2)	8 (18.2)
Poor	48 (8.0)	21 (5.2)	1 (2.3)
Duration of LBP, mean years (range)	15.2 (0.3-66)	13.5 (0.25-60)	11.0 (1-40)
Pain intensity, mean NRS score* (range)	6.1 (1-10)	6.2 (1-10)	6.2 (4-10)
Pain intensity score*, n (%)			
0	0	0	0
1-3	70 (11.7)	47 (11.7)	0
4	59 (9.9)	37 (9.2)	5 (11.4)
5	90 (15.1)	66 (16.5)	14 (31.8)
6	105 (17.6)	73 (18.2)	9 (20.5)
7	124 (20.7)	60 (15.0)	5 (11.4)
8-10	150 (25.1)	118 (29.4)	11 (25.0)

* Scale of 0 to 10, with higher scores indicating greater pain intensity.
LBP, low back pain; NRS, numeric rating scale.

Descriptive characteristics of PAL-I items from second quantitative administration.

Item # and content	Mean score (SD) [median]	Floor n (%)	Ceiling n (%)	Item-to-total score correlation	Alpha if item deleted
1. Limited your walking	1.20 (0.86) [1]	94 (23.4)	22 (5.5)	0.44	0.80
2. Limited your sitting	1.14 (0.80) [1]	84 (20.9)	18 (4.5)	0.37	0.81
3. Limited your standing	1.33 (0.78) [1]	57 (14.2)	18 (4.5)	0.50	0.80
4. Limited your lifting	1.50 (0.92) [2]	58 (14.5)	58 (14.5)	0.43	0.80
5. Limited your sleep	1.18 (0.77) [1]	79 (19.7)	9 (2.2)	0.39	0.81
6. Limited your social activities	0.95 (0.92) [1]	150 (37.4)	23 (5.7)	0.56	0.79
7. Limited your travelling	0.93 (0.97) [1]	156 (38.9)	28 (7.0)	0.46	0.80
8. Limited your climbing up or down stairs	1.23 (0.95) [1]	99 (24.7)	37 (9.2)	0.51	0.80
9. Limited your turning, twisting, or bending	1.47 (0.82) [1]	47 (11.7)	36 (9.0)	0.61	0.79

alleviate potential confusion among individuals who walk or hike as a form of exercise outside their usual daily activity.

3.6. Quantitative findings from second administration

A second quantitative data collection was conducted with 401 newly recruited subjects from the same commercial sample. All PAL-I items and response options were endorsed. Item mean scores ranged from 0.93 (limited your traveling; SD 0.97) to 1.50 (limited your lifting; SD 0.92) (**Table 5**). No data were missing. High floor effects were seen with 2 items (limited your travelling and limited your social activities). Only 2 PAL-I items were strongly correlated with each other: limited your walking and limited your standing (r = 0.65). Item-to-total correlations ranged from 0.56 to 0.82. Cronbach's alpha for the 9 items was 0.92, with alphas remaining at 0.91 to 0.92 if any item was deleted. The item threshold map (**Fig. 2**) showed that all items fit the model with no items exhibiting a disordered threshold.

3.7. Preliminary psychometric validation

A total of 45 patients with physician-diagnosed cLBP participated in the preliminary psychometric validation analyses (**Table 4**). All PAL-I items and response options were endorsed. Item mean scores ranged from 0.98 (limited your social activities; SD 0.78) to 1.51 (limited your sleep; SD 0.73). Test–retest reproducibility at 1 week was acceptable; the intraclass correlation coefficient was 0.88 (95% confidence interval 0.78-0.94). In convergent validity assessments, strong associations were seen between the PAL-I total score and the RMDQ (Pearson correlation coefficient 0.82), MOS-36 Physical Function (−0.71), and MOS-36 Bodily Pain

(−0.71). Known-groups assessments based on pain NRS tertiles showed that the PAL-I total scores were able to significantly discriminate between levels of pain (**Table 6**).

3.8. Cognitive assessment of paper and electronic equivalence

Eight patients participated in the cognitive assessment of paper and electronic versions of the instrument (**Table 1**). There was no indication that understanding of the instructions, items, or response options was affected by differences in mode of administration. Some formatting issues to be considered for both paper and electronic versions were noted by the patients, including how close response options were to the question item, how crowded text was to the left margin, and where a sentence broke at the end of a line of text.

4. Discussion

The PAL-I and its companion instrument to measure pain symptoms (PAL-S)[14] were developed in accordance with U.S. FDA guidance for PRO measures to be used for label claims of patient-reported improvements in symptoms and impacts for medications to treat cLBP. The PAL-I demonstrated content validity; ie, the items in the instruments are relevant and representative of the cLBP experience based on input from patients with cLBP. In addition, the PAL-I demonstrated 1-week test–retest reliability, convergent validity with relevant components of existing instruments, and appropriate item-level performance, scale structure, and scoring.
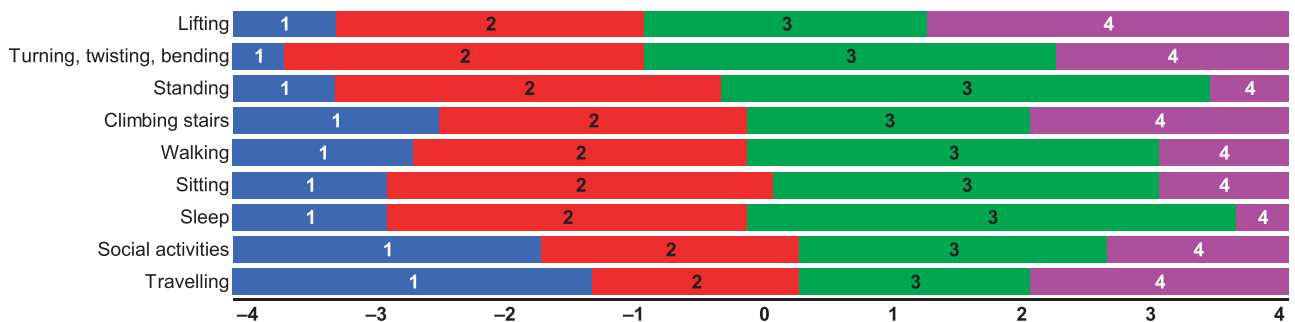
Figure 2. Item threshold map for items within the PAL-I Wave 2 web sample (n = 401).

Known-groups validity of the PAL-I: mean item scores (SD) stratified by NRS tertiles.

Item content	Second quantitative administration (N = 354)			Preliminary psychometric validation (N = 45)		
	NRS tertile 1, score 4-5 (n = 103)	NRS tertile 2, score 6-7 (n = 133)	NRS tertile 3, score 8-10 (n = 118)	NRS tertile 1, score 4-5 (n = 19)	NRS tertile 2, score 6-7 (n = 14)	NRS tertile 3, score 8-10 (n = 12)
1. Limited your walking	0.93 (0.74)	1.22 (0.86)	1.63 (0.81)	2.11 (0.94)	2.07 (0.73)	2.25 (0.87)
2. Limited your sitting	0.83 (0.60)	1.16 (0.71)	1.64 (0.82)	1.89 (0.74)	2.21 (0.58)	2.42 (0.90)
3. Limited your standing	1.08 (0.73)	1.32 (0.70)	1.78 (0.69)	2.32 (0.75)	2.07 (0.62)	2.42 (0.90)
4. Limited your lifting	1.20 (0.80)	1.57 (0.87)	2.00 (0.81)	2.16 (0.83)	2.43 (0.76)	3.00 (0.95)
5. Limited your sleep	0.91 (0.66)	1.25 (0.75)	1.55 (0.73)	2.37 (0.90)	2.64 (0.50)	2.58 (0.67)
6. Limited your social activities	0.49 (0.66)	0.95 (0.86)	1.61 (0.88)	1.68 (0.75)	2.00 (0.68)	2.42 (0.79)
7. Limited your travelling	0.54 (0.77)	0.87 (0.91)	1.56 (0.95)	1.74 (1.10)	2.00 (0.88)	2.50 (1.00)
8. Limited your climbing up or down stairs	0.82 (0.82)	1.24 (0.90)	1.79 (0.86)	2.16 (1.07)	2.36 (1.15)	2.92 (0.90)
9. Limited your turning, twisting, or bending	1.20 (0.81)	1.56 (0.71)	1.80 (0.79)	2.26 (0.81)	2.57 (0.51)	2.75 (0.62)
Total score	0.90 (0.51)	1.25 (0.59)	1.70 (0.61)	1.02 (0.57)	1.24 (0.45)	1.58 (0.63)

NRS, numeric rating scale (0-10 scale with higher scores indicating greater severity).

The PAL-I provides a single total impact score on a scale of 0 to 3, with higher scores indicating greater impacts due to cLBP. Patients score each of the 9 items with response options of "Not at all limited," (score = 0), "Limited a little," (score = 1), "Limited a lot," (score = 2), or "Did not do because of my LBP," (score = 3), or patients can opt-out for an item ("Did not do for other reasons"). The mean score for all scored items (excluding opt-out items) represents the single total impact score.

Although many instruments have been developed to assess pain from the patient perspective, a previous search revealed that none was fully compliant with the FDA guidance and none could be used for label claims.[18] In our review of existing instruments, the most common reason for failing to comply with FDA guidance was lack of patient input. Patient input is particularly important for pain assessments, as there are no measurable signs or laboratory tests to estimate or quantify pain severity, and physicians cannot easily or accurately assess the various dimensions of pain. For example, in emergency departments, where the most common presenting symptom is pain, physicians frequently rate pain as less severe than patients.[4,13] Similarly, considerable noncon-cordance between general practitioners' and patients' assessments of pain intensity has been observed for patients with chronic pain (Spearman correlation coefficient 0.20).[12] For the development of the PAL-I and the PAL-S, cohorts of patients with cLBP representing the spectrum of disease severity were recruited to elicit symptom and impact concepts that related specifically to cLBP. Although generic pain instruments are useful for clinical trials of medications to treat conditions associated with pain or to monitor treatment response in clinical practice, the PAL-I and PAL-S are designed for use in clinical trials to treat cLBP. Although not specifically developed for use in clinical practice, both of these measures may useful to assess a patient's status associated with their cLBP and providing helpful information to clinicians considering the next steps of care.

A limitation of this study was the small sample sizes for CE, cognitive interviews, and preliminary validation of the PAL-I. Saturation of concept, which was achieved with the second round of CE, suggested that a sufficient sample size was used to elicit key concepts for the PAL-I. Large sample sizes were used for quantitative administrations, but cLBP diagnosis was self-reported; however, patients with physician-confirmed cLBP participated in preliminary validation analyses. A strength of the study was the recruitment of patients across the spectrum of pain severity and representing both neuropathic and non-neuropathic pain.

Further studies are planned for the PAL-I and PAL-S. Although preliminary measurement properties were derived from the initial quantitative data collection, a formal validation study including sensitivity to treatment-related change has not yet been conducted. Further investigation of the influence of clinical history and potential differences between conditions (a noted limitation of this study) should also be explored. In addition, studies to develop the instruments for other cultures and other languages are planned. The PAL-I and PAL-S have the potential to provide between-study comparisons in clinical trials of cLBP treatments.

Conflict of interest statement

References

[1] Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G; Working Group on Guidelines for Chronic Low Back Pain. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006;15(suppl 2):S192–300.

[2] Balague F, Mannion AF, Pellise F, Cedraschi C. Non-specific low back pain. Lancet 2012;379:482–91.

[3] Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 2nd ed. London: Lawrence Erlbaum, 2007.

[4] Cakir U, Cete Y, Yigit O, Bozdemir MN. Improvement in physician pain perception with using pain scales. Eur J Trauma Emerg Surg 2017. doi: 10.1007/s00068-017-0882-7. [Epub ahead of print].

[5] Chou R, Qaseem A, Snow V, Casey D, Cross JT Jr, Shekelle P, Owens DK; Clinical Efficacy Assessment Subcommittee of the American College of Physicians, American College of Physicians, American Pain Society Low Back Pain Guidelines Panel. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007;147:478–91.

[6] Christoph A, Eerdekens MH, Kok M, Volkers G, Freynhagen R. Cebranopadol, a novel first-in-class analgesic drug candidate: first experience in patients with chronic low back pain in a randomized clinical trial. PAIN 2017;158:1813–24.

[7] Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, Lenderking WR, Cella D, Basch E; Ispor ePRO Task Force. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009;12:419–29.

[8] Freynhagen R, Baron R, Gockel U, Tolle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin 2006;22:1911–20.

[9] Freynhagen R, Tölle TR, Gockel U, Baron R. The painDETECT project—far more than a screening tool on neuropathic pain. Curr Med Res Opin 2016;32:1033–57.

[10] Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, Williams G, Smith E, Vos T, Barendregt J, Murray C, Burstein R, Buchbinder R. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014;73:968–74.

[11] Knezevic NN, Mandalia S, Raasch J, Knezevic I, Candido KD. Treatment of chronic low back pain—new approaches on the horizon. J Pain Res 2017;10:1111–23.

[12] Mantyselka P, Kumpusalo E, Ahonen R, Takala J. Patients' versus general practitioners' assessments of pain intensity in primary care patients with non-cancer pain. Br J Gen Pract 2001;51:995–7.

[13] Marquie L, Raufaste E, Lauque D, Marine C, Ecoiffier M, Sorum P. Pain rating by patients and physicians: evidence of systematic pain miscalibration. PAIN 2003;102:289–96.

[14] Martin ML, Blum SI, Liedgens H, Bushnell DM, McCarrier KP, Hatley NV, Ramasamy A, Freynhagen R, Wallace M, Argoff C, Eerdekens M, Kok M, Patrick DL. Mixed-methods development of a new patient-reported outcome instrument for chronic low back pain: part 1—the patient assessment for low back pain-symptoms (PAL-S). PAIN 2018;159: 1045–55.

[15] Muhr T. User's manual for ATLAS.ti 5.0. Berlin: ATLAS.ti Scientific Software Development GmbH, 2004.

[16] Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity–establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2–assessing respondent understanding. Value Health 2011; 14:978–88.

[17] Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, O'Neill R, Kennedy DL. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007;10(suppl 2): S125–137.

[18] Ramasamy A, Martin ML, Blum SI, Liedgens H, Argoff C, Freynhagen R, Wallace M, McCarrier KP, Bushnell DM, Hatley NV, Patrick DL. Assessment of patient-reported outcome instruments to assess chronic low back pain. Pain Med 2017;18:1098–110.

[19] United States Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282. Accessed June 22, 2017.

[20] van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, Koes B, Laerum E, Malmivaara A; Working Group on Guidelines for the Management of Acute Low Back Pain in Primary Care. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006;15(suppl 2):S169–191.

[21] Violante FS, Mattioli S, Bonfiglioli R. Low-back pain. Handb Clin Neurol 2015;131:397–410.