

# 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)

Mona L. Martin<sup>a,\*</sup>, Steven I. Blum<sup>b</sup>, Hiltrud Liedgens<sup>c</sup>, Donald M. Bushnell<sup>a</sup>, Kelly P. McCarrier<sup>a</sup>, Noël V. Hatley<sup>a</sup>, Abhilasha Ramasamy<sup>b</sup>, Rainer Freynhagen<sup>d</sup>, Mark Wallace<sup>e</sup>, Charles Argoff<sup>f</sup>, Mariëlle Eerdeken<sup>c</sup>, Maurits Kok<sup>c</sup>, Donald L. Patrick<sup>g</sup>

## Abstract

We describe the mixed-methods (qualitative and quantitative) development and preliminary validation of the Patient Assessment for Low Back Pain–Symptoms (PAL-S), a patient-reported outcome measure for use in chronic low back pain (cLBP) clinical trials. Qualitative methods (concept elicitation and cognitive interviews) were used to identify and refine symptom concepts and quantitative methods (classical test theory and Rasch measurement theory) were used to evaluate item- and scale-level performance of the measure using an iterative approach. Patients with cLBP participated in concept elicitation interviews (N = 43), cognitive interviews (N = 38), and interview-based assessment of paper-to-electronic mode equivalence (N = 8). A web-based sample of patients with self-reported cLBP participated in quantitative studies to evaluate preliminary (N = 598) and revised (n = 401) drafts and a physician-diagnosed cohort of patients with cLBP (N = 45) participated in preliminary validation of the measure. The PAL-S contained 14 items describing symptoms (overall pain, sharp, prickling, sensitive, tender, radiating, shocking, shooting, burning, squeezing, muscle spasms, throbbing, aching, and stiffness). Item-level performance, scale structure, and scoring seemed to be appropriate. One-week test–retest reproducibility was acceptable (intraclass correlation coefficient 0.81 [95% confidence interval, 0.61–0.91]). Convergent validity was demonstrated with total score and MOS-36 Bodily Pain (Pearson correlation –0.79), Neuropathic Pain Symptom Inventory (0.73), Roland-Morris Disability Questionnaire (0.67), and MOS-36 Physical Functioning (–0.65). Individual item scores and total score discriminated between numeric rating scale tertile groups and painDETECT categories. Respondent interpretation of paper and electronic administration modes was equivalent. The PAL-S has demonstrated content validity and is potentially useful to assess treatment benefit in cLBP clinical trials.

**Keywords:** Patient-reported outcome instrument, Chronic low back pain, Qualitative research

## 1. Introduction

Low back pain (LBP), generally described as pain, muscle tension, or muscle stiffness with or without radiation to the distal lower extremity,

is the second most common cause of disability among adults in the United States.<sup>1,2</sup> Although acute and subacute episodes (those lasting up to 3 months) are the most common presentations of LBP, chronic low back pain (cLBP) is associated with long-term pain and disability.<sup>8</sup>

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

<sup>a</sup> Health Research Associates, Inc, Mountlake Terrace, WA, USA, <sup>b</sup> Health Economics and Outcomes Research, Forest Research Institute, Jersey City, NJ, USA. Mr. Blum is now with the GlaxoSmithKline, Collegeville, PA, USA. Ms. Ramasamy is now with the Novo Nordisk, Plainsboro, NJ, USA, <sup>c</sup> Grünenthal GmbH, Aachen, Germany, <sup>d</sup> Zentrum für Anästhesiologie, Intensivmedizin, Schmerztherapie & Palliativmedizin, Benedictus Krankenhaus Tutzing and Technische Universität, München, Germany, <sup>e</sup> Department of Anesthesiology, University of California San Diego, San Diego, CA, USA, <sup>f</sup> Department of Neurology, Albany Medical College, Albany, NY, USA, <sup>g</sup> Department of Health Sciences, University of Washington, Seattle, WA, USA

\*Corresponding author. Address: 6505 216th St SW, Suite 105, Mountlake Terrace, WA 98043, USA. Tel.: 425-775-6565; fax: 425-775-6734. E-mail address: martin@hrainc.net (M.L. Martin).

PAIN 159 (2018) 1045–1055

Copyright © 2018 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the International Association for the Study of Pain. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

<http://dx.doi.org/10.1097/j.pain.0000000000001187>

Pain has been described as one of the most important domains to be assessed in LBP, along with back-specific function, generic health status, work disability, and patient satisfaction.<sup>1,10,19</sup> Pain is one of the best determinants of disability due to cLBP,<sup>4,9,20,31</sup> and is predictive of work resumption within the year after related short-term absence.<sup>5,28</sup> Although pain and function are interdependent parts of the patient experience, patients often seek care because of pain, and this variable therefore needs to be assessed at baseline and in response to treatment. Because the severity of cLBP symptoms is often related to the degree of impairment that patients experience, the assessment of this concept is an essential endpoint for clinical studies. Pain-related function/impairment is an outcome that should also be assessed.

Although various patient-reported outcome (PRO) instruments have been used in cLBP clinical trials, many present pain as a single overall concept, which may not fully capture the pain sensations of cLBP. The use of multidimensional pain assessments that encompass pain-related physical function and psychological factors is recommended to evaluate chronic pain.<sup>16,18</sup> We conducted a review of existing instruments that are currently used to collect patient-reported concepts of pain

and pain impacts in patients with cLBP.<sup>25</sup> We found that none of the instruments in current use was developed in accordance with the Food and Drug Administration (FDA) PRO guidance<sup>29</sup> and none could be used for label claims of patient-reported improvements in clinical trials.

The FDA PRO guidance<sup>29</sup> and other key publications<sup>23,24,27,30</sup> recommend identification of labeling goals for the clinical development program as the first step in developing clinical endpoints. The use of qualitative research to explore the patient experience allows these goals to be linked to specific PRO concepts that are relevant to the patient and the assessment of improvement in patient symptoms. This is particularly important in axial LBP, where the specific sensations of pain may be multiple but experienced in a single area.<sup>11</sup> Ultimately, an instrument with established content validity based on qualitative data from patients would be expected to successfully detect changes in symptoms in clinical studies of treatment benefit.

The objective of this report is to describe the process and results of the preliminary development of the Patient Assessment for Low Back Pain–Symptoms (PAL-S), a new symptom-based PRO measure, which is intended for use in assessing symptom change and treatment benefit in patients with cLBP. The development of the PAL-S was based on mixed-methods data collection among patients with cLBP; qualitative data were used to identify patient-relevant concepts to be included in the PRO instrument and to confirm respondent understanding, whereas quantitative data were used to evaluate the preliminary psychometric and measurement properties of the newly developed scale. Because of the iterative nature of the mixed-methods approach, the methods and results of this study are presented chronologically.

## 2. Methods

### 2.1. Study design and development steps for the patient assessment for low back pain–symptoms questionnaire

The development of the PAL-S followed the recommendations in the US FDA guidance for PRO development<sup>29,30</sup> and was in accordance with good research practices established by the International Society for Pharmacoeconomics and Outcomes Research PRO Good Research Practices Task Force for establishing content validity in newly developed PROs.<sup>23,24</sup> As shown in **Figure 1**, we used a mixed-methods approach, which iteratively combined both qualitative methods to generate and refine the concepts to be included in the PRO and quantitative methods to evaluate the concepts in the PRO.

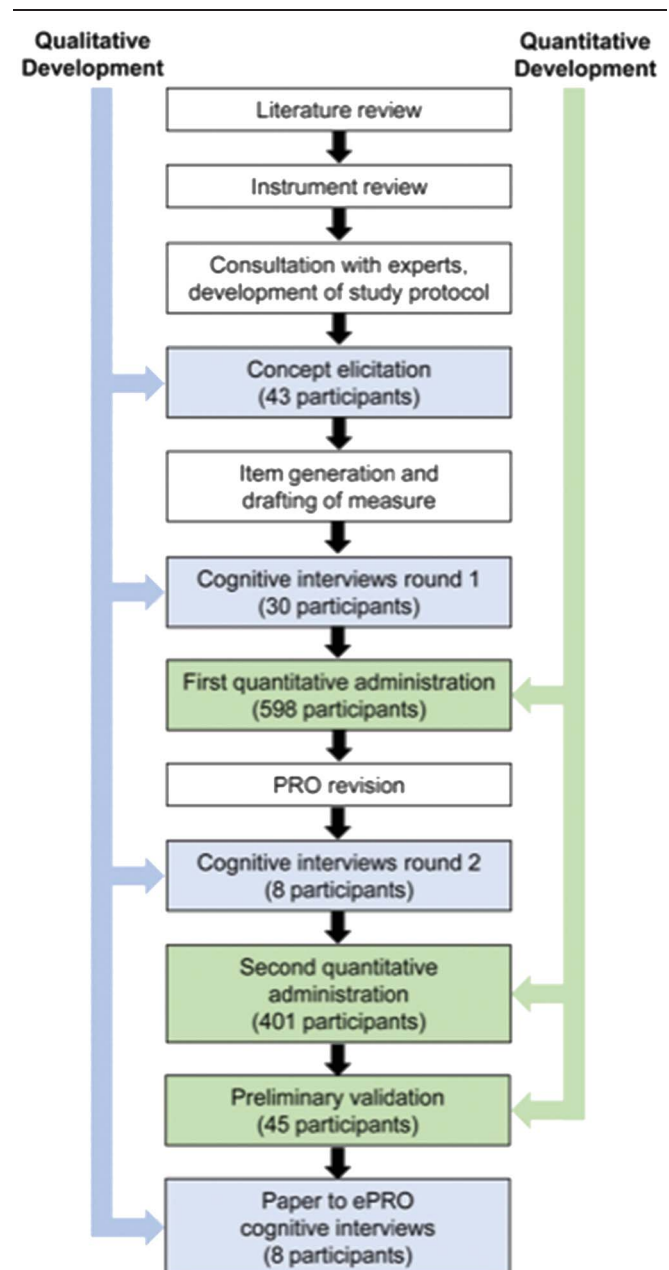
A review of LBP literature searched for symptom concepts and currently used PROs that were relevant to patients with cLBP. This search revealed a need for a PRO for use in clinical trials of cLBP.<sup>25</sup> Qualitative development of the instrument was based on expert clinical input, concept elicitation (CE) interviews, and cognitive interviews, and quantitative development was based on administration of the instrument to a large sample population (**Fig. 1**). Cognitive interviews were also conducted to assess the equivalency of the paper and electronic platforms of the PAL-S.<sup>15</sup>

### 2.2. Patient population and recruitment

Our recruitment strategy was designed to enroll a sample of patients across the spectrum of pain severity, including both patients with moderate to severe cLBP (similar to those who would be using the PRO assessments in future cLBP clinical trials) as well as patients with less severe pain. Patients with neuropathic and nonneuropathic pain experiences based on painDETECT criteria<sup>13</sup> were recruited. Adults

aged 18 to 80 years with a clinical diagnosis of LBP of nonmalignant origin with at least 3 months' duration were eligible. For CE, cognitive interviews, and preliminary psychometric validation, patients with a current pain score  $\geq 4$  on a 0 to 10-point numeric rating scale (NRS; moderate to severe pain) were eligible. For quantitative administrations, patients with a pain NRS score between 0 and 10 were eligible. Pain likely to be neuropathic LBP was identified based on a painDETECT<sup>13</sup> score  $>19$  during screening.<sup>14</sup> Overall, the population was intended to be diverse in terms of age, ethnicity, and other standard demographic characteristics.

Because the PAL-S was intended for use in global clinical trials, it was necessary for the content to reflect global patient experiences. Study participants were identified by clinicians at 5 clinical sites in the United States and at 2 market research facilities (MRFs) in Germany and the United Kingdom.



**Figure 1.** Steps in mixed-methods development of the PAL-S. Qualitative methods (blue) and quantitative methods (green) are shown for chronological steps in the development of the PAL-S.

Participants were screened by research staff at clinical sites and MRFs for eligibility, enrolled, and scheduled for qualitative interviews. This study was performed in accordance with Good Clinical Practice and applicable regulatory requirements. It was approved by the Quorum Review Institutional Review Board for sites in the United States; MRFs 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.

### 2.3. Concept elicitation interviews

The CE interview guide was designed to obtain both spontaneous and prompted patient input about the symptoms and impacts of cLBP. The initial questions were broad and open-ended, followed by probes to further describe the symptoms patients brought forward and to determine whether patients recognized other symptom descriptions common to patients with cLBP. These questions were followed by rating exercises on the full list of symptoms offered and endorsed. Contents of the interview guide were informed by expert panel input, the concepts identified in a review of patient-focused cLBP literature, and by the results of the initial focus group sessions with patients in the United States. In addition to the variability and impact of symptoms, the interview guide also explored in detail the specific language patients used to describe the unique sensations of pain they experienced due to cLBP.

All CE interviews were audio recorded and transcribed. All concepts mentioned by patients in the transcripts were assigned codes based on content, and organized into concept groups (based on similarity of content) using Atlas.ti software.<sup>21</sup> The coding framework provided an initial organization for grouping information and it was expanded throughout the process as new concept codes were identified in the transcripts.

### 2.4. Quality indicators for qualitative data

To assess saturation of concept (when no new information is forthcoming from interviews), transcripts were ordered chronologically and then grouped into quartiles of 10 and 11 transcripts. Newly established concept codes for each subsequent transcript group were compared with those derived from the preceding group. The absence of new concept codes in the last transcript group is interpreted as evidence that saturation was achieved. Interrater agreement was evaluated to identify the degree of consistency between the coders as they identified concepts and assigned codes. Interrater agreement was accomplished by dual coding 10% of the transcript database and comparing each pair of coded transcripts for differences.

### 2.5. Item generation

The goal in selecting the symptom concepts was to identify those most relevant to the patient experience, and important to clinicians in assessing changes in cLBP. Expert input, review of existing cLBP measures,<sup>25</sup> interview results, and specific patient language about the pain experience informed the development of the new PRO. The qualitative data from the patient interviews were reviewed by the development team (composed of pain specialists, PRO measurement scientists, and pharmaceutical company representatives) to select the most relevant symptom concepts for inclusion in the instrument. Draft items were then constructed to assess each selected concept at an appropriate reading level using patient-derived language and terminology.

## 2.6. Cognitive interviews (round 1) for instrument refinement

The draft measure was refined using a cognitive interview process to evaluate the accuracy and consistency of the patient comprehension of concepts presented in the draft items.<sup>24</sup> The cognitive interview process required patients to use a "think aloud" technique<sup>33</sup> to verbalize the thought process involved in responding to questions. Descriptors of pain were extensively defined and clarified. Patients were also asked about instrument instructions, response options, the recall period, and specific terminology to ensure appropriate understanding of the instrument.

### 2.7. First quantitative administration (pilot data collection)

The purpose of the first quantitative administration was to assess item performance and to identify items for potential revision. Classical test theory including evaluation of missing data, ceiling/floor effects, item-to-item correlations, item-to-total correlations, factor analysis, and reliability estimation was used for item reduction. Rasch measurement theory (RMT) analyses were used to assess item-level performance, and to examine the measurement model and scoring of the PAL-S instrument.

Data collection was conducted using an existing web-based panel through Ipsos Observer (<http://www.ipsos.com/observer/>). Individuals who had previously consented to join the Ipsos research panel and had previously reported cLBP received email invitations to complete the web-based survey. Web-based screening items were used to confirm eligibility, including: confirmation of self-report of previous 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. In addition to the PAL-S, other web-based questionnaire items assessed clinical variables (back pain severity, back pain location, pain movement, and sciatica/neuropathic pain assessment), treatment-related characteristics (currently being treated, length of time on treatment, medication type, satisfaction with current medication, and other nonmedication treatment), and demographic characteristics.

Descriptive statistics (sample size, frequency distribution, mean, median, range, and SD) and floor and ceiling effects were evaluated for the individual item responses for PAL-S items 1 to 13. Item-to-item correlation matrix (Pearson  $r$ ) was evaluated for each item, with coefficients greater than 0.70 potentially indicating a redundancy between the items.<sup>22</sup> Item-to-total correlations (bivariate Pearson  $r$ ) were examined for each item score against the total score (excluding the item of interest) with coefficients less than 0.40 potentially indicating nonassociations with the remaining items in the hypothesized scale.<sup>6</sup>

Rasch measurement theory analyses were used to examine whether each item exhibited appropriate psychometric scaling with the following criteria: (1) item response options were ordered; and (2) items formed a unidimensional construct. Items that did not fit the RMT model were flagged as candidates for item reduction or revision. The PAL-S items were assessed for model fit, using category probability curves (item characteristic curves) to identify items that did not demonstrate monotonically increased responses. For items that exhibited disordered thresholds (inconsistent responses), simulation analyses were conducted that collapsed response categories to assess potential improvements in item characteristics. To examine the consistency of the response pattern, a person-item distribution map was constructed. In this map, the distribution of the persons and items was displayed together on a logit scale with the most able persons and most difficult items on

**Table 1****Demographic and clinical characteristics of patients participating in concept elicitation and cognitive interviews.**

	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)

(continued on next page)

Table 1 (continued)

	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)
Pain intensity, mean NRS score (SD)*	6.7 (1.2)	7.0 (1.4)	7.0 (1.4)	6.0 (1.5)
Range	4-10	4-10	5-9	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.

NA, not available; NRS, numeric rating scale; NSAID, nonsteroidal anti-inflammatory drug.

one side and the least able persons and the easiest items on the other. The distance between items should not be more than 0.30 logits to avoid large gaps in the measurement.<sup>2</sup> Internal consistency of the PAL-S was assessed with Cronbach alpha.

### 2.8. Cognitive interviews (round 2) to test modifications

A revised version of the PAL-S instrument was evaluated in 2 waves of the second round of cognitive interviews (an additional 8 patients) (Fig. 1). Recruitment of these patients used the same eligibility criteria and procedures, and was conducted using similar cognitive interview methods as the initial qualitative development study.

### 2.9. Second quantitative administration

After completion of the cognitive interviews on the revised instrument, 2 separate US-based samples of adult patients with cLBP were recruited to test the revised questionnaire in confirmatory quantitative analyses. The first sample was a subset of 401 qualifying patients from the first quantitative administration (N = 598); this subset completed a web-based survey of the revised 14-item PAL-S. Results were analyzed to confirm item- and scale-level performance of the PAL-S using RMT analyses.

### 2.10. Preliminary psychometric validation

The second sample of patients to assess the revised questionnaire was a clinic-based cohort of physician-diagnosed cLBP patients recruited to complement the survey panel and to conduct preliminary psychometric validation evaluations. A total of 45 adults with cLBP were identified using patient records and completed the PAL-S, painDETECT, MOS-36, Roland-Morris Disability Questionnaire (RMDQ), and Neuropathic Pain Scale Inventory (NPSI) by paper. The painDETECT is a 9-item screening tool that was originally developed to distinguish neuropathic LBP (score of 19-38) from nonneuropathic LBP (score of 0-12).<sup>13</sup> It was given at the enrollment visit to categorize patients based on neuropathic/nonneuropathic pain. The MOS-36 is a multi-item scale that assesses 8 health concepts of which 2 were used in this study: (1) limitations in physical activities because of health problems, and (2) bodily pain.<sup>32</sup> MOS-36 scores range between 0 (most severe) and 100 (no disability). The RMDQ is a widely used health status measure designed to be completed by patients to assess physical disability due to LBP.<sup>26</sup> The 24-item RMDQ score ranges between 0 (no disability) and 24 (most severe). The NPSI is a self-administered questionnaire specifically designed to evaluate the different symptoms of

neuropathic pain.<sup>3</sup> Descriptors reflect spontaneous ongoing or paroxysmal pain, evoked pain (ie, mechanical and thermal allodynia/hyperalgesia), and dysesthesia/paresthesia. Scores range from 0 (no pain) to 10 (most severe).

Test-retest reproducibility was assessed with the intraclass correlation coefficient. Convergent validity was assessed using Pearson correlations with painDETECT, MOS-36, RMDQ, and NPSI. Known-groups validity was assessed by comparing PAL-S scores by pain NRS tertiles and painDETECT groups using an analysis of variance. Known-groups validity was also assessed with data provided by a subgroup of the sample population that participated in the second quantitative administration.

### 2.11. Cognitive interviews to assess paper to electronic equivalence

In conjunction with the clinic-based data collection, cognitive interviews were conducted with 8 participants to confirm that the intent and meaning of items, response options, and instructions were unaffected by the administration format.<sup>7</sup> Because it has been shown that under conditions of routine clinical practice, handheld computer questionnaires can give results equivalent to those obtained with a conventional paper questionnaire, further equivalence testing was not deemed necessary.<sup>17</sup>

## 3. Results

### 3.1. Concept elicitation results

Forty-three patients participated in CE interviews. The mean age was 48.6 years and 53.5% of patients were female (Table 1). Based on painDETECT scores, 32.5% of patients were unlikely to have neuropathic pain and 32.5% were likely to have neuropathic pain.

Evaluation of CE interview data indicated that saturation of concept was achieved by the end of the second transcript group. No new concepts were provided in the remaining 2 groups (20 patients). This finding provided evidence that the qualitative sample was robust enough to support complete elicitation of all meaningful concepts likely to be present in this study population across the 3 participating countries. The assessment of interrater agreement suggested a high degree of consistency in coding, as 82.5% to 86.9% agreement was observed between the 4 coders for the identification of concepts being expressed, and 97.2% to 99.2% agreement between coders regarding the assignment of specific concept codes to identified patient expressions.

As determined by number of patient expressions, the predominant symptom-related concepts were "unspecified cLBP pain,"

**Table 2****Summary of symptom concept code frequencies.**

Low back pain–symptom subdomains and concepts	No. (%) of 1342 total symptom expressions	No. (%) of 42 transcripts contributing to concept expression
Acute forms of pain	210 (16)	
Electric shocks/jolts	34 (2.5)	13 (31.0)
Excruciating pain	13 (1.0)	6 (14.3)
Sharp pain	85 (6.3)	22 (52.4)
Stabbing pain	66 (4.9)	20 (47.6)
Other acute pain forms	12 (0.9)	8 (19.0)
Chronic forms of pain	137 (10)	
Dull pain	40 (3.0)	18 (42.9)
Pressure/squeezing	50 (3.7)	21 (50.0)
Throbbing/pulsating pain	38 (2.8)	16 (38.1)
Other chronic pain forms	9 (0.7)	6 (14.3)
Radiating forms of pain	104 (8)	
Pulling pain	19 (1.4)	9 (21.4)
Shooting pain	63 (4.7)	20 (47.6)
Other radiating pain forms	22 (1.6)	12 (28.6)
Neuropathic forms of pain	318 (24)	
Burning pain	71 (5.3)	19 (45.2)
Heaviness	8 (0.6)	4 (9.5)
Hot/warm sensation	21 (1.6)	7 (16.7)
Numbness	90 (6.7)	27 (64.3)
Pins/needles	38 (2.8)	15 (35.7)
Tingling	45 (3.4)	22 (52.4)
Other neuropathic pain forms	45 (3.4)	17 (40.5)
General forms of pain	338 (25)	
Ache	88 (6.6)	19 (45.2)
Discomfort	14 (1.0)	6 (14.3)
Hurt	103 (7.7)	31 (73.8)
Sore	10 (0.7)	6 (14.3)
Tenderness	8 (0.4)	2 (4.8)
Unspecified pain	117 (8.7)	28 (66.7)
Other LBP symptoms	235 (18)	
Cramping	39 (2.9)	10 (23.8)
Muscle spasms	53 (3.9)	13 (31.0)
Stiffness	59 (4.4)	29 (69.0)
Swelling/inflammation	15 (1.1)	8 (19.0)
Tightness	36 (2.7)	18 (42.9)
Weakness	12 (0.9)	5 (11.9)
Other symptoms	21 (1.6)	11 (25.2)

LBP, low back pain.

“hurt,” “numbness,” and “ache” (**Table 2**). The most common spontaneously offered symptoms were numbness (reported by 51.2% of participants), burning (39.5%), and pain that was shooting (37.2%), stabbing (37.2%), and sharp (37.2%). The most bothersome symptoms were “excruciating pain,” “sharp pain,” “unspecified cLBP pain,” and “shooting pain.” The symptoms that patients rated as most severe were “unspecified cLBP pain,” “sharp pain,” “shooting pain,” “heaviness,” and “tightness.” Patients also described the most difficult symptoms to be “sharp pain,” “throbbing pain,” and “spasms.” “Unspecified” has been used to designate miscellaneous expressions of pain.

### 3.2. Item generation results

Item generation resulted in 23 symptom concepts being identified for inclusion in the PAL-S instrument (**Table 3**).

### 3.3. Cognitive interview round 1 results

A total of 30 patients participated in 4 waves during round 1 of the cognitive interviews. The mean age was 43.6 years and

56.7% were female (**Table 1**). Based on painDETECT scores, 45.0% were unlikely to have neuropathic pain and 45.0% were likely to have neuropathic pain.

Patient input from cognitive interviews led to the deletion of 9 items (eg, “dull pain,” “cold pain,” “itchy pain,”), modifications to the existing instructions (eg, “Please use the scale below...” changed to “Please rate your [symptom]...”), and modifications of the parenthetical descriptors of the various pain sensations. The remaining 14 items were used for quantitative testing. Cognitive interview data provided evidence from patients that the instrument was comprehensive, relevant to the cLBP experience, understandable, and easy to complete.

### 3.4. First quantitative administration results

A total of 598 patients participated in pilot data collection (first quantitative administration). The mean age was 55.5 years and 67.9% were female (**Table 4**). Based on the 11-point NRS pain scale, levels of pain intensity were well distributed.

**Table 3**  
**Low back pain concepts selected for PRO measurement.**

Targeted symptom concept	Actions on concepts and final item in PAL-S
Sharp (stabbing) sensation	Item 2: sharp (stabbing) back pain
Hot sensation	Item 9: burning back pain
Dull sensation	Removed after cognitive interview wave 2
Cold sensation	Removed after cognitive interview wave 3
Sensitive sensation	Item 4: sensitive (as if sunburned or raw to touch) back pain
Tender sensation	Item 5: tender (like a bruise) back pain
Itchy sensation	Removed after cognitive interview wave 3
Shooting sensation	Item 9: shooting back pain
Numb sensation	Removed after cognitive interview wave 3
Electrical sensation	Item 7: shocking (jolting) back pain
Tingling sensation	Item 3: prickling (tingling) back pain
Cramping sensation	Item 10: cramping (squeezing) back pain
Spasms	Item 11: back muscle spasms
Radiating sensation	Item 6: radiating (spreading) back pain
Throbbing sensation	Item 12: throbbing (pounding) back pain
Aching sensation	Item 13: aching (sore, nagging) back pain
Heavy sensation	Removed after cognitive interview wave 2
Stiffness	Item 14: stiffness (tight, less flexible) in back
Pulling sensation	Removed after cognitive interview wave 3
Unpleasant sensation	Removed after cognitive interview wave 2
Deep pain/surface pain	Removed after cognitive interview wave 2
Pain variability	Removed after cognitive interview wave 3
Overall pain intensity	Item 1: worst back pain

PAL-S, patient assessment for low back pain–symptoms; PRO, patient-reported outcome.

Patient Assessment for Low Back Pain–Symptoms items were endorsed in the first quantitative administration. Item mean scores ranged from 3.16 (please rate your worst pricking pain, SD 3.04) to 6.34 (worst back pain, SD 2.40). High floor effects (subjects indicating no symptom at all) were seen with 5 items (please rate your worst pricking, most sensitive, worst burning, worst cramping, and worst throbbing). Seven PAL-S item pairs were strongly correlated ( $>0.65$ ) with each other. Cronbach alpha for the 14 items was 0.94; alphas remained at 0.94 if any item were deleted. Results of the RMT analysis showed only 2 items with an ordered threshold: worst back pain and worst aching pain. The remaining items were disordered, most likely from the patients' inability to distinguish between the lowest levels of the scale (scores 0–3). The person–item threshold distributions suggested a somewhat narrow distribution that could potentially benefit from item modifications. Analytic simulations were conducted by collapsing the NRS responses into a 4-point categorical scale to allow visualization of the potential performance of a revised response scale. Results of the simulation suggested a 4-level verbal scale to be superior as the simulated response scale displayed ordered thresholds for all items.

Based on these results, changes were made to the PAL-S. No items were added or deleted, and no changes were made to the content of existing items. Response options for 13 of the items were changed from NRS format to a 4-level verbal rating scale. The verbal rating scales were tailored to each item to represent

responses of “not at all,” “slight,” “somewhat” (or moderate), or “very” (or severe). For example, the responses for muscle spasms would be: no muscle spasms, slight muscle spasms, moderate muscle spasms, or severe muscle spasms. Only the item “Please rate how bad your worst back pain was over the past 7 days” was retained in its existing format (0–10 NRS).

### 3.5. Cognitive interview round 2 results

Modifications made to the PAL-S based on item-level analyses were evaluated using a second round of cognitive interviews with 8 patients. The mean age was 47.2 years and 50.0% were female (Table 1). Based on the painDETECT score, 37.5% were unlikely to have neuropathic pain and 50.0% were likely to have neuropathic pain.

After the second round of cognitive interviews, additional changes were made to the PAL-S. To better facilitate consistency between paper and electronic administration formats, the instructions to “mark one box” were replaced with “select one” in each of the items using the verbal rating scale. The item assessing cramping/squeezing back pain was modified to minimize confusion with menstrual cramping expressed by some participants during the first quantitative administration. The main concept descriptor and the parenthetical descriptor in the item were swapped so that “squeezing” was presented as the primary descriptor for the item and “cramping” was provided as a parenthetical descriptor. With these changes, patients demonstrated understanding of the concepts in the PAL-S, confirmed the relevance of the concepts, and expressed no difficulty in providing answers.

### 3.6. Quantitative findings from second administration

From the initial web-based sample (N = 598), 401 patients participated in the second quantitative administration. The mean age was 55.3 years and 67.8% were female (Table 4). Based on painDETECT scores, 56.9% were unlikely to have neuropathic pain and 22.7% were likely to have neuropathic pain.

All PAL-S items and response options were endorsed. Mean item scores ranged from 0.98 (worst pricking pain, SD 1.02) to 2.19 (worst aching pain, SD 0.81) (Table 5). No data were missing. High floor effects were seen with 5 items (worst pricking pain, worst burning pain, worst squeezing pain, most sensitive pain, and worst shocking pain). A high ceiling effect (subjects indicating a 3) was observed for 1 item (worst aching pain). The highest item-to-item correlation ( $r = 0.65$ ) was seen between item 2 (worst sharp pain) and item 8 (worst shooting pain). Item-to-total correlations ranged from 0.54 to 0.71. Cronbach alpha for the 13 items was 0.91 with alphas remaining around 0.90 if any item was deleted. All items fit the RMT model, with no items exhibiting a disordered threshold. Results of the second quantitative administration showed the revisions to have improved the item-level performance of the PAL-S instrument, and the scale structure and scoring seemed to function correctly.

### 3.7. Preliminary psychometric validation

Forty-five patients with physician-diagnosed cLBP participated in the preliminary psychometric validation analyses. The mean age was 53.0 years and 52.3% were female (Table 4). Based on painDETECT scores, 31.8% were unlikely to have neuropathic pain and 29.5% were likely to have neuropathic pain.

All PAL-S items and response options were endorsed. PAL-S item mean scores ranged from 1.27 (worst pricking pain) to

**Table 4****Demographic and clinical characteristics of patients participating in quantitative analyses.**

	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/y (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.

2.42 (worst aching pain). The mean score for item 1 (how bad was your worst back pain) was 6.98 on the 0 to 10 NRS (SD 2.38). Test–retest reproducibility at 1 week was acceptable, with intraclass correlation coefficient of 0.81 (95% confidence interval: 0.61-0.91). In convergent validity assessments, strong associations were seen between the PAL-S total score and the MOS-36 Bodily Pain (−0.79), NPSI (0.73), RMDQ (0.67), and the MOS-36 Physical Functioning (−0.65). Known-groups validity was examined by comparing group mean values using an analysis of variance. Within the second quantitative administration, all individual items and the total score for the PAL-S were able to significantly discriminate ( $P < 0.001$ ) between NRS tertile groups (Table 6). Although fewer items attained actual significance (3 items), score trends

were similar when compared with the physician-diagnosed patients. The PAL-S total score was significantly different ( $P < 0.05$ ) (Table 6). painDETECT groups (scores <13, 13-18, and >18) were also examined (data not shown). In the second quantitative administration sample, all items and total score were significant ( $P < 0.001$ ). Within the physician-diagnosed sample, only one item (worst squeezing [cramping]) was not significant.

### 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



**Table 5**  
**Descriptive characteristics of PAL-S items from second quantitative administration.**

Item # and content	Mean score (SD) [median]	Floor n (%)	Ceiling n (%)	Item-to-item correlation >0.65	Item-to-total score correlation	Alpha if item deleted
1. How bad was your worst back pain (0-10 point scale*)	6.45 (2.24) [7]	4 (1.0)	30 (7.5)	Item 8	NA	NA
2. Worst sharp (stabbing)	1.64 (0.98) [2]	55 (13.7)	89 (22.2)		0.6616	0.8993
3. Worst pricking (pins and needles)	0.98 (1.02) [1]	175 (43.6)	40 (10.0)		0.6307	0.9005
4. Most sensitive (as if sunburned or raw to touch)	1.19 (1.06) [1]	139 (34.7)	54 (13.5)		0.6170	0.9012
5. Worst tender (like a bruise)	1.54 (1.02) [2]	75 (18.7)	82 (20.4)		0.6286	0.9006
6. Worst radiating (spreading)	1.58 (1.03) [2]	69 (17.2)	94 (23.4)		0.6533	0.8995
7. Worst shocking (jolting)	1.16 (1.00) [1]	124 (30.9)	48 (12.0)		0.6326	0.9005
8. Worst shooting	1.41 (1.05) [1]	97 (24.2)	75 (18.7)	Item 1	0.7090	0.8970
9. Worst burning	1.00 (1.02) [1]	169 (42.1)	41 (10.2)		0.6122	0.9013
10. Worst squeezing (cramping)	1.09 (1.04) [1]	149 (37.2)	50 (12.5)		0.5736	0.9031
11. Worst muscle spasms	1.31 (1.05) [1]	116 (28.9)	61 (15.2)		0.6245	0.9008
12. Worst throbbing (pounding)	1.25 (1.04) [1]	119 (29.7)	59 (14.7)		0.6297	0.9006
13. Worst aching (sore, nagging)	2.19 (0.81) [2]	8 (2.0)	170 (42.4)		0.5439	0.9041
14. Worst stiffness (tight, less flexible)	1.85 (0.87) [2]	27 (6.7)	99 (24.7)		0.5909	0.9023

\* Scale of 0 to 10, with higher scores indicating greater pain intensity.  
 NA, not applicable; PAL-S, patient assessment for low back pain-symptoms.

response options was affected by differences in modes of administration (paper vs electronic). However, formatting issues with both the paper and electronic versions were observed that made it easier or more difficult for patients to read the items thoroughly. Examples of formatting issues include 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. These features were all noted for future formatting on both paper and electronic versions.

#### 4. Discussion

The PAL-S is a newly developed PRO assessment tool designed to specifically reflect the pain sensations experienced by patients with cLBP. The qualitative evidence collected during this study supports the assessment of both the neuropathic and non-neuropathic sensations that patients with cLBP often report, and shows that these specific sensations are consistently described and recognized, and substantiates the content validity of this new measure. The variations of severity and bother observed among the specific concepts of pain underscore their importance in evaluating how patients feel and function, and are therefore relevant and important in the assessment of potential treatment benefit. The PAL-S is different from existing pain symptom measures in that it has been developed specific to cLBP and was developed in accordance with FDA guidelines to support product labeling<sup>29</sup> and with International Society for Pharmacoeconomics and Outcomes Research best practices.<sup>23,24</sup> Content validity was established based on qualitative evidence from CE focus groups, individual CE interviews, and cognitive interviews. Evidence of item-level performance was based on both classical test theory and RMT, and supported refinement of the measure during the mixed-methods process. Preliminary assessment of the measurement properties of the revised PAL-S showed acceptable reliability and validity.

The most predominant symptom-related concepts were unspecified cLBP pain, hurt, numbness, and ache. The most common spontaneously offered symptoms were numbness, burning pain, shooting pain, stabbing pain, and sharp pain. The most bothersome symptoms were excruciating pain, sharp pain, unspecified cLBP pain, and shooting pain. Unspecified cLBP was left as an overall pain NRS item but was not included in the calculation of the PAL-S total score. The unspecified cLBP item represents an overall general assessment that includes more specific types of pain and was therefore not a unique item in comparison with the more specific pain sensations. Although not included in the PAL-S total score, its usefulness as an independent overall pain descriptor is valid.

Further development and testing of the PAL-S is planned. A study will be conducted to formally assess additional psychometric characteristics of the PAL-S instrument, including construct (convergent) validity, test-retest reproducibility, internal consistency reliability, known-groups validity, and sensitivity to change. In addition, translation and linguistic validation activities are planned.

As with all qualitative studies, a limitation of the study was the relatively small sample sizes for CE, cognitive interviews, and the preliminary validation study. Notably, saturation of concept was achieved in the second wave of CE interviews, suggesting comprehensive capture of the patient experience with cLBP. Sample sizes for the quantitative administrations were large, but the diagnosis of cLBP was self-reported by the patients. A clinic-based sample of physician-diagnosed patients with cLBP was therefore recruited for the preliminary validation analyses. Notably, results of known-groups validity analyses based on NRS tertiles were similar between patients with self-reported cLBP and the cohort with physician-reported cLBP. Patients who participated in the study may not represent patients enrolled in clinical trials of therapies for

**Table 6****Known-groups validity of the PAL-S: 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. How bad was your worst back pain (0-10 point scale*)	5.33 (1.87)	6.77 (1.62)	8.21 (1.54)‡	6.68 (1.89)	8.79 (2.61)	9.08 (1.93)‡
2. Worst sharp (stabbing)	1.32 (0.89)	1.61 (0.84)	2.31 (0.82)‡	2.58 (0.84)	3.07 (1.07)	3.33 (0.99)
3. Worst prickling (pins and needles)	0.61 (0.76)	0.95 (0.96)	1.55 (1.15)‡	2.16 (0.90)	2.14 (1.10)	2.58 (1.17)
4. Most sensitive (as if sunburned or raw to touch)	0.72 (0.76)	1.31 (1.02)	1.75 (1.11)‡	1.95 (1.03)	2.43 (0.94)	2.75 (1.22)
5. Worst tender (like a bruise)	1.18 (0.87)	1.63 (1.01)	1.99 (0.97)‡	2.79 (0.86)	2.71 (0.91)	3.08 (1.17)
6. Worst radiating (spreading)	1.31 (0.97)	1.53 (0.93)	2.18 (0.95)‡	2.68 (0.89)	3.00 (1.18)	3.50 (0.91)‡
7. Worst shocking (jolting)	0.92 (0.99)	1.11 (0.86)	1.70 (1.02)‡	2.21 (0.86)	2.64 (1.22)	2.42 (1.38)
8. Worst shooting	1.04 (0.92)	1.44 (0.94)	2.03 (1.01)‡	2.42 (1.07)	2.57 (1.28)	2.83 (1.27)
9. Worst burning	0.67 (0.88)	0.90 (0.92)	1.60 (1.09)‡	2.00 (0.88)	2.50 (1.29)	3.08 (1.17)§
10. Worst squeezing (cramping)	0.81 (0.90)	0.98 (1.02)	1.69 (1.00)‡	2.16 (0.90)	2.64 (1.01)	2.67 (0.99)
11. Worst muscle spasms	0.96 (0.93)	1.27 (0.97)	1.93 (0.98)‡	2.26 (0.81)	2.57 (0.94)	3.50 (0.91)
12. Worst throbbing (pounding)	0.81 (0.78)	1.29 (0.97)	1.90 (1.04)‡	2.63 (0.96)	2.71 (1.14)	3.42 (0.79)
13. Worst aching (sore, nagging)	1.91 (0.77)	2.24 (0.81)	2.58 (0.70)‡	3.21 (0.63)	3.57 (0.85)	3.58 (0.67)
14. Worst stiffness (tight, less flexible)	1.63 (0.78)	1.89 (0.80)	2.19 (0.88)‡	2.84 (0.69)	3.00 (0.88)	3.42 (0.90)
Total score	1.07 (0.47)	1.40 (0.55)	1.95 (0.66)‡	1.45 (0.42)	1.74 (0.68)	2.09 (0.76)§

\* Scale of 0 to 10, with higher scores indicating greater pain intensity.

†Mean differences between three groups are significant ( $p \leq 0.001$ ).

‡Mean differences between three groups are significant ( $p \leq 0.01$ ).

§Mean differences between three groups are significant ( $p \leq 0.05$ ).

NRS, numeric response scale (0-10 scale with higher scores indicating greater severity); PAL-S, patient assessment for low back pain-symptoms.

cLBP, as clinical trials have rigid eligibility criteria. In addition, the inclusion of patients with self-reported cLBP and no supporting clinical information prevents a full characterization of the study population, including potentially important comorbidities. A strength of the study was the use of the web-based panel of patients, which represented the spectrum of pain severity and importantly represented patients with both neuropathic and nonneuropathic pain.

In conclusion, the PAL-S is a newly developed PRO instrument that is designed to assess treatment benefit for label claims in clinical trials for cLBP. The instrument was developed in accordance with US FDA PRO guidance, with cross-cultural patient input. The PAL-S reflects the specific symptoms of pain associated with cLBP, and is not a generic measure of pain in general. The mixed-methods development process reported here describes the new instrument and provides evidence of content validity for the PAL-S.

### Conflict of interest statement

M.L. Martin is an employee of Health Research Associates, Inc, which received funding from Grünenthal GmbH and Forest Research Institute to conduct the study. S.I. Blum reports personal fees from Forest Research Institute (a subsidiary of Forest Laboratories, Inc), and other support from Forest Laboratories, Inc

during the conduct of the study; personal fees and other support from GlaxoSmithKline, and personal fees and nonfinancial support from Patient Centered Outcomes Research Institute, outside the submitted work. H. Liedgens is a full-time employee of Grünenthal GmbH. D.M. Bushnell is an employee of Health Research Associates, Inc, which received funding from Grünenthal GmbH and Forest Research Institute to conduct the study. K.P. McCarrier is an employee of Health Research Associates, Inc, which received funding from Grünenthal GmbH and Forest Research Institute to conduct the study. N.V. Hatley is an employee of Health Research Associates, Inc, which received funding from Grünenthal GmbH and Forest Research Institute to conduct the study. R. Freynhagen reports personal fees from Astellas, Grünenthal GmbH, Lilly, Pfizer, Merck, Develco, Mitsubishi Tanabe Pharma, and Galapagos outside the submitted work. M. Wallace reports personal fees from Grünenthal GmbH, outside the submitted work. M. Eerdeken reports other support from Grünenthal GmbH, during the conduct of this study; other support from Grünenthal GmbH, outside the submitted work. M. Kok reports other support from Grünenthal GmbH, during the conduct of the study; other support from Grünenthal GmbH outside the submitted work. The remaining authors have no conflicts of interest to declare.

This study was supported by research funds from both Forest Research Institute and Grünenthal GmbH.

## Acknowledgments

Carla Asoytia (Health Research Associates, Inc) managed the interview process. Julia R. Gage (Gage Medical Writing, LLC; on behalf of Health Research Associates, Inc) provided assistance with editing the manuscript.

## Article history:

Received 9 August 2017

Received in revised form 24 January 2018

Accepted 1 February 2018

Available online 7 February 2018

## References

- [1] Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine (Phila Pa 1976)* 2000;25:3100–3.
- [2] Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 2nd ed. London: Lawrence Erlbaum, 2007.
- [3] Bouhassira D, Attal N, Fermanian J, Alchaar H, Gautron M, Masquelier E, Rostaing S, Lanteri-Minet M, Collin E, Grisart J, Boureau F. Development and validation of the neuropathic pain symptom inventory. *PAIN* 2004;108:248–57.
- [4] Campbell P, Foster NE, Thomas E, Dunn KM. Prognostic indicators of low back pain in primary care: five-year prospective study. *J Pain* 2013;14:873–83.
- [5] Cancelliere C, Donovan J, Stochkendahl MJ, Biscardi M, Ammendolia C, Myburgh C, Cassidy JD. Factors affecting return to work after injury or illness: best evidence synthesis of systematic reviews. *Chiropr Man Therap* 2016;24:32.
- [6] Clark LA, Watson D. Constructing validity: basic issues in objective scale development. *Psychol Assess* 1995;7:309–19.
- [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] da C Menezes Costa L, Maher CG, Hancock MJ, McAuley JH, Herbert RD, Costa LO. The prognosis of acute and persistent low-back pain: a meta-analysis. *CMAJ* 2012;184:E613–624.
- [9] Dunn KM, Jordan KP, Croft PR. Contributions of prognostic factors for poor outcome in primary care low back pain patients. *Eur J Pain* 2011;15:313–19.
- [10] Ferrer M, Pellise F, Escudero O, Alvarez L, Pont A, Alonso J, Deyo R. Validation of a minimum outcome core set in the evaluation of patients with back pain. *Spine (Phila Pa 1976)* 2006;31:1372–9.
- [11] Förster M, Mahn F, Gockel U, Brosz M, Freynhagen R, Tölle TR, Baron R. Axial low back pain: one painful area—many perceptions and mechanisms. *PLoS One* 2013;8:e68273.
- [12] Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, Castel LD, Kalsbeek WD, Carey TS. The rising prevalence of chronic low back pain. *Arch Intern Med* 2009;169:251–8.
- [13] Freynhagen R, Baron R, Gockel U, Tölle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. *Curr Med Res Opin* 2006;22:1911–20.
- [14] 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.
- [15] Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 2008;11:322–33.
- [16] Hawker GA. The assessment of musculoskeletal pain. *Clin Exp Rheumatol* 2017;107(suppl 35):8–12.
- [17] Junker U, Freynhagen R, Langler K, Gockel U, Schmidt U, Tölle TR, Baron R, Kohlmann T. Paper versus electronic rating scales for pain assessment: a prospective, randomised, cross-over validation study with 200 chronic pain patients. *Curr Med Res Opin* 2008;24:1797–806.
- [18] Kaiser U, Neustadt K, Kopkow C, Schmitt J, Sabatowski R. Core outcome sets and multidimensional assessment tools for harmonizing outcome measure in chronic pain and back pain. Vol. 4. Basel: Healthcare, 2016.
- [19] Mannion AF, Elfering A, Staerke R, Junge A, Grob D, Semmer NK, Jacobshagen N, Dvorak J, Boos N. Outcome assessment in low back pain: how low can you go? *Eur Spine J* 2005;14:1014–26.
- [20] Montgomery W, Vietri J, Shi J, Ogawa K, Kariyasu S, Alev L, Nakamura M. The relationship between pain severity and patient-reported outcomes among patients with chronic low back pain in Japan. *J Pain Res* 2016;9:337–44.
- [21] Muhr T. User's manual for ATLAS.ti 5.0. Berlin: ATLAS.ti Scientific Software Development GmbH, 2004.
- [22] Nunnally JC. Psychometric theory. 2nd ed. New York: McGraw-Hill, 1978.
- [23] 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 1—eliciting concepts for a new PRO instrument. *Value Health* 2011;14:967–77.
- [24] 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.
- [25] 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.
- [26] Roland M, Fairbank J. The Roland-Morris disability questionnaire and the Oswestry disability questionnaire. *Spine (Phila Pa 1976)* 2000;25:3115–24.
- [27] Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. *Value Health* 2009;12:1075–83.
- [28] Steenstra IA, Munhall C, Irvin E, Oranye N, Passmore S, Van Eerd D, Mahood Q, Hogg-Johnson S. Systematic review of prognostic factors for return to work in workers with sub acute and chronic low back pain. *J Occup Rehabil* 2017;27:369–81.
- [29] 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.
- [30] United States Food and Drug Administration. Clinical outcome assessment qualification program. 2017. Available at: <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm>. Accessed June 15, 2017.
- [31] Verkerk K, Luijsterburg PA, Heymans MW, Ronchetti I, Pool-Goudzwaard AL, Miedema HS, Koes BW. Prognosis and course of disability in patients with chronic nonspecific low back pain: a 5- and 12-month follow-up cohort study. *Phys Ther* 2013;93:1603–14.
- [32] 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.
- [33] Willis GB. Cognitive interviewing: a tool for improving questionnaire design. London: Sage Publications, 2005.