ORIGINAL ARTICLE

The assessment of symptoms and functional limitations in low back pain patients: validity and reliability of a new questionnaire

Martin Björklund · Jern Hamberg · Marina Heiden · Margareta Barnekow-Bergkvist

Abstract Many of the existing low back pain (LBP) questionnaires of function and symptoms have a content of different domains of disability presented as a single sum score, making it difficult to derive changes within a specific domain. The present study describes the development of a clinically derived back-specific questionnaire incorporating both a functional limitation and a symptom scale, with a further subdivision of the symptom scale in separate indices for severity and temporal aspects. The aims of the study were to assess the overall reliability and validity of the new questionnaire, named the Profile Fitness Mapping questionnaire (PFM). A total of 193 chronic LBP patients answered the PFM together with five validated criterion questionnaires. For the internal consistency of the questionnaires, the three indices of the PFM had the highest Cronbach's alpha (0.90–0.95) and all items had item–total correlations above 0.2. The correlation coefficients between the PFM and the back-specific criterion questionnaires ranged between 0.61 and 0.83, indicating good concurrent criterion validity. The best discriminative ability between patients with different pain severities was demonstrated by the functional limitation scale of the PFM. Well centered score distribution with no patient's score at the floor or the ceiling level indicates that the PFM has the potential to detect the improvement or worsening of symptoms and functional limitations in chronic LBP patients. Classification according to the International Classification of Functioning, Disability and health (ICF) of WHO revealed a high degree of homogeneous item content of the symptom scale to the domain of impairments, and of the functional limitation scale to the domain of activity limitations. The present study suggests that the PFM has a high internal consistency and is a valid indicator of symptoms and functional limitations of LBP patients. It offers the combination of a composite total score and the possibility of evaluations within specific domains of disability. Complementary evaluation of test–retest reliability and responsiveness to change is warranted.

Keywords Disability · Back pain · Condition-specific questionnaire · Outcome measure

Introduction

Low back pain (LBP) is a common condition that causes much individual suffering and a large burden on medical service and society [10]. Assessment and documentation of the patient's pain and other symptoms and functional status have become an essential part in understanding the impact of the LBP on the patient's life. Validated measures of symptoms and function are also necessary in clinical practice and trials in order to evaluate treatment efforts.

Pain is composed of different sensory and affective qualities that need to be assessed for a broader description of a patient's experience of pain. The presence and intensity of pain is a poor health outcome on its own [e.g., 30], and furthermore it correlates poorly with measures of physical functioning [26, 34]. Pain and other symptoms belong to the core outcome measures recommended for chronic pain clinical trials [9]. One usual shortcoming of pain measures is the lack of assessment of the temporal aspects of pain, such as the frequency of pain episodes [9].

M. Björklund (✉) · M. Heiden · M. Barnekow-Bergkvist
Centre for Musculoskeletal Research, University of Gävle,
Box 7629, 907-12 Umeå, Sweden
e-mail: mbb@hig.se

M. Björklund · J. Hamberg
Alfta Research Foundation, Box 94, 822-22 Alfta, Sweden

Evidence indicates that measures of pain frequency are valid and they represent a dimension of pain discerned from the intensity of pain [15].

The International Classification of Functioning, Disability and Health (ICF) of WHO [37] provides a biopsychosocial model of disability for diseases and disorders, divided into three perspectives of health—a bodily, personal and social perspectives. The ICF represents one of few existing reference frames that can be used as a tool to classify the content of questionnaires [13]. Condition-specific questionnaires for the assessment of functional status in patients with LBP (back-specific questionnaires) often have mixed contents that reflect constructs of both pain and physical functioning [13]. These constructs fall under the ICF domains of ''impairments'' and ''activity limitations'', which corresponds to bodily and personal perspectives, respectively. For some patients, the LBP is primarily manifested in decreased activity and a treatment effect may be revealed in increased activity with small changes in pain intensity [26]. For other patients, a desired level of activity is maintained during the episode of LBP at the expense of increased pain, and an improved status following successful treatment may show decreased pain intensity with little change in activity [9]. Hence, it is conceivable that questionnaires try to satisfy the need for the assessment of the combined effect on both pain and physical functioning. However, such composite measures, providing only an overall sum score, make it impossible to derive changes within a specific domain. If the patient's disability has improved in one domain, but worsened in another, a single sum score will fail to detect the changes—an effect referred to as item-masking score bias [23]. Therefore, much could be gained by having back-specific questionnaires with subscales of the specific domains of interest. The present study describes the development of a clinically derived back-specific questionnaire incorporating both a functional limitation and a symptom scale, with a further subdivision of the symptom scale in separate indices for severity and temporal aspects.

The aims of the study were to assess the overall validity and reliability of the new questionnaire, named the Profile Fitness Mapping questionnaire. This was done in a study on chronic LBP patients that answered the new questionnaire contemporaneously with five validated criterion questionnaires from the literature.

Methods

The concept of reliability reflects the degree to which test scores are influenced by random measurement errors, whereas the validity of a scale refers to the extent to which it measures what it intends to measure [5]. In this study, we assessed reliability and validity of a new questionnaire, named Profile Fitness Mapping questionnaire (PFM), by comparison with five validated criterion questionnaires from the literature that included four back-specific and one generic health questionnaires. A condition-specific measure may assess the aspects of relevance for the specific target group and may therefore be more responsive to treatment effects than a generic health questionnaire [11, 33]. However, an additional assessment of different aspects of health-related quality of life is important for all research on chronic pain including LBP [9]. Hence, the combination of condition-specific and generic measures in LBP research is recommended [9, 30].

For reliability, the concept of *internal consistency*, i.e., the extent to which similar questions give consistent responses, was used. For validity, the assessment encompassed components within the concepts of *concurrent criterion validity, construct validity, face validity* and *content validity* (see Statistical analyses for description).

Subjects

Patients referred to Alfta Rehab Center, a rehabilitation clinic in Alfta, Sweden, between 1994 and 1998 answered the PFM and the five criterion questionnaires on admission. In total, 1,040 LBP patients answered the questionnaires during this time period. The following inclusion criteria were set up: (i) diagnosis with a LBP disorder, categorized according to ICD-9 codes, on the basis of clinical history and an orthopedic physical examination; (ii) pain in the low back only, classified according to the letter of referral, the case record and the pain drawing [20, 25] made by each patient on admission; (iii) LBP at rest or associated with back strain for at least 1 year. Individuals were excluded if they had evidence of rheumatoid arthritis, cancer, a connective tissue disease, or a current infectious disease. Thus, 193 participants were selected for the purpose of validating the PFM. Note that the group of chronic LBP patients participating in the developmental process of the PFM (see below) was not included in the subject group of the study.

The study was performed after obtaining advisory pronouncement by the Regional Ethical Review Board in Uppsala, and informed consent from each subject.

Scale development

The PFM questionnaire was developed between 1992 and 1994 at the Alfta Rehab Center and consists of two back-specific scales, designed for the assessment of self-estimated symptoms and functional limitations. In this study, the expression of functional limitations refers to the activity limitations experienced because of the low back problems. The symptom scale contains two indices of separate aspects

of symptomatology. The functional limitation scale (one index) and the symptom scale (two indices) are presented in a self-administered form. The three indices constitute the fitness mapping of each individual, hence, the name of the questionnaire.

Three sources of information were used for the development process; (i) a group of 20 chronic LBP patients, (ii) the literature on the topic and (iii) an expert group of health professionals lead by one of the authors (J.H.). The expert group consisted of two orthopedic surgeons, one physiologist, one specialist in orthopedic medicine and six physiotherapists, all with extensive clinical experience of LBP patients.

The symptom scale

The purpose of the symptom scale of the PFM was to assess both the severity and the temporal aspects of the symptomatology of LBP patients. For that reason, the symptom scale was designed to measure both *the intensity* and *the frequency* of the symptoms, thus, yielding two separate indices.

Initially in the development of the scale, a list of symptoms commonly experienced by LBP patients was generated by asking the chronic LBP-patient group to rate all their symptoms in a severity order. The obtained list was thereafter revised by the expert group, whose task was both to complement the list and to control that symptoms noted by the patients were not overlapping. Four items were added—the impact of back pain on sleep, mood, sexual life and need for support (e.g. corset or cane). The revised list, comprising 27 items, formed the basis for the symptom scale of the PFM. For each item in the scale, the respondent is asked *how often* and *how much* he/she experiences the symptoms. All items are given six response alternatives (*how often*: range from 1 = never/very seldom, to 6 = very often/always; *how much*: range from 7 = nothing/none at all, to 12 = almost unbearable/unbearable, all/maximally). An average option was thus avoided and the response scale of six alternatives was judged sensitive enough. This is in line with the general recommendation of five to seven response alternatives for scales where intensity or frequency is to be judged [16].

All items were checked with the chronic LBP-patient group for face validity and the revised scale was pre-tested to avoid ambiguity of the questions and to ensure feasibility and comprehensibility. Any items not meeting these criteria were rewritten and tested again. Responses of the items were also checked for skewed distribution. In the final stage, a weighting of each individual item, based on the initial rating of the symptoms made by the chronic LBP-patient group, was determined. Symptoms with high severity rating were given higher weighting. Higher index scores reflect less symptoms/better health. See the Appendix.

The functional limitation scale

The main purpose of the functional limitation scale of the PFM was to assess how back problem affected the capability to perform an activity of daily life. First, the chronic LBP-patient group was asked to write down answers to the question: *What is difficult to do in your daily life because of your back problem?* The obtained list of the activity limitations of the patient group was discussed within the expert group with respect to the construct of interest. The expert group gave preference to simple physical activities over more complex situation-based tasks. The reason to this was the assumption, based on clinical observations, that limitations in simple activities often cause difficulties in more complex motor tasks, making the simple activities more valid for a general population of chronic LBP patients. Three items, suggested by the expert group, were added—perceived condition of the back, general health and likelihood of return to work. In the process of developing the functional limitation scale, items were continuously tested on the LBP patients to ensure comprehensibility. The revised list, comprising 28 elementary activities, formed the basis for the functional limitation scale of the PFM. All items were given six response alternatives (ranging from 1 = very good, no problem, very satisfying, very likely, to 6 = very bad, very difficult/impossible, very dissatisfying, very unlikely).

The same procedure for face validity was carried out as for the development of the symptom scale (see above). In the final stage, a weighting of each individual item was determined based on the question: *What would it cost to lose the capability to perform this activity, to lose this function?* Both the opinions of the expert group and of the chronic LBP-patient group were taken into account for the conclusive weighting. Higher index scores reflect better function/better health. See the Appendix.

Criterion questionnaires

The reliability and validity of the PFM were assessed by comparison with four back-specific and one generic health questionnaires from the literature, thus serving as criterion questionnaires. The criterion questionnaires all belong to the most widely used and validated self-administered questionnaires of the literature [9, 21, 27]. *The Aberdeen low back pain disability scale* [30] includes 19 items of how the pain affects activities like self-care, walking, sitting, standing, sport, housework, resting, bending and sleep. Higher scores

reflect poorer health. The scale is considered reliable [21] and well validated and thus recommended for use without further validation studies [13]. *The low-back outcome score* [12] comprises 13 weighted items about pain, rest, treatment, consumption of painkillers, work and activities of daily living. Items of pain and active activities are weighted more than treatments and rest required, which in turn are weighted more than passive activities. Higher scores reflect better function. *The low-back outcome score* has passed an extensive validation process, is reliable and is recommended as a short general assessment for backache, pain medication, ability to work and leisure activities [for references, see 21, 22]. *The Waddell disability index* [35] is a nine-item scale on basic physical activities of daily living commonly restricted by low-back pain, comprising walking, sitting, standing, lifting, sex life, traveling, sleeping, dressing and social life. Higher scores reflect more disability. *The Waddell disability index* is considered well validated and recommended for the assessment of functional status and disability on LBP patients [13]. *The Roland–Morris disability questionnaire* [29] contains 24 statements selected from the 136-item sickness impact profile [4] and measures disability of the natural history of LBP through aspects of daily living like sleeping, appetite, self-care, walking, lifting, work, dressing, housework and resting. Each statement gives one point, meaning that higher scores reflect more disability. *The Roland–Morris disability questionnaire* is considered thoroughly validated, have acceptable reliability and is recommended and referred to as a tool of choice in the assessment of the severity of disability caused by back pain [21, 22, 27, 31]. *The short form health survey*, SF-36 [36], is the most commonly used generic measure of health-related quality of life. It provides an indicator across eight dimensions of health and well being: physical functioning; physical role limitations; pain; general health; vitality; social functioning; emotional role limitations; mental health. Higher scores reflect better health status. *The SF-36* has proved reliable and valid over several different conditions, among those LBP [c.f. 11, 19]. Therefore, *the SF-36* is also recommended as an outcome measure for LBP research [8].

The back-specific criterion questionnaires were translated into Swedish by the expert group, supervised and checked by a bilingual physician whose native language was English. The translated Swedish versions were back-translated to English by a second bilingual physician, and discrepancies between the back-translated version and the original English version were discussed with the translators before the final wording was approved.

Statistical analyses

The indices of the questionnaires were normalized (expressed as a percentage of the maximum score) for comparison purposes, and questionnaires with the calculation principle of *the lower the score- the better the health/function* (i.e., *the Aberdeen low back pain disability scale, the Roland–Morris disability questionnaire* and *the Waddell disability index*) were transformed to the principle of *the higher the score—the better the health/function* by subtracting the normalized indices from one. If a question was omitted by a respondent, then the total score of the scale was adjusted by removing the maximum score for that question from the denominator before calculating the percentage. Omitted questions in *SF-36* were handled according to the manual (SF-36$^{©}$ Health Survey: Manual and Interpretation Guide). Each questionnaire form having omitted items with a maximum sum score exceeding 50% of the total maximum score of the scale, or that had more than half of the items omitted, was considered non-valid.

All analyses were performed using SPSS 13.0 for Windows (SPSS, Chicago, IL, USA). Data are presented in the text as the mean and standard deviation. *The internal consistency* of the PFM was assessed by Cronbach's alpha [7], reflecting the strength of the relationship between items, and by item–total correlations, relating scores of the individual questions to the total scores. For comparison, Cronbach's alpha was also calculated for each of the five criterion questionnaires, and was considered to be excellent above 0.8, adequate between 0.7 and 0.8 and low or inadequate below 0.7 [24]. Questions with item–total correlation below 0.2 were considered non-representative for the index [32]. An estimate of *the concurrent criterion validity* of the PFM was determined by calculating the correlations between the PFM index scores and the scores of the criterion questionnaires. *The concurrent criterion validity* was considered to be good if $r \geq 0.60$, moderate if r was between 0.30 and 0.59, and weak if $r < 0.30$ [2, 13]. Complementary to the correlation analyses, which only assess the associations, the limits of agreement (mean $\pm$ 2 SD of the normalized difference scores of the PFM indices and the criterion questionnaires) were determined [1]. *The construct validity* of the PFM was evaluated by comparing its ability to discriminate between patients with different pain severities (as measured by the consumption of painkillers on the worst day during the last 2 weeks) to that of the criterion questionnaires, in one-way analyses of variance (*discriminant validity*). Patients were grouped by painkiller consumption into four categories: 0 pills, 1–3 pills, 4–8 pills, and >8 pills. Post hoc tests between categories were performed with Tukey's HSD. Further, the information content of each PFM item was estimated by calculating the maximum response frequency. Items with maximum response frequency higher than 80% were considered to have limited information content [32]. *Face* and *content validity* of the PFM, i.e., item's relevance and adequacy for the intended use, could partly be judged by the developmental process, but these aspects were

also assessed by investigating the completeness of item responses, the distribution of scores, and the magnitude of floor and ceiling effects estimated by counting the number of minimum and maximum possible scores in the PFM. Finally, the content of the PFM scales were classified according to the ICF [37]. In all statistical tests, $P < 0.05$ was considered to be significant.

Results

A total of 193 patients consisting of 81 men and 112 women participated in the study. The mean age was 43.4 ± 10.5 and 45.4 ± 10.8 years for the men and women, respectively. Socio-demographic data for these patients are shown in Table 1. The number of questionnaires was reduced for *the SF-36* ($n = 161$) and *the Aberdeen low back pain disability scale* ($n = 147$) due to that these criterion questionnaires were incorporated first during 1995. Percentage non-valid forms for the PFM scales were 6% for the intensity index, 4% for the frequency index and 2% for the function index. Non-valid forms for the criterion questionnaires were as follows: *the Aberdeen low back pain disability scale* 14%, *the low-back outcome score* 6%, *the Waddell disability index* 2% and for the scales of *the SF-36* physical functioning 10%, physical role limitations 14%, pain 11%, general health 13%, vitality 11%, social functioning 11%, emotional role limitations 16% and mental health 11%.

Internal consistency

The calculations of Cronbach's alpha revealed excellent internal consistency ($\alpha > 0.80$) among the PFM items (The symptom scale; frequency index: $\alpha = 0.90$ and intensity index: $\alpha = 0.91$. The functional limitation scale; function index: $\alpha = 0.95$). For the criterion questionnaires, excellent internal consistency was seen for *the Roland–Morris disability questionnaire* ($\alpha = 0.86$) and for five of the eight dimensions of *the SF-36*. Adequate consistency ($\alpha = 0.70$–0.80) was seen for two dimensions of *the SF-36* (physical role limitations and general health), for *the Aberdeen low back pain disability scale* ($\alpha = 0.78$) and for *the low-back outcome score* ($\alpha = 0.80$). Low internal consistency ($\alpha < 0.70$) was seen for the dimension of vitality of *the SF-36*, for and for *the Waddell disability index* ($\alpha = 0.69$).

Spearman's correlations between PFM items and the total scores, all three PFM scales pooled, are shown in Fig. 1. All items had item–total correlations above 0.2. The average item–total correlations for the PFM scales were in general high (the symptom scale; intensity index: 0.49 ± 0.10 and frequency index: 0.50 ± 0.09. The functional limitation scale; function index: 0.63 ± 0.09). The

functional limitation scale had higher item–total correlation than the symptom scale (2-sample t test: $P < 0.001$).

Concurrent criterion validity

The correlation coefficients between the PFM index scores and the scores of the criterion questionnaires are shown in

Table 1 Selected socio-demographic characteristics of the study sample

Variables	n (%)	Missing
Gender		
Woman	112 (58.0)	
Man	81 (42.0)	
Marital status		3
Single	41 (21.6)	
Cohabit	146 (76.8)	
Widow/widower	3 (1.6)	
Children		3
Yes	162 (85.3)	
No	28 (14.7)	
Socio-economic classification		6
White-collar worker	48 (25.7)	
Blue-collar worker	113 (60.4)	
Self-employed	8 (4.3)	
Housewife	4 (2.1)	
Early retirement/long-term unemployed	14 (7.5)	
Cigarette smoking		29
No	112 (68.3)	
Yes, 1–14 cigarettes per day	36 (22.0)	
Yes, ≥ 15 cigarettes per day	16 (9.8)	
Alcohol consumption		19
No	59 (33.9)	
Sometimes	95 (54.6)	
Every week	19 (10.9)	
Every day	1 (0.6)	
Feeling of unhappiness/discomfort		33
Yes	30 (18.8)	
At work	9	
At home	14	
In leisure time	10	
No	130 (81.2)	
Medication		
No	34 (17.6)	
Yes		
Anti-inflammatories	54 (28.0)	
Muscle relaxants	52 (26.9)	
Non-opioid analgesics	67 (34.7)	
Anti-depressives	7 (3.6)	

The percentage shown within the parenthesis is based on those who answered the question
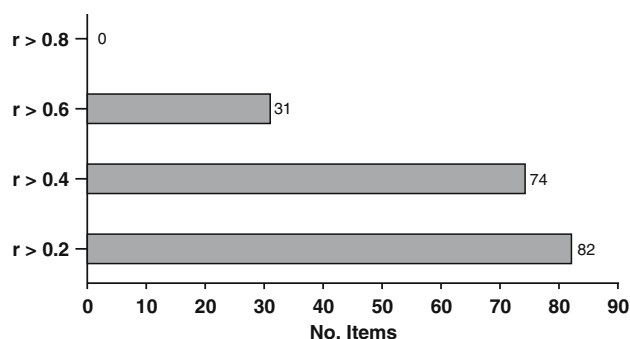
Fig. 1 The number of item–total correlations (Spearman's rank correlation), reaching different levels of correlation, for the PFM scales pooled (27 + 27 + 28 = 82 items)

Table 2. The correlation coefficients between the total score of the PFM and the back-specific criterion questionnaires all exceeded 0.60, indicating good concurrent criterion validity, whereas the correlations between the PFM total score and the eight dimensions of *the SF-36* ranged between 0.25 and 0.59. For the back-specific questionnaires, the highest correlation was obtained for *the Aberdeen low back pain disability scale* ($r = 0.83$ for the PFM total score), while the highest correlation between the PFM and *the SF-36* was seen for the dimension of physical functioning ($r = 0.61$ for the PFM function index).

The limits of agreement between the PFM scales and all criterion questionnaires can be seen in Fig. 2. The figure indicates that also the agreement was highest between the PFM and *the Aberdeen low back pain disability scale*.

Construct validity

One-way analyses of variance were used to compare the ability of the questionnaires to discriminate between patients with different pain severities, as measured by the consumption of painkillers on the worst day during the last 2 weeks. Patients were categorized in the following groups; *none* (23.8%), *less than four tablets* (32.5%), *between four and eight tablets* (31%), *more than eight tablets* (12.7%). The analyses revealed that the PFM, as well as the included back-specific criterion questionnaires, had the ability to discriminate patients based on painkiller consumption. Also seven out of the eight dimensions of *the SF-36* had significant discriminative ability. Post hoc tests (Tukey's HSD) between the four categories of the consumption of painkillers (six comparisons) showed that the PFM function index had the highest number of significant comparisons (four of six). See Table 3.

The maximum response frequency, i.e., the response frequency for the response alternative that was most frequently chosen, for the items in the PFM is shown in Fig. 3. With values ranging from 21.8 to 65.8%, none of the items exceeded the limit 80%, and therefore could be considered sensitive enough to discriminate between different levels of back symptoms and functional limitations [32].

Face and content validity

The developmental procedure of the PFM assured clinical face validity in so far that the questions were derived

Table 2 Bivariate correlations (Pearson's correlation coefficient) between the PFM index scores and the scores of the criterion questionnaires

	PFM symptom scale		PFM function index	PFM total
	Intensity index	Frequency index		
Aberdeen low back pain disability scale	0.70***	0.74***	0.80***	0.83***
Low-back outcome score	0.54 [S]***	0.52 [S]***	0.75 [S]***	0.67 [S]***
Waddell disability index	0.62 [S]***	0.62 [S]***	0.68 [S]***	0.71 [S]***
Roland–Morris disability questionnaire	0.55***	0.51***	0.65***	0.61***
SF-36 (physical functioning)	0.48***	0.46***	0.61***	0.59***
SF-36 (physical role limitations)	0.16 [S]	0.14 [S]	0.30 [S]***	0.25 [S]**
SF-36 (pain)	0.40 [S]***	0.34 [S]***	0.60 [S]***	0.49 [S]***
SF-36 (general health)	0.30***	0.24**	0.42***	0.37***
SF-36 (vitality)	0.48***	0.39***	0.39***	0.47***
SF-36 (social functioning)	0.39 [S]***	0.29 [S]**	0.38 [S]***	0.39 [S]***
SF-36 (emotional role limitations)	0.28 [S]**	0.23 [S]**	0.24 [S]**	0.27 [S]**
SF-36 (mental health)	0.51***	0.38***	0.43***	0.49***

Spearman's rank correlation is denoted by S

* Significant at 5% level, ** significant at 1% level, *** significant at 0.1% level

Fig. 2 Limits of agreement (mean ± 2 SD) of the normalized difference scores of the PFM indices and the criterion questionnaires. A narrow and well centered line indicate high agreement
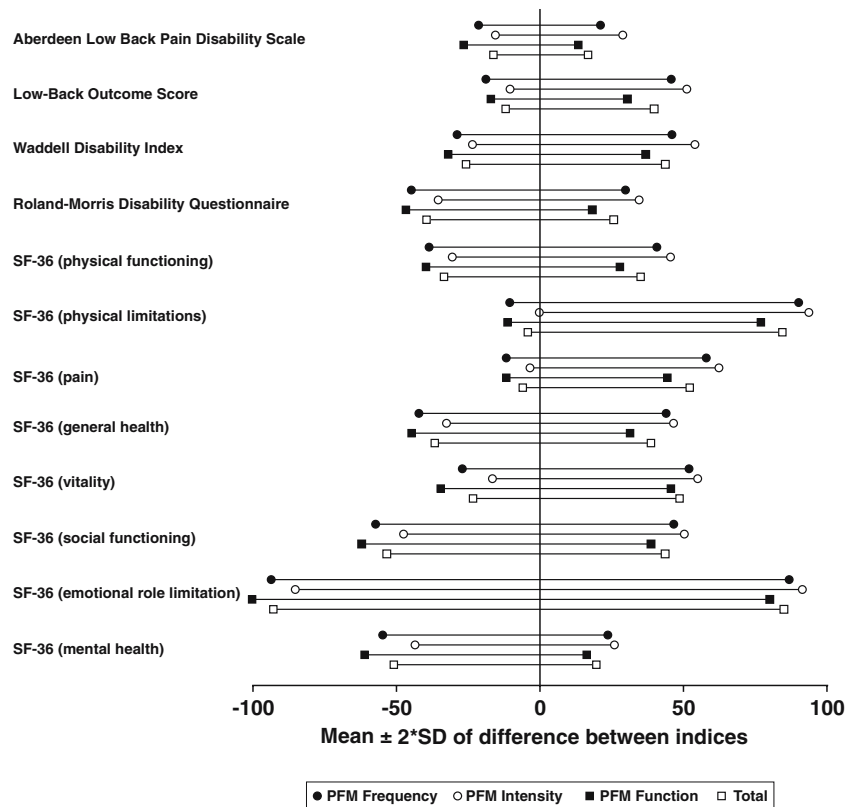
from a chronic LBP-patient group in co-operation with an expert group that had extensive clinical experience of LBP patients (see ''Methods'' section). Further analyses within these concepts of validity included completeness of item responses, which are shown in Fig. 4. The figure shows that item 27 of the symptom scale had by far the highest rate of omissions followed by item 26 (43 and 13%, respectively, for the intensity index, and 34 and 11%, respectively, for the frequency index). Missing data for the rest of the items, i.e., item 1–25, was evenly distributed (mean % omission for item 1–25 was 7 and 5% for the intensity and the frequency index, respectively). Missing data for the functional limitation scale of the PFM was consistently lower than those for the symptom scale, with a peak omission rate of item 28 of 9%.

The distribution of scores and the magnitude of floor and ceiling effects give an indication of the adequacy of the questionnaire for the tested sample. Figure 5 shows the score distribution of the PFM. The scores of the PFM indices were normally distributed and well centered with a mean for the PFM total normalized score of 49.98 ± 14.25, and no person scored lowest or highest possible score in any of the indices. For the back-specific criterion questionnaires, the scores of *the Waddell disability index* and *the low-back outcome score* showed non-normal distributions.

The content of the PFM scales, classified according to the domains of disability of the ICF, is shown in Table 4. All items of the symptom scale except item 27 could be classified to the domain of ''impairments—limitations at body level''. Item 27, a question about the need for support, could not be classified according to ICF, thus the designation of ''other''. The majority of the items of the functional disability scale were classified to the domain of ''activity limitations—limitations at personal level''. Items 25 and 28, questions about the capability to manage and return to work, were classified to the domain of ''participation restriction—limitation at societal level'', and the more general questions of well being—items 26 and 27—could not be classified (''others'').

Discussion

The results of the present study show that the new questionnaire, PFM, was comparable or better than the criterion questionnaires for the reliability and validity aspects explored. Cronbach's alpha of the PFM scales was higher than for any of the back-specific criterion questionnaires and all items of the PFM had item–total correlations clearly above 0.20 [32]. The internal consistency for the criterion questionnaires was in line with previous investigations [21, 27].

For the different aspects of validity examined, promising psychometric properties were disclosed for the PFM. Well-centered score distribution with no patient's score at the floor or the ceiling level indicates that the PFM has the potential to detect the improvement or worsening of symptoms and functional limitations in chronic LBP patients [3]. Also, the response alternatives chosen by the respondents were well dispersed and no item was near the limit of 80% of maximum response frequency (Fig. 3), set by Streiner and Norman [32] to specify the limit for when a question no longer adds to the information content. The correlation between the back-specific criterion questionnaires and the PFM total score and the scores of the function index, respectively, were good, although the correlation coefficients of the two symptom indices of the PFM were somewhat lower (Table 2). This seems logical since the contents of the back-specific criterion questionnaires are dominated by the ICF "activity limitation" domain, and only represent the "impairment" domain, withholding pain and symptoms, to a lesser extent [13]. In contrast, the correlations to the different domains of the SF-36 appear more incoherent with regard to the clearly lower correlations seen between the pain dimension of the SF-36 and the PFM symptom indices, compared with the PFM function index (Table 2). The back-specific questionnaire showing the highest correlation with the PFM was the Aberdeen low back pain disability scale (Table 2), which has a more equal mix of the ICF domains of "impairment" and "activity limitation" compared with the other criterion questionnaires [13].

On the whole, all questionnaires had the ability to discriminate between patients with different pain severities, but a perfect discrimination (i.e., the ability to differentiate between all four categories of patients tested by six post hoc comparisons) was not achieved by any of the questionnaires (Table 3). The best discriminative ability (four of six possible) was demonstrated by the functional limitation scale of the PFM, whereas the scales conceptually closer to symptom and pain measures (i.e., the symptom scale of the PFM and the pain domain of the SF-36) had

Table 3 Discrimination between patients ($n = 126$) with different pain severities, tested with one-way analyses of variance (Kruskal–Wallis one-way analysis of variance is denoted by "K")

	General ability to discriminate	Significant post hoc tests
PFM symptom scale		
Intensity index	$P = 0.008$	2
Frequency index	$P < 0.001$	3
PFM function index	$P < 0.001$	4
PFM total	$P < 0.001$	3
Low-back outcome score	$P < 0.001^K$	3
Waddell disability index	$P = 0.005^K$	2
Roland–Morris disability questionnaire	$P < 0.001$	3
SF-36 (physical functioning)	$P = 0.015$	2
SF-36 (physical role limitations)	$P = 0.033^K$	2
SF-36 (pain)	$P = 0.006^K$	2
SF-36 (general health)	$P = 0.025$	2
SF-36 (vitality)	$P = 0.109$	0
SF-36 (social functioning)	$P = 0.024^K$	1
SF-36 (emotional role limitations)	$P = 0.013^K$	1
SF-36 (mental health)	$P < 0.001$	3

Pain severity is measured by the consumption of painkillers on the worst day during the last 2 weeks, divided into four categories: *0 pills*, *1–3 pills*, *4–8 pills*, and *>8 pills*. Post hoc tests between categories were performed with Tukey's HSD (six comparisons between the four categories). The information of painkiller consumption was retrieved from the Aberdeen low back pain disability scale, which is the reason for the removal of this scale from the table
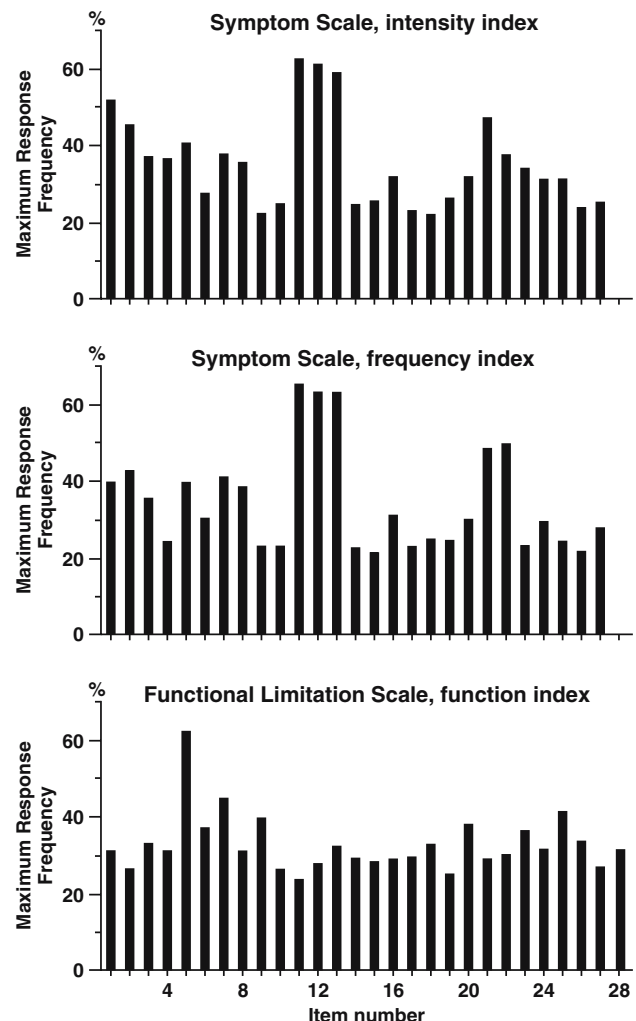
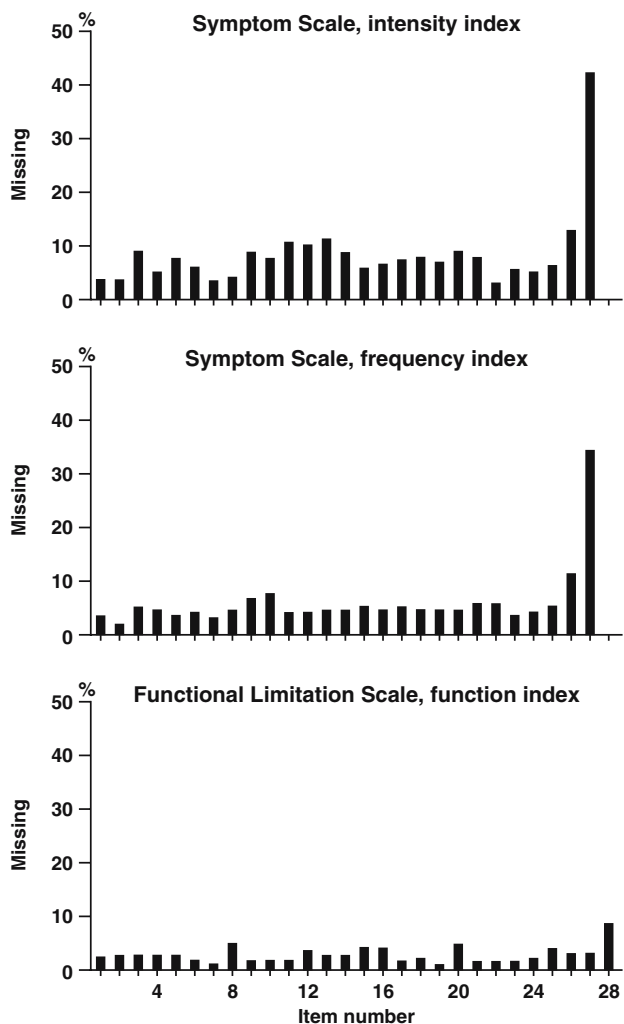**Fig. 3** Maximum response frequency for the PFM questionnaire

Fig. 4 Completeness of item responses for the PFM questionnaire

two to three significant post hoc comparisons. This may seem contradictory to the expected, but illustrates the complexity of pain severity and behavior [26, 30, 34]. One might speculate that limitations in daily activities, i.e., when people notice that they no longer manage their normal activities because of the LBP, are more important determinants for increasing analgesic consumption than mere changed symptoms. An interesting parallel of differing predictive ability between function and pain was shown by Lackner and Carosella [18]. In a population of 100 chronic back pain patients, the self-rated function predicted actual lifting performance significantly better than the pain ratings [18].

The PFM uses different weighting on items, primarily based on the judgments from the chronic LBP-patient group (see ''Methods'' section and ''Appendix''). Weighting based on the importance that patients place on various activities is suggested to improve the today's standard of questionnaires, where often no rationale is presented for the different weights [22].

An item with a high rate of omission probably suffers from ambiguity, incomprehensibility or may not be suitable for use in the general population of the specific group of interest [3, 17]. Item 27 of the symptom scale of the PFM was distinguished by a clearly higher omission rate than the rest of items (Fig. 4), in spite of the efforts during the developmental process of the PFM to avoid ambiguity (see ''Methods'' section). We suggest this item to be removed from the scale. This would not only decrease ambiguity of the scale, but also result in a homogeneous item content of the symptom scale with regard to the ICF disability domain of ''impairments—limitations at body level'' (Table 4).

Fig. 5 Distribution of normalized scores of the PFM questionnaire
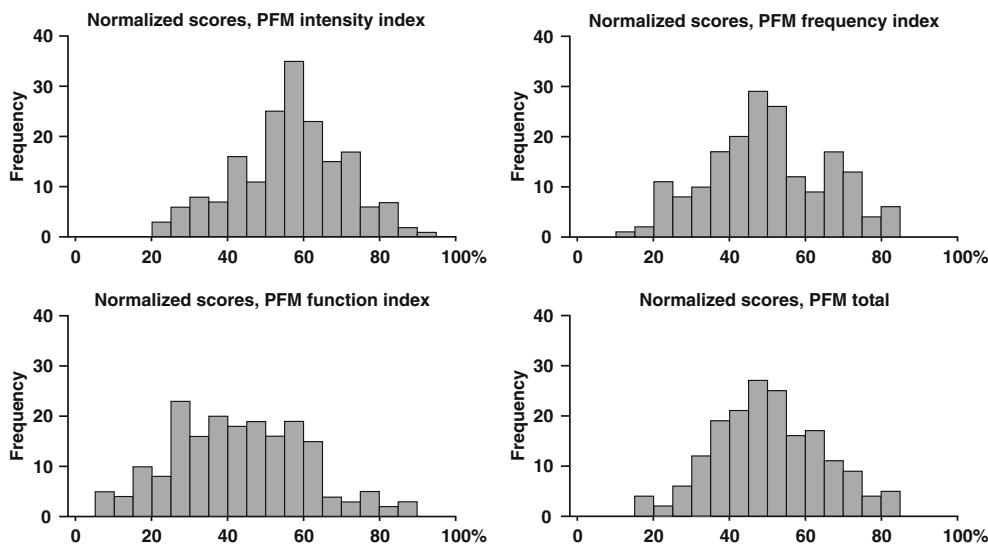
Table 4 The content of the profile fitness mapping scales classified according to domains of disability in the International Classification of Functioning, Disability and Health: ICF. Short version [37]

Disability			Others (not covered in the ICF)
Impairments (limitations at body level)	Activity limitations (limitations at the personal level)	Participation restriction (limitations at societal level)	
The symptom scale			
Ch 1: Mental functions, items 17–19, 21, 24–25			Item 27
Ch 2: Sensory functions and pain, items 2–3,5–7, 14, 22–23			
Ch 5: Functions of the digestive, metabolic and endocrine systems, items 12–13			
Ch 6: Genitourinary and reproductive functions, items 11, 26			
Ch 7: Neuromusculoskeletal and movement-related functions, items 1, 4, 8–10, 15–16, 20			
The functional limitation scale			
	Ch 4: Mobility, items 1–8, 11–24	Ch 8: Major life areas, items 25, 28	Items 26–27
	Ch 5: Self-care, items 9–10		

Ch in the table refers to chapters in the ICF book

Scales with defined main concepts and content will have an advantage in future LBP research in the interpretation and comparison of results. Back-specific questionnaires with items belonging to all three ICF perspectives of health like *the low-back outcome score* may, on one hand, be preferred because the question flow can be harmonized and similar questions can be united in one [23]. On the other hand, mixing items that focus on different domains in one sum score increases the risk of item masking score bias [23], might cause problems in interpretation and comparison [3] and reduces the possibility to link treatment interventions to the intended areas of outcome [13]. Therefore, sub-scores focusing on separate dimensions is recommended [23]. This is achieved in the PFM questionnaire, where the content of the symptom scale correspond to ''impairments'' and the functional limitation scale assesses ''activity limitations'' and to a minor extent ''participation restriction''. A possibility of development for the PFM functional limitation scale would be to further refine the scale to the exclusive domain of ''activity limitation'', which would require the elimination of the general questions of well being (items 26–27) as well as the work-related questions (items 25 and 28).

In 2004, Cieza and co-workers presented an ICF core set for LBP in order to define the typical spectrum of problems in functioning of patients with LBP [6]. The ICF classification of the PFM items (Table 4) is in good correspondence with the ICF core set for LBP, which further supports the construct and content validity of the PFM [6].

The mean score on *the Roland–Morris disability questionnaire* of our sample of in-care LBP patients was rather low, 10.35 ± 5.06, compared with a mean score of 11.4 in 230 LBP patients in family practice [29]. In the same study by Roland and Morris [29], a score of 9.4 (8.3–10.5, 95% confidence limits) corresponded to ''moderate pain''. Mean scores of *the Roland–Morris disability questionnaire* in other studies on LBP patients show similar or higher values as those acquired in the present study [for references, see 28]. Thus, as spread of pain, in contrast to localized pain, is usually a sign of increasing disability [14], one possible reason for the relatively low degree of disability in our sample could be that only those who had exclusively pain in the low back were included and patients with any kind of radiating pain or pain in other body parts were rejected. Therefore, the sample used in the present study might be considered representative also for primary care patients who normally have lower severity grade than in-care chronic LBP patients.

Important limitation of the study includes the lack of retest data of the sample, which makes the evaluation of reliability incomplete. Also, the assessment of validity of the PFM needs to be complemented by determination of responsiveness to change.

Conclusion

The present study suggests that the PFM has a high internal consistency and is a valid indicator of symptoms and

functional limitations of LBP patients. The PFM seems well adapted to the sample of chronic LBP patients studied in that it had a balanced score distribution with no floor or ceiling effect in any of the indices. Moreover, the PFM offers the combination of a composite total score and the possibility of evaluations within specific domains of disability. Further, the findings of this study justify additional evaluation of the PFM in which test–retest reliability and responsiveness to change with the determination of minimal clinically important difference appears most warranted.

Appendix: The profile fitness mapping questionnaire

The symptom scale

The Symptom Scale

Do you feel/experience:	How often? Answer 1-6	How much? Answer 7-12
1. Stiffness in your back?	☐	☐
2. Soreness in your back?	☐	☐
3. Swelling in your back?	☐	☐
4. Tension in your back?	☐	☐
5. Cracking sound in your back?	☐	☐
6. Numbness in your leg?	☐	☐
7. Tiredness in your back?	☐	☐
8. Weakness in your back?	☐	☐
9. Crick in your back?	☐	☐
10. Sudden loss of control of your back?	☐	☐
11. Problems with urination?	☐	☐
12. Problems emptying the bowels?	☐	☐
13. Problems with your stomach?	☐	☐
14. Having a crooked back?	☐	☐
15. Limping during walking?	☐	☐
16. Disturbance of balance?	☐	☐
17. Irritability, short tempered?	☐	☐
18. Stressed?	☐	☐
19. Depressed?	☐	☐
20. Fumblingness in your legs/feet?	☐	☐
21. Anxiety?	☐	☐
22. Backache during activity?	☐	☐
23. Backache during resting?	☐	☐
Does your back problems affect:		
24. Your sleep?	☐	☐
25. Your mood?	☐	☐
26. Your sexlife?	☐	☐
Do you use, because of back problems:		
27. Support, e.g. corset? Cane? Underline which!	☐	☐

Note! Fill in **both** columns!

How often? Answer with one of the figures 1-6
1. Never/very seldom
2. Seldom
3. Rather seldom
4. Rather often
5. Often
6. Very often/always

How much? Answer with one of the figures 7-12
7. Nothing/none at all
8. Weak/little
9. Rather weak/rather little
10. Rather strong/rather much
11. Strong/much
12. Almost unbearable/unbearable, all/maximally

The functional limitation scale

The Functional Limitation Scale

Because of your **low back pain** problems, how do you manage to:	How? Answer 1-6
1. Stand?	☐
2. Walk?	☐
3. Sit?	☐
4. Lay down?	☐
5. Run?	☐
6. Carry?	☐
7. Lift?	☐
8. Throw?	☐
9. Put on and take off a sweater?	☐
10. Put on and take off socks?	☐
11. Bend your back forward?	☐
12. Bend your back backward?	☐
13. Sidebend your back to the right?	☐
14. Sidebend your back to the left?	☐
15. Turn your back to the right?	☐
16. Turn your back to the left?	☐
17. Walk upstairs?	☐
18. Walk downstairs?	☐
19. Squat down?	☐
20. Jump with both feet together?	☐
21. Lift your right leg, when lying down?	☐
22. Lift your left leg, when lying down?	☐
23. Lift your right leg, when sitting?	☐
24. Lift your left leg, when sitting?	☐
25. Do your work?	☐
What do you say about:	
26. The condition of your back?	☐
27. Your general health?	☐
28. Return to work?	☐

Answer with one of the figures 1-6
1. Very good, no problem, very satisfying, very likely.
2. Good, easily, satisfying, likely.
3. Rather good, rather easy, rather satisfying, rather likely.
4. Rather bad, rather difficult. rather dissatisfying, rather unlikely.
5. Bad, difficult, dissatisfying, unlikely.
6. Very bad, very difficult/impossible very dissatisfying very unlikely.

Method of score calculation for the Profile Fitness Mapping scales

The table shows the weighting and maximum score of each item in the Profile Fitness Mapping scales, and the calculation of scores for each index.

Frequency (f) is the answer on *how often* the symptom is felt (six-point scale from 1 = never/very seldom, to 6 = very often/always). Intensity (i) is the answer on *how much* the symptom is felt (six-point scale from 7 = nothing/none at all, to 12 = almost unbearable/unbearable, all/maximally). The answers of the functional limitation scale (fl) range from 1 = very good, no problem, very satisfying, very likely, to 6 = very bad, very difficult/impossible, very dissatisfying, very unlikely.

The result of each index is expressed as the percentage of the maximum score, where 100% is the best possible result. Adjustments due to omitted questions are done by removing the maximum score for those questions from the denominator before calculating the percentage.

The symptom scale (s)					The functional limitation scale (fl)			
$Item_s$	Weight (W_s)	Score frequency index	Score intensity index	Max score	$Item_{fl}$	Weight (W_{fl})	Score function index	Max score
1	2.4	$(6-f_1)*W_s$	$(12-i_1)*W_s$	12	1	3	$(6-fl_1)*W_{fl}$	15
2	2	$(6-f_2)*W_s$	$(12-i_2)*W_s$	10	2	3	$(6-fl_2)*W_{fl}$	15
3	1.6	$(6-f_3)*W_s$	$(12-i_3)*W_s$	8	3	4	$(6-fl_3)*W_{fl}$	20
4	2	$(6-f_4)*W_s$	$(12-i_4)*W_s$	10	4	3	$(6-fl_4)*W_{fl}$	15
5	1.2	$(6-f_5)*W_s$	$(12-i_5)*W_s$	6	5	2	$(6-fl_5)*W_{fl}$	10
6	2.4	$(6-f_6)*W_s$	$(12-i_6)*W_s$	12	6	1.6	$(6-fl_6)*W_{fl}$	8
7	3	$(6-f_7)*W_s$	$(12-i_7)*W_s$	15	7	1.6	$(6-fl_7)*W_{fl}$	8
8	3	$(6-f_8)*W_s$	$(12-i_8)*W_s$	15	8	1.2	$(6-fl_8)*W_{fl}$	6
9	4	$(6-f_9)*W_s$	$(12-i_9)*W_s$	20	9	2	$(6-fl_9)*W_{fl}$	10
10	3	$(6-f_{10})*W_s$	$(12-i_{10})*W_s$	15	10	2.4	$(6-fl_{10})*W_{fl}$	12
11	5	$(6-f_{11})*W_s$	$(12-i_{11})*W_s$	25	11	3	$(6-fl_{11})*W_{fl}$	15
12	5	$(6-f_{12})*W_s$	$(12-i_{12})*W_s$	25	12	3	$(6-fl_{12})*W_{fl}$	15
13	2	$(6-f_{13})*W_s$	$(12-i_{13})*W_s$	10	13	3	$(6-fl_{13})*W_{fl}$	15
14	3	$(6-f_{14})*W_s$	$(12-i_{14})*W_s$	15	14	3	$(6-fl_{14})*W_{fl}$	15
15	4	$(6-f_{15})*W_s$	$(12-i_{15})*W_s$	20	15	3	$(6-fl_{15})*W_{fl}$	15
16	3	$(6-f_{16})*W_s$	$(12-i_{16})*W_s$	15	16	3	$(6-fl_{16})*W_{fl}$	15
17	2	$(6-f_{17})*W_s$	$(12-i_{17})*W_s$	10	17	2	$(6-fl_{17})*W_{fl}$	10
18	2	$(6-f_{18})*W_s$	$(12-i_{18})*W_s$	10	18	2.4	$(6-fl_{18})*W_{fl}$	12
19	2	$(6-f_{19})*W_s$	$(12-i_{19})*W_s$	10	19	2	$(6-fl_{19})*W_{fl}$	10
20	2.4	$(6-f_{20})*W_s$	$(12-i_{20})*W_s$	12	20	1.6	$(6-fl_{20})*W_{fl}$	8
21	2	$(6-f_{21})*W_s$	$(12-i_{21})*W_s$	10	21	2	$(6-fl_{21})*W_{fl}$	10
22	8	$(6-f_{22})*W_s$	$(12-i_{22})*W_s$	40	22	2	$(6-fl_{22})*W_{fl}$	10
23	8	$(6-f_{23})*W_s$	$(12-i_{23})*W_s$	40	23	2	$(6-fl_{23})*W_{fl}$	10
24	3.6	$(6-f_{24})*W_s$	$(12-i_{24})*W_s$	18	24	2	$(6-fl_{24})*W_{fl}$	10
25	3.6	$(6-f_{25})*W_s$	$(12-i_{25})*W_s$	18	25	7	$(6-fl_{25})*W_{fl}$	35
26	4	$(6-f_{26})*W_s$	$(12-i_{26})*W_s$	20	26	4	$(6-fl_{26})*W_{fl}$	20
27	2.4	$(6-f_{27})*W_s$	$(12-i_{27})*W_s$	12	27	4	$(6-fl_{27})*W_{fl}$	20
					28	7	$(6-fl_{28})*W_{fl}$	35

References

1. Altman DG (1991) Practical statistics for medical research. Chapman and Hall, London
2. Andresen EM (2000) Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil 81:S15–S20
3. Atroshi I, Gummesson C, Andersson B, Dahlgren E, Johansson A (2000) The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: reliability and validity of the Swedish version evaluated in 176 patients. Acta Orthop Scand 71:613–618
4. Bergner M, Bobbitt RA, Kressel S, Pollard WE, Gilson BS, Morris JR (1976) The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. Int J Health Serv 6:393–415
5. Brown FG (1983) Principles of educational and psychological testing, 3rd edn. Holt, Rinehart and Winston, New York
6. Cieza A, Stucki G, Weigl M, Disler P, Jackel W, van der Linden S, Kostanjsek N, de Bie R (2004) ICF core sets for low back pain. J Rehabil Med July:69–74
7. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16:297–334
8. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. Spine 23:2003–2013
9. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, Kerns RD, Stucki G, Allen RR, Bellamy N, Carr DB, Chandler J, Cowan P, Dionne R, Galer BS, Hertz S, Jadad AR, Kramer LD, Manning DC, Martin S, McCormick CG, McDermott MP, McGrath P, Quessy S, Rappaport BA, Robbins W, Robinson JP, Rothman M, Royal MA, Simon L, Stauffer JW, Stein W, Tollett J, Wernicke J, Witter J (2005) Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 113:9–19
10. Ekman M, Jonhagen S, Hunsche E, Jonsson L (2005) Burden of illness of chronic low back pain in Sweden: a cross-sectional, retrospective study in primary care setting. Spine 30:1777–1785
11. Garratt AM, Klaber Moffett J, Farrin AJ (2001) Responsiveness of generic and specific measures of health outcome in low back pain. Spine 26:71–77
12. Greenough CG, Fraser RD (1992) Assessment of outcome in patients with low-back pain. Spine 17:36–41

13. Grotle M, Brox JI, Vollestad NK (2005) Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. Spine 30:130–140

14. Jacob T, Zeev A (2006) Are localized low back pain and generalized back pain similar entities? Results of a longitudinal community based study. Disabil Rehabil 28:369–377

15. Jensen MP, Karoly P (2001) Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R (eds) Handbook of pain assessment. Guilford Press, New York, pp 15–34

16. Kjellberg A (1989) Att ställa frågor om arbetsmiljön: en kort handledning i konstruktion av frågeformulär [Swedish]. Arbetsmiljöinstitutet, Förlagstjänst, Solna

17. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI (1996) The Quebec back pain disability scale: conceptualization and development. J Clin Epidemiol 49:151–161

18. Lackner JM, Carosella AM (1999) The relative influence of perceived pain control, anxiety, and functional self efficacy on spinal function among patients with chronic low back pain. Spine 24:2254–2260

19. Lansky D, Butler JB, Waller FT (1992) Using health status measures in the hospital setting: from acute care to 'outcomes management'. Med Care 30:MS57–MS73

20. Margolis RB, Chibnall JT, Tait RC (1988) Test–retest reliability of the pain drawing instrument. Pain 33:49–51

21. Müller U, Duetz MS, Roeder C, Greenough CG (2004a) Condition-specific outcome measures for low back pain. Part I: validation. Eur Spine J 13:301–313

22. Müller U, Roeder C, Dubs L, Duetz MS, Greenough CG (2004b) Condition-specific outcome measures for low back pain. Part II: scale construction. Eur Spine J 13:314–324

23. Müller U, Röder C, Greenough CG (2006) Back related outcome assessment instruments. Eur Spine J 15(Suppl 1):S25–S31

24. Nunally JC (1967) Psychometric theory. McGraw-Hill, New York

25. Ohnmeiss DD (2000) Repeatability of pain drawings in a low back pain population. Spine 25:980–988

26. Rainville J, Ahern DK, Phalen L, Childs LA, Sutherland R (1992) The association of pain with physical activities in chronic low back pain. Spine 17:1060–1064

27. Rocchi MB, Sisti D, Benedetti P, Valentini M, Bellagamba S, Federici A (2005) Critical comparison of nine different self-administered questionnaires for the evaluation of disability caused by low back pain. Eura Medicophys 41:275–281

28. Roland M, Fairbank J (2000) The Roland–Morris disability questionnaire and the Oswestry disability questionnaire. Spine 25:3115–3124

29. Roland M, Morris R (1983) A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine 8:141–144

30. Ruta DA, Garratt AM, Wardlaw D, Russell IT (1994) Developing a valid and reliable measure of health outcome for patients with low back pain. Spine 19:1887–1896

31. Stratford PW, Binkley JM, Riddle DL (2000) Development and initial validation of the back pain functional scale. Spine 25:2095–2102

32. Streiner GL, Norman DR (1990) Health measurement scales: a practical guide to their development and use. Oxford University Press, Oxford

33. Suarez-Almazor ME, Kendall C, Johnson JA, Skeith K, Vincent D (2000) Use of health status measures in patients with low back pain in clinical settings: comparison of specific, generic and preference-based instruments. Rheumatology (Oxf) 39:783–790

34. Turk DC (2002) Clinical effectiveness and cost-effectiveness of treatments for patients with chronic pain. Clin J Pain 18:355–365

35. Waddell G, Main CJ (1984) Assessment of severity in low-back disorders. Spine 9:204–208

36. Ware JE Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473–483

37. WHO (2001) International Classification of Functioning, Disability and Health: ICF. Short version. World Health Organization, Geneva