

REVIEW ARTICLES

# Measurement properties of patient-specific instruments measuring physical function

J.A. Barten<sup>a,b,\*</sup>, M.F. Pisters<sup>a,b</sup>, P.A. Huisman<sup>a,b</sup>, T. Takken<sup>b,c</sup>, C. Veenhof<sup>a</sup>

<sup>a</sup>Netherlands Institute for Health Services Research (NIVEL), P.O. Box 1568, 3500 BN Utrecht, The Netherlands

<sup>b</sup>Program of Physiotherapy Sciences, School of Clinical Health Sciences, University Medical Center Utrecht, The Netherlands

<sup>c</sup>Child Development and Exercise Center, University Medical Center Utrecht, The Netherlands

Accepted 10 December 2011; Published online 27 March 2012

## Abstract

**Objective:** To identify patient-specific self-assessment instruments, which measure physical function in patients with musculoskeletal disorders and to evaluate the descriptive properties and the psychometric qualities of these instruments.

**Study Design and Setting:** After a systematic search, included instruments were evaluated psychometrically by the checklist “quality criteria for measurement properties of health status instruments.”

**Results:** Twenty-three studies were included, referring to 12 instruments. Nine different versions of the Patient-Specific Functional Scale (PSFS) were identified. The practical elaboration of the different versions of the PSFS varied widely. None of the instruments were tested on all psychometric quality criteria of the checklist. The PSFS described by Cleland et al. was most extensively investigated and obtained exclusively positive scores. Overall, construct validity, reliability, and responsiveness were evaluated most frequently.

**Conclusion:** The descriptive properties and psychometric quality of patient-specific instruments measuring physical function are only partly investigated. The PSFS was the most investigated instrument: nine different versions have been evaluated psychometrically. The version of Cleland et al. was most extensively investigated, obtained exclusively positive scores following the quality criteria by Terwee et al., and could be recommended for clinical use therefore. Future research will be necessary to confirm the psychometric quality of patient-specific instruments measuring physical function in patients with musculoskeletal disorders. © 2012 Elsevier Inc. All rights reserved.

**Keywords:** Physical function; Musculoskeletal disorders; Patient-specific instruments; Systematic review; Measurement properties; Patient specific functional scale

## 1. Introduction

Musculoskeletal disorders are one of the major health care problems facing the Dutch population. Low back pain is the most prevalent disorder, with a point prevalence of 24.1% in the total population [1]. Frequently, patients with musculoskeletal disorders are faced with disability, which limits them in performing activities of daily living. Disability entails high economic, societal, and personal cost [2,3]. To diminish disability, nonpharmacological treatments (such as rehabilitation or physical therapy) are focused on both a patient's physical functioning and/or his context, including his psychological and social functioning [4,5]. Considering many contextual factors that determine disability are common across musculoskeletal disorders and even relevant to any chronic health condition, especially a patient's physical functioning makes the difference between the one and the other patient with musculoskeletal disorders [5]. To

assess a patients' level of physical functioning and to evaluate the effect of interventions in the clinical encounter, high-quality measurements are necessary [6].

Physical functioning can be assessed in different ways. Firstly, general measurement instruments, such as the MOS 36-Item Short Form Health Survey (SF-36) (physical function subscale) [7–9], are often applied. Secondly, disease-specific tools are used, for example, the Neck Disability Index [10]. Both general and disease-specific tools are instruments with content completely set, irrespective of a patient's health problem, request for help, or nature of complaints. Data produced by these so-called “fixed-item” tools are convenient and relatively simple to categorize and compare across patients and settings [11].

However, fixed-item tools are often difficult to interpret on an individual patient level. These tools do not consider patients' preferences and variability in performance on particular activities [12]. For example, the ability to perform gardening will be of low relevance to a patient who does not own a garden or such as the ability to climb stairs

\* Corresponding author. Tel.: +31-30-2729746; fax: +31-30-2729729.  
E-mail address: d.barten@nivel.nl (J.A. Barten).

**What's new?**

- Twelve instruments were developed to objectify patient-specific function in patients with musculoskeletal disorders.
- The descriptive properties and psychometric quality of these instruments are only investigated to a limited extent.
- The Patient-Specific Functional Scale (PSFS) is the most widely investigated questionnaire.
- The PSFS by Cleland et al. obtained the most positive scores and could be recommended for clinical use.

will not be relevant to a patient who always takes the elevator.

The interest in so-called patient-specific outcomes, which address each patient's priorities in outcome assessment, is increasing in clinical practice and research [12,13]. In contrast to fixed-item instruments, patient-specific instruments can identify relevant issues on an individual level and allow the evaluation to focus on what is important to each patient [12]. Similar to fixed-items instruments, patient-specific instruments have limitations. The question arises to what extent the outcomes are comparable between patients because of the individualized content. The application of statistical techniques is therefore questionable. In addition, floor effects may occur as patients will choose difficult tasks as "most important impaired activities" [11].

Nevertheless, because of the increase of a "patient-based concept" in health policy, individualized outcome tools will become more and more important. Several patient-specific self-assessment instruments (self-administered and interview based) have been developed in patients with musculoskeletal disorders [12]. However, a complete overview including a psychometric quality assessment of available patient-specific self-assessment instruments concerning physical function is lacking.

### 1.1. Objective

To identify available patient-specific self-assessment instruments, which measure physical function in patients with musculoskeletal disorders and to evaluate the descriptive properties and the psychometric qualities of these instruments.

## 2. Method

### 2.1. Search strategy

An extensive search strategy was conducted in the electronic databases PubMed (1966–December 2011), CINAHL

(1982–December 2011), and EMBASE (1988–December 2011). The search strategy was built on four elements: (1) outcome assessment, (2) patient-specific character of outcome assessment, (3) outcome dimension physical function, and (4) psychometric qualities. The search strategy was formulated in PubMed and adapted for use in other databases (see Appendix A on the journal's web site at [www.elsevier.com](http://www.elsevier.com)). Additionally, reference lists of all relevant articles were screened to include potential articles.

### 2.2. Selection criteria

The following inclusion criteria were used:

(1) The instrument is a questionnaire, a rating scale or an (semistructured) interview; (2) The instrument minimally measures the dimension "physical function"; (3) The instrument is applied to patients with musculoskeletal disorders; (4) The instrument has a patient-specific character; (5) Investigating the measurement properties of the instrument was the main aim of the study; (6) Publications describing different versions of the same test with different items were both included; (7) For practical reasons, only articles published in English, German, French, or Dutch were considered for inclusion; (8) Only full-length published articles were included.

An article was excluded if: (1) The instrument was a performance-based test; (2) The instrument was a different language version of an original instrument.

### 2.3. Study selection procedure

The study selection process was performed in two stages. The first selection on titles and subsequently on abstracts was independently performed by two reviewers (J.a.B., P.a.H.). The second step comprised reviewing full-text articles against the mentioned inclusion criteria. Disagreements concerning selection and inclusion of studies were resolved by discussion. A third reviewer (M.f.P.) was consulted in case of persisting disagreement.

### 2.4. Data extraction

Study characteristics (authors, title, and year of publication) and descriptive characteristics of the instrument were extracted from the selected articles, including name of the instrument, target population, size of the population in which the instrument was applied, purpose of the instrument, outcome domain(s), referral to a specific time period, mode of administration (self-administered or interview based), mode of selecting items, number of scales, number of items per scale, response options, range of scores, and time to administer.

### 2.5. Psychometric quality assessment

Psychometric quality was determined using the standardized checklist "quality criteria for measurement

properties of health status questionnaires” [14] (see Appendix B on the journal’s web site at [www.elsevier.com](http://www.elsevier.com)). This checklist, developed by Terwee et al., is based on the criteria of the “Scientific Advisory Committee (SAC) of the Medical Outcomes Trust [15].” In contrast to the SAC list, the list of Terwee et al. [14] mentions explicit criteria for what constitutes good measurement properties. All measurement properties were rated as “+” (positive), “?” (doubtful), “–” (negative), or “0” (no information available). The following measurement properties were evaluated.

#### 2.5.1. Content validity

Content validity examines the extent to which the domain of interest is comprehensively represented by the items in the instrument [16]. This term was operationalized by describing the measurement aim of the instrument, the target population, the measurement concept, the way of item selection, and the interpretability of the items [17,18].

#### 2.5.2. Internal consistency

Internal consistency is a measure of the extent to which items in an instrument (sub)scale are correlated, thus measuring the same construct. Factor analysis should have been applied to determine whether the items of a scale measure the same construct. In addition, Cronbach’s alpha should have been calculated as a measure of internal consistency. An alpha of  $\geq 0.70$  was considered acceptable [19].

#### 2.5.3. Criterion validity

Criterion validity refers to the extent to which scores on a particular instrument are related to a gold standard. Positive evaluations were given when the gold standard was convincingly described as a real gold standard and if the correlation with the gold standard was at least 0.70.

#### 2.5.4. Construct validity

Construct validity is a measure of the extent of which scores on a particular instrument relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured [17,18]. Construct validity was considered adequate if specific hypotheses were defined regarding the relationships with other measures of physical function and if  $\geq 75\%$  of these hypotheses were confirmed.

#### 2.5.5. Reproducibility

Reproducibility concerns the amount of which repeated measurements in stable persons provide similar results. Reproducibility can be divided into two aspects: reliability and absolute agreement. Reliability refers to the extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error). Intraclass correlation coefficient (ICC) (continuous data) or weighted Cohen’s Kappa (discrete data) were regarded as adequate measures [18]. A value of 0.70 was used as a minimum standard [19].

Agreement describes the extent to which the scores on repeated measures are similar to each other (absolute measurement error). Bland & Altman Limits of Agreement (LOA) and the smallest detectable change (SDC) were considered adequate measures of agreement [19–22]. A positive rating was assigned if the minimal important change (MIC) was outside the LOA or if the SDC was smaller than the MIC [14].

#### 2.5.6. Responsiveness

Responsiveness was defined as the ability of an instrument to detect clinically important changes over time in the concept to be measured [23–25]. Responsiveness could be determined in many different ways [14]. However, considering responsiveness as a measure of longitudinal validity, responsiveness is independent of the treatment effect. It was considered adequate if the SDC was smaller than the MIC, if the responsiveness ratio (RR) was  $\geq 1.96$  [26] or if the area under the “receiver operating characteristics” (ROC) curve was  $\geq 0.70$  [23].

#### 2.5.7. Floor and ceiling effects

Floor and ceiling effects were considered to be present if more than 15% of the respondents achieved the lowest or highest possible score, respectively [27]. A positive rating was assigned if floor and ceiling effects were absent.

#### 2.5.8. Interpretability

Interpretability is defined as the extent to which one can assign qualitative meaning to quantitative scores [28]. To assess interpretability, means and standard deviations (SDs) of relevant groups should have been presented. In addition, the MIC should have been defined. Interpretability was scored positively if mean scores and SD were presented of at least four subgroups of patients and if the MIC was defined.

Psychometric quality assessment was conducted by two reviewers independently (J.a.B., M.f.P.). When disagreement was found between the two reviewers, the measurement quality, which was subject of disagreement, was discussed. A third reviewer (C.V.) was consulted in case of persisting disagreement.

### 2.6. Overall quality

To obtain an overall score for psychometric quality of the identified instruments, the number of positive ratings out of the total rated items for each instrument was counted.

## 3. Results

### 3.1. Study selection procedure

The literature search identified 1,617 unique articles. After the selection procedure, 23 studies were included, referring to 12 different instruments (Fig. 1) [29–71]. The full

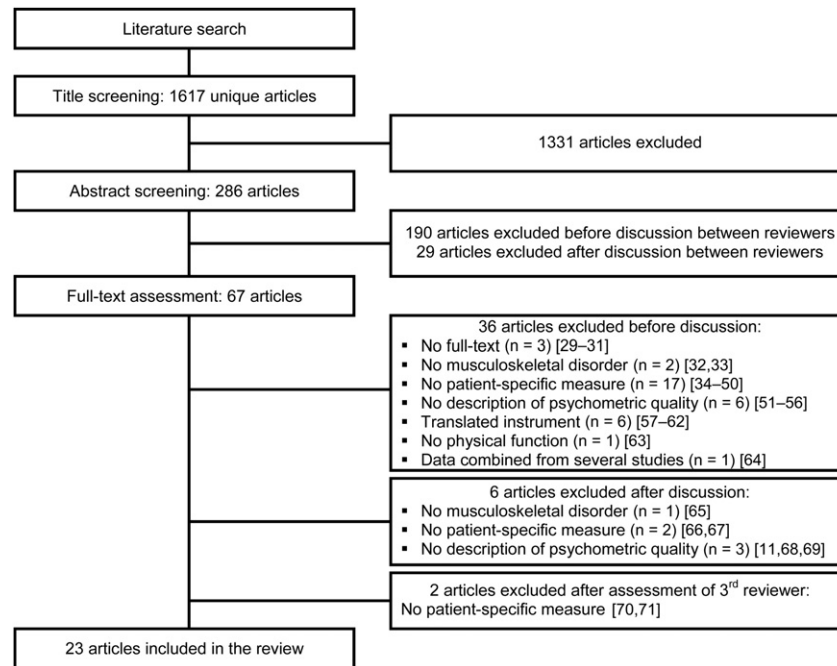


Fig. 1. Selection procedure.

names of the investigated instruments are presented in Table 1. The percentage of agreement between the two reviewers amounted to 88 after stage 1 and 75 after stage 2 of the selection procedure.

### 3.2. Description of the instruments

Three instruments, derived from 11 articles, were found which represented minor variations of the same construct measured, namely: patients rate their most important disabilities and evaluate these disabilities over time. The most common example to measure this construct is the Patient-Specific Functional Scale (PSFS). Nine different versions of the PSFS were included in this review [72–80]. These versions differed in target population, mode of administration (interview based vs. self-administrated), and scoring method. The Severity of the Main Complaints Scale (SMCS) [81] measures the same construct, as well as the McMaster Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR) does. However, the MACTAR additionally investigates patients' general, emotional, and social health status [82,83].

More extended versions of the PSFS rate, besides the most important impaired activities, the severity of these impairments and/or the frequency of the impaired activities in daily life. Hereby, two measurement tools offer patients the possibility to define their most important impaired activities [84,85]; four measurement tools permit patients to choose impaired activities out of a predefined list [12,86–88] and also four measurement tools score severity and importance of all activities named in a predefined list [13,89–91].

Three studies described instruments which were derived from existing instruments: the Patient-specific Arm, Shoulder and Hand questionnaire (PS-DASH) [90], the Individualized Health Assessment Questionnaire Disability Index (I-HAQ-DI) [12], and the Individualized Western Ontario and McMaster Universities Osteoarthritis Index (I-WO MAC) [88]. The original DASH, HAQ, and WOMAC are extensively psychometrically evaluated and widely used in clinical practice.

Most identified instruments are used to evaluate changes over time [13,72,73,75–78,81,82,84–91]. Only the PSFS described by Westaway et al. and the I-HAQ-DI have a discriminative purpose [12,79].

Table 1. Full names of the included instruments

Abbreviation	Full name
<i>COPM</i>	Canadian Occupational Performance Measure
<i>I-HAQ-DI</i>	Individualized Health Assessment Questionnaire Disability Index
<i>IMAS</i>	Individualized Milliken Activities of Daily Living Scale
<i>I-PSI</i>	Interviewer-administrated Patient-Specific Index
<i>I-WOMAC</i>	Individualized Western Ontario and McMaster Universities Osteoarthritis Index
<i>MACTAR</i>	McMaster Toronto Arthritis Patient Preference Disability Questionnaire
<i>PSA</i>	Patient-Specific Approach
<i>PSAQ</i>	Patient-Specific Activity Questionnaire
<i>PS-DASH</i>	Patient-Specific Disabilities of the Arm, Shoulder and Hand questionnaire
<i>PSFS</i>	Patient-Specific Functional Scale
<i>SMCS</i>	Severity of the Main Complaints Scale
<i>S-PSI</i>	Self-Reported Patient-Specific Index



All descriptive characteristics of the included measurement tools are presented in Table 2.

### 3.3. Psychometric quality assessment

Psychometric properties of the included patient-specific instruments are presented in Table 3. None of these instruments has been tested on all psychometric quality criteria of the checklist. One instrument has been tested positively on four items [73], nine on two items [13,72,78–80,84,85,87,91,92], five on one item [74,77,81,82,88], and six instruments have not been evaluated positively on any of the items [12,75,76,86,89,90].

#### 3.3.1. Content validity

None of the instruments scored positively on content validity, unless the measurement aim, the concept being measured, and the target population were mostly described well. Content validity was often assessed as “unknown” because it was unclear whether both the patient and an investigator or an expert were involved in item selection.

#### 3.3.2. Internal consistency

Two instruments (I-WOMAC and I-HAQ-DI) [12,88] were assessed on internal consistency. Although Cronbach's alpha was 0.93 and 0.87, respectively, internal consistency was scored as “doubtful” because confirmatory factor analysis was not performed.

#### 3.3.3. Criterion validity

The self-reported version of the Patient-Specific Index (Self-Reported Patient-Specific Index [S-PSI]) was positively related to the interviewbased version of the Patient-Specific Index (Interviewer-administrated Patient-Specific Index [I-PSI]) (Pearson product moment correlation [ $r$ ] = 0.78) and therefore obtained a positive score on criterion validity [91]. The Individualized Milliken Activities of Daily Living Scale (IMAS) instrument was not related to a convincing gold standard and was scored as “doubtful” [89]. Criterion validity was not assessed for the remaining instruments.

#### 3.3.4. Construct validity

Ten studies presented specific hypotheses regarding the strength and direction of expected correlations with other measurement tools, which are supposed to measure physical function [13,72–74,78–80,82,84,90]. Eight of these 10 studies obtained a positive score. Two studies obtained a negative score because less than 75% of the hypotheses were confirmed [79,90]. The I-WOMAC and the I-HAQ-DI were assessed on construct validity without hypotheses testing but with comparison with other original instruments [12,88]. In nine studies, construct validity was not assessed.

#### 3.3.5. Reproducibility—reliability

The PSFS described by Chatman et al., Cleland et al., Stratford et al., and Westaway et al., as well as the I-WO-

MAC, the I-PSI, the S-PSI, and the Patient-Specific Approach (PSA) described by Rollman et al. seem to be reliable instruments to assess patient-specific physical function ( $0.72 \leq \text{ICC} \leq 0.92$ ) [13,72,73,78,79,85,88,91]. The PSFS described by Young et al. [80] showed an ICC of 0.17 and was rated negatively as a consequence. One study used Pearson correlations to express reliability ( $r = 0.91$ ) and was therefore scored as “doubtful” [89].

#### 3.3.6. Reproducibility—agreement

Two instruments obtained positive scores on agreement [73,79]. In both cases, the SDC was smaller than the MIC. The PSFS by Stratford et al. [78] presented a standard error (SE) of measurement of 0.41 but did not define an MIC and subsequently obtained a “doubtful” score.

#### 3.3.7. Responsiveness

Information on responsiveness lacked in four studies [74,89–91]. The PSFS by Cleland et al. indicated an AUC of 0.99, whereas the PSFS by Stewart and Maher and Young et al. indicated both an AUC of 0.71 [73,77,80]. The Canadian Occupational Performance Measure (COPM) showed an AUC of 0.79, the SMCS showed an AUC of 0.82, the Patient-Specific Activity Questionnaire (PSAQ) of 0.75, and the PSA described by Rollman et al. presented an AUC of 0.91 [85–87]. An AUC  $\geq 0.70$  is required for a positive score. The PSFS by Pengel was also scored positively, based on a Guyatt RR of 2.1 [76]. Conforming to the criteria of Terwee et al., the remaining studies used an inadequate design to assess responsiveness, such as a sensitivity-to-change coefficient, the standardized response mean, and the SE.

#### 3.3.8. Floor and ceiling effects

The PSAQ and the PSFS by Chatman et al. [72] were evaluated with respect to floor and ceiling effects. The PSAQ scored positively on this item because only six percent of the respondents had the lowest possible score [87]. Chatman et al. [72] did not indicate a specific percentage and was therefore assessed as “doubtful.”

#### 3.3.9. Interpretability

None of the studies met all criteria concerning interpretability. Because at least four subgroups lacked mean scores and SDs, as well as information about the MIC, interpretability was scored as “no information available” in all studies, as well as information about the MIC, interpretability was scored as “no information available” in all studies.

## 4. Discussion

Twenty-three studies referring to 12 different instruments, which investigated psychometric properties of patient-specific instruments measuring physical function

**Table 2.** Description of the patient-specific instruments

Measurement tool	Purpose	Target population	Size <sup>a</sup>	Outcome domain(s)	Mode of administration	Mode of item selection	Number of scales	Number of items per scale	Response options	Range of scores	Time to administer
COPM [84,92]	Evaluative	Various musculoskeletal disorders (small part neuromuscular disorders)	237	Physical functioning (performance)	Interview based	Patient	2	Unlimited	0–10	0–100	30–45 min
I-HAQ-DI [12]	Discriminative	Rheumatoid arthritis	370	Physical functioning (capacity)	Self-administrated	Investigator + patient	2	5	0–3	0–45	?
IMAS [89]	Evaluative	Postsurgical traumata upper extremity	45	Physical functioning (capacity and performance)	Self-administrated	Patient	6	Meal: 8 Hygiene: 9 Dressing: 8 Manipulation: 9 Cleaning: 7 Else: 6	5 + 3	47–235	10 min
I-PSI [13]	Evaluative	Total hip replacement	74	Physical functioning (capacity and performance)	Self-administrated	Investigator + patient	4	6, 7, 15, 28	7	0–100 percentile score	16 ± 5.5 min
I-WOMAC [88]	Evaluative	Hip/knee osteoarthritis	1,218	Physical functioning (performance)	Self-administrated	Investigator + patient	1	5	0–100	0–500	?
MACTAR [82,83]	Evaluative	Rheumatoid arthritis, Chronic low back pain	155, 100	Physical functioning (performance), health status	Interview based	Patient	2	Trans: 0–10 Status: 5–9	1–7	11–47	15 min
PSA by Beurskens et al. [86]	Evaluative	Low back pain	150	Physical functioning (performance)	Self-administrated	Patient	1	3	0–100	0–300	?
PSA by Rollman et al. [85]	Evaluative	Temporomandibular disorders	132	Physical functioning (performance)	Self-administrated	Patient	1	1	0–100	0–100	?
PSAQ [87]	Evaluative	Low back pain	201	Physical functioning (performance)	Self-administrated	Patient	1	3	0–10	0–30 per activity	?
PS-DASH [90]	Evaluative	Carpal tunnel release	98	Physical functioning (performance), arm-specific symptoms	?	Patient	2	30	0–5	0–100 percentile score	?
PSFS by Chatman et al. [72]	Evaluative	Knee dysfunction	38	Physical functioning (performance)	Interview based	Patient	1	0–5	0–10	0–10	?

(Continued)

**Table 2**  
Continued

Measurement tool	Purpose	Target population	Size <sup>a</sup>	Outcome domain(s)	Mode of administration	Mode of item selection	Number of scales	Number of items per scale	Response options	Range of scores	Time to administer
PSFS by Cleland et al. [73]	Evaluative	Cervical radiculopathy	38	Physical functioning (performance)	?	Patient	1	3	0–10	0–10	?
PSFS by Gross et al. [74]	Evaluative	Workers' compensation claimants with musculoskeletal disorders	294	Physical functioning (performance)	Self-administrated	Patient	1	3	0–30	0–10	?
PSFS by McMillan and Binhammer [75]	Evaluative	Hand injuries	81	Physical functioning (performance)	Self-administrated	Patient	1	0–5	1	0–5	?
PSFS by Pengel [76]	Evaluative	Low back pain	156	Physical functioning (performance)	?	Patient	1	3	0–10	0–30	?
PSFS by Stewart and Maher [77]	Evaluative	Whiplash	132	Physical functioning (performance)	Self-administrated	Patient	1	3	0–10	0–10	?
PSFS by Stratford et al. [78]	Evaluative	Low back pain	63	Physical functioning (performance)	Interview-based	Patient	1	0–5	0–10	0–50	?
PSFS by Westaway et al. [79]	Discriminative	Neck pain	31	Pain Physical functioning (capacity)	Interview-based	Patient	2	P: 2 F: 3	0–10	0–10	?
PSFS by Young et al. [80]	Evaluative	Cervical radiculopathy	165	Physical functioning (performance)	?	Patient	1	3	0–10	0–10	?
SMCS [81]	Evaluative	Low back pain	81	Physical functioning (performance)	?	Patient	1	3	0–100	0–300	?
S-PSI [91]	Evaluative	Total hip replacement	113	Physical functioning (capacity and performance)	Self-administrated	Patient	2	≥24	7	48–336	18 min

P, pain; F, function; Trans, transitional part; Status, status part.

<sup>a</sup> Size of the population in which the instrument was evaluated.

**Table 3.** Psychometric quality of the patient-specific instruments

Measurement tool	Content validity	Internal consistency	Criterion validity	Construct validity	Reproducibility			Floor and ceiling effects	Interpretability	No. of positive ratings/total no. of ratings
					Agreement	Reliability	Responsiveness			
COPM [84,92]	0	0	0	+	0	0	+	0	0	2/2
I-HAQ-DI [12]	0	?	0	?	0	0	?	0	0	0/3
IMAS [89]	0	0	?	0	0	?	0	0	0	0/2
I-PSI [13]	0	0	0	+	0	+	?	0	0	2/3
I-WOMAC [88]	0	?	0	?	0	+	?	0	0	1/4
MACTAR [82,83]	?	0	0	+	0	0	?	0	0	1/3
PSA by Beurskens et al. [86]	0	0	0	0	0	0	?	0	0	0/1
PSA by Rollman et al. [85]	0	0	0	0	0	+	+	0	0	2/2
PSAQ [87]	0	0	0	0	0	0	+	+	0	2/2
PS-DASH [90]	0	0	0	–	0	0	0	0	0	0/1
PSFS by Chatman et al. [72]	0	0	0	+	0	+	?	?	0	2/4
PSFS by Cleland et al. [73]	0	0	0	+	+	+	+	0	0	4/4
PSFS by Gross et al. [74]	0	0	0	+	0	0	0	0	0	1/1
PSFS by McMillan and Binhammer [75]	0	0	0	0	0	0	?	0	0	0/1
PSFS by Pengel [76]	0	0	0	0	0	0	+	0	0	0/1
PSFS by Stewart and Maher [77]	0	0	0	0	0	0	+	0	0	1/1
PSFS by Stratford et al. [78]	0	0	0	+	?	+	?	0	0	2/4
PSFS by Westaway et al. [79]	0	0	0	–	+	+	?	0	0	2/4
PSFS by Young et al. [80]	0	0	0	+	0	–	+	0	0	2/3
SMCS [81]	0	0	0	0	0	0	+	0	0	1/1
S-PSI [91]	0	0	+	0	0	+	0	0	0	2/2

+, Positive assessment; ?, doubtful assessment; –, negative assessment; 0, no information available.

in patients with musculoskeletal disorders were identified. The PSFS is the most described tool, with nine different variations. Extended versions of the PSFS were developed to additionally prioritize the patient-mentioned impaired activities. In none of the instruments, all measurement properties proposed by Terwee et al. [14] have been evaluated. The PSFS described by Cleland et al. [73] achieved the highest score with four positively scored items. The PSAQ, the PSFSs by Cleland et al., Gross et al., and Stewart and Maher, the SMCS, the S-PSI, the COPM, and the PSA described by Rollman et al. obtained positive scores on all measurement properties which have been evaluated [73,74,77,81,84,85,87,91,92].

One of the treatment goals in patients with musculoskeletal disorders who are disabled in performing activities of daily living is to improve physical functioning. Assessing impaired activities and evaluating them over time is an adequate method to meet this goal. Nevertheless, the practical elaboration of this method varies across the different instruments. Some instruments only identify impaired activities and rank them with respect to severity [72–76,78–82,85], whereas other instruments additionally investigate the importance of the performance on the impaired activities

[12,84,88,90] or the frequency of the impaired activities in daily life [86,87].

The differences in practical elaboration become even more clear in the variations between different versions of the PSFS. Nine identified studies comprised different versions of the PSFS [72–80]. Although all these instruments are designated as “PSFSs,” the scoring method and mode of administration vary between the measurement tools. This hampers comparison between different patients or conditions.

The method of identifying impaired activities differed between the identified instruments. Most instruments allow patients to formulate their own impaired activities. On the other hand, the SMCS, PSAQ, PS-DASH, IMAS, I-WOMAC, I-HAQ, I-PSI, S-PSI, and MACTAR use predefined lists of potential activities [12,13,81–83,87–91]. The advantage of the application of predefined lists is the facilitated possibility for comparison with and between different populations and settings. Furthermore, change over time can be indicated easier in disabilities with potential for improvement. On the contrary, the application of predefined lists entails the risk of missing important activities [11].



Psychometric quality was extracted using a checklist developed by Terwee et al. [14]. In total, 189 items were assessed. One instrument achieved four of eight positive scores [73]; six instruments noted no positive score [12,75,76,86,89,90]. Hereby, the overall psychometric quality of patient-specific instruments, which measure physical function seems to be low. However, only three items obtained an insufficient score. The vast majority (75%) has been assessed as “no information available.” The relative short existence of patient-specific measures might explain this phenomenon. Nevertheless, despite the scarcity of psychometric data concerning patient-specific instruments, construct validity, reliability, and responsiveness were investigated in more than half of the studies. Construct validity was positively assessed in eight of 12 studies in which this quality was evaluated [13,72–74,78,80,82,84]. It can be concluded that patient-specific instruments seem to measure the same construct as disease-specific or generic physical function tools. Eight instruments appeared to be reliable [13,72,73,78,79,85,88,91]. However, in case of evaluative tools, responsiveness is possibly a more appropriate property than reliability. The most adequate approach for evaluating responsiveness is still unclear [25]. Therefore, it is not surprising that we found many different ways in which responsiveness had been determined. Conforming to the criteria of Terwee et al. [14], eight studies used an adequate method (ROC curve or Guyatts RR) [73,76,77,80,81,85,87,92]. Nine studies used inadequate methods, such as the standardized response mean and the effect size. However, the standardized response mean and effect size are widely administered in psychometric research to assess responsiveness. Therefore, the question arises whether this measurement property is not judged too strictly.

Floor and ceiling effects have been poorly investigated yet. However, patients may indicate very difficult activities as their most impaired activities and, as a consequence, rate these activities with the lowest possible score. The likelihood of improvement of these impossible activities is small, even as the improvement in rating [11]. Therefore, more solid research on floor and ceiling effects of patient-specific instruments is needed.

Some limitations of this systematic review should be acknowledged. First, patient-specific instruments are developed to enable clinicians to measure changes in activities, which really matter to individual patients. However, to determine the psychometric quality of patient-specific instruments, evaluation should occur in clinical practice in addition to evaluation in population studies. Moreover, this review was limited to studies with a primary goal to determine psychometric qualities; studies in which several measurement tools were applied to detect differences in physical function might provide complementary evidence concerning measurement properties. Third, the checklist developed by Terwee et al. is not a gold standard to determine psychometric quality of an instrument. It is just a tool

to check whether the properties are clear and systematically presented. Thereby, the criteria to obtain a positive score are very stringent sometimes. For example, content validity is exclusively supposed to be positive if the measurement aim the target population, the concepts that are being measured, the item selection, and the involvement in item selection were sufficiently described. Missing just one of these criteria induce a “doubtful” score. A final remark concerning the checklist of Terwee et al. is the unavailability to determine the overall best instrument.

Nevertheless, based on the current evidence, the PSFS described by Cleland et al. [73] could be recommended to use in clinical practice to monitor a patients’ progress in a patient-specific way. This instrument has been evaluated most comprehensively and obtained just positive scores. In future, more studies, performed in clinical practice, are needed to confirm the psychometric quality of patient-specific instruments in general and the PSFS described by Cleland et al. in particular.

In conclusion, although 12 instruments have been developed to measure patient-specific function in patients with musculoskeletal disorders, the descriptive properties and psychometric quality of these instruments are only partly investigated. The PSFS is the most investigated instrument: nine different versions have been evaluated psychometrically. The version of Cleland et al. [73] was most extensively investigated, obtained exclusively positive scores following the quality criteria by Terwee et al. [14], and could be recommended for clinical use therefore. Future research will be necessary to confirm the psychometric quality of patient-specific instruments measuring physical function in patients with musculoskeletal disorders.

## Appendix

### Supplementary material

Supplementary material can be found, in the online version, at [10.1016/j.jclinepi.2011.12.005](http://dx.doi.org/10.1016/j.jclinepi.2011.12.005).

## References

- [1] Picavert H, Schouten J. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC(3)-study. *Pain* 2003;102:167–78.
- [2] Dunlop D, Semanik P, Song J, Manheim L, Shih V, Chang R. Risk factors for functional decline in older adults with arthritis. *Arthritis Rheum* 2005;52:1274–82.
- [3] Miller M, Rejeski W, Reboussin B, Ten Have T, Ettinger W. Physical activity, functional limitations, and disability in older adults. *J Am Geriatr Soc* 2000;48:1264–72.
- [4] American College of Sports Medicine, Chodzko-Zajko W, Proctor D, Fiatarone Singh M, Minson C, Nigg C, et al. American College of Sports Medicine position stand. Exercise and physical activity for older adults. *Med Sci Sports Exerc* 2009;41:1510–30.
- [5] Weigl M, Cieza A, Cantista P, Reinhardt J, Stucki G. Determinants of disability in chronic musculoskeletal health conditions: a literature review. *Eur J Phys Rehabil Med* 2008;44:67–79.

- [6] Tugwell P, Bombardier C, Buchanan W, Goldsmith C, Grace E, Hanna B. The MACTAR Patient Preference Disability Questionnaire—an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. *J Rheumatol* 1987;14:446–51.
- [7] McHorney C, Ware J, Raczek A. The MOS 36-Item Short-Form Health Survey (SF-36), II: psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247–63.
- [8] McHorney C, Ware J, Lu R, Donald Sherbourne C. The MOS 36-item short-form health survey (SF-36): III. Tests of data-quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;32:40–66.
- [9] Ware J, Sherbourne C. The MOS 36-Item Short-Form Health Survey (SF-36), I: conceptual framework and item selection. *Med Care* 1992;30:473–83.
- [10] Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol* 1991;14:409–15.
- [11] Jolles BM, Buchbinder R, Beaton DE. A study compared nine patient-specific indices for musculoskeletal disorders. *J Clin Epidemiol* 2005;58:791–801.
- [12] Seror R, Tubach F, Baron G, Guillemin F, Ravaud P. Measure of function in rheumatoid arthritis: individualized or classical scales? *Ann Rheum Dis* 2010;69:97–101.
- [13] Wright J, Young N. The patient-specific index: asking patients what they want. *J Bone Joint Surg Am* 1997;79:974–83.
- [14] Terwee C, Bot S, de Boer M, van der Windt D, Knol D, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- [15] Scientific Advisory Committee of the Medical Outcome Trust. Assessing health status and quality of life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193–205.
- [16] Guyatt G, Feeny D, Patrick D. Measuring health related quality of life. *Ann Intern Med* 1993;118:622–9.
- [17] Kirshner B, Guyatt B. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27–36.
- [18] Streiner D, Norman G. *Health Measurement Scales; a practical guide to their development and use*. 4th ed. Oxford, UK: Oxford University Press; 2008.
- [19] Nunnally J. In: *Psychometric theory*. 2nd ed. New York: McGraw-Hill; 1978.
- [20] Beckerman H, Roebroeck M, Lankhorst G, Becher J, Bezemer P, Verbeek A. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res* 2001;10:571–8.
- [21] Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;307–10.
- [22] de Vet H, Bouter L, Bezemer P, Beurskens A. Reproducibility and responsiveness of evaluative outcome measures. *Int J Technol Assess Health Care* 2001;17:479–87.
- [23] Deyo R, Centor R. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 1986;39:897–906.
- [24] Mokkink L, Terwee C, Knol D, Stratford P, Alonso J, Patrick D, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22.
- [25] Terwee C, Dekker F, Wiersinga W, Prummel M, Bossuyt P. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003;12:349–62.
- [26] Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171–8.
- [27] McHorney C, Tarlov A. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293–307.
- [28] Lohr K, Aaronson N, Alonso J, Burnam M, Patrick D, Perrin E, et al. Evaluating quality of life and health status instruments: development of scientific review criteria. *Clin Ther* 1996;18:979–92.
- [29] Carpenter L, Baker GA, Tyldesley B. The use of the Canadian occupational performance measure as an outcome of a pain management program. *Can J Occup Ther* 2001;68:16–22.
- [30] Hefford C, Lodge S, Elliott K, Abbott JH. Measuring patient-specific outcomes in musculoskeletal clinical practice: a pilot study. *N Z J Physiother* 2008;36:41–8.
- [31] Sterling M, Brentnall D. Patient specific functional scale. *Aust J Physiother* 2007;53:65.
- [32] Juniper EF, Buist AS, Cox FM, Ferrie PJ, King DR. Validation of a standardized version of the Asthma Quality of Life Questionnaire. *Chest* 1999;115:1265–70.
- [33] Nakawatase Y, Taru C, Tsutou A, Shiotani H, Kido Y, Ohara T, et al. Development of an evaluation scale for self-management behavior related to physical activity of type 2 diabetic patients. *Diabetes Care* 2007;30:2843–8.
- [34] Akai M, Doi T, Fujino K, Iwaya T, Kurosawa H, Nasu T. An outcome measure for Japanese people with knee osteoarthritis. *J Rheumatol* 2005;32:1524–32.
- [35] Alexander M, Franko OI, Makhni EC, Zurakowski D, Day CS. Validation of a modern activity hand survey with respect to reliability, construct and criterion validity. *J Hand Surg Eur Vol* 2008;33:653–60.
- [36] Alonso J, Lamarca R, Marti V. The pain and function of the hip (PFH) scale: a patient-based instrument for measuring outcome after total hip replacement. *Orthopedics* 2000;23:1273–7.
- [37] Bago J, Perez-Gruoso FJ, Les E, Hernandez P, Pellise F. Minimal important differences of the SRS-22 Patient Questionnaire following surgical treatment of idiopathic scoliosis. *Eur Spine J* 2009;18:1898–904.
- [38] Davis SE, Chung KC. Validity and responsiveness of the Jebsen-Taylor Hand Function Test. *J Hand Surg Am* 2010;35:30–7.
- [39] Dawson J, Doll H, Boller I, Fitzpatrick R, Little C, Rees J, et al. The development and validation of a patient-reported questionnaire to assess outcomes of elbow surgery. *J Bone Joint Surg Br* 2008;90B:466–73.
- [40] Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *J Rheumatol* 2009;36:2061–6.
- [41] Goodacre L, Smith J, Meddis D, Goodacre J. Development and validation of a patient-centred Measure of Activity Limitation (MAL) in rheumatoid arthritis. *Rheumatology* 2007;46:703–8.
- [42] Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. *J Bone Joint Surg Am* 1998;80:1132–45.
- [43] Moe RH, Garratt A, Slatkowsky-Christensen B, Maheu E, Mowinckel P, Kvien TK, et al. Concurrent evaluation of data quality, reliability and validity of the Australian/Canadian Osteoarthritis Hand Index and the Functional Index for Hand Osteoarthritis. *Rheumatology* 2010;49:2327–36.
- [44] Van Der Wees P, Hendriks E, Van Beers H, Van Rijn R, Dekker J, De Bie R. Validity and responsiveness of the ankle function score after acute ankle injury. *Scand J Med Sci Sports* 2010.
- [45] de Laat F, Rommers G, Geertzen J, Roorda L. Construct validity and test-retest reliability of the questionnaire rising and sitting down in lower-limb amputees. *Arch Phys Med Rehabil* 2011;92:1305–10.
- [46] El Miedany Y, El Gaafary M, El Arousy N, Ahmed I, Youssef S, Palmer D. Patient reported outcomes in ankylosing spondylitis: development and validation of a new questionnaire for functional impairment and quality of life assessment. *Clin Exp Rheumatol* 2011;29:801–10.
- [47] Gabel C, Melloh M, Burket B, Michener L. Lower Limb Functional Index: development and clinimetric properties. *Phys Ther* 2011.
- [48] Katz P, Radvanski D, Allen D, Buyske S, Schiff S, Nadkarni A. Development and validation of a short form of the valued life activities disability questionnaire for rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2011;63:1667–71.

- [49] Thorborg K, Holmich P, Christensen R, Petersen J, Roos E. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. *Br J Sports Med* 2011;45:478–91.
- [50] Ornetti P, Dougados M, Paternotte S, Logeart I, Gossec L. Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis. *Ann Rheum Dis* 2011;70:740–6.
- [51] Andrew WD, Jane KS, Sebastian JP, Rajkumar S, Bennetts K. Performance problems of patients with chronic low-back pain and the measurement of patient-centered outcome. *Spine (Phila Pa 1976)* 2004;29:87–93.
- [52] Bearon LB, Crowley GM, Chandler J, Robbins MS, Studenski S. Personal functional goals: a new approach to assessing patient-centered outcomes. *J Appl Gerontol* 2000;19:326–44.
- [53] Gossec L, Dougados M, Rincheval N, Balanescu A, Boumpas DT, Canadello S, et al. Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Ann Rheum Dis* 2009;68:1680–5.
- [54] Hefford C, Lodge S, Elliott K, Abbott JH. Measuring patient-specific outcomes in musculoskeletal clinical practice: a pilot study. *N Z J Physiother* 2008;36:41–8.
- [55] Marx RG, Jones EC, Atwan NC, Closkey RF, Salvati EA, Sculco TP. Measuring improvement following total hip and knee arthroplasty using patient-based measures of outcome. *J Bone Joint Surg Am* 2005;87A:1999–2005.
- [56] Pouchot J. Toward individualized quality-of-life assessment. *Jt Bone Spine* 2000;67:83–5.
- [57] Berendes T, Pilot P, Willems J, Verburg H, te SR. Validation of the Dutch version of the Oxford Shoulder Score. *J Shoulder Elbow Surg* 2010;19:829–36.
- [58] Bizzini M, Gorelick M. Development of a German version of the knee outcome survey for daily activities. *Arch Orthop Trauma Surg* 2007;127:781–9.
- [59] Costa LOP, Maher CG, Latimer J, Ferreira PH, Ferreira ML, Pozzi GC, et al. Clinimetric testing of three self-report outcome measures for low back pain patients in Brazil: which one is the best? *Spine* 2008;33:2459–63.
- [60] Kjekken I, Dagfinrud H, Uhlig T, Mowinckel P, Kvien TK, Finset A. Reliability of the Canadian Occupational Performance Measure in patients with ankylosing spondylitis. *J Rheumatol* 2005;32:1503–9.
- [61] Kohlmann T, Bullinger M, Kirchberger-Blumstein I. The German version of the Nottingham Health Profile (NHP): method of translation and psychometric validation. *Soz Präventivmed* 1997;42:175–85.
- [62] Wollmerstedt N, Kirschners S, Faller H, König A. Reliability, validity and responsiveness of the German Short Musculoskeletal Function Assessment Questionnaire in patients undergoing surgical or conservative inpatient treatment. *Qual Life Res* 2006;15:1233–41.
- [63] Cott CA, Teare G, McGilton KS, Lineker S. Reliability and construct validity of the client-centred rehabilitation questionnaire. *Disabil Rehabil* 2006;28:1387–97.
- [64] Hall AM, Maher CG, Latimer J, Ferreira ML, Costa LO. The patient-specific functional scale is more responsive than the Roland Morris disability questionnaire when activity limitation is low. *Eur Spine J* 2011;20:79–86.
- [65] Berghuis-Kelley D, Scherer S. Research corner. Outcome measures in cardiopulmonary physical therapy: use of the Patient Specific Functional Scale. *Cardiopulm Phys Ther J* 2007;18:21–3.
- [66] Amjadi SS, Maranian PM, Paulus HE, Kaplan RM, Ranganath VK, Furst DE, et al. Validating and assessing the sensitivity of the Health Assessment Questionnaire-Disability Index-derived Short Form-6D in patients with early aggressive rheumatoid arthritis. *J Rheumatol* 2009;36:1150–7.
- [67] Dawson J, Coffey J, Doll H, Lavis G, Cooke P, Herron M, et al. A patient-based questionnaire to assess outcomes of foot surgery: validation in the context of surgery for hallux valgus. *Qual Life Res* 2006;15:1211–22.
- [68] Clinch J, Tugwell P, Wells G, Shea B. Individualized functional priority approach to the assessment of health related quality of life in rheumatology. *J Rheumatol* 2001;28:445–51.
- [69] Sanchez K, Papelard A, Nguyen C, Jousse M, Rannou F, Revel M, et al. Patient-Preference Disability Assessment for Disabling Chronic Low Back Pain. *Spine* 2009;34:1052–9.
- [70] En MC, Clair DA, Edmondston SJ. Validity of the Neck Disability Index and Neck Pain and Disability Scale for measuring disability associated with chronic, non-traumatic neck pain. *Man Ther* 2009;14:433–8.
- [71] Hoving JL, O'Leary EF, Niere KR, Green S, Buchbinder R. Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. *Pain* 2003;102:273–81.
- [72] Chatman AB, Hyams SP, Neel JM, Binkley JM, Stratford PW, Schomberg A, et al. The patient-specific functional scale: measurement properties in patients with knee dysfunction. *Phys Ther* 1997;77:820–9.
- [73] Cleland JA, Fritz JM, Whitman JM, Palmer JA. The reliability and construct validity of the neck disability index and patient specific functional scale in patients with cervical radiculopathy. *Spine* 2006;31:598–602.
- [74] Gross DP, Battie MC, Asante AK. The Patient-Specific Functional Scale: validity in workers' compensation claimants. *Arch Phys Med Rehabil* 2008;89:1294–9.
- [75] McMillan CR, Binhammer PA. Which outcome measure is the best? Evaluating responsiveness of the Disabilities of the Arm, Shoulder, and Hand Questionnaire, the Michigan Hand Questionnaire and the Patient-Specific Functional Scale following hand and wrist surgery. *Hand (N Y)* 2009;4:311–8.
- [76] Pengel LHM. Responsiveness of pain, disability, and physical impairment outcomes in patients with low back pain. *Spine (Phila Pa 1976)* 2004;29:879–83.
- [77] Stewart M, Maher CG. Responsiveness of pain and disability measures for chronic whiplash. *Spine (Phila Pa 1976)* 2007;32:580–5.
- [78] Stratford P, Gill C, Westaway M, Binkley J. Assessing disability and change on individual patients: a report of a patient specific measure. *Physiotherapy Canada* 1995;47:258–63.
- [79] Westaway MD, Stratford PW, Binkley JM. The Patient-Specific Functional Scale: validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther* 1998;27:331–8.
- [80] Young IA, Cleland JA, Michener LA, Brown C. Reliability, construct validity, and responsiveness of the neck disability index, patient-specific functional scale, and numeric pain rating scale in patients with cervical radiculopathy. *Am J Phys Med Rehabil* 2010;89:831–9.
- [81] Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 1996;65:71–6.
- [82] Verhoeven A, Boers M, van der Lindern S. Validity of the MACTAR questionnaire as functional index in a rheumatoid arthritis clinical trial. *J Rheumatol* 2000;27:2801–9.
- [83] Sanchez K, Papelard A, Nguyen C, Bendeddouche I, Jousse M, Rannou F, et al. McMaster-Toronto Arthritis Patient Preference Disability Questionnaire sensitivity to change in low back pain: influence of shifts in priorities. *PLoS One* 2011;6:e20274.
- [84] Dedding C, Cardol M, Eyssen IC, Dekker J, Beelen A. Validity of the Canadian Occupational Performance Measure: a client-centred outcome measurement. *Clin Rehabil* 2004;18:660–7.
- [85] Rollman A, Naeije M, Visscher CM. The reproducibility and responsiveness of a patient-specific approach: a new instrument in evaluation of treatment of temporomandibular disorders. *J Orofac Pain* 2010;24:101–5.
- [86] Beurskens AJ, de Vet HC, Koke AJ, Lindeman E, van der Heijden GJ, Regtop W, et al. A patient-specific approach for measuring functional status in low back pain. *J Manipulative Physiol Ther* 1999;22:144–8.
- [87] Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 2007;10:S94–105.

- [88] Seror R, Tubach F, Baron G, Falissard B, Logeart I, Dougados M, et al. Individualising the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) function subscale: incorporating patient priorities for improvement to measure functional impairment in hip or knee osteoarthritis. *Ann Rheum Dis* 2008;67:494–9.
- [89] Seaton MK, Groth GN, Matheson L, Feely C. Reliability and validity of the Milliken Activities of Daily Living Scale. *J Occup Rehabil* 2005;15:343–51.
- [90] Vranceanu AM, Kadzielski J, Hwang R, Ring D. A patient-specific version of the Disabilities of the Arm, Shoulder, and Hand Questionnaire. *J Hand Surg Am* 2010;35:824–6.
- [91] Wright J. Evaluating the outcome of treatment. Shouldn't we be asking patients if they are better? *J Clin Epidemiol* 2000;53:549–53.
- [92] Eysen IC, Steultjens M, Oud T, Bolt E, Maasdam A, Dekker J. Responsiveness of the Canadian Occupational Performance Measure. *J Rehabil Res Dev* 2011;48:517–28.