Austin Publishing Group

Research Article

Construct Validity and Validity to Change of the Patient-Specific Functional Scale in Patients with Shoulder and Low Back Pain: A Clinimetric Study

Kromer TO[1,5]**, Saner J**[2,5]**, Sieben JM**[3 4,5] **and Bastiaenen CHG**[4,5,6]*****

[1]Faculty of Health, Safety, Society, Furtwangen University, Germany
[2]School of Health Professions, Institute of Physiotherapy, Zurich University of Applied Sciences, Switzerland
[3]Department of Anatomy and Embryology, Maastricht University, The Netherlands
[4]Research Line Functioning & Rehabilitation, Maastricht University, The Netherlands
[5]Caphri Research Institute, Research Line Functioning & Rehabilitation, Maastricht University, The Netherlands
[6]Department of Epidemiology, Maastricht University, The Netherlands

***Corresponding author:** Caroline H.G. Bastiaenen, Associate Professor, Research Program Functioning and Rehabilitation, Department of Epidemiology, P. Debyeplein 1, 6229 HA Maastricht, The Netherlands

Abstract

Background: Patient-specific and condition-specific measures are widely used in clinical practice and research to measure disability or change over time. While condition-specific outcome measures comprise a range of restrictions generally relevant for all patients, the Patient-Specific Functional Scale measures restrictions chosen by the individual patient.

Objectives: Based on the hypothesis that patient-specific and condition-specific scales deliver comparable results when used on group level. The aim of this study was to test for floor and ceiling effects, to evaluate construct validity and validity to change of the Patient-Specific Functional Scale when compared to condition-specific outcome measures. For this purpose, two datasets from patients with shoulder pain and low back pain were analyzed.

Methods: Patient-Specific Functional Scale scores were compared to the Shoulder Pain and Disability Index and the Roland Morris Disability Questionnaire at 4 time-points using stem-and-leaf-plots and correlations using Pearson's r. Hypothesis-driven correlation levels for data interpretation were predefined, with r ≥0.75=high, r ≥0.5=moderate, r ≥0.25=low.

Results: Patient-Specific Functional Scale floor effects were comparable to condition-specific outcome measures in both samples. At none of the time-points did the Patient-Specific Functional Scale correlate with the condition-specific outcome measures in the expected manner.

Conclusion: Hypotheses regarding expected ranges of correlation between the Patient-Specific Functional Scale and the condition-specific outcome measures for construct validity and validity to change were not met. While the use of the Patient-Specific Functional Scale in a clinical context has its advantages, the measure is not recommended for group-level evaluations.

Keywords: Patient-centered outcome; Validity; Subacromial pain syndrome; Low back pain

Abbreviations

BL: Baseline; CSOM: Condition-Specific Outcome Measures; GPE: Global Perceived Effect Scale; LBP: Low Back Pain; MCI: Movement Control Impairment; NDI: Neck Disability Index; NRS: Numeric Rating Scale; ODI: Oswestry Disability Index; Pearson's r: Pearson's Correlation Coefficient; PSFS: Patient-Specific Functional Scale; RMDQ: Roland & Morris Disability Questionnaire; ROC: Receiver Operating Characteristics Curve; SD: Standard Deviation; SPADI: Shoulder Pain and Disability Index; SPS: Subacromial Pain Syndrome; T1-T3: Follow-up Time Points.

Background

For a therapist, it is essential to ascertain whether improvements in body function or structure also lead to increased activity and participation levels. Therefore, the use of assessment tools which can reflect the actual status or degree of restriction and which can measure patient's change over time is of crucial importance. Improvements in body functions and structures are predominantly assessed through physical testing; activities and participation are commonly measured using questionnaires. Scores gathered using these measurement tools also allow comparison at a group level and enable patients, therapists and researchers to "measure" the impact of a disease, the progression over time or the effect of an intervention. However, since questionnaires often contain very specific items related to certain activities, it is possible that some items will not be relevant to all patients in the target group. By that, the importance of the individual items could vary between patients. Moreover, a "prefixed" item set may not include activities that are of importance to individual patients. Therefore, patients may be required to score questions that are only partly relevant to them. As a result, these standard questionnaires might not adequately reflect a patient's individual restrictions or the change in these restrictions over time. In an attempt to solve this problem, the Patient-Specific Functional Scale (PSFS) was developed with the intention to monitor a patient's progress based on relevant restrictions chosen by the individual himself [1]. The PSFS is comprised of 1 to 5 activities; each activity is

rated on an 11-point Numeric Rating Scale (NRS) from 0 (impossible to do) to 10 (no difficulties at all). The PSFS is easy to administer, and takes about five minutes to complete. However, the PSFS also has been used in the past by researchers to determine the current state of function and the development of activity restrictions over time on an average group level. By choosing this approach, researchers have moved away from the originally intended individual focus of the instrument and applied the PSFS to situations for which it was not developed or validated. From a test-theoretical perspective, there are numerous problems in deviating from the original construct. Firstly, the interpretation of an average score across self-selected activities by individuals is a challenge. For researchers and clinicians who are familiar with interpreting data on a clearly defined aspect of disability, it is tempting to interpret outcomes using the same approach; but in fact, one is averaging different constructs. Another problem is that floor or ceiling effects could occur if a patient chooses either lightly activities with scores at the lower end of the scale or severely restricted activities with scores at the upper end of the scale. In the first case, it is difficult to detect a positive development and in the second case to detect a negative development over time; this may affect results for validity to change analysis to a certain degree. Problems may also occur when the initially chosen activities become increasingly irrelevant as a problem as time gone by, due to either the patient's improved condition or reduction in complaints or because of seasonal effects, when the activity becomes more and more irrelevant during follow up as for example snow shoveling in spring. Dependent on the activities chosen by the individual patient it could also be that outcomes in the PSFS indicate higher or lower disability levels for that patient compared to Condition-Specific Outcome Measures (CSOMs) and that scorings on the PSFS may differ significantly more between patients than their corresponding outcomes on a CSOM, where all patients rate the same standardized set of items. Despite these problems, which have not yet been adequately realized or addressed, several researchers have investigated the psychometric properties of the PSFS on a group level for a variety of musculoskeletal conditions. Results have been formulated as "promising", since the PSFS has been reported as having good construct validity, discriminant validity, and responsiveness [2-4]. Based on these results we think that testing psychometric properties and comparisons at a group level can be justified by defining the PSFS as an instrument assessing "activity restriction based on items selected by an individual patient" as the overarching construct. We hypothesized that specific musculoskeletal disorders (in our example subacromial shoulder pain and low back pain) lead to specific activity restrictions and specific pain patterns. CSOMs in our case the Shoulder Pain and Disability Index (SPADI) and the Roland & Morris Disability Questionnaire (RMDQ), summarize these typical activities and include a range of tasks from easy to more difficult. These questionnaires were designed to include items that cover the whole range of items assumed relevant for a patient group, although not every item may be of equal importance to each individual patient. Therefore, we assume that many activity restrictions chosen by individual patients for their PSFS can be traced back or are closely related to items listed in the CSOMs. If this is the case, the PSFS could be approached as a construct and, because of the hypothesized close association between the operationalization of both types of measurement, assumed to deliver a relatively high correlation with the CSOM, especially in a cross-sectional analysis.

However, there is also the possibility that the PSFS measures a different dimension not covered in the COSM. Taking these arguments into consideration, the aims of this paper are threefold: In two groups of patients, suffering from either Subacromial Pain Syndrome (SPS) or Low Back Pain (LBP), and using the CSOM as an external standard comparator: 1) To test for possible floor and ceiling effects of the PSFS; 2) To evaluate its construct validity compared to the CSOM, and 3) To assess the ability of the PSFS to detect changes over time with reference to an external anchor [5].

Methods

Data were used from two different datasets collected during randomized controlled trials investigating effects of physiotherapy interventions in a patient group with SPS and a second group with LBP in primary care. A detailed description of the inclusion processes, applied treatments and primary analyses can be found in the published study protocols [6,7] and trial results [8-11]. Ethical approval was granted by the ethics committee of the Ludwig-Maximilians-University Munich, Germany (project-no. 018-10) for the SPS trial, and the Swiss Ethics Committee granted ethical approval (KEK-ZH-NR: 2010-0034/5) for the LBP trial. All patients in each trial gave informed consent. Datasets of the two samples were analyzed independently of each other.

Dataset 1 - SPS patients

Participants were recruited through referral for physiotherapy due to shoulder complaints. After baseline assessment, they were randomly assigned to either an intervention or a control group. The intervention group received exercise therapy plus manual therapy, while the control group received only exercise therapy. Baseline characteristics of the 90 participants included in the trial are presented in Table 1.

The primary outcome measure was the SPADI, a shoulder-specific, self-reported questionnaire measuring pain and disability [12]. SPADI sub-scales for pain (items 1 to 5) and function (items 6 to 13) are scored from 0 to 100, with higher scores reflecting higher pain or disability levels. Total SPADI score was calculated by averaging scores of the two sub-scales. The SPADI has been shown to be valid and highly sensitive [12,13]. The German version of the SPADI has also been shown to have excellent reliability and internal consistency [14]. The PSFS [1,15] was also applied. Patients were instructed to choose 3 activities important to them, in which they were impaired, and to rate their ability to perform those on an 11-point NRS from 0 (impossible to do) to 10 (fully capable). The average score across all activities was calculated. For reasons of standardization, the PSFS has been rescaled in this paper, so that 0 now means "no difficulties at all" and 10 means "impossible to do", in accordance with the other outcome measures used in this analysis. All measurement instruments were applied at Baseline (BL), after 5 weeks (T1), 12 weeks (T2), and at one year follow-up (T3).

Dataset 2 - LBP patients

A total of 106 patients with LBP, defined as pain persisting for longer than six weeks and with no radiating symptoms below the knee, were included in the original LBP trial. Eligible patients presented with LBP in combination with defined complaints associated with Movement Control Impairment (MCI). Other inclusion criteria were

Table 1: Baseline demographic data and results for SPS and LBP patients initially included in the original trials (mean (SD) if not otherwise stated).

	SPS (n=90)	LBP (n=106)
Age in years	51.8 (11.2)	41.6 (14.1)
Gender (female) %, n	51.1, 46	37.4, 40
Duration of the current episode in weeks	33.9 (42.8)	--
Overall duration of complaints in years	8.7 (12.7)	10.0 (11.0)
SPADI/RMDQ total score	40.4 (17.0)	8.7 (3.3)
PSFS average score	6.0 (1.7)	5.7 (1.6)
GCPS total score	--	27.8 (10.4)
GCPS sub-score disability	--	12.4 (7.6)
FABQ total score	32.7 (17.4)	32.2 (14.7)

SD: Standard Deviation; SPADI: Shoulder Pain and Disability Index; RMDQ: Roland Morris Disability Questionnaire; PSFS: Patient-Specific Functional Scale; FABQ: Fear Avoidance Beliefs Questionnaire; GCPS: Graded Chronic Pain Scale (Total Score: 70; Pain Intensity: 0-30, Disability: 0-40); SPS: Subacromial Pain Syndrome; LBP: Low Back Pain.

a score of at least two positive out of six movement tests (representing MCI) and a minimal level of disability of 5 points on the RMDQ [7,16]. Participants were randomly allocated either to an intervention group that received an individual complaint-specific exercise program, or to a control group that received general exercise therapy. Baseline characteristics are presented in Table 1.

Primary outcome measure was the PSFS [1,15,17]. Patients received the same instructions as in the shoulder trial. A secondary outcome was the RMDQ, which measures LBP-related disability. It consists of 24 dichotomous questions to be answered with either "yes" or "no", with a "yes" score meaning high disability. Reliability was shown to be high and construct and internal validity to be good, also for the German version [18-20]. All outcomes were measured at baseline (BL), at 9-12 weeks (T1), 6 months (T2), and at one year follow-up (T3).

An overview and description of all outcome measures for both datasets are provided in Table 2. For this study, we decided only to include those patients with complete data regarding variables relevant to our analysis.

Data analysis and hypotheses

Floor and ceiling effects: Since patients may be greatly restricted in their activities at the start of treatment, they may also have high scorings for their chosen activities on the PSFS. Because high or low scores at any time point could have influenced the measurement properties of the outcomes, data were checked for possible floor and ceiling effects by using stem-and-leaf-plots at every measurement point before validity to change was investigated. Floor and/or ceiling effects were assumed when more than 15% of values were within 10% of the highest and/or lowest possible scores.

Hypothesis 1: There are no floor or ceiling effects at any measurement point.

Construct validity: To test aspects of construct validity of the PSFS we calculated correlations between the PSFS and SPADI or RMDQ, respectively, at every measurement point using Pearson's r. A high correlation was defined as r ≥0.75, a moderate correlation as r ≥0.5 and a low correlation as r ≥0.25. High correlations were

expected for baseline scorings because both measurements should reflect the status of disability. For the consecutive time points, T1, T2, and T3 progressively decreasing correlations were expected: from high at baseline to low at T3, especially in patients showing good improvement. This also could be because patients were not allowed to change the initially chosen PSFS activities over the 1-year follow up period, an application of the PSFS often used in research and clinical practice nowadays. In consequence, we expected that these activities would become increasingly irrelevant for patients as their health status improved over time.

Hypothesis 2a: There is a high correlation between the PSFS and CSOMs (RMDQ/SPADI) at baseline.

Hypothesis 2b: The correlation between the PSFS and the CSOMs (RMDQ/SPADI) measured at every follow-up point in a cross-sectional independent way (T1, T2, and T3) is lower than the correlation of the preceding point: r-values will decrease from high at baseline to moderate, and to low at T3.

Validity to change: To test the ability of the PSFS to detect change over time we calculated correlations between the change scores in the SPADI/RMDQ and the change scores in the PSFS for the following intervals: BL to T1, T1 to T2, and T2 to T3. We expected that the correlation between change scores would be acceptable in the short term, but would diverge over the longer term. Therefore, our third hypothesis was:

Hypothesis 3: PSFS change scores show high correlations with both the SPADI and the RMDQ between BL to T1, moderate correlations between T1-T2, and low correlations between T2-T3. The CSOMs are used as external anchors.

Results

Participants

Complete datasets were available for 87 SPS-participants (96.7%), and for 60 LBP-participants (56.6%). Characteristics of both samples are described in Table 3.

Floor and ceiling effects (hypothesis 1)

For the PSFS no evidence of ceiling effects at any measurement point was found. Floor effects were found in the SPS sample at T2 and T3, with 32% (n=28) and 52.9% (n=46), respectively, within 10% of the lowest possible score. However, at T2 these participants had a mean (SD) SPADI score of 3.7 (3.6) points with only three participants scoring 10 points or higher. At T3, the mean SPADI score was 2.5 (3.6) with again only three participants scored 10 points or higher. In the LBP sample floor effects were found at T1, T2, and T3, increasing from 21.6% (n=13), 36.7% (n=22) to 43.3% (n=26), respectively. As seen in the SPS sample, the average scores of these patients on the RMDQ were also comparably low. Results are summarized in Table 4.

Construct validity (hypotheses 2a and 2b)

The PSFS correlated well with SPADI and RMDQ at baseline. However, correlation coefficients were below our predefined cut-off level of r ≥0.75. Our hypothesis 2a, therefore, had to be rejected. Correlations for the time-points T1, T2 and T3 showed a progressive increase: we found the strongest correlations at T3, with r = 0.90 in the SPS and r = 0.74 in the LBP sample. This development was completely contrary to our hypothesis 2b. Based on these results, hypothesis 2b

Table 2: Outcome measures used in the two trials.

Outcomes measures	Dimension	Scale	Scorings
a) SPS			
Shoulder Pain and Disability Index (SPADI)	SPS - related pain & activity limitations	0-100, continuous 0-100, continuous	Items 1-5 scored on a 100mm VAS Items 6-13 scored on a 100mm VAS Mean of item scores. Higher scores mean higher pain/disability.
Patient-Specific Functional Scale (PSFS)	SPS - related disability	0-10, continuous	11 point visual numeric rating scale (end descriptors of 0 = impossible to do, 10 = no difficulties at all)
b) LBP			
Patient-Specific Functional Scale (PSFS)	LBP - related activity limitations	0-10, continuous	11 point visual numeric rating scale (end descriptors of 0 = impossible to do, 10 = no difficulties at all)
Roland-Morris Disability Questionnaire (RMDQ)	LBP - related disability	0-24, continuous	Dichotomous questions (yes = with disability, no=no disability); Scores 0 – 24 (minimal enrolment to trial RMDQ = 5)

SPS: Subacromial Pain Syndrome; LBP: Low Back Pain; VAS: Visual Analogue Scale.

Table 3: Baseline demographic data and results for SPS and LBP samples included in this analysis (mean (SD) if not otherwise stated).

	SPS (n=87)	LBP (n=60)
Age in years	52.0 (11.4)	41.8 (13.9)
Gender (female) %, n	49.4, 43	38.3, 23
Duration of the current SPS episode in weeks	33.6 (43.5)	--
Overall duration of complaints in years	8.6 (12.9)	9.1 (10.4)
SPADI/RMDQ total score	41.0 (17.0)	8.7 (3.3)
PSFS average score	6.0 (1.6)	5.7 (1.6)
GCPS total score	--	26.5 (10.1)
GCPS sub-score disability	--	11.1 (7.4)
FABQ total score	32.0 (17.2)	29.3 (13.8)

also had to be rejected. Detailed results are displayed in Table 5.

Validity to change (hypothesis 3)

Here we expected decreasing correlation coefficients for change scores over time. However, results did not support this hypothesis. Instead of decreasing correlation coefficients, we found alternating

patterns that varied between samples. The SPS r-values varied in an down-up sequence, with the strongest correlation for the change score between T2 and T3. In contrast, the LBP sample showed an up-down pattern. Here, a very low correlation was found between T2 and T3. Results are displayed in Table 5.

Discussion

The aim of this paper was to test the following hypotheses regarding the PSFS when compared to well-established CSOMs: no floor and ceiling effects, acceptable construct validity and validity to change over time. Two samples from different populations were analyzed and similar results were found. Four measurement points were incorporated in the analysis whereby it was possible to present the development of the relationship between the PSFS and the CSOMs over the period of one year. Other than for floor and ceiling effects results showed opposite effects than we hypothesized, resulting in the rejection of our hypotheses regarding construct validity and validity to change. These results demonstrate that the overarching construct defined in the introduction must be doubted. The PSFS certainly does not reflect change on a group level in the same way that CSOMs do. The development over time of the correlations

Table 4: Floor and ceiling effects. n= (%) of patients scoring within the highest and lowest 10% of the PSFS.

	BL	T1	T2	T3
SPS sample (n= 87)				
PSFS ≤1	N=0 (0%)	N=13 (14.9%)	N=28 (32.0%)	N=46 (52.9%)
SPADI mean (SD) score*	--	--	3.7 (3.6)	2.5 (3.6)
PSFS >9	N=5 (5.8%)	N=0 (0%)	N=2 (2.3%)	N=2 (2.3%)
LBP sample (n = 60)				
PSFS ≤1	N=0 (0%)	N=13 (21.6%)	N=22 (36.7%)	N=26 (43.3%)
RMDQ mean (SD) score*	--	1.3 (1.6)	1.8 (3.0)	1.5 (1.8)
PSFS > 9	N=1 (1.8%)	N=0 (0%)	N=0 (0%)	N=0 (0%)

*If more than 15% scored in the lower or higher 10% range of the scale, the mean and (SD) of the SPADI/RMDQ score of this sub-sample is displayed. SD: Standard Deviation; SPADI: Shoulder Pain and Disability Index; RMDQ: Roland Morris Disability Questionnaire; PSFS: Patient-Specific Functional Scale; SPS: Subacromial Pain Syndrome; LBP: Low Back Pain.

Table 5: Correlations (Pearson`s r) between PSFS and condition-specific disability scores.

Outcome measure	BL	T1		T2		T3	Change BL-T1	Change T1-T2	Change T2-T3
	SPS & LBP	5 weeks[a] SPS	9-12 weeks[a] LBP	12 weeks SPS	6 months LBP	12 months[b] SPS & LBP			
SPADI	0.406	0.663		0.794		0.903	0.572	0.417	0.661
RMDQ	0.301		0.523		0.651	0.743	0.497	0.542	0.176

BL: Baseline; r: Pearson's Correlation Coefficient; a: Post Intervention; b: Final Follow Up; SPADI: Shoulder Pain and Disability Index; RMDQ: Roland Morris Disability Questionnaire; PSFS: Patient-Specific Functional Scale; SPS: Subacromial Pain Syndrome; LBP: Low Back Pain.

between PSFS and CSOMs has led us to conclude that the underlying constructs are different and, therefore, should not be used for the same purpose. Although the PSFS has been used in several studies as a secondary outcome measure to analyze longitudinal development of activity restrictions on a group level [8,11,21-23] and seemed to perform well for this purpose, our data suggest that the underlying construct remains unclear. Therefore, we cannot recommend the use of the PSFS without taking into account that the underlying construct is besides different from CSOM also unclear to interpret on a group level, at the moment [24].

Other authors also have investigated validity aspects of the PSFS. Hall et al. [25] investigated responsiveness of RMDQ and PSFS in patients with LBP and attested both outcome measures an "acceptable" responsiveness. In a first step, they calculated correlations between RMDQ/PSFS and the Global Perceived Effect scale (GPE), which they used as an external reference standard for change. In a second step, they used a receiver operating characteristics (ROC) curve to assess responsiveness. This methodological approach provides information about the relationship between RMDQ/PSFS and GPE as an external indicator of change. Although GPE is often used for this purpose, GPE (change) scores might be more a reflection of the current health status than of true change and GPE might, therefore, be insufficient to serve as a valid external reference for change [26]. Furthermore, when using ROC it is necessary to dichotomize the external change scores, which leads to a loss of information on the magnitude of change [27]. Our purpose was to analyze whether the PSFS was responsive relative to our CSOMs. The different approach to Hall et al. [25] may explain the difference in conclusions drawn from these results.

Thoomes-de Graaf et al. [28] tested convergent validity between the PSFS and the Neck Disability Index (NDI) in patients with neck pain; they assumed that both tools measure the same construct of "activity limitations" and expected to find a strong baseline correlation between them. However, results showed only a moderate correlation.

Abbott and Schmitt [2] investigated concurrent validity (which would be defined as construct validity in the absence of a gold standard according to Mokkink et al. [5]) and validity to change of the PSFS in a sample with mixed acute and chronic musculoskeletal disorders. According to our classification system for the correlation coefficient, they found a moderate correlation between PSFS and CSOMs at baseline in the subgroup with upper extremity disorders and a low correlation in the subgroup with LBP. Interestingly these correlations were stronger at 6-month follow-up, which were similar to the outcomes in our samples. The moderate correlations found for change scores between baseline and follow-up at 6 months were also similar to ours (between baseline and our last measurement point at 1 year), although the use of different time frames may complicate this comparison. Similar results regarding construct validity were also found by Heldmann et al. [29] in LBP-patients. They used the Oswestry Disability Index (ODI) as reference measure and found low correlations at baseline and moderate correlations at follow-up. The fact that we obtained similar results to those of Abbott et al. [2] and Heldmann et al. [29] for patients with LBP and SPS has led us to the assumption that results could be independent of the condition specific population of interest. This assumption is underpinned by similar results found for lower extremity and neck disorders in the study by Abbott et al. [30], and by results from Stratford et al. [31]

who also found low correlations between change scores of PSFS and CSOM used in patients who underwent total knee arthroplasty. In our opinion, it is more likely that the similarity of results of the mentioned studies and our study is mainly based on a basic difference between the constructs of the PSFS and CSOMs.

In another study, Mannberg-Bäckman et al. [32] investigated the validity and sensitivity to change of the PSFS in a group of patients after proximal humerus fracture. Although 62% of the chosen items in the PSFS were also represented in the CSOM, correlation between the two instruments was low, whereas sensitivity to change was high in both instruments. Resnik et al. [33] investigated responsiveness of several CSOMs and the PSFS in patients undergoing rehabilitation for upper limb prosthesis. The PSFS showed the largest effect sizes from all outcomes analyzed. This difference between PSFS and other outcomes was most obvious for long-term results, where PSFS still showed significant changes while most of the other instruments did not. The authors concluded that the PSFS was one of the outcome measures most responsive to change. Heldmann et al. [29] also concluded that the PSFS is more responsive than the Oswestry Disability Index (ODI) because they found higher effect sizes for the PSFS. However, effect sizes might be rather a reflection of quantitative properties in the study population (it is merely a measure focusing on individual changes) than necessarily of the validity of the instrument itself. Therefore, these results may alternatively suggest that the PSFS measures a different construct than CSOMs and that larger effect can be reached when using the PSFS instead of a CSOM, but do not clarify the construct underlying the PSFS. Originally, the PSFS was developed to address individual activity restrictions, to prioritize treatment goals, and, last but not least, to keep patients motivated by working on personally important goals. We would agree with using the PSFS for these purposes. We also would agree with using the PSFS for risk assessment [34] and to monitor short-term effects in individual patients. The PSFS also may help to redirect focus towards value driven treatment goals chosen by the individual patients themselves and consequently on function and ability rather than pain and disability. Based on our data, we believe that the PSFS does not provide a clinically meaningful unidimensional scale comparable on a group level to a CSOM, because it aggregates not only heterogeneous functional items in one scale but also averages item-dependent scores. This makes it difficult to compare outcomes of the two types of measures on a group level in patients.

Limitations

Both of our samples contained mainly patients with chronic complaints and our results could be different from studies that investigate samples including acute patients. Furthermore, correlations based on the Pearson's r measure the closeness to a linear relationship but the relationship between two measures may be close but non-linear which we did not analyze. Our values set for the Pearson's correlation coefficient for the acceptance of the stated hypotheses were high and subjective to a certain degree; this led to a more conservative interpretation of the results. Our results for construct validity were mainly based on convergent validity; based on existing data it was not possible for us to analyze divergent validity.

Conclusions

Our hypotheses of the expected ranges of correlation between

the PSFS and the CSOM for construct validity and validity to change had to be rejected. While the use of the PSFS in a clinical context has its advantages, the measure is not recommended to assess the development of pathology or syndromes or to compare between patients on a group level.

References

1. Stratford PW, Gill C, Westaway MD, Binkley JM. Assessing disability and change on individual patients: a report of a patient specific measure. Physiother Can. 1995; 47: 258-263.

2. Abbott JH, Schmitt JS. The Patient-Specific Functional Scale was valid for group-level change comparisons and between-group discrimination. J Clin Epidemiol. 2014; 67: 681-688.

3. Hefford C, Abbott JH, Arnold R, Baxter GD. The patient-specific functional scale: validity, reliability, and responsiveness in patients with upper extremity musculoskeletal problems. J Orthop Sports Phys Ther. 2012; 42: 56-65.

4. Rosengren J, Brodin N. Validity and reliability of the Swedish version of the Patient Specific Functional Scale in patients treated surgically for carpometacarpal joint osteoarthritis. J Hand Ther. 2013; 26: 53-60.

5. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010; 63: 737-745.

6. Kromer TO, de Bie RA, Bastiaenen CH. Effectiveness of individualized physiotherapy on pain and functioning compared to a standard exercise protocol in patients presenting with clinical signs of subacromial impingement syndrome. A randomized controlled trial. BMC Musculoskelet Disord. 2010; 11: 114.

7. Saner J, Kool J, Sieben JM, Luomajoki H. Movement control exercises versus general exercise to reduce disability in patients with low back pain and movement control impairment. A randomized controlled trial. BMC Musculoskeletal Disorders. 2011; 12: 207.

8. Kromer TO, de Bie RA, Bastiaenen CH. Physiotherapy in patients with clinical signs of shoulder impingement syndrome: a randomized controlled trial. J Rehabil Med. 2013; 45: 488-497.

9. Kromer TO, de Bie RA, Bastiaenen CHG. Effectiveness and costs of physiotherapy in patients with shoulder impingement syndrome: 1-year follow up of a randomized controlled trial. Journal of Rehabilitation Medicine. 2014; 46: 1029-1036.

10. Saner J, Kool J, Sieben JM, Luomajoki H, Bastiaenen CHG, De Bie RA. A tailored exercise program versus general exercise for a subgroup of patients with low back pain and movement control impairment: A randomized controlled trial with one-year follow up. Manual Therapy. 2015; 20: 672-679.

11. Saner J, Sieben JM, Kool J, Luomajoki H, Bastiaenen CHG, De Bie RA. A tailored exercise program versus general exercise for a subgroup of patients with low back pain and movement control impairment: Short-term results of a randomized controlled trial. Journal of Bodyworks & Movement Therapies. 2016; 20: 189-202.

12. MacDermid JC, Solomon P, Prkachin K. The Shoulder Pain and Disability Index demonstrates factor, construct and longitudinal validity. BMC Musculoskelet Disord. 2006; 7: 12.

13. Beaton DE, Richards RR. Measuring function of the shoulder. A cross-sectional comparison of five questionnaires. J Bone Joint Surg Am. 1996; 78: 882-890.

14. Angst F, Goldhahn J, Pap G, Mannion AF, Roach KE, Siebertz D, et al. Cross-cultural adaptation, reliability and validity of the German Shoulder Pain and Disability Index (SPADI). Rheumatology (Oxford). 2007; 46: 87-92.

15. Horn KK, Jennings S, Richardson G, Vliet DV, FHeffod C, Abbott JH. The patient-specific functional scale: psychometrics, clinimetrics, and application as a clinical outcome measure. J Orthop Sports Phys Ther. 2012; 42: 30-42.

16. Luomajoki H, Kool J, de Bruin ED, Airaksinen O. Movement control tests of the low back; evaluation of the difference between patients with low back pain and healthy controls. BMC Musculoskelet Disord. 2008; 9: 170.

17. Hall AM, Maher CG, 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.

18. Roland M, Morris R. 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 (Phila Pa 1976). 1983; 8: 141-144.

19. Wiesinger GF, Nuhr M, Quittan M, Ebenbichler G, Wolfl G, Fialka-Moser V. Cross-cultural adaptation of the Roland-Morris questionnaire for German-speaking patients with low back pain. Spine (Phila Pa 1976). 1999; 24: 1099-1103.

20. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine (Phila Pa 1976). 2000; 25: 3115-3124.

21. Costa LO, Maher CG, Latimer J, Hodges PW, Herbert RD, Refshauge KM, et al. Motor control exercises for chronic low back pain: a randomized placebo-controlled trial. Physical Therapy Journal. 2009; 89: 1275-1286.

22. Brennan KL, Allen BC, Maldonado YM. Dry Needling Versus Cortisone Injection in the Treatment of Greater Trochanteric Pain Syndrome: A Noninferiority Randomized Clinical Trial. J Orthop Sports Phys Ther. 2007; 47: 232-239.

23. Ferreira G, Stieven F, Araujo F, Wiebusch M, Rosa C, Plentz R, et al. Neurodynamic treatment did not improve pain and disability at two weeks in patients with chronic nerve-related leg pain: a randomised trial. Journal of Physiotherapy. 2016; 62: 197-202.

24. Kyte DG, Calvert M, van der Wees PJ, ten Hove R, Tolan S, Hill JC. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. Physiotherapy. 2015; 101: 119-125.

25. Hall MA, Maher CG, Latimer J, Ferreira ML, Costa LOP. 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.

26. Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, De Vet HCW, Hancock MJ. Global perceived effect scale provides reliable assessments of health transition in people with musculuskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol. 2010; 63: 760-766.

27. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000; 53: 459-468.

28. Thoomes-de Graaf M, Fernandez-De-Las-Penas C, Cleland JA. The content and construct validity of the modified patient specific functional scale (PSFS 2.0) in individuals with neck pain. Journal of Manual & Manipulative Therapy. 2020; 28: 49-59.

29. Heldmann P, Schöttker-Köninger T, Schäfer A. Cross-cultural Adaption and Validity of the "Patient Specific Functional Scale. International Journal of Health Professionals. 2015; 2: 73-82.

30. Abbott JH, Schmitt JS. Minimum important difference for the patient-specific functional sclae, 4 region-specific outcome measures, and the numeric pain rating scale. J Orthop Sports Phys Ther. 2014; 44: 560-564.

31. Stratford P, Kennedy DM, Wainwright AV. Assessing the Patient-Specific Functional Scale's Ability to Detect Early Recovery Following Total Knee Arthroplasty. Phys Ther. 2014; 94: 838-844.

32. Mannberg-Bäckman S, Stråt S, Ahlström S, Brodin N. Validity and sensitivity to change of the Patient Specific Functional Scale used during rehabilitation following proximal humeral fracture. Disabil Rehabil. 2016; 35: 487-492.

33. Resnik L, Borgia M. Responsiveness of outcome measures for upper limb prosthetic rehabilitation. Prosthet Orthot Int. 2016; 40: 96-108.

34. Hinami K, Alkhalil A, Chouksey S, Chua J, Trick WE. Clinical significance of physical symptom severity in standardized assessments of patient reported outcomes. Qual Life Res. 2016; 25: 2239-2243.