ELSEVIER

The Spine Journal ■ (2013) ■

THE SPINE JOURNAL

Clinical Study

The Spine Functional Index: development and clinimetric validation of a new whole-spine functional outcome measure

Charles P. Gabel, MPhD, PT[a,*], Markus Melloh, MD, DMedSc, MPH, MBA[b], Brendan Burkett, PhD[a], Lori A. Michener, PhD, PT, ATC[c]

[a]*Faculty of Science, Health and Education, Centre for Healthy Activities, Sport and Exercise, University of the Sunshine Coast, Sippy Downs Dve, Sippy Downs, Sunshine Coast Qld, 4556 Australia*

[b]*Western Australian Centre for Medical Research, Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Perth, Australia*

[c]*Department of Physical Therapy, Virginia Commonwealth University, 821 W Franklin St, Richmond, VA 23284, USA*

Received 14 March 2012; revised 11 July 2013; accepted 28 September 2013

Abstract

BACKGROUND CONTEXT: Most spine patient-reported outcome measures are divided into neck and back subregions. This prevents their use in the assessment of the whole spine. By contrast, whole-spine patient-reported outcome measures assess the spine from cervical to lumbar as a single kinetic chain. However, existing whole-spine patient-reported outcomes have been critiqued for clinimetric limitations, including concerns with practicality.

PURPOSE: To develop the Spine Functional Index (SFI) as a new whole-spine patient-reported outcome measure that addressed the limitations of existing whole-spine questionnaires; and to determine the SFI's clinimetric and practical characteristics concurrently with a recognized criterion, the Functional Rating Index (FRI).

STUDY DESIGN: Observational cohort study within 10 physical therapy outpatient clinics.

PATIENT SAMPLE: Spine-injured patients were recruited from a convenience sample referred by a medical practitioner to physical therapy. A pilot study (n=52, 57% female, age 47.6±17.5 years) followed by the main study (n=203, 48% female, age 41.0±17.8 years) that had an average symptom duration of less than 5 weeks.

OUTCOME MEASURES: Spine Functional Index, FRI, and Numerical Rating Scale (NRS).

METHODS: The SFI was developed through three stages: 1) item generation, 2) item reduction with an expert panel and patient focus group, and 3) pilot field testing to provide provisional clinimetric properties and sample size requirements and to determine suitability for a larger study. Participants completed the SFI, FRI, and NRS every 2 weeks for 6 weeks, then every 4 weeks until discharge or study completion at 6 months. Responses were assessed to provide individual psychometric and practical characteristics for both patient-reported outcomes, with the overall performance evaluated by the Measurement of Outcome Measures and Bot clinimetric assessment scales.

RESULTS: The SFI demonstrated a high criterion validity with the FRI (Pearson r=0.87, 95% confidence interval [CI]), equivalent internal consistency (α=0.91), and a single-factor structure. The SFI and FRI demonstrated suitable reliability (intraclass correlation coefficient$_{2,1}$=0.97:0.95), responsiveness (standardized response mean=1.81:1.68), minimal detectable change with 90% CI (6.4%:9.7%), Flesch scale reading ease (64%:47%), and user errors (1.5%:5.3%). The clinimetric performance was higher for the SFI on the Measurement of Outcome Measures (96%:64%) and on the Bot scale (100%:75%).

CONCLUSIONS: The SFI demonstrated sound clinimetric properties with lower response errors, efficient completion and scoring, and improved responsiveness and overall clinimetric performance compared with the FRI. These results indicated that the SFI was suitable for functional outcome

FDA device/drug status: Not applicable.

Author disclosures: ***CPG:*** Nothing to disclose. ***MM:*** Nothing to disclose. ***BB:*** Nothing to disclose. ***LAM:*** Nothing to disclose.

* Corresponding author. PO Box 760, Coolum Beach, Queensland 4573, Australia. Tel.: (61) 7-5446-1022; fax: (61) 7-5471-7022.

E-mail address: cp.gabel@bigpond.com (C.P. Gabel)

1529-9430/$ - see front matter
http://dx.doi.org/10.1016/j.spinee.2013.09.055

measurement of the whole spine in both the research and clinical settings.

Keywords: Outcome; Clinimetrics; Spine; Back; Neck; Measurement; Kinetic-chain

Introduction

Patients with pain or symptoms that arise from the spine may be evaluated with patient-reported outcome measures to determine their functional status [1–3]. These patient-reported outcome measures can be regional, designed to assess a region of the body or it can be specific to a single joint, condition, or disease. When assessing the functional status of patients with musculoskeletal conditions of the upper or lower limbs, a regional patient-reported outcome measure may be preferred, as practicality is improved without compromising the essential psychometrics properties [4,5]. However, when assessing the spine, patient-reported outcome measures remain distinctly divided into back [2] and neck [6]. Few whole-spine patient-reported outcome measures are recommended because of documented problems with either or both psychometric and practical characteristics [2]. Another measurement option is a generic patient-reported outcome, such as the Short Form 36 health survey or the EuroQol. These generic patient-reported outcomes can be applied to all types of patients, regardless of their diagnosis or health problem [1]. However, these generic patient-reported outcomes have demonstrated reduced responsiveness over time because they do not contain sufficient items that are specific to the region, joint, condition, or disease being assessed [7]. Consequently, these generic tools are less suited to measure regional musculoskeletal conditions [4,5], including spine related conditions for both the back [4] and neck [8].

The adoption of the single kinetic chain concept for whole-spine patient-reported outcomes was first proposed by Williams et al. [9]. Justifications supporting this concept included pathophysiological grounds, as the etiology for many mechanical nonspecific spinal problems remains unknown; coexisting regions, as presenting symptoms often occur in multiple, interconnected spinal areas; and improved practicality, as one tool would provide measurement for all spinal areas [5,10]. It has been recommended that a whole-spine patient-reported outcome be developed, particularly one that demonstrates acceptable clinimetric properties and performance, and subsequently compared with specific subregion spine patient-reported outcomes for the back and neck [9–11]. The development and validation of a new whole-spine patient-reported outcome requires two phases: 1) initial development and evaluation of clinimetrics that includes concurrent validation with an existing whole-spine patient-reported outcome and 2) subsequent concurrent validation with advocated criteria in separate subregions and condition-specific back and neck populations. This study's purpose was Phase 1.

There are at least 43 back-specific patient-reported outcomes with 13 that can be used to evaluate responsiveness to change [2]. Among these, the Oswestry Disability Index and Roland Morris Disability Questionnaire are the most commonly advocated [2,12]. For the neck, at least 13 patient-reported outcomes have been developed [13], but there is limited agreement on which ones should be advocated [6,8]. Five patient-reported outcomes purport validity for the whole-spine: the Functional Rating Index (FRI) [10], the Bournemouth Questionnaire [14], the Extended Aberdeen Spine Pain Scales [9], the Pain Disability Questionnaire [12], and the Core Outcome Measures Index [3]. However, further testing is required of these whole-spine tools because none have demonstrated an adequate factor structure through either Rasch analysis or factorial analysis [2] and the capacity to measure the whole-spine as a single kinetic chain [15]. Of these five patient-reported outcomes, the FRI is advocated most strongly because of its preferred administrative practicality and level of independent research on comparative clinimetric properties for both low back [16] and neck pain [8]. Consequently, the FRI is the optimal choice as a criterion measure ahead of the other four available whole-spine patient-reported outcomes when developing a new whole-spine patient-reported outcome and in preference to generic patient-reported outcomes such as the Short Form 36 or EuroQol.

The development of each of these five whole-spine tools has attempted to address the need for a single whole-spine tool. The initial three were questioned because of poor methodology in development, practicality, factor analysis, and validation [2]. For example, the Pain Disability Questionnaire is not spine specific, nor does it account for acute situations because it is for ‘‘chronic disabling musculoskeletal disorders’’ [12]. The 11-item Core Outcome Measures Index has separate neck and back versions and is designed to measure patients after operative procedures within secondary and tertiary settings. Completion involves several scoring techniques with computerized input [3], and independent validation as a whole-spine measure is still required. Both the Aberdeen [9] and Pain Disability Questionnaire [12] have dual-factor structures that limit their validity as a single summated score and consequently, are less than optimal measure [15]. The remaining three patient-reported outcomes have even less research in this aspect because they have not had their factor structure determined by the recommended maximum likelihood extraction method [17]. Consequently, a whole-spine patient-reported outcome is needed that has been appropriately developed [18], represents a single kinetic chain, has a single factor structure, and appropriate clinimetric properties for both the back and neck.

A patient-reported outcome must be clinically practical, effective, efficient, and validated with a recognized criterion standard [19]. The Spine Functional Index (SFI) (Fig. 1) was developed to comply and satisfy these requirements. The aim of this study was to describe the development of the SFI; determine the psychometric, practical, and factor structure characteristics in a general spinal population; and compare the SFI with a whole-spine criterion measure, the FRI [10].

Materials and methods

A prospective observational study was completed in two phases (Fig. 2):

1. SFI development in three stages
2. SFI validation in a symptomatic spine cohort

Phase 1: development of the Spine Functional Index

The established three-stage development process used item generation, item reduction, and field testing [15,18] (Fig. 2).

Stage 1: item generation

Electronic databases, PubMed, Cinahl, Embase, and Pedro, from 1980 to 2010 were searched by the primary author (CPG) with key words "outcomes," "self-report," "function," "disability," "impairment," "spine," "neck," "back," "thoracic," "cervical," and "lumbar." An additional search included clinicians and researchers for unpublished questionnaires. This produced 129 patient-reported outcomes. A four-person peer-panel was formed, comprising an occupational therapist, physical therapist with

SPINE FUNCTIONAL INDEX (SFI) **DATE:** ______________

NAME: ____________________ **INJURY:** _______________ ☐ **Neck** ☐ **Mid Back** ☐ **Lower Back**

PLEASE COMPLETE: Your spine may make it difficult to do some things you normally do. This list contains sentences people use to describe themselves with such problems. Think of yourself over the last few days. **If an item describes you, mark the box. If not, leave the box blank. If an item partly describes you, Use a Half (½) Mark.**

DUE TO MY SPINE:

___ 1. I stay at home most of the time.
___ 2. I change position frequently for comfort.
___ 3. I avoid heavy jobs (e.g. cleaning, lifting more than 5kg or 10lbs, gardening, etc).
___ 4. I rest more often.
___ 5. I get others to do things for me.
___ 6. I have the pain / problem almost all the time.
___ 7. I have difficulty lifting and carrying (e.g. bags, shopping up to 5kg or 10lbs).
___ 8. My appetite is now different.

___ 9. My walking or normal recreation or sporting activity is affected.
___ 10. I have difficulty with normal home or family duties and chores.
___ 11. I sleep less well.
___ 12. I need assistance with personal care (e.g. washing and hygiene).
___ 13. My regular daily activities (work, social contacts) are affected.
___ 14. I am more irritable and / or bad tempered.
___ 15. I feel weaker and / or stiffer.
___ 16. My transport independence is affected (driving, public transport).

___ 17. I require assistance or am slower with dressing.
___ 18. I have difficulty moving in bed.
___ 19. I have difficulty concentrating and / or reading.
___ 20. My sitting is affected.
___ 21. I have difficulty getting in and out of chairs.
___ 22. I only stand for short periods of time.
___ 23. I have difficulty squatting and / or kneeling down.
___ 24. I have trouble reaching down (e.g. pick-up things, put on socks).
___ 25. I go up stairs slower or use a rail.

SFI SCORE: To score the upper part - add the marked boxes:

☐ **TOTAL (SFI points)** **100 Scale: 100 – (TOTAL x 4) =** ☐ **%**

MDC (90% CI): Neck = 6.9% or 1.7 SFI points; **Mid and Lower Back** = 5.9% or 1.5 SFI points;

All spine = 6.5 % or 1.6 SFI points Change less than this may be due to error.

Fig. 1. Spine Functional Index. MDC, minimal detectable change; CI, confidence interval.

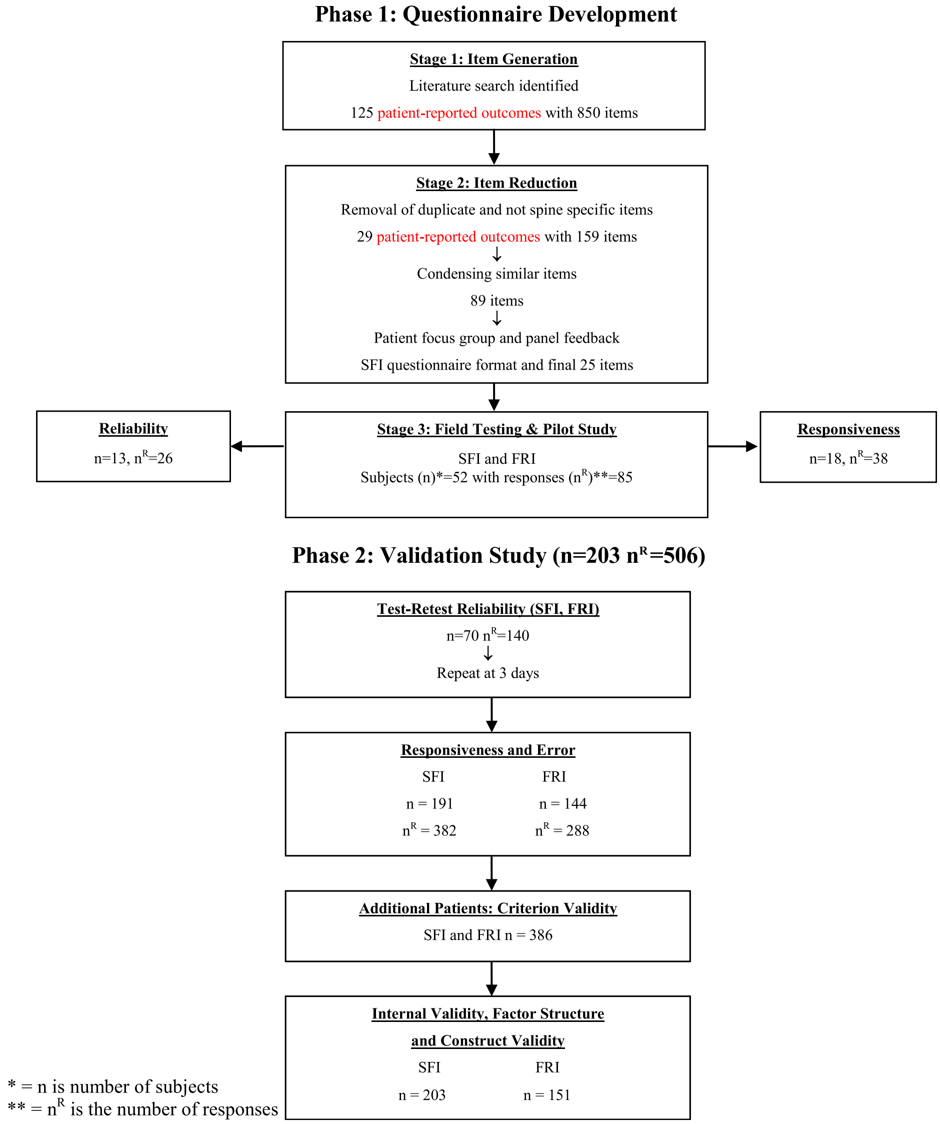

Fig. 2. Flow chart of SFI development and validation. SFI, Spine Functional Index; FRI, Functional Rating Index.

spine-specific postgraduate qualifications, general practitioner physician, and occupational medicine physician with spine-specific consultancy work. The panel used consensus opinion that required a three-vote minimum [20,21] to review and shorten the list to 29 patient-reported outcome tools, with 850 items that were directly cited in each of the patient-reported outcomes and relevant to the spine injuries. The list was reduced to 409 items by the panel through binning and winnowing methodology that removed duplicate and nonapplicable items [22,23].

Stage 2: item reduction

The 409 items were reduced in five separate stages (2a–e) by the panel. Stage 2a reduced the list to 159 items by pooling items with a common construct (eg, ‘‘sitting,’’ ‘‘sit in a chair,’’ ‘‘sit on a stool,’’ and so on, were collapsed to ‘‘sitting’’). Stage 2b classified [18] items using the World Health Organisation-International Classification of Functioning (WHO-ICF) [24] codes from the ICF Browser [25]: b=body functions, s=body structures, d=activities and participation, and e=environmental factors [26]. Stage 2c reduced the 159 items to 89 by combining the ICF codes to common descriptive construct titles (eg, ‘‘stairs’’ and ‘‘ladders’’ became ‘‘code d4551-climbing’’). Stage 2d reduced the list to 74 by grouping and deletion (eg, ‘‘dressing’’ and ‘‘putting on pants’’ were retained but ‘‘fastening clothing’’ was deleted). Stage 2e further combined items via consensus of importance and relevance to achieve the final 25 items, 15 general and 10 spine-specific. The stems for each question were formulated: ‘‘Due to my spine: I have difficulty/problems…;’’ or ‘‘I stay/change/avoid/get others...’’

To ensure that current best practice epidemiologic standards were met, each question’s final wording was achieved through peer panel consensus, then given to two focus groups for feedback and relevance for face and content validity [18]: a spine symptoms patient focus group (n=10, three cervical, three thoracic, and four lumbar) and the four-person author group that included a physical therapist and an orthopedic surgeon, both with extensive experience in the spine, a biomechanist, and a physical therapist with extensive clinimetric research experience. The 10-person patient focus group and the 4-person author panel supplemented the initial item reduction process performed by the "expert panel." The focus groups were provided with the final 25 items list and the list of the 49 items excluded in Stage 2d. The mixed methods, semistructured interview process [27] was used to determine if the 25 items should be changed and if any of the 49 excluded items should be reinstated or included within the final item list. The "Isikawa" qualitative methodological process [28] was used to supplement the consensus agreement from both the patient and author focus groups and the expert panel. The format and three-item response option, "Yes," "No," and "Half" [15,29], were selected.

Stage 3: field testing

A pilot investigation enrolled 52 participants who provided a total of 85 responses (n^R). This ensured n=52 baseline responses and an additional 33 responses: 13 for reliability (n=13; n^R=26) and 20 for responsiveness, where two participants completed an additional third set of responses (n=18; n^R=38) (Fig. 2). This allowed for a preliminary assessment of floor and ceiling effects, sampling method practicality, and sample size calculations.

Sample size

From the pilot study, minimum samples were determined for an 80% chance of detecting actual difference with 15% attrition (p<.05) [30]. This compared favorably with previous FRI investigations [10,31] for concurrent validity (n=106), reliability (n=56), responsiveness (n=84), and predictive ability through construct validity (n=168).

Phase 2: validation of the SFI in a cohort population

Design

A single stage, prospective observational study analyzed concurrent SFI and FRI responses. Each participant was classified by subregion (cervical, thoracic, or lumbar), where the percentage noted ensured proportional reliability and responsiveness representation [15,18].

Setting and participants

Participants who complained of spinal pain or symptoms (n=203, responses=506) were consecutively recruited from 10 Australian physical therapy clinics. Inclusion criteria were referral by a medical practitioner for musculoskeletal spine condition or symptoms. Exclusion criteria were pregnancy, red flag signs, younger than 18 years, and English language difficulty. Symptoms and classifications of spinal diagnoses represent the entire spinal region, as described in Table 1.

Participants completed both the SFI and FRI patient-reported outcomes, however, the number of FRI responses (n=173; responses=386) was reduced because of a misunderstanding with one participating clinic that returned only the SFI responses. Participants receiving ongoing treatment were remeasured every 2 weeks for 6 weeks, then every 4 weeks until discharge. Status was classified as acute at 0 to 6 weeks, subacute at 6 to 12 weeks, and chronic beyond 12 weeks. Pooled responses assessed criterion validity, distribution, and missing responses. Participants also completed an 11-point global numerical rating scale (NRS) of perceived present overall status [32,33], where subjects rate their status on a scale from 0 to 10 (0=worst possible, 10=normal). The global NRS was used as an external criterion measure of clinical change, by calculating the difference in global perceived present status over time.

Questionnaires

The FRI [10] is a single page patient-reported outcome that contains 10 items, each rated on a five-point Likert scale incorporating visual and descriptive response options. Five items on the FRI are common to the Oswestry Disability Index and the Neck Disability Index, with three additional Oswestry Disability Index items, one Neck Disability Index item, and a new "pain" item [2]. The raw score of the FRI is multiplied by 2.5 to generate a 0% to 100% score on the FRI (100%=no disability). One missing response is permitted.

The SFI is a single page 25-item patient-reported outcome, with a three-point Likert scale response option for each item. The scores from the 25 items are tallied for the sum, the sum is multiplied by four and then subtracted from 100 to generate a 0% to 100% score (100%=no disability). Two missing responses are permitted.

An 11-point global NRS (0=worst possible, 10=normal or fully recovered) was used to reflect the individual perceived global functional status and act as an external criterion.

Data analysis: psychometric characteristics

Distribution and normality were assessed from the baseline histogram inspection and one-sample Kolmogorov-Smirnov tests (significance >0.05) [30]. *Internal consistency* used baseline Cronbach alpha (α=0–1.00) calculations with an optimal value recommended as 0.90 to 0.95 [18,30]. *Test-retest reliability* was assessed through the intraclass correlation coefficients Type 2,1 and expressed with 95% CI using scores on the patient-reported outcome from acute/subacute participants at baseline and again on Day 3 during a nontreatment period. Participants rating on the global NRS of perceived overall status at baseline and on Day 3 provided the reference criterion to determine change. Only those participants who had a change of 0 to ±1 were entered into analysis for test-retest reliability (n=70) [15].

Responsiveness was assessed using the effect size and the standardized response mean statistics [18]. Participants

Table 1
Participant demographics for SFI: Stage 1, pilot and Stage 2, validation

	Stage 1	Stage 2
Demographic data	Development and pilot validation	Validation (main study)
Participants (n)	52	SFI=203, FRI=173
Responses (n)	85	SFI=506, FRI=386
Age (y)	47.6±17.5	41.0±17.8
Gender (% female)	57.1%	48.0%
Injury		
Duration (wks)	71.5±103.0	4.6±8.4
Time range (wks)	1–300 (1 outlier=1,575 wks)	1–45 (1 outlier=520 wks)
Subregion and diagnosis*		
Cervical	n=24 (46%)	n=96 (47%)
Whiplash		32%
Joint		29%
Disc		7%
Soft tissue		10%
Nerve root and neural		3%
Other (nonspecific, pain, and so on)		19%
		100%
Thoracic	n=4 (8%)	n=48 (24%)
Joint		29%
Disc		6%
Soft tissue		23%
Fracture		2%
Costovertebral		2%
Other (nonspecific, pain, and so on)		38%
		100%
Lumbar	n=24 (46%)	n=101 (50%)
Joint		30.5%
Disc		29.5%
Soft tissue		5%
Nerve root or neural		1%
Postoperative		3%
Sacropelvic		3%
Osteoarthritis		4%
Other (nonspecific, pain, etc.)		24%
		100%
Multiarea	Nil	n=46 (23%)

SFI, Spine Functional Index; FRI, Functional Rating Index.
* Subregion % values include multiarea individuals within each of their symptomatic regions making the total >100%.

were classified by subregion with repeated measures analyzed (n=191 for the SFI; n=144 for the FRI) for acute at 2 weeks, subacute at 4 weeks, and chronic at 6 weeks. This accounted for variations in healing and therapists interventions [15]. There were participants who received no follow-up or early discharge (SFI, n=12; FRI, n=7). The global NRS score of a change of 2.0 or more was the cutoff used to define patient-rated clinical change. *Error score* was determined with the minimal detectable change (MDC) with 90% CI (MDC_{90}) using the standard error of the measurement formula and the intraclass correlation coefficients. *Minimal clinically important difference (MCID)* was calculated using an anchor-based method, with the anchor of patient-rated change determined from the global numeric rating of change. Patients were classified as improved or deteriorated if they had a minimum change of 2.0 or more points on the global NRS between baseline and follow-up [18,33,34]. Consequently, the MDC appears as a statistically and clinically appropriate MCID [35].

Validity was assessed for *face* and *content* through focus groups, panel feedback, and readability scores [36]; and for *criterion* through Pearson r coefficient (n=386). *Construct validity* used discriminant validity with the external criterion global NRS of perceived self-rated change of health status of 2.0 or more points [34]. Additionally, an a priori paired *t* test statistical difference was required between baseline and repeated test groups mean scores to categorize subjects as improved or deteriorated when calculating the MCID. *Factor analysis* used baseline SFI and FRI data with loading suppression at 0.30 and varimax rotation for maximum likelihood extraction [30], which required assumptions of normality. Factor extraction had three a priori requirements: scree-plot "point of inflection;" eigenvalue of more than1.0; and variance of 10% or more [30].

Data analysis: practical characteristics, readability, and summary performance

Practicality considered nine areas [15,36], with five being self-evident: 1) self-administered; applicable across a variety of 2) conditions; 3) severity levels; 4) relevance to defined populations; and 5) single-page length. The remaining four areas were determined through focus groups for interviews for ease of understanding and completion; questionnaire completion time; therapist scoring time from three separate scores averaged from each clinic; and missing responses as percentages of total responses (SFI, n=506; FRI, n=386). *Readability* used the Flesch-Kincaid grade scales (range, 0–12, optimum <7) and reading ease (optimum >60%) calculated from word-processing software. Summary performance used the "Measurement of Outcome Measures" scale that evaluated 25 essential properties [5]; and the "Bot" scale that evaluated 12 items [36]. The "Bot" cutoff classifications were adjusted [15,29] for "time to administer" at 3 minutes and "readability and comprehension" determined by the Flesch-Kincaid scale cutoffs [15]. Significance was set at $p<.05$.

Results

Participant demographics are reported in Table 1.

Psychometric properties

Characteristics of internal consistency, reliability, responsiveness, and error score are summarized in Table 2 and construct validity in Table 3.

Distribution and normality were demonstrated through the Kolmogorov-Smirnov test (SFI=1.163, significance =0.87; FRI=1.18, significance=0.87) with identical SFI

Table 2
Methodological characteristics of SFI and FRI criteria

	Reliability (ICC)	Internal consistency	Error score		Responsiveness			Missing responses
Stage 2	Rxx	Alpha	SEM	MDC_{90}	SD_{100}	ES	SRM	Percentage
SFI	0.972	0.911	2.76	6.44	24.80	1.25	1.81	1.5%
FRI	0.948	0.908	4.14	9.66	22.67	1.23	1.68	5.3%

Rxx, test-retest reliability coefficient; alpha, Cronbach alpha; SFI, Spine Functional Index; FRI, Functional Rating Index; ICC, intraclass correlation coefficient; SEM, standard error of the measurement; MDC_{90}, minimal detectable change (90% CI); SD_{100}, standard deviation at baseline (100% scale); ES, effect size; SRM, standard response mean.

and FRI baseline score ranges (0%–98%). The SFI histogram shape was preferred particularly in the upper 90% to 100% interval that contained 15 (7.5%) responses compared with the FRI with a single response (2%). The ‘‘half mark’’ option was used by 57% of participants at baseline and in 43% of all responses. The baseline scores by subregion were comparable between the SFI and FRI apart from the multiarea group (Table 4).

Criterion validity was high (Pearson r=0.85) between the SFI and FRI scores. *Construct validity* through *discriminant validity* was demonstrated for the a priori criterion (Table 3). The subregion mean scores were different for both patient-reported outcomes and between both patient-reported outcomes, though the cervical, thoracic, and multiarea groups were of a similar value. However, none were statistically significant apart from the multiarea group (p<.001).

Factor analysis was suitable, as the correlation matrix Kaiser-Meyer-Oklin value was 0.912 and Bartlett test of sphericity significant (p<.001). A unidimensional structure was indicated for both patient-reported outcomes as the three a priori criteria were met with second point screeplot inflection and one eigenvalue of more than 1.0, where variance was greater than 10% (SFI=33.4%, FRI=55.6%). The SFI had six more factors with eigenvalues of more than 1.0 but with variance of lower than 10% that accounted for 30.5%. Both patient-reported outcomes had four factors with eigenvalues between 0.5 and 1.0, with the remaining factors all below a 0.5 eigenvalue.

Practical characteristics

Completion time was SFI=122±37 seconds and FRI=84±23 seconds; scoring time was SFI=16±4 seconds and FRI=27±13 seconds. The FRI required a computational aid, and with one missing response the scoring time increased to 53±19 seconds. Combined completion and scoring was SFI=138±41 seconds and FRI=137±39 seconds. *Missing responses* were less than 1.5% for the SFI and 5.3% for the FRI. *Readability* for the SFI was grade=7, reading ease=64% and for the FRI grade=7, reading ease=47.2%. *Summary performance* on the Measurement of Outcome Measures was SFI=96%, FRI=64% and on the ‘‘Bot’’ for the SFI=12/12 or 100%, FRI=9/12 or 75%.

Table 3
Construct validity comparing means for baseline, repeated scores, and difference for the SFI and FRI (n=113)

Tools	Baseline mean	Repeat test mean*	Difference in means
SFI	43.6±24.8	17.3±19.1	31.0±17.1
FRI	43.0±22.7	17.1±18.6	28.0±16.7

SFI, Spine Functional Index; FRI, Functional Rating Index.
* Repeated measures were made after a period of known natural healing: acute after 2 weeks; subacute after 4 weeks; and chronic after 12 weeks, all with a p value <.0001 for the *t*-statistic measures.

Discussion

The SFI was developed using a structured methodology. It demonstrated acceptable psychometric properties, a single factor structure, and strong practical characteristics in patients with spinal pain and symptoms of the cervical, thoracic, and lumbar spine. Compared with the FRI, by visual comparison of the results, the SFI had equal or preferable psychometric properties of reliability, validity, responsiveness, and error. The summary performance scores of practical characteristics on the Measurement of Outcomes Measure and Bot scales showed high scores for the SFI. The SFI was demonstrated to be capable of assessing functional status at a single point in time and change over time to determine the effectiveness of treatment interventions. The practical characteristics of short scoring times, low missed responses, and reading ease will reduce both the patient and administrative burden.

The SFI has a three-point response format that was used by participants 57% of the time. This response format provided a simple scoring format within a stable equally spaced scale [37]. This also enabled sound individual interpretation for the psychological perspective of an item's ‘‘presence,’’ ‘‘absence,’’ or an ‘‘intermediate position’’ [38] as opposed to a dichotomous response option.

Normalized SFI distribution and subregion scores in this cohort of patients presenting to physical therapy demonstrated no floor or ceiling tendency. The FRI had more missing responses at the higher levels of functional loss, indicating reduced measurement capacity. This measurement capacity of the SFI may improve the ability to discriminate change throughout the scale range. Internal consistency, test-retest reliability, and responsiveness values for the SFI were acceptable and comparable with the FRI. The SFI demonstrated lower error values (standard error of the measurement and MDC_{90}) that may allow for improved

Table 4
Mean baseline scores by subregion for SFI and FRI

	Stage 2	
Subregions	SFI	FRI
Cervical	41.1±25.9 n=96 (47%), with Cx only n=78 (38%), Cx, Tx n=16, Cx, Lx n=2	39.2±21.6 n=67 (39%), with Cx only n=53 (31%), Cx, Tx n=13, Cx, Lx n=2
Thoracic	41.2±25.8 n=48 (24%), with Tx only n=8 (4%), Cx, Tx n=16, Tx, Lx n=24	38.2±23.4 n=38 (22%), with Tx only n=4 (2%), Cx, Tx n=13, Tx, Lx n=21
Lumbar	46.8±26.1 n=101 (50%) with Lx only n=75 (37%), Cx, Lx n=2, Tx, Lx n=24	45.0±21.4 n=90 (52%) with Lx only n=67 (39%), Cx, Lx n=2, Tx, Lx n=21
Multiarea	41.0±26.8 n=42 (21%) with Cx, Tx n=16, Cx, Lx n=2, Tx, Lx n=24	34.0±20.3 n=36 (21%) with Cx, Tx n=13, Cx, Lx n=2, Tx, Lx n=21
All, average for all data	44.8±26.1 n=203	41.4±17.2 n=173

Cx, cervical; Tx, thoracic; Lx, lumbar; SFI, Spine Functional Index; FRI, Functional Rating Index.
Note: Multiarea included participants who reported symptoms in more than one area.

sensitivity for detecting change over time in the assessment of intervention effectiveness that may otherwise not show a valid effect [39]. Moreover, this may subsequently reduce the number needed to treat [40].

Responsiveness of the SFI in a cohort of patients undergoing physical therapy treatment was acceptable and comparable with the FRI despite the higher diversity in baseline impairment [41,42]. As an observational study in a cohort of patients undergoing physical therapy care, other influences on responsiveness may have been present. These include variation in interventions provided, follow-up duration (as responsiveness is less over a shorter follow-up period), and baseline severity (as acute and chronic patients change at different rates) [34]. These variables were attempted to be minimized by using the concurrent testing methodology. Factor analysis demonstrated a single-factor structure and consistent variance levels for both the SFI and FRI. This study is the first to report the FRI factor structure.

Limitations and strengths of the study

One limitation of this study was the recruitment of patients presenting for care at physical therapy outpatient clinics only. Consequently, results cannot be generalized to inpatient or community settings. Patients referred to physical therapy most likely represent the midrange of spine conditions. The study's strength was the prospective, multicenter investigation that included patients from each spinal region with varied degrees of severity and duration that represented both the general and work-injured populations with a large variation in diagnoses (Table 1). Furthermore, 191 subjects were available for the responsiveness sample, measuring these subjects on repeated occasions over time. This facilitated their measurement throughout the severity spectrum as indicated by the suitable levels of distribution within the histogram, including the least affected level at the point of discharge.

Implications for further research

The high SFI and FRI criterion validity implied generalizability to populations where the FRI has been validated or compared with other spine related patient-reported outcomes. 'This includes the Oswestry Disability Index, Roland Morris Disability Questionnaire, and Neck Disability Index. However, independent investigations are required where spine subregion patient-reported outcomes are concurrently compared through repeated measures on diagnoses, such as whiplash, acute, and chronic low back and neck pain. The SFI had several factors that accounted for substantial variance. This suggests that shortening to perhaps 10 items may be possible. This may improve practicality and reduce both respondent and clinician burden. A confirmatory factor analysis should be considered.

Conclusions

The SFI is a practical patient-reported outcome for measurement of spine-related patient status and change over time. Compared with the FRI, an advocated whole-spine patient-reported outcome, the SFI had comparable and sometimes improved psychometric and practical characteristics and overall performance. The findings of this study indicated the SFI is a viable patient-reported outcome for measuring whole-spine functional status in both the clinical and research settings.

Acknowledgments

Research support was provided by the University of the Sunshine Coast. We would like to thank all participating patients, general practitioners, and therapists for their time and effort. The study was approved by the Human Research Ethics Committee of the University of the Sunshine Coast.

References

[1] Garratt A. Patient reported outcome measures in trials, Editorial. BMJ 2009;338:2597.

[2] Cleland JA, Gillani R, Bienen EJ, Sadosky A. Assessing dimensionality and responsiveness of outcomes measures for patients with low back pain. Pain Pract 2011;11:57–69.

[3] Mannion AF, Porchet F, Lattig F, et al. The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 2009;18:374–9.

[4] Garratt AM, Klaber MJ, Farrin AJ. Responsiveness of generic and specific measures of health outcome in low back pain. Spine 2001;26:71–7.

[5] Gabel CP, Michener L, Burkett B, Neller A. The Upper Limb Functional Index (ULFI): development and determination of reliability, validity and responsiveness. J Hand Ther 2006;19:328–49.

[6] van der Velde G, Beaton D, Hogg-Johnston S, et al. Rasch analysis provides new insights into the measurement properties of the neck disability index. Arthritis Rheum 2009;61:544–51.

[7] Suarez-Almazor ME, Kendall C, Johnson JA, et al. Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. Rheumatology (Oxford) 2000;39:783–90.

[8] Rebbeck T, Sindhusake D, Cameron ID, et al. A prospective cohort study of health outcomes following whiplash associated disorders in an Australian population. Inj Prev 2006;12:93–8.

[9] Williams N, Wilkinson C, Russell IT. Extending the Aberdeen Back Pain Scale to include the whole spine: a set of outcome measures for the neck, upper and lower back. Pain 2001;94:261–74.

[10] Feise RJ, Menke JM. Functional Rating Index. A new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. Spine 2001;26:78–86.

[11] Gabel CP, Burkett B, Yelland M. Balancing fidelity and practicality in short version musculoskeletal outcome measures. Phys Ther Rev 2009;14:221–5.

[12] Anagnostis C, Gatchel RJ, Mayer TG. The pain disability questionnaire: a new psychometrically sound measure for chronic musculoskeletal disorders. Spine 2004;29:2290–302; discussion 2303.

[13] Resick D. Subjective outcome assessments for cervical spine pathology: a narrative review. J Chiro Med 2005;3:113–34.

[14] Bolton JE, Humphrey's BK. The Bournemouth Questionnaire: a short-form comprehensive outcome measure. II. Psychometric properties in neck pain patients. J Manipulative Physiol Ther 2002;25:141–8.

[15] Gabel CP, Michener L, Melloh M, Burkett B. Modification of the Upper Limb Functional Index to a three-point response improves clinimetric properties. J Hand Ther 2010;23:41–52.

[16] Chansirinukor W, Maher CG, Latimer J, Hush J. Comparison of the functional rating index and the 18-item Roland-Morris Disability Questionnaire: responsiveness and reliability. Spine 2005;30:141–5.

[17] Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods 1999;4:272–99.

[18] Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. Oxford: Oxford University Press, 2008.

[19] Liang MH, Jette AM. Measuring functional ability in chronic arthritis: a critical review. Arthritis Rheum 1981;24:80–6.

[20] Kirshner B, Guyatt GH. A methodological framework for assessing health indices. J Chronic Dis 1985;38:27–36.

[21] Kopec JA, Sayre EC, Davis AM, et al. Assessment of health-related quality of life in arthritis: conceptualization and development of five item banks using item response theory. Health Qual Life Outcomes 2006;4:33.

[22] Kopec JA. Measuring functional outcomes in persons with back pain: a review of back-specific questionnaires. Spine 2000;25:3110–4.

[23] Patient-Reported Outcome Measurement Information Systems (PROMIS) (2010) Version 1.0 Item Banks Volume. Available at: http://www.nihpromis.org/science/ItemClassification. Accessed August 20, 2011.

[24] World Health Organisation (WHO). International Classification of Functioning, disability and health (ICF). 2001. Available at: www.who.int/icidh. Accessed July 31, 2009.

[25] World Health Organisation. The ICF Browser. 2010. Available at: http://apps.who.int/classifications/icfbrowser/. Accessed January 12, 2010.

[26] Escorpizo R, Stucki G, Cieza A, et al. Creating an interface between the international classification of functioning, disability and health and physical therapist practice. Phys Ther 2010;90:1053–63.

[27] Johnson RB, Onwuegbuzie AJ, Turner LA. Toward a definition of mixed methods research. J Mix Methods Res 2007;1:112.

[28] Ishikawa K, Loftus JH. Introduction to quality control. Tokyo: 3A Corporation, 1990.

[29] Gabel CP, Melloh M, Burkett B. The Lower Limb Functional Index: development and validation of the clinimetric properties and practical characteristics. Phys Ther 2012;92:98–110.

[30] Field A. Discovering statistics using SPSS. 2nd ed. London: SAGE Publications Ltd, 2005.

[31] Childs JD, Piva SR. Psychometric properties of the functional rating index in patients with low back pain. Eur Spine J 2005;14:1008–12.

[32] Bowling A. Just one question: if one question works, why ask several? J Epidemiol Community Health 2005;59:342–5.

[33] Ostelo RW, Deyo R, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 2008;33:90–4.

[34] Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine 2005;30:1331–4.

[35] Copay A, Glassman SD, Subach BR, et al. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J 2008;8:968–74.

[36] Bot SD, Terwee CB, van der Windt DA, et al. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis 2004;63:335–41.

[37] Krosnick JA. The handbook of questionnaire design. New York: Oxford University Press, 1991.

[38] Albarracin D, Johnson BT, Zanna MP. The handbook of attitudes. Hillsdale, NJ: Erlbaum, 2005.

[39] Stratford PW, Riddle DL. Assessing sensitivity to change: choosing the appropriate change coefficient. Health Qual Life Outcomes 2005;3:23.

[40] Moore RA, Smugar SS, Wang H, et al. Numbers-needed-to-treat analyses—do timing, dropouts, and outcome matter? Pooled analysis of two randomized, placebo-controlled chronic low back pain trials. Pain 2010;151:592–7.

[41] Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum, 1988.

[42] Liang MH, Fossel AH, Larson MG. Comparison of five health status instruments for orthopaedic evaluation. Med Care 1990;28:632–42.