Mokhtarinia *et al.*
BMC Musculoskeletal Disorders (2025) 26:370
https://doi.org/10.1186/s12891-025-08604-0

BMC Musculoskeletal Disorders

RESEARCH

Open Access

A shortened 10-item Spine Functional Index: clinimetric properties indicate a reliable, responsive and practical measure

Hamid Reza Mokhtarinia[1,2*], Antonio Cuesta-Vargas[3], Almir Vieira Dibai-Filho[4], Markus Melloh[5] and Agnieszka Bejer[6,7]

Abstract

Background To assess the 10-item Spine Functional Index (SFI-10) clinimetric properties in a general musculoskeletal disorder (MSD) spine population. Ascertain the psychometric characteristics' consistency with the developmental study findings for structural and criterion validity, internal consistency, and floor/ceiling effect; establish the longitudinal characteristics for test–retest reliability, responsiveness, construct validity, and error scores; and clarify practical characteristics of readability, missing responses, and time/errors for completion/scoring related to administrative burden.

Methods A longitudinal study of deidentified spine MSD patients ($n = 1317$, 53.4% female, age = 18-91yrs, av = 49.5 ± 16.4yrs; neck = 36.5%, mid-back = 8.4%, low back = 56.0%, multi-site = 0.3%) who completed the SFI-10, the Patient Specific Functional Scale (PSFS), and Numerical Rating Scales for Global-function (G-NRS) and Pain (P-NRS). Structural validity used factor analysis, exploratory (EFA) and confirmatory (CFA), plus Rasch analysis. Criterion validity used Spearman's correlation coefficient (r) between the SFI-10 and criteria (PSFS, G-NRS and P-PRS) scores, and construct validity ($n = 91$, known-groups independent t-test). Internal consistency used Cronbach's alpha (α) and floor/ceiling effects were determined. Subgroups determined reliability ($n = 104$, intraclass correlation coefficient, $ICC_{2.1}$); error ($n = 171$) through the standard error of measurement (SEM) and minimum detectable change (MDC_{90}). Responsiveness ($n = 171$) was calculated using effect-size (ES), standard response mean (SRM), and area under the curve (AUC); and interpretability through the minimal clinically important difference (MCID). Practicality ($n = 16$) clarified missing responses, readability, and time/errors for completion/scoring.

Results The SFI-10's structural validity was unequivocally one-dimensional from EFA and verified by CFA with acceptable fit-indices (chi-square/df = 2.88, CFI = 0.981, TLI = 0.975, RMSEA = 0.061), and supported by Rasch analysis (PSR = 0.79, Infit = 0.678–1.216, Outfit = 0.604–1.279, Item-difficulties = -1,215–2.488). Criterion validity varied from high (G-NRS, r = 0.60) and moderate (PSFS, r = 0.43) to low-inverse (P-NRS, r = -0.24). Internal consistency was strong (α = 0.84) and no floor/ceiling effects were present. Reliability was excellent ($ICC_{2.1} = 0.97$), responsiveness substantial (ES = 1.54; SRM = 1.64; AUC = 0.89), and measurement error robust (SEM = 3.84; $MDC_{90} = 8.98\%$, MDIC = 10.4%), with construct validity confirmed ($p < 0.001$). Practicality showed no missing responses, completion/scoring errors < 1%, excellent readability (Grade = 5.1, Ease = 74.1%), short completion (39.2 ± 10.3 s) and scoring times (8.5 ± 1.8 s).

*Correspondence:
Hamid Reza Mokhtarinia
hrmokhtarinia@yahoo.com
Full list of author information is available at the end of the article

Conclusions The SFI-10 demonstrates sound measurement properties in a general physiotherapy outpatient MSD spine population for both psychometric and practical characteristics. Further investigation in culturally diverse settings that include both inpatients and community settings with whole-spine and regional-spine criteria is required.

Keywords Spine, Musculoskeletal, Questionnaires, Psychometrics, Practicality, SFI- 10

Background

Musculoskeletal disorders (MSD) affecting the spine have significant repercussions for the global adult population. Functional loss, disability, and reduced social participation cause subsequent increases in global demands on the health and welfare support systems of the individual, society, and governments [1]. Measuring function and disability is critical for clinicians and epidemiological researchers who monitor the course of any spine-patient's condition and assess the effectiveness of the interventions employed [2]. Currently, disagreement remains between the advocated regional-spine static-patient-reported outcome measures (PROMs) and multi-region conditions [3, 4]. Further, consensus is lacking on which instruments be utilized for specific diagnoses or procedures, which has led to inconsistency in measurement approaches and a reduced capacity to amalgamate and analyze collected data [2]. To overcome these issues, the whole-spine static-PROM was initiated some 25 years ago with the concept of a continuous functional kinetic-chain being applied [5, 6]. More recently several dynamic-PROMs have become available that are computer and internet based using Item Response Theory (IRT) and Computerized Adaptive Testing (CAT), e.g. as PROMIS-PF (https://www.healthmeasures.net) or Classic Test Theory (CTT) and algorithm driven Computerized Decision Support Systems (CDSS), e.g. Advise Rehab (https://www.adviserehab.com). To date, four whole-spine static-PROMs have been proposed, the Extended Aberdeen Spine Pain Scale (EASPS) [6], Functional Rating Index (FRI) [7], Spine Functional Index (SFI- 25) [4], and the SFI- 10 [8], the SFI- 25's shortened 10-item version. A whole-spine systematic review was completed in 2016 and included the initial three PROMs with the advocation of the FRI and SFI- 25, where the former, with 10-items, had greater practicality [5].

The SFI- 25 was developed and E-published in 2013 prior to print in 2019 [4], while the SFI- 10 was developed and published in 2024 [8]. The SFI- 25 has been translated, cross-culturally adapted and validated in eight languages [9–16], and shown to have acceptable responsiveness properties for the evaluating of health status change in chronic neck pain patients post-intervention [17]. The SFI- 25 and SFI- 10 resolved the earlier issues of the EASPS [6], that included excess length [5] and lack of COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) compliance, and the FRI [7], with inadequate representation of essential whole-spine item-constructs [3, 5]. The SFI- 10 implemented the improvements advocated for the SFI-25 in the translation and cross-cultural studies [8] by clarifying the one-dimensional factor structure [9–12] and shortening the number of item-questions [13–15]. These initiatives reduce the key barriers to PROM adoption [18], namely administrative burden [14, 19] and the transition from regional-spine to whole-spine PROMs [3, 5, 7]. This in turn facilitates the adoption of routine clinical use, repeated measurements [19], and the retention of cultural transferability [14, 17].

The SFI- 10 retained the essential 60:40 biopsychosocial ratio of general to region-specific item-questions [20, 21] (E-Appendice). It also retained high criterion validity with regional, general, and condition-specific PROMs, where findings approximated those of the SFI- 25 and were found preferable to the FRI [8]. However, full scientific support is lacking for clinical and research use as the critical longitudinal measurement properties are yet to be established. Therefore, this study aimed considered the SFI- 10 clinimetric properties to: 1) verify the psychometric cross-sectional characteristics of structural and criterion validity, internal consistency, and floor/ceiling effects established by the SFI- 10 development study 2) establish longitudinal characteristics of test–retest reliability, responsiveness, construct validity, and error scores; and 3) determine the practical characteristics of readability, missing responses, and administrative burden through time/errors for completion/scoring.

Methods

Study design and ethical aspects

This longitudinal study collected SFI- 10 responses from deidentified MSD spine patients seen in clinical settings across Australia, either under primary contact or referred by a medical practitioner. Certified physiotherapists administered and recorded the outcome measures during routine care using a structured digital clinical decision support system [22]. Nine psychometric and four practical characteristics were analyzed [18, 23]. All research procedures were approved by the Ethical Committee of Universidade Federal do Maranhão under protocol number 4.284.203.

Subjects

A total of 1,317 deidentified spine patients were included in the study, with subgroups analyzed for test–retest reliability ($n = 104$), responsiveness ($n = 171$), construct validity ($n = 91$), and practicality ($n = 16$) (Table 1). These subgroups were identified based on data completeness and availability of follow-up or comparative measures required for each analysis, rather than random sampling. Inclusion criteria were: being under the care of a physiotherapist with a diagnosed musculoskeletal spinal condition, symptom duration of more than two weeks, pain intensity greater than 3 (on a numerical rating scale), age over 18 years, and adequate English-language proficiency [9, 13, 14]. These criteria were verified and recorded at the point of care by certified physiotherapists using a structured digital clinical decision support system. *Exclusion criteria* were: English-language difficulty, age < 18 yrs pregnancy, and red-flags [24].

Measures

The SFI- 10, is a 10-item shortened-version of the SFI- 25 whole-spine PROM [4] published in 2024 [8]. It is used to evaluate functional status and limitations in the activity level and health of patients with spine disorders [8]. Scoring has a 3-point response option of: 'Yes' (score = 1), 'Partly/Sometimes' (score = 1/2) and 'No' (score = 0) that creates a 0–10'raw score' from summation of all item responses. The final score, on a 0–100% scale (100% = 'normal' or 'preinjury function', 0% = 'worst possible'), is calculated with the formulae: [100—(Raw Score × 10)].

The Patient Specific Functional Scale (PSFS) [25] and the 11-point *Numerical Rating Scales (NRS)* for self-perceived *Global function (G-NRS)* [26] and *Pain (P-NRS)* [27], are well-established and have been described in previous publications. The G-NRS measures patients' overall perception of their functional status on a scale from 0 (no disability) to 10 (maximum possible disability). It is commonly employed to assess changes in patient-rated global functional status over time and has demonstrated good reliability, validity, and responsiveness. Pain intensity (P-NRS) was recorded at baseline to quantify self-reported pain severity.

Psychometric characteristics

Face and content validity were not required, being previously established in the development study [8], with the individual item-questions ratified in translated and cultural adaptation studies of the original SFI- 25 [9–16].

Structural validity: Structural validity ($n = 1317$) was determined through factorial analysis, including both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), as well as Rasch analysis [28]. The EFA used polychoric correlation matrix and robust diagonally weighted least squares (RDWLS) extraction method with a factor loading threshold of > 0.40 to ensure adequate item retention. The retained factors were defined through parallel analysis (AP) with a random exchange of observed data and robust promin rotation [29]. Model adequacy was assessed using the Kaiser–Meyer–Olkin (KMO) measure (> 0.70) and Bartlett's sphericity test ($p < 0.05$), calculated using Factor software (version 12.01.02, Universitat Rovira I Virgili, Spain). The three a-priori criteria that were met included an inflection at the second Eigen-point of the scree plot, one Eigenvalue > 1.0, and variance > 10%. The CFA model used the following fit indices: chi-square/degrees of freedom (chi-square/df < 3), root mean square error of approximation (RMSEA < 0.08), comparative fit index (CFI > 0.90), and Tucker-Lewis index (TLI > 0.90), all calculated using R-Studio software with Lavaan and SemPlot packages [30, 31].

Rasch analysis considered: 'Person Abilities' and 'Item Difficulties' (preferred mean = 0.00); targeting analysis;

Table 1 Demographics for all study participants

Prospective	n	Age x̄	Age *SD*	Female n	Female %	Cx	Tx	Lx	Multi
Advise Rehab Data Base	1317	49.5	16.4	703	53.4%	468 35.5%	111 8.4%	738 56.1%	4 0.3%[a]
Reliability	104	44.1	9.6	49	47.1%	28 26.9%	16 15.4%	65 62.5%	5 4.8%*
Responsiveness	171	49.1	8.7	95	55.6%	53 31.0%	15 8.8%	102 59.6%	1 0.6%*
Construct Validity	91	50.6	18.7	59	4.8%	30 33.3%	8 8.8%	53 58.2%	0
Practicality	16	46.5	9.9	8	50%	5 31.3%	3 18.8%	6 37.5%	2 12.5%

Pain intensity, as measured by the point Numerical Rating Scales (P-NRS(, had a mean score of 5.18 ± 2.60 at *SFI* indicates Spine Functional Index, *n* number, *x̄* Mean, *SD* Standard deviation, *%* percent, *Cx* Cervical spine, *Tx* Thoracic spine, *Lx* Lumbar spine

[a] Subregion % values include multi-area individuals within each of their symptomatic regions making total > 100% baseline

Personal separation reliability (PSR:cut-off > 0.70); one-dimensionality (Martin-Löf test:$p >$ 0.05), and Principle Component Analysis (PCA) of Rasch-residuals Eigenvalues (cut-off:Linacare's value < 2.0); infit-outfit statistics (range 0.5–1.5 [32]); item characteristic curves (ICCs); and thresholds proximity (three-response options crossover, with item difficulties ordering); Wright-Mapping (for item spacing and redundancy); and Rasch corrected raw-scores (for person ability), and Differentuial Item Functioning (DIF: negligible when the chi-square test is not significant or the change in Nagelkerke R^2 is < 0.035) from software R (version 4.1.2) [33] with eRm and WrightMap packages [34].

Construct validity through hypothesis testing ($n =$ 1317) assessed the SFI- 10's correlation with criteria PROMs, including the PSFS, G-NRS, and P-NRS. The a priori hypothesis was that the SFI- 10 would show strong positive correlations with the PSFS, G-NRS, and P-NRS, based on its intended use as a comprehensive measure of spinal function. This was performed concurrently within the'Advise Rehab' online completion for all participants. The Spearman r correlation coefficient (SCC) was used for non-normally distributed data, with agreement levels classified as: excellent (± 0.90–1); high (± 0.50–0.89); moderate (± 0.30–0.49); and low (< ± 0.29) [35].

Construct validity through hypothesis testing ($n =$ 91, from the full responsiveness subgroup $n =$ 171) was established using the'known groups' method. We tested a priori hypothesis that the SFI- 10 would discriminate significantly between patients classified as 'symptomatic' at baseline and those in the 'recovered known group', defined by a conservative cut-off of 75% recovered. As hypothesized, a significant diference in SFI- 10 scores was confirmed between these two groups ($p <$ 0.001), supporting the construct validity of measure.

Internal consistency ($n =$ 1317) used Cronbach's alpha (high:$\alpha >$ 0.7, item-redundancy:$\alpha \geq$ 0.95 [23]).

Floor and ceiling effects ($n =$ 1317) were determined by the percentage frequency for the highest and lowest scores at a 15% cut-off [23].

Test–retest reliability ($n =$ 104, average days between measures 5.9 ± 2.7, range 2–11) used the intra-class correlation coefficient ($ICC_{2.1}$) during a period of 'no change' as determined by the difference in both the G-NRS and PSFS of ± 10% ($n =$ 104, low-back = 65; mid-back = 28; neck = 28; multi-area = 5). The levels of agreement were determined as excellent = 0.90, good = 0.75–0.89, moderate = 0.50–0.74, and poor < 0.50 [35].

Additionally, to evaluate the test–retest reliability of individual items in the questionnaire, Weighted Kappa with quadratic weights was employed. Each item was analyzed separately, and the kappa values were interpreted as follows: values < 0.20 indicate poor agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and > 0.81 almost perfect agreement [36, 37].

Error score (responsiveness subgroup, $n =$ 171) used the 'standard error of the measurement' (SEM) and the 90 th percentile 'minimum detectable change' (MDC_{90}) [23].

Responsiveness ($n =$ 171, average days from baseline to repeat measure = 56.6 ± 85.7, range = 2–498, mean = 28, with non-normal distribution, Shapiro-Wilke test) detected relevant construct change over time [38] through repeated measures [23] that had a > 20% change in perceived patient-rated global-status (G-NRS ≥ ± 2 [39, 40]). To determine responsiveness this study used 'Effect Size' (ES:cut-off > 1.2) [41], 'Standard Response Mean' (SRM:cufoff > 1.5) [41], and 'Receiver Operating Curves' (ROC) using the 'Area Under the Curve' (AUC:cut-off > 0.70) [42].

Interpretability was assessed using the minimal clinically important difference (MCID). It was calculated from the available responsiveness data through an anchor-based method that required a change > 30% above the baseline average value [43] (i.e. G-NRS > ± 3).

Practical characteristics

Readability used the Flesch-Kincaid grade-score (cut-off: < 6 th-grade) and reading-ease (> 60%) [44]. *Missing responses* were calculated as a percentage of completed baseline responses with an acceptable level of < 1%. *Administrative burden* was determined from *completion time* in a paper and digital Tablet-PC format ($n =$ 16, female = 50%, age = 46.5 ± 19.9, with separate Cx, Tx, Lx, and multi-area conditions, see Table 1) with 50% completing the paper format first then the digital version at > 2 days apart, giving a total of $n =$ 32 measures of SFI-10 completions, 16 in each format with the average of the two completion times used. S*coring time* was determined from four therapists' time-averages for the 16 completed paper version questionnaires ($n =$ 64), while the digital version was determined instantaneously within the 'Advise Rehab' software, with an acceptable total completion/scoring time being < 2 min [14, 19]. Completion and scoring errors were determined by incomplete or erroneous questionnaires and differences in the scores between the paper and digital versions.

Statistical analysis

The sociodemographic data and questionnaire scores used mean ($\bar{x}$) and standard deviation (SD) in SPSS version 17 at significance $p <$ 0.05. The Kolmogorov–Smirnov and Shapiro-Wilke tests verified data distribution. Factorial and Rasch analyses were performed to assess the structural validity and unidimensionality of the SFI- 10, ensuring that the scale measures a single

construct and behaves consistently across different samples. Sample size minimums were determined from previous SFI- 25 [9–16] and FRI [3, 7, 45] studies for an 80% chance of detecting characteristics with a 15% attrition ($p < 0.05$) [35]. These indicated the sample minimums for validity ($n > 225$), reliability ($n > 90$), and responsiveness ($n > 168$).

Results

All samples had non-normal distributions.

Psychometric characteristics

Structural validity of the SFI- 10 was verified where the EFA (Fig. 1) identified a one-dimensional structure ($n =$ 1317:KMO = 0.87:Bartlett's test-$p < 0.05$), with the three a-priori requirements of a scree-plot inflection at the second point, and only one Eigenvalue > 1.0, with variance of 62.14% explained by the first factor.

The CFA unequivocally confirmed the one-dimensional structure with fit-indices (Table 2) and appropriate factor loadings (> 0.40) between the SFI- 10 domains and all items (Fig. 2).

Rasch analysis demonstrated adequate model fit where 'Person Abilities' ($\bar{x} = -0.07$, SD = 1.21, Min = − 2.71, Max = 3.03) and 'Item Difficulties' ($\bar{x} = -0.05$, SD = 1.0, Min = − 1.21, Max = 2.49). The range of person abilities covered all item difficulties, while the range of item difficulties covered about 80% of person ability scores. These findings indicate that the questionnaire effectively captures the functional range of our sample, with minimal ceiling or floor effects. Therefore, the targeting analysis supports the appropriateness of the instrument for assessing functional spinal status in this population.

Table 2 Structural validity of the SFI- 10, confirmation from CFA

	Chi-square/df	CFI	TLI	RMSEA (90% CI)
$n = 1317$	2.88	0.981	0.975	0.061 (0.053, 0.120)

df indicates degrees of freedom, *CFI* Comparative fit index, *TLI* Tucker–Lewis index, *RMSEA* (90% CI), Root Means Square Error Of Approximation (90% Confidence Interval), *n* number

Additionally, PSR scores (0.79) exceeded the cut-off (> 0.70). The one-dimensionality hypothesis (Martin-Löf test) was accepted ($p > 0.50$), the Rasch-residuals PCA demonstrated cut-off compliance (1.45 < 2.0), and Infit-Outfit statistics (range = 0.68/1.22–0.60/1.28) were within the required range 0.5–1.5 (Table 3). The Wright Map item-spacing and redundancy were acceptable, though not ideal due to some excess-spacing, but overall supported the selected item-shortening methodology. The ICCs (Infit-Outfit statistic) and Thresholds approximated a common point.

The Rasch corrected raw scores were completed (range: 0–20). Table 4 presents the conversion of raw scores to Rasch-corrected scores, enabling clinicians to interpret functional ability in patients with spinal conditions based on their questionnaire responses. Overall, the Rasch-analysis indicated the SFI- 10 preserved the critical Rasch model-fit.

Table 5 presents the DIF analysis for sex and age groups. None of the items showed substantial DIF (all R^2 changes < 0.02), indicating that the questionnaire functions equitably across demographic groups. This suggests that the scale does not unfairly favor specific subgroups based on sex or age.

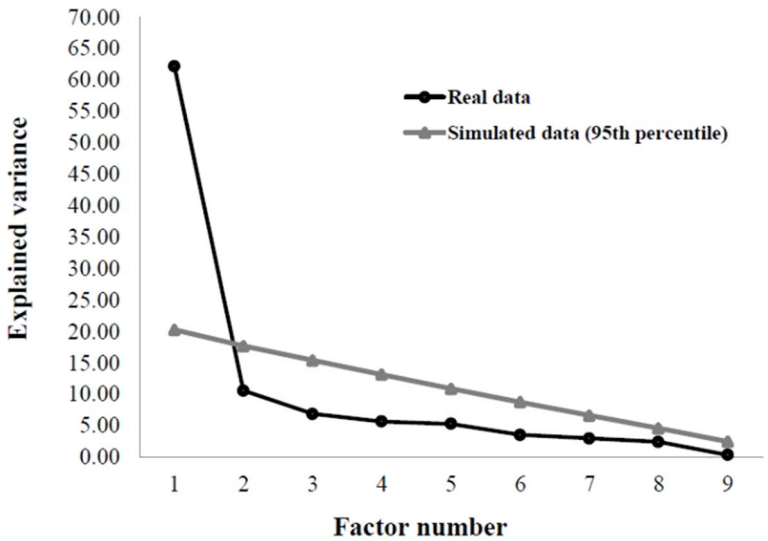

Fig. 1 SFI- 10 EFA Scree Plot ($n = 1317$), inflection at point #2 indicates one-dimension

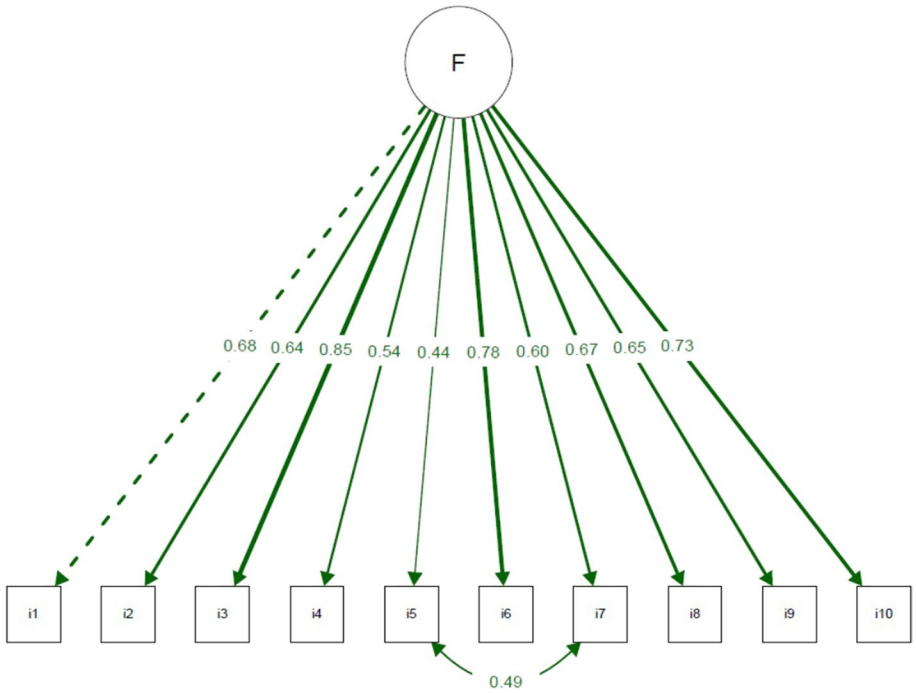

Fig. 2 Factor loadings from CFA for SFI- 10 ($n = 1317$)

Table 3 Structural validity of the SFI- 10, determination from Rasch analysis ($n = 1317$)

Item SFI- 10	Outfit	Infit	Item difficulties
1. Avoid Heavy Jobs	0.908	0.891	− 0.879
2. Pain/Problem	1.004	0.976	− 1.215
3. Duties/Chores	0.604	0.678	− 0.266
4. Sleep	1.279	1.216	− 0.547
5. Personal Care	0.858	0.841	2.488
6. Daily Activities	0.810	0.836	− 0.461
7. Dressing	0.856	0.858	1.120
8. Sitting	0.972	1.005	− 0.406
9. Stand	1.097	1.092	0.095
10. Reach/Bend Down	0.846	0.872	− 0.410

Construct validity through hypothesis testing demonstrated moderate correlation between the SFI- 10 and PSFS ($r = 0.43$), high correlation with the G-NRS ($r = 0.60$), and low-inverse correlation with the P-NRS ($r = -0.24$), consistent with our a priori hypotheses.

Construct validity through hypothesis testing was further supported by the Wilcoxon paired test ($p < 0.001$), which confirmed our a priori hypothesis that the SFI-10 would distinguish significantly between the 'symptomatic' group (baseline mean$n = 43.63 \pm 24.87$, median$n = 40$) and the 'recovered' known group (cut-off = 75% recovered, mean$n = 86.54 \pm 8.46$, median$n = 85$). The effect

Table 4 Rasch-corrected scoring

Raw score	Person ability	Rasch-corrected score
0	− 3.41	0.00
1	− 2.71	1.94
2	− 2.06	3.71
3	− 1.67	4.78
4	− 1.38	5.58
5	− 1.14	6.24
6	− 0.93	6.83
7	− 0.73	7.37
8	− 0.55	7.88
9	− 0.36	8.39
10	− 0.18	8.89
11	0.01	9.42
12	0.21	9.97
13	0.43	10.58
14	0.68	11.25
15	0.96	12.04
16	1.30	12.98
17	1.72	14.14
18	2.27	15.63
19	3.03	17.74
20	3.85	20.00

Table 5 Differential Item Functioning (DIF) analysis for sex and age groups

Sex (men-women) group						
Item	**Uniform DIF**			**Non-uniform DIF**		
	p	**Nagelkerke R^2 change**	**Interpretation**	***p***	**Nagelkerke R^2 change**	**Interpretation**
1	0.01	0.003	Negligible DIF	0.808	0.000	Negligible DIF
2	0.21	0.001	Negligible DIF	0.200	0.001	Negligible DIF
3	< 0.001	0.005	Negligible DIF	0.048	0.001	Negligible DIF
4	0.003	0.006	Negligible DIF	0.469	0.000	Negligible DIF
5	0.471	0.000	Negligible DIF	0.665	0.000	Negligible DIF
6	0.018	0.002	Negligible DIF	0.018	0.002	Negligible DIF
7	0.096	0.002	Negligible DIF	0.611	0.000	Negligible DIF
8	0.501	0.000	Negligible DIF	0.136	0.001	Negligible DIF
9	0.661	0.000	Negligible DIF	0.313	0.001	Negligible DIF
10	< 0.001	0.011	Negligible DIF	0.004	0.004	Negligible DIF
Age (≤ 50 vs > 50) group						
Item	**Uniform DIF**			**Non-uniform DIF**		
	p	**Nagelkerke R^2 change**	**Interpretation**	***p***	**Nagelkerke R^2 change**	**Interpretation**
1	0.012	0.003	Negligible DIF	0.961	0.000	Negligible DIF
2	0.759	0.000	Negligible DIF	0.250	0.001	Negligible DIF
3	0.002	0.002	Negligible DIF	0.695	0.000	Negligible DIF
4	0.970	0.000	Negligible DIF	0.057	0.002	Negligible DIF
5	0.042	0.004	Negligible DIF	0.692	0.000	Negligible DIF
6	0.086	0.001	Negligible DIF	0.865	0.000	Negligible DIF
7	0.080	0.002	Negligible DIF	0.066	0.002	Negligible DIF
8	0.001	0.006	Negligible DIF	0.659	0.000	Negligible DIF
9	0.032	0.003	Negligible DIF	0.040	0.002	Negligible DIF
10	0.282	0.000	Negligible DIF	0.368	0.000	Negligible DIF

Table 6 Item-total statistics

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
SF1	8.7115	21.775	.554	.820
SF2	8.5885	22.720	.482	.827
SF3	8.9954	20.492	.705	.805
SF4	8.8633	22.528	.423	.834
SF5	9.8975	24.785	.365	.837
SF6	8.9058	21.177	.613	.814
SF7	9.5809	22.406	.526	.823
SF8	8.9294	21.658	.533	.823
SF9	9.1678	21.734	.502	.826
SF10	8.9271	21.162	.599	.816

size ($r = 0.38$) indicated a moderate effect, reinforcing the practical significance of these findings.

Internal consistency was excellent ($\alpha = 0.83$), with item-total correlations ranging from 0.365 to 0.705, and Cronbach's α values for individual items if deleted ranging from 0.805 to 0.837 (Table 6).

There was no floor/ceiling effect as the minimum (floor = 0%) scored ($n = 46$, 1.7%) and maximum (ceiling = 100%) scored ($n = 118$, 4.3%) were well below the 15% cut-off for extreme responses, as visualized in the Fig. 3. The distribution of the total scores was approximately normal, indicating that the SFI- 10 captured a broad range of responses without any significant concentration at either extreme.

The Weighted Kappa coefficients for individual questionnaire items ranged from 0.616 to 0.837, indicating substantial to perfect agreement across all items (Table 7). These results suggest good test–retest reliability for the self-report measure.

For the overall questionnaire score, $ICC_{2.1}$ value was 0.972 ($p < 0.001$), demonstrating a strong association between the two administrations (Table 8). *Measurement error, typically associated with individuals having stable symptoms, was robust in the responsiveness subgroup ($n = 171$, baseline $\bar{x} = 34.6\% \pm 24.3\%$) with SEM = 3.84 and MDC90 = 8.98%* (Table 8). *This measurement error calculation is more applicable to individuals with stable symptoms, while changes over time in the responsiveness subgroup were assessed separately.*

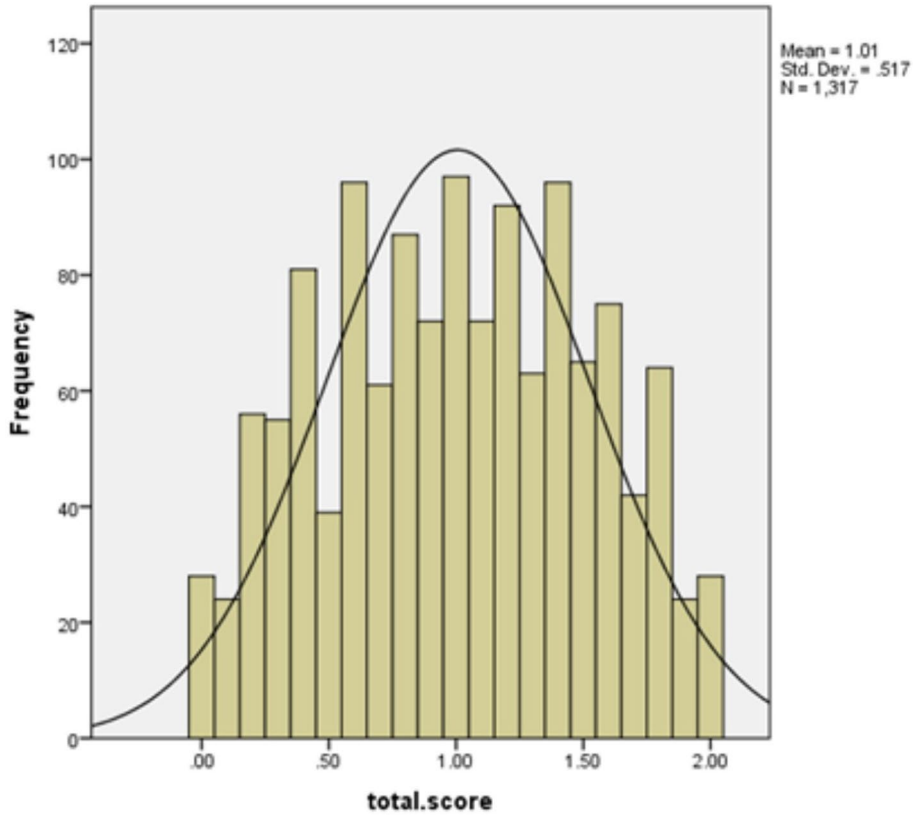

Fig. 3 Histogram to visualize the entire distribution of the data

Table 7 Weighted Cohen's kappa analysis for individual items of the SFI- 10 ($n = 104$)

Parameter	κ	95% CI		Agreement
SF1	0.653	0.527	0.779	Substantial
SF2	0.616	0.484	0.747	Substantial
SF3	0.678	0.557	0.798	Substantial
SF4	0.657	0.534	0.780	Substantial
SF5	0.837	0.726	0.968	Perfect
SF6	0.796	0.697	0.895	Perfect
SF7	0.650	0.503	0.797	Substantial
SF8	0.654	0.533	0.774	Substantial
SF9	0.675	0.558	0.792	Substantial
SF10	0.751	0.597	0.825	Perfect

κ Weighted Cohen's kappa, *CI* confidence interval

R*esponsiveness* was high (ES = 1.54; SRM = 1.64; AUC = 0.894) (Table 8).

Interpretability was calculated using the MCID from the responsiveness subgroup ($n = 171$), with an MCID of 10.4% (Table 8). During the follow-up period, participants were under the care of a physiotherapist, receiving treatments related to their musculoskeletal spine condition. However, detailed information about specific interventions and their potential impact on the outcomes was not available for this study.

Practical characteristics

Optimal practicality was demonstrated as there were no *missing responses,* favorable *readability* (5.1; Ease = 74.1%), and excellent *administration time* (47.7 ± 10.1 s) from *completion* (39.2 ± 10.3 s, range = 24–63 s) and *scoring* (8.5 ± 1.8 s, range = 6–13 s), with *errors* < 1% (Table 8).

Discussion

The study aims were achieved with all nine psychometric and four practical SFI- 10 characteristics demonstrated as adequate. This indicated a valid and reliable instrument for assessing function and symptoms in subacute/chronic spine patients. Critically, the biopsychosocial 60:40 ratio for general: regional items [20, 21] (E-Appendice) was verified and the established SFI- 10 [8] one-dimensional structural validity confirmed, along

Table 8 Psychometric and Practical characteristics of the SFI- 10

Psychometric Characteristics										
	Construct validity	Internal consistency	Test–retest reliability	Error score		Responsiveness Interpretability				
SFI- 10	Wilcoxin	Alpha	$ICC_{2.1}$	SEM	MDC_{90}	ES	SRM	AUC		MCID
	$n = 91$	$n = 1317$	$n = 104$	$n = 171$		$n = 171$		$n = 171$		
	$p < 0.001$	0.834	0.972	3.84%	8.98%	1.54	1.64	0.894	10.4%	
Practical Characteristics										
SFI- 10	Readability (Flesch Kincaid)		Administration (time in seconds)			Missing Responses				
	Grade	Ease	Complete	Score	Total	Percentage				
	5.1	74.1	39.2 ± 10.3	8.5 ± 1.8	47.7 ± 10.1	< 1%				

SFI indicates Spine Functional Index, $ICC_{2.1}$ Test–retest reliability intra-class correlation coefficient, Alpha, Cronbach's Alpha, *SEM* Standard Error of the Measurement, MDC_{90} Minimal Detectable Change, (90% Confidence Interval), *ES* Effect Size, *SRM* Standard Response Mean, *AUC* Area Under the Curve in the receiver operating curve (ROC)

with criterion validity, internal consistency, and without floor/ceiling effects. The unequivocal one-dimensionality improved on the ambiguity of the seven previous SFI- 25 EFA findings [4, 9, 10, 12–14, 16], and concurred with the those of the appropriate CFA and Rasch analysis in two previous SFI- 25 studies [12, 16] (E-appendix Table 9). This is preferable to the questionable factor structure determined for the FRI [3, 45] and the spine-regional ODI, RMQ [28, 46] and NDI [47], which have been shown to have limitations in their applicability across multiple spinal regions. In contrast, the SFI- 10 is a comprehensive tool that provides a more reliable measure of overall spinal function, addressing the full range of spinal conditions, not just a single region. Additionally, its one-dimensional structure and strong psychometric properties make it a more appropriate tool for assessing patients with sub-acute and chronic spine conditions, offering better sensitivity to change over time and a broader scope for evaluating treatment outcomes.

Criterion validity through hypothesis testing exceeded the a-priori hypothesis for correlation with the G-NRS, paralleling finding from the EQ- 5D in the extracted-Polish-SFI- 10 data [8], and demonstrating stronger associations compared to previous findings from the Polish SFI- 25 [14] and Spanish-SFI- 25 [9], despite differences in the underlying constructs of the EQ- 5D and G-NRS. Similarly, the PSFS and P-NRS followed trends established in earlier spine regional studies including the NDI [48], ODI [28, 46] and RMQ [49]. However, the criteria r values were below their respective within-study a-priori cut-offs and those of the extracted-Polish-SFI- 10 and SFI- 10 development-study [8]. They also correlated less than those found in three previous SFI- 25 studies [4, 14, 15]. These variations are likely attributed to either or both the sample size and cultural differences. However, it must also be noted that the PSFS uses diverse patient-selected items with no incorporated broad-general or spine-specific transferable items, which impedes direct comparative data-pooling analysis, and further that pain, as represented by the P-NRS, is neither related nor interchangeable with function [50]. The weak negative correlation between SFI- 10 and pain intensity (NRS) further supports the validity of this measure, indicating that spinal function is influenced by multiple biopsychosocial factors beyond pain alone. Future research may explore the relationship between pain intensity, symptom duration, and functional assessment outcomes in longitudinal studies. The inclusion of NRS scores further strengthens the psychometric evaluation of SFI- 10, as it provides a standardized measure of baseline pain intensity (4.72 ± 1.92). Consequently, specific whole-spine comparative cross-sectional and longitudinal PROM investigation is required. Future research could further explore measurement invariance in different cultural and linguistic contexts to confirm the applicability of SFI- 10 across diverse populations.

The internal consistency concurred with the SFI-10 development study and the range determinedin all earlier SFI- 25 [9–15], FRI [45, 51], and spine-regional NDI [48], ODI, and RMQ [46] findings (E-Appendix Table 9). Additionally, the SFI- 10 had no floor/ceiling effect, consistent with the SFI- 25 [4, 11, 13, 14, 16] and the FRI [7, 52].

The newly established SFI- 10 psychometric properties included construct validity, determined with the 'known-group' methodology at the conservative G-NRS cut-off of 75% recovered. While the SFI- 10 provides a continuous measure of spinal function applicable across multiple spinal regions, future research should further explore its known-groups validity by comparing patients with varying levels of symptom severity, such as those stratified by pain intensity or disability classification. This was consistent with previous SFI- 25 findings that used this methodology [4, 14], and other studies with structural and cross-cultural validity and hypothesis testing [9, 12, 13]. Similar validity levels were found with the PROM-shortened QuickDASH [53] and NDI- 5 [47, 48, 54] which confirmed this methodology retained multiple domains. This was preferable to a 'Single Assessment Numeric Evaluation (SANE)' [26] which provides no broad representation of the multiple constructs necessary in spine-related disorders. However, consensus is limited on whether pain or function is the optimal recovery criterion as 'problem duration' and 'length of follow-up' are not related to recovery cut-off scores. The magnitudes of the determined criterion validity also support construct validity [55], which, in this study, exceeded those of other whole-spine and most advocated spine-regional PROMs, including the NDI [47, 48, 54] and RMQ [46].

The newly established test–retest reliability was comparable to or exceeded that found in the SFI- 25 cross-cultural adaption studies [9–15], and the accepted FRI [45, 51] and spine-regional standards [47, 48, 54]. Similarly, responsiveness and error (SEM and MDC_{90}), which are partially dependent on reliability [41], were comparable to the original SFI- 25 [4, 9, 10, 13, 14] and FRI [5, 45], and the accepted spine-regional RMDQ, ODI [28, 46] and NDI [48, 54] findings.

The practical characteristics exceeded all other published whole-spine PROMs [14, 45, 51] and spine-regional neck [47, 48, 54, 56] or LBP [28, 46] PROMs. The SFI- 10 Flesh-Kincade readability (Grade = 5.1,

Ease = 74.1%) exceeded the cut-off requirements which improved comprehension and reduced cognitive load; which in-turn improves accuracy and reduces errors [18]. Future research should investigate the applicability of SFI- 10 in populations with varying levels of English proficiency and consider validated translations to assess potential linguistic influences on patient responses.

The FRI, ODI and NDI each have poorer readability [14] which requires increased cognitive-demand. However, with more response options (5–6 versus SFI- 10 = 3) they have higher discriminant capacity, but these responses are influenced by pain-expectations [57]. Consequently, these legacy PROMs administration-time and completion-errors increase [18] which reduces clinimetric quality and the capacity for accurate statistical analysis [57]. This was reinforced by the SFI- 10 showing no completion/scoring errors by virtue of the online CDSS format, and < 1% in the paper completed practicality findings, along with the combined completion/scoring time < 60 s (E-Appendix Table 9). A noted trend during SFI- 10 administration was that each participant's completion-times and therapist's scoring-times consistently decreased with test-repetition, indicating reduced burden as PROM familiarity increased.

The results of the DIF analysis indicate that none of the items demonstrated significant DIF across sex or age groups, reinforcing the fairness and applicability of the questionnaire across different populations. This finding supports the cross-cultural validity of the measure and its suitability for diverse patient groups.

Study limitations and strengths

Study limitations included potential bias in patient selection as all recruitment was from primary contact sources within allied health outpatient centers, consequently, investigation of inpatient and community settings is required. The MCID analysis using 'estimated' comparative methodology based on G-NRS change values of ≥ 30% [43], which approximate the MDC_{90} rather than reflecting clinical relevance for the individual patient. Thus, analysis of patient-specific change to clarify the MCID for the individual patient and SFI- 10 user is required. Criterion validity with a 'gold standard' whole-spine PROM, such as the FRI or SFI- 25, was not possible as neither PROM were concurrently available within the 'Advise Rehab' software platform [22], and independent prospective validation should be performed. Additionally, while this study applied both EFA and CFA on the full dataset to maintain statistical power and model stability, future research may consider the independent evaluation of EFA and CFA using randomly split samples to further validate the factor structure of the SFI- 10. Although the EFA identified a strong unidimensional structure with over 60% variance explained, future research could explore whether additional subdimensions exist to further enhance the construct validity of the SFI- 10. Randomly splitting the sample into separate exploratory and confirmatory analyses may provide additional robustness and reduce potential overfitting concerns. Though completion/scoring errors were established by the development study as < 1% for the SFI- 10 paper-version [8], further prospective investigation is required as all data in this study came from an error-free digital source, and the completed/scored paper-version responses samples are insufficient for error determination. However, no errors were found in the completed paper versions and errors are anticipated at < 1% based on previous SFI- 25 studies [4, 14].

Study strengths included the large sample size, where the analyzed psychometric characteristics cut-offs exceeded the COSMIN minimum requirements, and that the SFI- 10 characteristics were almost exclusively preferable to the findings for other shortened MSD PROMs, whole-spine, and spine-regional PROMs.

Further research should consider prospective investigations that concurrently use the SFI- 10 with whole-spine and spine-regional PROMs in different language and patient groups, and with specific conditions such as whiplash, spinal stenosis, and post-operative circumstances.

Conclusions

This longitudinal study demonstrates the SFI- 10 has sound clinimetric properties including an unequivocal one-dimensional factor structure for general MSD spine-patients in a physiotherapy outpatient setting. These properties equate to or exceed those recognized for whole-spine and spine-regional PROMs. These findings suggest generalizability for outpatient settings and a preference over existing whole-spine and spine-regional static-PROMs in clinical and research settings, however further investigation of inpatient and community settings is required. Additionally, prospective investigation with criteria-PROMs in separate cultural and condition-specific settings is required to clarify these findings and expand the determination of MCID and responsiveness. Such studies could then be considered for systematic reviews of spine PROMs that include the SFI- 10 and published SFI- 25 studies.

Appendix

Table 9 Clinimetric properties of the SFI- 10 and SFI- 25: psychometric and practical characteristics

Study	Psychometric Characteristics											Practical Characteristics					
	Structural Validity		Reliability	Internal Consistency	Error score		Interpretability	Responsiveness				Readability (Flesch Kincaid)		Administration (time in seconds)			Missing Responses
	EFA[b]	CFA	$ICC_{2.1}$	Alpha	SEM	MDC_{90}	MCID	SD_{100}	ES	SRM	AUC	Grade	Ease	Complete	Score	Total	Percentage
This Study SFI- 10 (*n* = 1317)	*n* = 1317		*n* = 104	*n* = 1371	*n* = 171			*n* = 171									
	1	1	0.97	0.84	3.84%	8.98%	10.4%	0.22	1.54	1.64	0.894	5.1	74.1	37.8 ± 7.3	9.8 ± 2.1	47.5 ± 10.7	< 1%
SFI- 10 Development (*n* = 505)	*n* = 505																
	1	1		0.80													
Original SFI- 25 Study																	
SFI- 25 (*n* = 203) [4]	1		0.97	0.911	2.76%	6.44%		24.80	1.25	1.81		7.0	64.0	122 ± 37	16 ± 4	138 ± 41	1.5%
FRI (*n* = 143) [4]	1		0.95	0.908	4.14%	9.66%		22.67	1.23	1.68		7.0	47.2	84 ± 23	27 ± 13	111 ± 36	5.3%
Other SFI- 25 Studies																	
SFI- 25 Spain (*n* = 226) [9]	1		0.96	0.845	2.81%	[a]6.56%-6.89%		-	-								Nil
SFI- 25 Turkey (*n* = 285) [10]	1		0.85	0.93	2.96%	[a]6.91%-7.12%											Nil
SFI- 25 China (*n* = 265) [12]	1	1	0.96	0.91													< 1%
SFI- 25 Korea (*n* = 60) [11]	-		0.94	0.88													
SFI- 25 Iran (*n* = 224) [13]	1		0.96	0.88	2.52%	4.58%-[a]5.88%					-						
SFI- 25 Poland (*n* = 225) [14]			0.975	0.90	3.14%	7.33%					-			229	27	256	< 1%
Greek (*n* = 60) [15]			0.96	0.89													
Brazilian (*n* = 194) [16]	1	1	0.83	0.86	9.08%	25.25%											

Detectable Change, (90% Confidence Interval), *MCID* Minimal clinically important difference, SD100, Standard deviation at baseline (100% scale), *ES* Effect Size, *SRM* Standard Response Mean, *AUC* Area Under the Curve

SFI indicates Spine Functional Index, ICC2.1, Test–retest reliability intra-class correlation coefficient, Alpha, Cronbach's Alpha, *SEM* Standard Error of the Measurement, MDC90, Minimal

[a] MDC recalculated by Polish Study [14]

[b] Eigenvalues > 1.0 but variance < 10%

Abbreviations

SFI- 25	Spine Functional Index, 25-items
PROM	Patient reported outcome measure
SFI- 10	Spine Functional Index, 10-items
N	Number
FRI	Functional Rating Index
ICC	Intra-class correlation coefficient
α	Cronbach's alpha
ES	Effect size
SRM	Standard response mean
AUC	Area under the curve
SEM	Standard error of the measurement
MDC	Minimum detectable change
Df	Degrees of freedom
CFI	Comparative fit index
TLI	Tucker-Lewis index
RMSEA	Root means square error of approximation
PSR	Personal separation reliability
sec	Seconds
MSD	Musculoskeletal disorder
COSMIN	COnsensus-based Standards for the selection of health status Measurement Instruments
EFA	Exploratory factor analysis
CFA	Confirmatory factor analysis
PSFS	Patient Specific Functional Scale
G-NRS	Global [function] Numerical Rating Scale
P-NRS	Pain Numerical Rating Scale
K-S	Kolmogorov–Smirnov statistic
SCC	Spearman's correlation coefficient
MCID	Minimal clinically important difference
ROC	Receiver operating curves
RDWLS	Robust diagonally weighted least squares
AP	Parallel analysis
KMO	Kaiser–Meyer–Olkin test
PCA	Principle component analysis
$\bar{x}$	Mean
SD	Standard deviation
SPSS	Statistical Package for the Social Sciences

Acknowledgements
The authors wish to acknowledge the late Dr. Charles Philip Gabel for his significant contributions to the initial conceptualization and early draft of this study. Although he was not involved in the final stages or submission of this manuscript, his foundational input was instrumental in shaping the work. The authors also thank Łukasz Deryło, MSc (Statistics), Independent Statistician, Cracow, Poland, for his assistance with statistical analysis.

Clinical trial number
Not applicable.

Authors' contributions
All authors contributed to the study conception and design. Material preparation and data collection were conducted by all six authors. Data analysis was performed by A.C., A.V., H.R.M., and A.B., with assistance from the acknowledged contributor L.D. The original draft of the manuscript was prepared by M.M. Review and editing were carried out by all authors, including M.M., A.C., A.V., H.R.M., and A.B., who contributed to subsequent manuscript revisions and the final draft. All authors read and approved the final manuscript.

Funding
No funding has been received by any agency in relation to this research.

Data availability
The data that support the findings of this study are not publicly available due to privacy and ethical restrictions. However, the datasets are available from the corresponding author [HRM] upon reasonable request. The data were collected from multiple sources across four countries and are stored securely on a local system. Requests for access can be made via email to the corresponding author and will be reviewed in accordance with ethical and privacy considerations. email: hrmokhtarinia@yahoo.com, hr.mokhtarinia@uswr.ac.ir.

Declarations

Ethics approval and consent to participate
This study was approved by the Ethics Committee of the Universidade Federal do Maranhão, São Luís, Maranhão, Brazil (approval protocol number 4.284.203). Written informed consent was obtained from all participants prior to completing the questionnaires. The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and relevant national guidelines and regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.
ZSF

Author details
[1]Department of Ergonomics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. [2]Neuromusculoskeletal Rehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. [3]Department of Physiotherapy, Faculty of Health Science at the University of Malaga, IBIMA, Málaga, Spain. [4]Postgraduate Program in Physical Education, Federal University of Maranhao, Sao Luis, Maranhao, Brazil. [5]School of Health Sciences, Zurich University of Applied Sciences, Winterthur, Switzerland. [6]Faculty of Health Sciences and Psychology, Collegium Medicum, University of Rzeszów, Rzeszów, Poland. [7]The Holy Family Specialist Hospital, Rudna Mała 600, Głogów, Głogów Małopolski 36 - 060, Poland.

Received: 3 January 2025 Accepted: 31 March 2025
Published online: 16 April 2025

References

1. Hjollund NH, Valderas JM, Kyte D, Calvert M. Health data processes: a framework for analyzing and discussing efficient use and reuse of health data with a focus on patient-reported outcome measures. J Med Internet Res. 2019;21(5):e12412.
2. Guzman JZ, Cutler HS, Connolly J, Skovrlj B, Mroz TE, Riew KD, Cho SK. Patient-reported outcome instruments in spine surgery. Phila Pa 1976. 2016;41(5):429–37.
3. Menke JM. The functional rating index: twenty years of invalid measurement. Spine. 2022;47(7):574–81.
4. Gabel CP, Melloh M, Burkett B, Michener LA. The Spine Functional Index: development and clinimetric validation of a new whole-spine functional outcome measure. Spine J. 2019;19(2):e19-27 Epub 2013 Oct 2025.
5. Leahy E, Davidson M, Benjamin D, Wajswelner H. Patient-Reported Outcome (PRO) questionnaires for people with pain in any spine region. A systematic review. Man Ther. 2016;22(22):22–30.
6. 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(3):261–74.
7. 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(1):78–86.
8. Gabel CP, Cuesta-Vargas A, Dibai-Filho AV, Mokhtarinia HR, Melloh M, Bejer A. Developing a shortened spine functional index (SFI-10) for patients with sub-acute/chronic spinal disorders: a cross-sectional study. BMC Musculoskelet Disord. 2024;25(1):236.
9. Cuesta-Vargas AI, Gabel CP. Cross-cultural adaptation, reliability and validity of the Spanish version of the spine functional index. Health Qual Life Outcomes. 2014;12(96):1.
10. Tonga E, Gabel CP, Karayazgan S, Cuesta-Vargas AI. Cross-cultural adaptation, reliability and validity of the Turkish version of the spine functional index. Health Qual Life Outcomes. 2015;13:30.
11. In TS. The reliability and validity of the Korean version of the spine functional index. J Phys Ther Sci. 2017;29(6):1082–4.
12. Zhou X, Xu X, Fan J, Wang F, Wu S, Zhang Z, Yang Y, Li M. Wei1 X: Cross-cultural validation of simplified Chinese version of spine functional index. Health Qual Life Outcomes. 2017;15:203.

13. Mokhtarinia HR, Hosseini A, Maleki-Ghahfarokhi A, Gabel CP, Zohrabi M. Cross-cultural adaptation, validity, and reliability of the Persian version of the spine functional index. Health Qual Life Outcomes. 2018;16(1):95.
14. Bejer A, Kupczyk M, Kwaśny J, Majkut A, Moskal K, Niemiec M, Gabel CP. Cross-cultural adaptation and validation of the Polish version of the Spine Functional Index. Eur Spine J. 2019;29(6):1424–34.
15. Georgoudis G, Sotiropoulos S, Skouras A, Likourgia I, Retalis E, Karameri M. Musculoskeletal functional assessment: cross-cultural adaptation, validity, and reliability of the Greek version of the Spine Functional Index (SFI). Geneva: World Congress of Physiotherapy; 2019.
16. Freitas DWN. Tradução, adaptação transcultural e validação do Spine Functional Index para o português brasileiro (Portuguese: Translation, cross-cultural adaptation and validation of the Spine Functional Index into Brazilian Portuguese). Masters by Research, Thesis dissertation. Maranhão: Universidade Federal do Maranhão; 2023. https://tedebc.ufma.br/jspui/handle/tede/tede/4688.
17. Hessam M, Narimisa M, Monjezi S, Saadat M. Responsiveness and minimal clinically important changes to physical therapy interventions of Persian versions of copenhagen neck functional disability index, neck bournemouth questionnaire and spine functional index questionnaires in people with chronic neck pain. Physiother Theory Pract. 2024;40(8):1751–8.
18. Long C, Beres LK, Wu AW, Giladi AM. Patient-level barriers and facilitators to completion of patient-reported outcomes measures. Qual Life Res. 2022;31(6):1711–8.
19. Lam KC, Harrington KM, Cameron KL, Valier ARS. Use of patient-reported outcome measures in athletic training: common measures, selection considerations, and practical barriers. J Athl Train. 2019;54(4):449–58.
20. Gabel CP, Melloh M, Burkett B, Michener LA. The Lower Limb Functional Index: development and validation of the clinimetric properties and practical characteristics. Phys Ther. 2012;92(1):98–110.
21. International Classification of Functioning, Disability and Health (ICF) www.who.int/icidh.
22. Advise Rehab. Advise Rehabilitation, decision support software. 3rd ed. Coolum Beach: Advise Rehabilitation Pty Ltd; 2013.
23. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1171–9.
24 Khalilian-Ekrami N, Amirshakeri B, Ghanavati T, Mokhtarinia HR, Gabel CP. Cross-cultural adaptation, reliability, and validity of the Persian version of the Lower Limb Functional Index. Musculoskelet Sci Pract. 2022;62:102626.
25. Stratford P, Gill C, Westaway M, Brinkley J. Assessing disability and change on individual patients: a report of a patient specific measure. Physiother Can. 1995;47(4):258–63.
26. Bowling A. Just one question: if one question works, why ask several? J Epidemiol Community Health. 2005;59(5):342–5.
27. Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, Fainsinger R, Aass N, Kaasa S. European Palliative Care Research Collaborative E: Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage. 2011;41(6):1073–93.
28 Cook CE, Garcia AN, Wright A, Shaffrey C, Gottfried O. Measurement properties of the Oswestry Disability Index in recipients of lumbar spine surgery. Spine (Phila Pa 1976). 2021;46(2):E118–25.
29. Lorenzo-Seva U, Ferrando PJ. Robust Promin: a method for diagonally weighted factor rotation. Lib Rev Peru Psicol. 2019;25(1):99–106.
30. Hu L-t. Bentler PM: fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychol Methods. 1998;3(4):424.
31. Montoya AK, Edwards MC. The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. Educ Psychol Measur. 2021;81(3):413–40.
32. Linacre M. Teaching rasch measurement. Rasch Meas Trans. 2017;31(2):1630–1.
33. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
34. Mair P, Hatzinger R. Extended Rasch modeling: The eRm package for the application of IRT models in R. J Stat Softw. 2007;20(9):1–20.
35. Field A. Discovering Statistics using SPSS. 3rd ed. London: SAGE Publications Ltd; 2009.
36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.
37. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360–3.
38. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. 4th ed. Oxford: Oxford University Press; 2008.
39. Ostelo RW, Deyo R, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, De Vet HC. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine. 2008;33(1):90–4.
40. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine. 2005;30(11):1331–4.
41. Liang MH, Fossel AH, Larson MG. Comparison of five health status instruments for orthopaedic evaluation. Med Care. 1990;28:632–42.
42. de Vet HCW, Terwee CB, Mokkink LB. Measurement in Medicine A Practical Guide. London: Cambridge Press; 2011.
43. Sandvall B, Okoroafor UC, Gerull W, Guattery J, Calfee RP. Minimal clinically important difference for PROMIS physical function in patients with distal radius fractures. J Hand Surg Am. 2019;44(6):454-459.e451.
44. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel. In Research Branch report. Memphis: Naval Air Station; 1975. pp.8–75.
45. Ansari NN, Komesh S, Naghdi S, Fakhari Z, Alaei P. Responsiveness of minimal clinically important change for the Persian functional rating index in patients with chronic low back pain. Asian Spine J. 2019;13(1):111–8.
46. Jenks A, Hoekstra T, van Tulder M, Ostelo RW, Rubinstein SM, Chiarotto A. Roland-Morris disability questionnaire, Oswestry Disability Index, and Quebec back pain disability scale: which has superior measurement properties in older adults with low back pain? J Orthop Sports Phys Ther. 2022;52(7):457–69.
47. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, Goldsmith CH, MacDermid JC. Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther. 2009;39(5):334–50.
48 Barreto FS, Avila MA, Pinheiro JS, Almeida MQG, Ferreira CSB, Fidelis-de-Paula-Gomes CA, Dibai-Filho AV. Less is more: five-item neck disability index to assess chronic neck pain patients in Brazil. Spine (Phila Pa 1976). 2021;46(12):E688–93.
49. Williams RM, Myers AM. Support for a shortened Roland-Morris Disability Questionnaire for patients with acute low back pain. Physio Can. 2001;53:60–6.
50. Pathak A, Wilson R, Sharma S, Pryymachenko Y, Ribeiro DC, Chua J, Abbott JH. Measurement properties of the patient-specific functional scale and its current uses: an updated systematic review of 57 studies using COSMIN guidelines. J Orthop Sports Phys Ther. 2022;52(5):262–75.
51. Feise RJ, Menke JM. Functional rating index: literature review. Med Sci Monit. 2010;16(2):RA25-36.
52 Ansari NN, Feise RJ, Naghdi S, Mohseni A, Rezazadeh M. The functional rating index: reliability and validity of the persian language version in patients with neck pain. Spine (Phila Pa 1976). 2012;37(14):E844-848.
53. Beaton DE, Wright JG, Katz JN, Group UEC. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005;87(5):1038–46.
54 Walton DM, MacDermid JC. A brief 5-item version of the neck disability index shows good psychometric properties. Health Qual Life Outcomes. 2013;11:108.
55. Pontes-Silva A, Avila MA, Fidelis-de-Paula-Gomes CA, Dibai-Filho AV. The short-form neck disability index has adequate measurement properties in chronic neck pain patients. Eur Spine J. 2021;30(12):3593–9.

56. van der Velde G, Beaton D, Hogg-Johnston S, Hurwitz E, Tennant A. Rasch analysis provides new insights into the measurement properties of the neck disability index. Arthritis Rheum. 2009;61(4):544–51.
57. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. J Clin Epidemiol. 2018;95:73–93.

Publisher's Note