PTJ: Physical Therapy & Rehabilitation Journal | *Physical Therapy*, 2024;104:1–8
https://doi.org/10.1093/ptj/pzad113
Advance access publication date August 22, 2023
Original Research

Reliability, Measurement Error, Responsiveness, and Minimal Important Change of the Patient-Specific Functional Scale 2.0 for Patients With Nonspecific Neck Pain

Erik Thoomes, PT, MSc[1,2,*], Joshua A. Cleland, PT, PhD[3], Deborah Falla, PT, PhD[1], Jasper Bier, PT, PhD[4,5], Marloes de Graaf, PT, PhD[2,4]

[1]Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, UK
[2]Research Department, Fysio-Experts, Hazerswoude, The Netherlands
[3]Department of Physical Therapy, Tufts University School of Medicine, Boston, Massachusetts, USA
[4]Department of Manual Therapy, Breederode University of Applied Science, Rotterdam, The Netherlands
[5]Department of General Practice, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

*Address all correspondence to Dr Thoomes at: erikthoomes@gmail.com

Abstract

Objective. The Patient-Specific Functional Scale (PSFS) is a patient-reported outcome measure used to assess functional limitations. Recently, the PSFS 2.0 was proposed; this instrument includes an inverse numeric rating scale and an additional list of activities that patients can choose. The aim of this study was to assess the test–retest reliability, measurement error, responsiveness, and minimal important change of the PSFS 2.0 when used by patients with nonspecific neck pain.

Methods. Patients with nonspecific neck pain completed a numeric rating scale, the PSFS 2.0, and the Neck Disability Index at baseline and again after 12 weeks. The Global Perceived Effect (GPE) was also collected at 12 weeks and used as an anchor. Test–retest measurement was assessed by completion of a second PSFS 2.0 after 1 week. Measurement error was calculated using a Bland–Altman plot. The receiver operating characteristic method with the anchor (GPE) functions as the reference standard was used for calculating the minimal important change.

Results. One hundred patients were included, with 5 lost at follow-up. No floor and ceiling effects were reported. In the test–retest analysis, the mean difference was 0.15 (4.70 at first test and 4.50 at second test). The ICC (mixed models) was 0.95, indicating high agreement (95% CI = 0.92–0.97). For measurement error, the upper and lower limits of agreement were 0.95 and −1.25 points, respectively, with a smallest detectable change of 1.10. The minimal important change was determined to be 2.67 points. The PSFS 2.0 showed satisfactory responsiveness, with an area under the curve of 0.82 (95% CI = 0.70–0.93). There were substantial to high correlations between the change scores of the PSFS 2.0 and the Neck Disability Index and GPE (0.60 and 0.52, respectively; $P < .001$).

Conclusion. The PSFS 2.0 is a reliable and responsive patient-reported outcome measure for use by patients with neck pain.

Keywords: Measurement Qualities, Neck, Patient Reported Outcome Measure, Patient Specific Functional Scale, Responsiveness

Received: March 3, 2023. **Revised:** June 15, 2023. **Accepted:** July 24, 2023

Introduction

The Patient-Specific Functional Scale (PSFS)[1] is a questionnaire widely used to subjectively assess "activity limitations."[2,3] This patient-reported outcome measure (PROM) was designed to provide a comparison of a patient's activity levels on specific functional tasks in comparison to before their complaints started. A recent study assessed the construct validity of the original PSFS and revealed that patients with neck pain preferred an inverse way of answering the questions then is traditionally used.[4] Specifically, patients preferred rating their responses using a numeric rating scale (NRS)[5] where 0 represents "no difficulty" and 10 represents "impossible to perform the activity" which is opposite of the original PSFS.[6,7] Patients stated that this response option was more "logical" and easier to answer.[4] This led to developing the PSFS 2.0 with an NRS along with a list of proposed activity limitations should patients find it difficult to name 3 activities themselves.[8]

Conflicting evidence remains regarding the psychometric properties of the original PSFS when used for patients with various neck pain disorders.[8,9] A systematic review indicates moderate evidence supporting high test–retest reliability for the original PSFS among patients with cervical radiculopathy[9] (based upon 1 study with a small sample size of 38 patients[10]). However, a more recent study found evidence of poor reliability (165 patients).[11] It should be noted that most reliability studies of the original PSFS were performed on small samples ($n = 31$ and $n = 38$) of patients with neck pain.[8,12] A study assessing the minimal important difference of the original PSFS in patients with neck pain reported that a change of 2.3 was both a "small change" and a "large change," and a change of 0.6 was a "medium change," making it difficult to interpret the results.[13] One study in patients with cervical radiculopathy reported adequate responsiveness of the original PSFS.[11] A recent systematic review on the psychometric properties of the original PSFS concluded that this instrument was valid, reliable, and responsive in populations with neck dysfunction.[9] However, this systematic review also reported that "although the use of the (original) PSFS as an outcome measure is increasing in physical therapist practice, there are gaps in the research literature regarding its validity, reliability, and responsiveness in many health conditions."[9]

The PSFS 2.0 has been validated on patients with neck pain and showed a high correlation with the Neck Disability Index (NDI),[10] similar to the original PSFS.[8]

The aim of this study was to assess the test–retest reliability and measurement error of the PSFS 2.0 and to assess the responsiveness and minimal important change in patents with nonspecific neck pain.

Methods

Design

This reliability and responsiveness study is part of a cohort study, Cervical Range of Motion Measurements (CROMM-study),[4,14] including patients with nonspecific neck pain treated in a physical therapist setting. The Medical Ethic Center in Rotterdam approved the study (MEC-2018-129). The study was registered in the Netherlands Trial Register NTR7463. To ensure methodological rigor, the Guidelines for Reporting Reliability and Agreement Studies guideline was adhered to Kottner et al,[15] as well as the relevant sections of the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) reporting guideline for studies on measurement properties.[16] Informed written consent was obtained from all patients. No funding was received for this work.

Participants

Patients were recruited from a primary care physical therapist clinic between July 2018 and January 2019. Patients with nonspecific neck pain were eligible if they were more than 18 years old, adequately understood Dutch, and were classified as having grade I or II pain as described by the Neck Pain Task Force.[17] Patients were excluded in the presence of serious pathology (such as infection, cancer, fracture, or rheumatoid arthritis) and previous cervical surgery. A minimum number of 100 patients were recruited for the assessment of construct validity, and 10 of the 100 patients were recruited for the assessment of content validity as done in previous studies.[18,19]

Baseline Measurement

All participants received an automated email questionnaire that included demographic characteristics as well as the NRS, PSFS 2.0, and NDI in this particular order. All forms were available online, using Limesurvey, a free online survey tool.

Patients were informed about the purpose to test the test–retest reliability of the questionnaire and that a second, follow-up questionnaire would be sent in 1 week.

NRS

Neck pain in the past 24 hours was measured using an NRS, where 0 represents "no pain" and 10 represents "the worst pain possible."[20,21] The minimal detectable change has been reported to be approximately 4 points.[20,21] The NRS seems to be the most appropriate measure to assess the pain intensity[22,23] and is recommend in clinical practical guidelines for neck pain.[24,25] We could not identify studies examining the content validity, construct validity, reliability, or responsiveness of the NRS for patients with neck pain.

NDI

The NDI is designed to measure "activity limitations" during activities of daily living in patients with neck pain and was derived from the Oswestry Disability Index for measuring activity limitations in individuals with low back pai.[26,27] The 10 items of the NDI have 6 response categories (range = 0–5; total score range = 0–50).[27] No floor or ceiling effects have been detected.[27–30] The content validity is poor.[31] Hypothesis testing has shown that the NDI has a positive correlation with instruments measuring pain and/or physical functioning ($r = 0.53$–0.70)[27,29,32,33] and can detect differences in scores between subgroups (eg, same work status vs altered work status).[29,34] There is moderate evidence for responsiveness of the NDI (area under the curve [AUC] = 0.79; 95% CI = 0.68–0.89) using the Global Rating of Change as a comparator.[34] The NDI is recommended in English[24,35] and in Dutch (as it is reliable, valid, and responsive).[36,37]

PSFS 2.0

The PSFS 2.0 was designed as a functional outcome scale to measure "activity limitations."[1]The PSFS 2.0 is based on the concept of generating a list of problems specific for each patient instead of having patients check a general list of their most commonly encountered problems. The PSFS 2.0 allows each patient to nominate any activity that he or she may

be having difficulty with. Patients were asked to identify 3 important activities they were unable to perform or were having difficulty with because of their neck problem.[4,6,8,25]

An example list was provided to assist in either formulating activities or checking if the mentioned activities were indeed the most important ones. Patients were asked to score their "activity limitations," where 0 represents "no difficulty" and 10 represents "impossible to perform the activity."[4,6,7] An average PSFS 2.0 score was calculated. Higher scores indicate a higher level of activity limitation. The PSFS 2.0 has been shown to have good content validity.[4]

Test–Retest Measurement

All patients received an automated follow-up email requesting that they complete a second PSFS 2.0 and the General Perceived Effect (GPE) scale 1 week after the first assessment. If possible, reasons were recorded for not replying to the retest measurement. The time interval of 1 week between both measurements was chosen to minimize recall bias as well as progression bias and is often considered appropriate.[38] Patients received usual care during this time.[25]

GPE Scale

The GPE scale is a 7-point Likert scale asking if the patient's condition has improved or deteriorated since the start of treatment ("Could you please state the amount of change concerning your recovery compared to when you first started treatment?"). This scale ranges from "worse than ever" to "completely recovered" (completely recovered, much improved, slightly improved, no change, slightly worse, much worse, and worse than ever). The GPE scale has been shown to have good test–retest reliability and correlates well with changes in pain and disability.[39] Despite controversy about the role of global rating items, the GPE scale has frequently been used as an anchor in responsiveness studies.[40–44]

Follow-Up Measurement

All patients received an automated email including the NRS, PSFS 2.0, NDI, and GPE scale 12 weeks after their baseline measures. Within this period, the patient received physical therapist treatment for 1 or more sessions; therapy sessions were not standardized but were tailored to the individual.[25]

Analyses

All statistical analyses were performed with SPSS version 29 (SPSS Statistics for Windows version 29.0; IBM Corp, Armonk, NY, USA). Handling of missing items on the NDI was performed as previously described; if a patient did not complete 1 or more questions, the average of all other items was added to the completed items.[45] All data were checked for normality using a stem and leaf plot, Q–Q plot, and box and whisker plot. Nonparametric tests were used if data were not normally distributed. Descriptive statistics were used to calculate frequencies and summarize the data. Results for continuous variables are presented as mean and SD or, if variables were not normally distributed, as median and interquartile range. Summaries for continuous variables are expressed as mean and SD.

Floor and Ceiling Effects

Frequencies are presented as means and SDs for normally distributed data or as medians and interquartile ranges when the data were not normally distributed. If more than 15% of the responders at baseline or at the follow-up after 12 weeks achieved the highest or lowest possible scores on the PSFS 2.0, then we considered this result a sign of floor or ceiling effects.[46]

Test–Retest

Patients with a stable GPE were included in this analysis (ie, slightly improved, no change, or slightly worse on the GPE), and differences were assessed between patients who were included and those who were not included using a chi-square test or *t*-test. We used the test–retest data to determine whether or not there were systematic errors (although they reduce the validity but affect the accuracy of the measurement but do not affect the reliability [because they are always the same]) by using an analysis of variance test. We presented a Bland–Altman plot to visually illustrate systematic errors. The ICC agreement was used in case of systematic errors; otherwise, consistency was used to calculate the test–retest reliability of the PSFS 2.0—that is, the extent to which the same test results are obtained for repeated assessments when no real change is expected in the intervening period (7 days). The ICC can range from 0.00 (no stability/agreement) to 1.00 (perfect stability/agreement).[47] An ICC of 0.70 is considered to be acceptable.[47,48]

Measurement Error

The limits of agreement were calculated using a Bland–Altman plot with the mean and SDs of the differences between 2 (test–retest) measurements.[49] The resulting graph is a scatterplot *xy*, in which the *y*-axis shows the difference between the 2 paired measurements (A and B) and the *x*-axis represents the average of these measures [(A + B)/2], so that the difference of the 2 paired measurements is plotted against the mean of the 2 measurements. It is recommended that 95% of the data points lie within ±2 SDs of the mean difference. Additionally, the smallest detectable change (SDC) was calculated as $1.96 \times SD_{difference}$[48] to assess the change beyond measurement error. Ideally, the minimal important change should be higher than the SDC.[50]

Minimal Important Change

The minimal important change is the smallest change in a score within the construct being measured that patient's perceive as important. The minimal important change is a threshold for a minimal within-person change over time above that patients perceive themselves to be changed in an important way. If all patients have their individual threshold of what they consider a minimal important change, the minimal important change can be conceptualized as the mean of these individual thresholds.[51,52] The minimal important change does not refer to thresholds for changes that are considered more than minimal (eg, a mean change in patients who reported to be "much better" is not a minimal important change). Next, the minimal important change is not a minimal detectable change (MDC, also referred to as SDC). The MDC is the smallest change in score than can be detected statistically with some degree of certainty (eg, 95% or 90%), based on the standard error of measurement or limits of agreement from a test–retest reliability design.[53]

We used the receiver operating characteristic (ROC) method, with the PSFS 2.0 as the diagnostic test and the anchor (GPE) functions as the reference standard for calculating the minimal important change. The anchor

distinguishes patients considered recovered from patients with "no important change." The instrument's sensitivity is the proportion of patients who were considered recovered according to the anchor that are correctly identified as such by the PSFS 2.0. Specificity is the proportion of patients with no important change that is correctly identified as such by the PSFS 2.0. The minimal important change is defined as the optimal ROC cutoff point, meaning the point on the ROC curve nearest to the upper left corner.[46]

We used the GPE anchor and considered patients as recovered when they answered that they were completely recovered or much improved and as not importantly improved when they answered slightly improved, no change, or slightly worse.[40,41,54]

Responsiveness

Responsiveness was assessed using the area under the ROC curve (AUC) and hypothesis testing. The GPE has a high level of face validity and is considered to be a suitable criterion to measure change.[46] The AUC was calculated to assess the ability of the PSFS 2.0 to discriminate between patients who are considered improved and not importantly changed according to the GPE, using a similar anchor as described in the interpretability section.[46] A benchmark that was previously used to establish that outcome measures are useful in discriminating patients who were improved from those who were not improved was set at an AUC of 0.70.[48]

Some have expressed concerns about the reliability and validity of the GPE in measuring change.[39,55] We also chose to test specific hypotheses. Hypothesis testing for responsiveness was based on the concept that the correlation between the change score of related constructs (GPE and NDI) must be moderate. Hypothesis testing was analyzed using the Pearson correlation coefficient in case of a normal distribution of the data; otherwise, a Spearman correlation coefficient was used. Correlation coefficients between the PSFS 2.0 change score and the change score of the NDI and the GPE were expected to be above 0.50.[46] Correlations were rated as follows: $r < 0.30$ as low/insignificant, $0.30 \leq r < 0.45$ as moderate. $0.45 \leq r < 0.60$ as substantial, and $r \geq 0.60$ as high.[56]

Results

A total of 100 consecutive patients agreeing to participate were included at baseline. Five patients were excluded from those who were originally included in the study due to loss to follow-up for the final analysis, and 19 participants were unable to respond within 1 week for the test–retest analysis due to personal time-constraint reasons. The mean age of the patients was 52.6 (SD = 14.5) years, and 75% were female. Demographic characteristics of the 100 included patients as well as the 81 patients included in the test–retest analysis are reported in Table 1.

Floor and Ceiling Effects

No floor and ceiling effects were reported at baseline in the original group ($n = 100$; 3.0%) (Fig. 1) or the stable test–retest group ($n = 81$; 8.6%) (Fig. 2).

As was to be expected, in the follow-up analysis after 12 weeks, there was a floor effect for the proportion of patients who had completely recovered ($n = 29$; 30.5%) but no ceiling effect ($n = 1$; 1.1%) (Suppl. Fig. S1).

Test–Retest

A total of 81 patients had data for the test–retest analysis as they were considered to be stable. Patients who were selected did not differ significantly in baseline characteristics from patients who were not selected (Tab. 1). Analysis of variance revealed a systematic difference between the first and second measures. The mean difference was 0.15 (4.70 at first test and 4.50 at second test). The ICC (mixed models) was 0.95, indicating high agreement (95% CI = 0.92–0.97).

Measurement Error

The upper and lower limits of agreement were 0.95 and −1.25 points, respectively (Suppl. Fig. S2), with an SDC of 1.10.

Minimal Important Change

Table 2 shows the frequency of change scores on the GPE, and Figure 3 illustrates the anchor-based distribution of the percentage change scores for the improved and unchanged groups. The minimal important change was determined to be 2.67 and therefore higher than the SDC.

Responsiveness

Figure 4 presents the ROC curve generated for the PSFS 2.0 based on the GPE as an anchor. The PSFS 2.0 showed satisfactory responsiveness, with an AUC of 0.82 (95% CI = 0.70–0.93). There were substantial to high correlations between the change scores of the PSFS 2.0 and the NDI and GPE (0.60 and 0.52, respectively; $P < .001$) and high correlations between the GPE and the NDI (0.63; $P < .001$).

Discussion

This is the first study assessing the test–retest reliability, measurement error, responsiveness, and minimal important change of the PSFS 2.0 when used by patients with nonspecific neck pain. The PSFS 2.0 is reliable, responsive, and shows no sign of floor and ceiling effects. Additionally, the PSFS 2.0 showed substantial to high correlations with the change score of the NDI, and the GPE in predicting improvement in patient status versus no change. As the minimal important change in our study is higher than the SDC, this implies that the PSFS 2.0 is able to detect important change and distinguish it from measurement error at an individual level on the basis of single measurements in a patient population with neck pain.[53] Additionally, the test–retest reliability results indicate that patients with nonspecific neck pain will have similar scores on the PSFS 2.0 with different administrations over time.

In clinical practice, PROMs are most often used to determine progress (outcomes) of individual patients.[57] However, the most common reason for not using PROMs is that they can be too time consuming for patients to complete (43%) and for clinicians to analyze, calculate, and score (30%). Moreover, some PROMs are too difficult for patients to complete independently (29.1%).[57] Several researchers have started to validate shortened or abbreviated versions of existing PROMs.[58–62] The PSFS 2.0 is an example of a PROM which is easy to complete, takes little time to complete, and is simple for clinicians to analyze and interpret.

The PSFS 2.0 is considered similar to the original PSFS[13] and is also focused on activity limitations, asking patients to identify 3 important activities that they are unable to perform or were having difficulty with because of their neck problem. The PSFS 2.0, however, has an inverse scoring system from

Table 1. Demographic Characteristics of the Included Patients[a]

Characteristic	Complete Cohort ($N = 100$)[b]	Test–Retest Cohort ($n = 81$)[b]	*P*
Sex assigned at birth, no. (%) men	25 (25.0)	18 (22.2)	<.05
Age, y, mean (SD)	52.6 (14.5)	52.8 (14.5)	<.01
Duration of neck pain, wk, median (IQR)	20.0 (8.0–100.0)	24.0 (10–178)	<.05
Acute/subacute	45 (45.0)	33 (40.7)	<.05
Chronic	55 (55.0)	48 (59.3)	<.05
History of neck pain	79 (79.0)	63 (77.8)	<.05
Ability to work despite neck pain			
No, completely unable	1 (1.0)	1 (1.2)	<.01
No, but I do not work at all	22 (22.0)	19 (23.5)	<.05
Yes, it's possible to perform my ordinary work activities	60 (60.0)	48 (59.3)	<.05
Yes, but I have to adjust my work	17 (17.0)	13 (16.0)	<.05
NDI baseline percentage score, mean (SD)	12.0 (6.1)	12.6 (6.0)	<.01
Initial pain on NRS, mean (SD)	4.7 (2.4)	4.9 (2.3)	<.01
PSFS 2.0 total score, median (IQR)	4.5 (3.3–6.0)	4.7 (3.3–6.3)	<.01

[a]IQR = interquartile range; NDI = Neck Disability Index; NRS = numeric rating scale; PSFS = Patient-Specific Functional Scale. [b]Data are reported as numbers (percentages) of patients unless otherwise indicated.

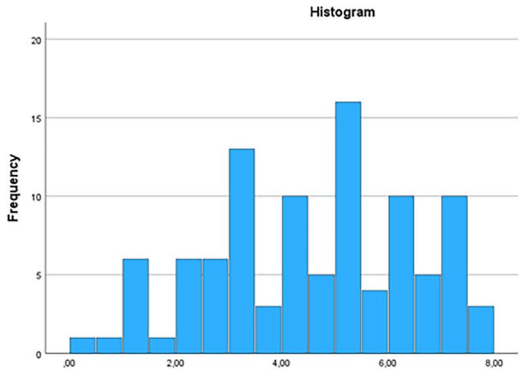

Figure 1. Histogram of number of average scores on the Patient-Specific Functional Scale 2.0 in the original group.

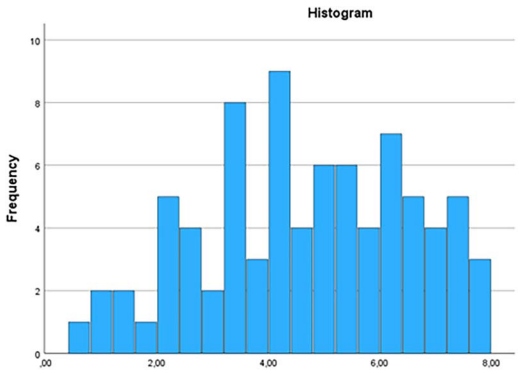

Figure 2. Histogram of number of average scores on the Patient-Specific Functional Scale 2.0 in the retest group.

0 to 10, where 0 represents no difficulty and 10 represents impossible to perform the activity. In addition, the PSFS 2.0 provides the patients with a list of examples to assist in either formulating activities or checking if the mentioned activities were indeed the most important ones.

In a previous study on the PSFS 2.0, patients with neck pain reported the PSFS 2.0 to be appropriate and easy to understand and showed an explicit preference for the PSFS 2.0 version.[4] They also indicated the inverse scoring system made more sense to them. The above-mentioned study also

Table 2. Frequency of Change Scores on the Global Perceived Effect Scale

Rating	Frequency (*n*)	%	Cumulative %
Completely recovered	29	30.5	30.5
Much improved	43	45.3	75.8
Slightly improved	14	14.7	90.5
No change	8	8.4	98.9
Slightly worse	0	0	98.9
Much worse	1	1.1	100
Worse than ever	0	0	100
Total	95	100	

reported a substantial correlation between the PSFS 2.0 and the NDI and a significant difference between known groups. Based on the findings, the PSFS 2.0 possess adequate content and construct validity and is deemed it to be acceptable for patients with nonspecific neck pain.[4]

Although the original PSFS is applicable to all patients with upper extremity[5] and other musculoskeletal disorders,[9] we examined the measurement properties of the PSFS 2.0 specifically in patients with nonspecific neck pain. The measurement properties of the PSFS 2.0 need to be examined in patients with other musculoskeletal disorders before the scale could be recommended for use in these populations.

The high ICC test–retest values in our study (0.95) are comparable to those in previously performed studies of the psychometric qualities of the original PSFS, with ICC values ranging from 0.82 to 0.98.[8,10,63,64] The measurement error of 0.95 to −1.25 is also in line with outcomes reported in a recent systematic review on the measurement properties of the PSFS standard error of measurement values of the PSFS were ≤ 1 for most reported conditions, except for cervical radiculopathy, where it was 1.5. The reported SDC values ranged from 1.5 points to approximately 3 points, but these were not consistently lower than minimal important change values.[65] They also reported the PSFS to be more responsive than the NDI.[10,65] The responsiveness determined in the current study (0.82) is also comparable to the responsiveness reported for patients with neck pain ranging from 0.71 to 0.99 in the above-mentioned systematic review (7 studies, 650 participants). The minimal important change value determined in the current study is comparable to a previously reported minimal important change of 2.00.[66]

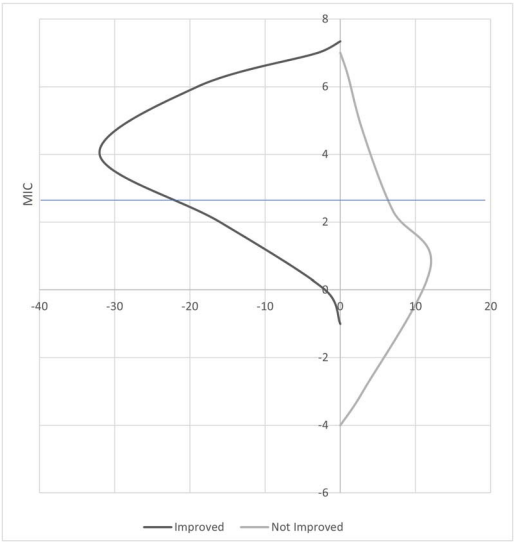

Figure 3. Visual anchor-based distribution of minimal important change (MIC) scores. Shown is the distribution of change scores on the Patient-Specific Functional Scale (PSFS) 2.0 of patients who reported an important improvement ($n = 72$) (left) compared with those who reported no important change ($n = 23$) (right) on the anchor (Global Perceived Effect scale). The left lower quadrant below the line represents the misclassified patients who felt importantly improved but were not classified as such by the PSFS 2.0 change score. The right upper quadrant represents the patients who were misclassified as they considered themselves not importantly improved but, according to their PSFS 2.0 change score, were classified as importantly improved. The blue horizontal line represents the minimal important change value.

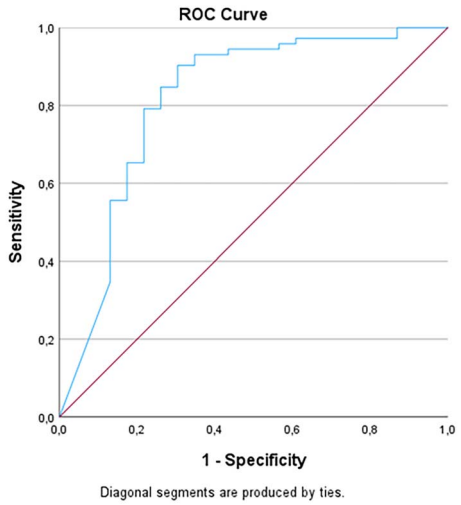

Figure 4. Receiver operating characteristic (ROC) curve for the change scores of the Patient-Specific Functional Scale 2.0. Diagonal segments are produced by ties.

Methodological Considerations

The sample size ($n = 100$), with very little loss to follow-up ($n = 5$) is one of the strengths of our study, as is the adherence to both the COSMIN reporting guideline for studies on measurement properties as well as the Guidelines for Reporting Reliability and Agreement Studies to ensure methodological rigor.

The set order in which participants received and answered the questionnaires was due to limitations in the Limesurvey software and this could be seen as a limitation. However, we do not expect biased interaction, since participants were first required to report their own individual top 3 activity limitations in the PSFS 2.0. Only if they were unable to report 3 were they directed to the list of examples in the PSFS 2.0, and only after completing the PSFS 2.0 in full were they exposed to the 10 activities on the NDI.

Conclusion

The PSFS 2.0 is a reliable and responsive PROM in patients with nonspecific neck pain. The minimal important change is higher than the SDC, suggesting that the PSFS 2.0 can detect important change and distinguish it from measurement error at an individual level on the basis of single measurements in a patient population with neck pain.

Author Contributions

Erik Thoomes (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Software [equal], Supervision [equal], Validation [equal], Visualization [equal], Writing—original draft [lead], Writing—review & editing [lead]), Joshua A. Cleland (Formal analysis [equal], Writing—original draft [equal], Writing—review & editing [equal]), Deborah Falla (Formal analysis [equal], Investigation [equal], Methodology [equal], Supervision [equal], Validation [equal], Writing—original draft [equal], Writing—review & editing [equal]), Jasper Bier (Formal analysis [equal], Validation [equal], Writing—original draft [equal], Writing—review & editing [equal]), and Marloes de Graaf (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Software [equal], Supervision [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal])

Ethics Approval

The Medical Ethic Center in Rotterdam approved this study (MEC-2018-129).

Funding

There are no funders to report for this study.

Clinical Trial Registration

This study was registered in the Netherlands Trial Register (NTR7463).

Data Availability

Data are available from the authors upon request.

Disclosures

The authors completed the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported no conflicts of interest.

References

1. Stratford P, Gill C, Westaway M, Binkley J. Assessing disability and change on individual patients: a report of a patient-specific measure. *Physiother Can*. 1995;47:258–263. https://doi.org/10.3138/ptc.47.4.258.
2. Macdermid JC, Walton DM, Côté P et al. Use of outcome measures in managing neck pain: an international multidisciplinary

survey. *Open Orthop J*. 2013;7:506–520. https://doi.org/10.2174/1874325001307010506.
3. Maissan F, Pool J, Stutterheim E, Wittink H, Ostelo R. Clinical reasoning in unimodal interventions in patients with nonspecific neck pain in daily physiotherapy practice, a Delphi study. *Musculoskelet Sci Pract*. 2018;37:8–16. https://doi.org/10.1016/j.msksp.2018.06.001.
4. Thoomes-de Graaf M, Fernández-De-Las-Peñas C, Cleland JA. The content and construct validity of the modified patient specific functional scale (PSFS 2.0) in individuals with neck pain. *J Man Manip Ther*. 2020;28:49–59. https://doi.org/10.1080/10669817.2019.1616394.
5. Nazari G, Bobos P, Lu Z, Reischl S, MacDermid JC. Psychometric properties of patient-specific functional scale in patients with upper extremity disorders. A systematic review. *Disabil Rehabil*. 2022;44:2958–2967. https://doi.org/10.1080/09638288.2020.1851784.
6. Koehorst ML, Trijffel VAN, Lindeboom R. Evaluative measurement properties of the patient-specific functional scale for primary shoulder complaints in physical therapy practice. *J Orthop Sports Phys Ther*. 2014;44:595–603. https://doi.org/10.2519/jospt.2014.5133.
7. Beurskens AJ, Vet DE, Koke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain*. 1996;65:71–76. https://doi.org/10.1016/0304-3959(95)00149-2.
8. Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther*. 1998;27:331–338. https://doi.org/10.2519/jospt.1998.27.5.331.
9. Horn KK, Jennings S, Richardson G, van Vliet D, Hefford C, Abbott JH. The patient-specific functional scale: psychometrics, clinimetrics, and application as a clinical outcome measure. *J Orthop Sports Phys Ther*. 2012;42:30–D17. https://doi.org/10.2519/jospt.2012.3727.
10. Cleland JA, Fritz JM, Whitman JM, Palmer JA. The reliability and construct validity of the neck disability index and patient specific functional scale in patients with cervical radiculopathy. *Spine (Phila Pa 1976)*. 2006;31:598–602. https://doi.org/10.1097/01.brs.0000201241.90914.22.
11. Young IA, Cleland JA, Michener LA, Brown C. Reliability, construct validity, and responsiveness of the neck disability index, patient-specific functional scale, and numeric pain rating scale in patients with cervical radiculopathy. *Am J Phys Med Rehabil*. 2010;89:831–839. https://doi.org/10.1097/PHM.0b013e3181ec98e6.
12. Cleland JA, Childs JD, Whitman JM. Psychometric properties of the neck disability index and numeric pain rating scale in patients with mechanical neck pain. *Arch Phys Med Rehabil*. 2008;89:69–74. https://doi.org/10.1016/j.apmr.2007.08.126.
13. Abbott JH, Schmitt J. Minimum important differences for the patient-specific functional scale, 4 region-specific outcome measures, and the numeric pain rating scale. *J Orthop Sports Phys Ther*. 2014;44:560–564. https://doi.org/10.2519/jospt.2014.5248.
14. Thoomes-de Graaf M, Thoomes E, Fernández-de-las-Peñas C, Plaza-Manzano G, Cleland JA. Normative values of cervical range of motion for both children and adults: a systematic review. *Musculoskelet Sci Pract*. 2020;49:102182. https://doi.org/10.1016/j.msksp.2020.102182.
15. Kottner J, Audigé L, Brorson S et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64:96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002.
16. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Qual Life Res*. 2021;30:2197–2218. https://doi.org/10.1007/s11136-021-02822-4.
17. Guzman J, Hurwitz EL, Carroll LJ et al. A new conceptual model of neck pain: linking onset, course, and care: the bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *J Manip Physiol Ther*. 2009;32:S17–S28. https://doi.org/10.1016/j.jmpt.2008.11.007.
18. Mokkink LB, Terwee CB, Knol DL et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22. https://doi.org/10.1186/1471-2288-10-22.
19. Terwee CB, Prinsen CAC, Chiarotto A et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27:1159–1170. https://doi.org/10.1007/s11136-018-1829-0.
20. Pool JJ, Ostelo RWJG, Hoving JL, Bouter LM, de Vet HCW. Minimal clinically important change of the neck disability index and the numerical rating scale for patients with neck pain. *Spine (Phila Pa 1976)*. 2007;32:3047–3051. https://doi.org/10.1097/BRS.0b013e31815cf75b.
21. Kovacs FM, Abraira V, Royuela A et al. Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. *BMC Musculoskelet Disord*. 2008;9:43. https://doi.org/10.1186/1471-2474-9-43.
22. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain*. 1986;27:117–126. https://doi.org/10.1016/0304-3959(86)90228-9.
23. Hjermstad MJ, Fayers PM, Haugen DF et al. 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 Manag*. 2011;41:1073–1093. https://doi.org/10.1016/j.jpainsymman.2010.08.016.
24. Blanpied PR, Gross AR, Elliott JM et al. Neck pain: revision 2017. *J Orthop Sports Phys Ther*. 2017;47:A1–a83. https://doi.org/10.2519/jospt.2017.0302.
25. Bier JD, Scholten-Peeters WGM, Staal JB et al. Clinical practice guideline for physical therapy assessment and treatment in patients with nonspecific neck pain. *Phys Ther*. 2018;98:162–171. https://doi.org/10.1093/ptj/pzx118.
26. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980;66:271–273.
27. Vernon H, Mior S. The neck disability index: a study of reliability and validity. *J Manip Physiol Ther*. 1991;14:409–415.
28. En MC, Clair DA, Edmondston SJ. Validity of the neck disability index and neck pain and disability scale for measuring disability associated with chronic, non-traumatic neck pain. *Man Ther*. 2009;14:433–438. https://doi.org/10.1016/j.math.2008.07.005.
29. Riddle DL, Stratford PW. Use of generic versus region-specific functional status measures on patients with cervical spine disorders. *Phys Ther*. 1998;78:951–963. https://doi.org/10.1093/ptj/78.9.951.
30. Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B. Using the neck disability index to make decisions concerning individual patients. *Physiother Can*. 1999;51:107.
31. Ailliet L, Knol DL, Rubinstein SM, de Vet HCW, van Tulder MW, Terwee CB. Definition of the construct to be measured is a prerequisite for the assessment of validity. The neck disability index as an example. *J Clin Epidemiol*. 2013;66:775–782.e2 quiz 782 e1-2. https://doi.org/10.1016/j.jclinepi.2013.02.005.
32. Hains F, Waalen J, Mior S. Psychometric properties of the neck disability index. *J Manip Physiol Ther*. 1998;21:75–80.
33. Hoving JL, O'Leary EF, Niere KR, Green S, Buchbinder R. Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. *Pain*. 2003;102:273–281. https://doi.org/10.1016/S0304-3959(02)00406-2.
34. Young BA, Walker MJ, Strunce JB, Boyles RE, Whitman JM, Childs JD. Responsiveness of the neck disability index in patients with mechanical neck disorders. *Spine J*. 2009;9:802–808. https://doi.org/10.1016/j.spinee.2009.06.002.
35. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic

review. *Qual Life Res*. 2012;21:659–670. https://doi.org/10.1007/s11136-011-9965-9.
36. Schellingerhout JM, Heymans MW, Verhagen AP, de Vet HC, Koes BW, Terwee CB. Measurement properties of translated versions of neck-specific questionnaires: a systematic review. *BMC Med Res Methodol*. 2011;11:87. https://doi.org/10.1186/1471-2288-11-87.
37. Jorritsma W, de Vries GE, Dijkstra PU, Geertzen JHB, Reneman MF. Neck pain and disability scale and neck disability index: validity of Dutch language versions. *Eur Spine J*. 2012;21:93–100. https://doi.org/10.1007/s00586-011-1920-5.
38. Streiner DL, Norman GR. *Health Measurement Scales A Practical Guide to the Development and Use*. Oxford, UK: Oxford University Press; 2008.
39. Kamper SJ, Ostelo RW, Knol DL, Maher CG, Vet HC DE, Hancock MJ. Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol*. 2010;63:760–766.e1. https://doi.org/10.1016/j.jclinepi.2009.09.009.
40. Weenink JW, Braspenning J, Wensing M. Patient reported outcome measures (PROMs) in primary care: an observational pilot study of seven generic instruments. *BMC Fam Pract*. 2014;15:88. https://doi.org/10.1186/1471-2296-15-88.
41. Luijsterburg PA, Verhagen AP, Ostelo RWJG et al. Physical therapy plus general practitioners' care versus general practitioners' care alone for sciatica: a randomised clinical trial with a 12-month follow-up. *Eur Spine J*. 2008;17:509–517. https://doi.org/10.1007/s00586-007-0569-6.
42. Jorritsma W, Dijkstra PU, de Vries GE, Geertzen JHB, Reneman MF. Detecting relevant changes and responsiveness of neck pain and disability scale and neck disability index. *Eur Spine J*. 2012;21:2550–2557. https://doi.org/10.1007/s00586-012-2407-8.
43. Soer R, Reneman MF, Vroomen PCAJ, Stegeman P, Coppes MH. Responsiveness and minimal clinically important change of the pain disability index in patients with chronic back pain. *Spine (Phila Pa 1976)*. 2012;37:711–715. https://doi.org/10.1097/BRS.0b013e31822c8a7a.
44. Demoulin C, Ostelo R, Knottnerus JA, Smeets RJEM. What factors influence the measurement properties of the Roland-Morris disability questionnaire? *Eur J Pain*. 2010;14:200–206. https://doi.org/10.1016/j.ejpain.2009.04.007.
45. Vernon DC. *The Neck Disability Index (NDI). An Informal "Blurb" from the Author*. Accessed January 27, 2023. http://www.chiro.org/forms/NDI_Explain.shtml.
46. Vet HC DE, Terwee CB, Mokkink LB, Knol DL. *Practical Guides to Biostatistics and Epidemiology. Measurement in Medicine*. Cambridge, UK: Cambridge University Press; 2011.
47. Nunally JC. *Psychometric Theory*. 3rd ed. New York: McGraw-Hill; 1994.
48. Terwee CB, Bot SDM, de Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.
49. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–160. https://doi.org/10.1177/096228029900800204.
50. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59:1033–1039. https://doi.org/10.1016/j.jclinepi.2005.10.015.
51. Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. *J Clin Epidemiol*. 2017;83:90–100. https://doi.org/10.1016/j.jclinepi.2016.12.015.
52. Terluin B, Eekhout I, Terwee CB, de Vet HCW. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. *J Clin Epidemiol*. 2015;68:1388–1396. https://doi.org/10.1016/j.jclinepi.2015.03.015.
53. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge, UK: Cambridge University Press; 2011.
54. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001;94:149–158. https://doi.org/10.1016/S0304-3959(01)00349-9.
55. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol*. 1997;50:869–879. https://doi.org/10.1016/S0895-4356(97)00097-8.
56. Burnand B, Kernan WN, Feinstein AR. Indexes and boundaries for "quantitative significance" in statistical decisions. *J Clin Epidemiol*. 1990;43:1273–1284. https://doi.org/10.1016/0895-4356(90)90093-5.
57. Jette DU, Halbert J, Iverson C, Miceli E, Shah P. Use of standardized outcome measures in physical therapist practice: perceptions and applications. *Phys Ther*. 2009;89:125–135. https://doi.org/10.2522/ptj.20080234.
58. Thoomes-de Graaf M, Scholten-Peeters W, Karel Y, Verwoerd A, Koes B, Verhagen A. One question might be capable of replacing the shoulder pain and disability index (SPADI) when measuring disability: a prospective cohort study. *Qual Life Res*. 27:401–410.
59. Verwoerd AJH, Luijsterburg PAJ, Timman R, Koes BW, Verhagen AP. A single question was as predictive of outcome as the Tampa scale for Kinesiophobia in people with sciatica: an observational study. *J Phys*. 2012;58:249–254. https://doi.org/10.1016/S1836-9553(12)70126-1.
60. Austin DC, Torchia MT, Werth PM, Lucas AP, Moschetti WE, Jevsevar DS. A one-question patient-reported outcome measure is comparable to multiple-question measures in total knee arthroplasty patients. *J Arthroplast*. 2019;34:2937–2943. https://doi.org/10.1016/j.arth.2019.07.023.
61. O'Connor CM, Ring D. Correlation of single assessment numeric evaluation (SANE) with other patient reported outcome measures (PROMs). *Arch Bone Jt Surg*. 2019;7:303–306.
62. Torchia MT, Austin DC, Werth PM, Lucas AP, Moschetti WE, Jevsevar DS. A SANE approach to outcome collection? Comparing the performance of single- versus multiple-question patient-reported outcome measures after total hip arthroplasty. *J Arthroplast*. 2020;35:S207–S213. https://doi.org/10.1016/j.arth.2020.01.015.
63. Yalçinkaya G, Kara B, Arda MN. Cross-cultural adaptation, reliability and validity of the Turkish version of patient-specific functional scale in patients with chronic neck pain. *Turk J Med Sci*. 2020;50:824–831. https://doi.org/10.3906/sag-1905-91.
64. Nakamaru K, Aizawa J, Koyama T, Nitta O. Reliability, validity, and responsiveness of the Japanese version of the patient-specific functional scale in patients with neck pain. *Eur Spine J*. 2015;24:2816–2820. https://doi.org/10.1007/s00586-015-4236-z.
65. Pathak A, Wilson R, Sharma S et al. 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;52:262–275. https://doi.org/10.2519/jospt.2022.10727.
66. Sharma S, Palanchoke J, Abbott JH. Cross-cultural adaptation and validation of the Nepali translation of the patient-specific functional scale. *J Orthop Sports Phys Ther*. 2018;48:659–664. https://doi.org/10.2519/jospt.2018.7925.