Musculoskeletal Science and Practice 79 (2025) 103372

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

journal homepage: www.elsevier.com/locate/msksp

Reliability, construct validity, responsiveness and minimum clinically important difference of the numeric pain rating scale and shoulder pain and disability index in patients with subacromial pain syndrome

Ian Young [a,b,*], James Dunning [b,c], James Escaloni [b,d], Filippo Maselli [e], Joshua Prall [b,f], Firas Mourad [g], César Fernández-de-las-Peñas [h]

[a] Tybee Wellness & Osteopractic, Tybee Island, GA, USA
[b] American Academy of Manipulative Therapy Fellowship in Orthopaedic Manual Physical Therapy, Montgomery, AL, USA
[c] Montgomery Osteopractic Physical Therapy & Acupuncture, Montgomery, AL, USA
[d] Wellward Regenerative Medicine, Lexington, KY, USA
[e] Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy
[f] Lebanon Valley College, Department of Physical Therapy, Annville, PA, USA
[g] Luxembourg Health & Sport Sciences Research Institute, Differdange, Luxembourg
[h] Department of Physical Therapy, Occupational Therapy, Rehabilitation and Physical Medicine, Universidad Rey Juan Carlos, Alcorcón, Madrid, Spain

ARTICLE INFO

Keywords:
Shoulder pain
Psychometric
MCID
MDC
Rotator cuff tendinopathy

ABSTRACT

Background: The numeric pain rating scale (NPRS) and shoulder pain and disability index (SPADI) are commonly used patient-reported outcome measures (PROMs) in patients with rotator cuff tendinopathy. To date, there are gaps in the evidence supporting the clinimetric properties of these PROMs for patients treated with subacromial pain syndrome (SAPS).

Methods: A clinimetric analysis (n = 145) was performed to examine the reliability, construct validity, responsiveness, interpretability, minimal detectable change (MDC_{95}) and minimum clinically important difference (MCID) of the NPRS and SPADI for "improved" (global rating of change from +3 to +7) and "much-improved" (global rating of change from +5 to +7) patients at 3-months follow-up.

Results: The NPRS (ICC: 0.86; 95 %CI, 0.33–0.96) and SPADI (ICC: 0.79; 95 %CI 0.12–0.94) exhibited good reliability and excellent responsiveness (NPRS: area under the curve (AUC) = 0.96, 95 %CI 0.92–0.99; SPADI: AUC = 0.90, 95 %CI 0.84–0.95) in this patient population. Both outcomes demonstrated strong construct validity (Pearson's r; $p < 0.001$). The MDC_{95} was a 1.7- and 20.5-point change for the NPRS and SPADI, respectively. For the NPRS, the MCID was a 1.5-point change in the "improved" group and a 2.5-point change in the "much improved" group. For the SPADI, the MCID was an 18-point or 50 % change for the "improved" group, and a 25-point or 70 % change in the "much improved" group.

Conclusions: The NPRS and SPADI demonstrated sound clinimetric properties in patients with SAPS. The MCID exceeded measurement error in the "much improved" group. Diagnosis, type of intervention, level of improvement, and measurement error should be considered when applying the MCID.

1. Introduction

Patients presenting with shoulder pathology or mechanical dysfunction exhibit various levels of pain and disability, usually depending on the established diagnosis (McClure and Michener, 2015). Gauging and understanding appropriate levels of improvement in patients with shoulder pathology allows the clinician to set appropriate goals after an intervention (Desmeules et al., 2025). Rotator cuff tendinopathy (RC-tendinopathy) encompasses three of the more common shoulder diagnoses, including non-specific shoulder pain (NSSP), rotator cuff disorders (RCD), and subacromial pain syndrome (SAPS) (Desmeules et al., 2025; Littlewood et al., 2019; Lewis, 2016). The

* Corresponding author. Tybee Wellness & Osteopractic, 8 Horsepen Point Drive, Tybee Island, 31328, USA.
E-mail addresses: tybeewellness@gmail.com (I. Young), drjamesdunning@gmail.com (J. Dunning), james.escaloni.pt@gmail.com (J. Escaloni), masellifilippo76@gmail.com (F. Maselli), prall@lvc.edu (J. Prall), firasmourad@me.com (F. Mourad), cesar.fernandez@urjc.es (C. Fernández-de-las-Peñas).

https://doi.org/10.1016/j.msksp.2025.103372
Received 18 April 2025; Received in revised form 25 June 2025; Accepted 27 June 2025
Available online 28 June 2025

numeric pain rating scale (NPRS) and shoulder pain and disability index (SPADI) are commonly used patient-reported outcome measures (PROMs) in non-surgical patients with RC-tendinopathy (Desmeules et al., 2025; Roy et al., 2009; Furtado et al., 2022; Huang et al., 2015; Mintken et al., 2009; Michener et al., 2011; Thoomes-de et al., 2017; Dabija and Jain, 2019; Buchbinder et al., 2020; Dabija et al., 2019). However, inadequate clinimetric methodology for reliability, measurement error, and responsiveness has been suggested with these and other commonly used PROMs in RC-tendinopathy (Furtado et al., 2022; Huang et al., 2015). Clinimetric studies supporting the reliability, using intraclass correlation coefficients ($ICC_{2,1}$) of the NPRS and SPADI are scarce in this patient population; nevertheless, it has been reported to be 0.74 for the NPRS (Mintken et al., 2009), and ranges from 0.66 to 0.95 for the SPADI (Roy et al., 2009; Spanou et al., 2020). The minimum clinically important difference (MCID) ranges from 1.1 to 6.3 for the NPRS (Desmeules et al., 2025; Mintken et al., 2009; Michener et al., 2011; Dabija and Jain, 2019; Tashjian et al., 2009) and 8 to 20 points for the SPADI (Thoomes-de et al., 2017; Dabija and Jain, 2019; Buchbinder et al., 2020; Spanou et al., 2020; Ekeberg et al., 2010; Williams et al., 1995). Notably, this wide range of threshold values for the MCID gives a high degree of ambiguity to the selection of an appropriate change score for clinical use or during the design of clinical trials (Desmeules et al., 2025; Dabija and Jain, 2019). Furthermore, it has been recommended that the intervention, outcome measure and patient population should be similar to the setting in which the MCID cut-off was previously established (Draak et al., 2019). Acquiring more recent data to compare and contrast clinimetric outcomes of different shoulder diagnostic and treatment subcategories should also be addressed for future clinical trials. Therefore, the purpose of this study was to comprehensively examine the reliability, construct validity, and responsiveness of the NPRS and SPADI in patients with SAPS that were successfully treated with outpatient physical therapy. The results of this analysis may benefit clinicians and the design of future clinical trials seeking to use these outcome measures with a similar patient population, intervention, and follow-up time frames.

2. Methods

This study is a secondary analysis of a large North American multicenter randomized clinical trial (Dunning et al., 2021) that investigated the effects of two different physical therapy interventions in patients with SAPS. Consecutive individuals with SAPS (N = 145) from 14 outpatient physical therapy clinics in twelve U.S. states were screened for eligibility and recruited over a 22-month period (June 2017 to April 2019). In the original trial, patients were randomized to receive cervicothoracic spine and upper rib thrust manipulation and electrical dry needling or nonthrust peripheral joint/soft tissue mobilization, exercise, and electrotherapy (Dunning et al., 2021). To be eligible, patients had to report a primary complaint of anterolateral shoulder pain lasting longer than 6 weeks, have a positive Neer impingement test (Tate et al., 2010), and/or a positive Hawkins-Kennedy test (Tate et al., 2010). In addition, patients had to report 1 or more of the following symptoms: (1) a painful arc with active shoulder elevation(Tate et al., 2010), 2) pain with resisted shoulder external rotation at 90° of abduction (Tate et al., 2010), or (3) pain with resisted shoulder abduction in the empty-can test position (Tate et al., 2010). The inclusion and exclusion criteria are previously described in detail elsewhere (Dunning et al., 2021). The NPRS (2) and SPADI (1) were collected in all patients at baseline, 1 week, 4 weeks, and 3 months. Perceived recovery using the Global Rating of Change Scale (GROC)(Jaeschke et al.) was collected at all follow-up points. To investigate the clinimetric properties of all the outcome measures, both the manipulation/dry needling group and mobilization/exercise group completing the trial were collapsed into a single cohort for this secondary analysis using the 3-month follow-up.

2.1. Outcome measures

The NPRS was used to capture the patient's level of shoulder pain. Patients were asked to indicate the intensity of their current pain level using an 11-point scale, ranging from 0 (no pain) to 10 (worst pain imaginable) (Jensen et al., 1986). Reliability (ICC) of the NPRS in patients with NSSP has been reported to be 0.74 (Mintken et al., 2009). The minimal detectable change (MDC) has been reported to be 2.1 points (Mintken et al., 2009), while the MCID ranges from 1.1 to 2.2 points (Mintken et al., 2009; Michener et al., 2011) in NSSP/RCD.

The SPADI is a commonly used instrument for assessing self-rated disability in patients with NSSP and RCD (Roy et al., 2009; Breckenridge and McAuley, 2011). The SPADI contains 13 items that assess two domains; a 5-item subscale that measures pain and an 8-item subscale that measures disability. A mean is taken of the two subscales to give a total score out of 100, higher score indicating greater impairment or disability (Breckenridge and McAuley, 2011). The reliability (ICC) of the SPADI ranges from 0.66 to 0.95 in patients with NSSP/RCP (Roy et al., 2009). The minimal detectable change (MDC) was found to be from 19.7 to 22.5 points (Thoomes-de et al., 2017; Ekeberg et al., 2010), whilst the MCID was found to be 8 to 20 points in patients with NSSP/RCD (Thoomes-de et al., 2017; Ekeberg et al., 2010; Williams et al., 1995; Paul et al., 2004) Notably, the clinimetric properties of the SPADI have not yet been adequately established in a large sample of patients specifically diagnosed with and treated for SAPS.

Patients also completed a 15-point Global Rating of Change (GROC) scale described by Jaeschke et al. (Jaeschke et al.) to rate their own perception of improved function. The scale ranges from −7 (a very great deal worse) to 0 (about the same) to +7 (a very great deal better). The MCID for the GROC has not been specifically reported, but scores of +4 and + 5 have typically been indicative of moderate changes in patient status (Jaeschke et al.) Scores of +3 to +5 are commonly used to identify "improved" versus "stable" patients, +5 to +7 to identify that are "much improved", and 0 to indicate "no change" (Young et al., 2010, 2018, 2019, 2025a).

2.2. Data analysis

We categorized patients into 4 mutually exclusive groups at the 3-month follow-up based on their GROC scores. Those scoring 0 were considered to have "no change", those scoring ranging from −2 to +2 were considered clinically "stable'' (minimal to no change); those scoring +3 to +7 were considered clinically "improved" (at least somewhat better), and those scoring +5 to +7 were considered clinically "much improved" (at least a good deal better). Similar categorization has been used in previous clinimetric studies (Thoomes-de et al., 2017; Young et al., 2019; Young et al., 2025b).

Test-retest reliability was examined for the NDI and NPRS using "stable" patients. Intraclass $ICC_{2,1}$ for the NPRS and SPADI were calculated for the two groups of patients who were classified as having "no change" (GROC = 0, n = 13) by comparing scores at the initial examination with those at the 3-month follow-up. The "unchanged" group was used in this analysis, as it may be more representative of the true (no-change) test-retest reliability over time (Young et al., 2018). The $ICC_{2,1}$ was calculated according to procedures described by Shrout and Fleiss (Shrout and Fleiss, 1979; Koo and Li, 2016). Values < 0.50 indicate poor reliability, while values between 0.50 and 0.75, between 0.75 and 0.90, and >0.90 denote moderate, good, and excellent agreement, respectively (Koo and Li, 2016).

Construct validity of the NPRS and SPADI was examined by comparing the change in outcome scores for the "stable'', "improved", and "much improved" groups using separate, two-way analyses of variance (ANOVA) for repeated measures at baseline and reevaluation. We hypothesized that "stable'' patients in each group would have NPRS and SPADI intake values that had very little change, whereas patients classified in the improved categories would demonstrate a significant

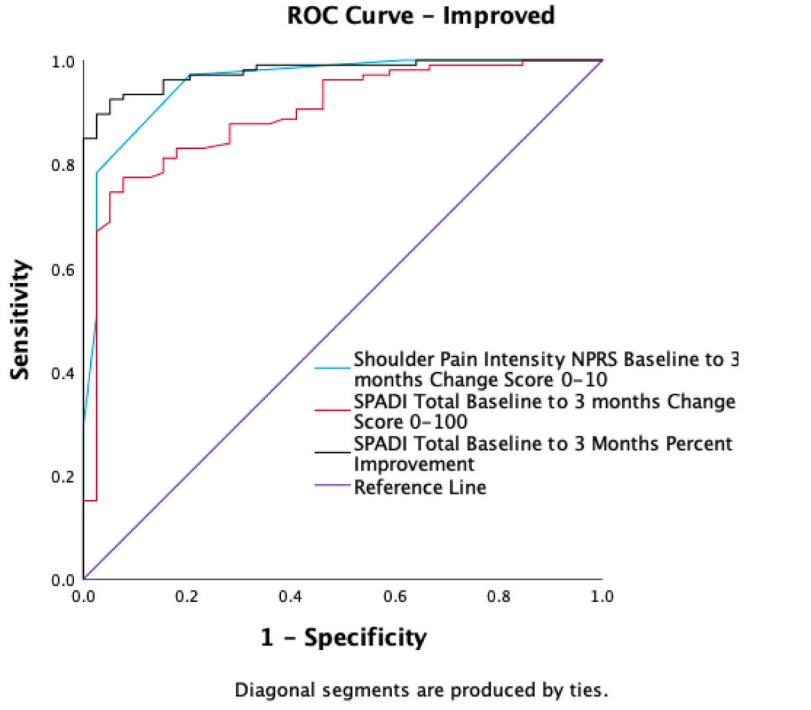
change in values. This would be represented by a significant group × time interaction. Pearson's correlation (r) was also calculated to examine the linear association between the outcome measures (NPRS, SPADI) and the criterion measure used for perceived improvement (GROC). Values between 0.30 and 0.49 indicate a moderate correlation, while values between 0.50 and 1.0 indicate a strong correlation (Schober et al., 2018).

The responsiveness and interpretability of the NPRS and SPADI was assessed using the clinically "stable", and "improved" groups at the 3-month follow-up point. Receiver operator characteristic (ROC) curves (Hanley and McNeil, 1982; Nahm, 2022) were constructed by plotting sensitivity values (true-positive rate) on the y axis and 1-specificity values (false-positive rate) on the x axis for each level of change score. Separate ROC curves were constructed for improved and much-improved groups of the NPRS and SPADI (Figs. 1–2). Additionally, the percentage of change score for the SPADI was also included in the analysis. The area under the curve (AUC) and the 95 % CI were obtained as a method for determining the ability of each measure to distinguish improved patients from stable patients in each category. An AUC of 0.50 indicates that the measure has no diagnostic accuracy beyond chance, whereas a value of 1 suggests perfect accuracy (Hanley and McNeil, 1982; Nahm, 2022). MCID, the smallest difference that patients perceive as beneficial, was calculated by identifying the point on the ROC curve nearest to the upper left-hand corner (maximizing sensitivity and specificity values), which is considered to be the best cutoff score for distinguishing improved and stable patients and much improved and stable patients (Hanley and McNeil, 1982; Nahm, 2022; Beaton et al., 2001). Sensitivity and specificity values for the selected cutoff scores were also calculated. These procedures were also performed for the percentage change of the SPADI. The MDC, the amount of change that must be observed before the change can be considered to have exceeded measurement error, was calculated by determining the standard error of measurement (SEM) for the NPRS and SPADI in the "no-change" group (n = 13) (Beaton et al., 2001). The SEM was estimated using the formula (SD*$\sqrt{}$ (1- reliability coefficient) where SD is the pooled standard deviation in the group "no change" group. The SEM was multiplied by 1.96 to determine the 95 % CI (MDC_{95}) (Swets, 1988). This value was multiplied by the square root of 2 to account for the errors taken with repeated measurements (Swets, 1988).

3. Results

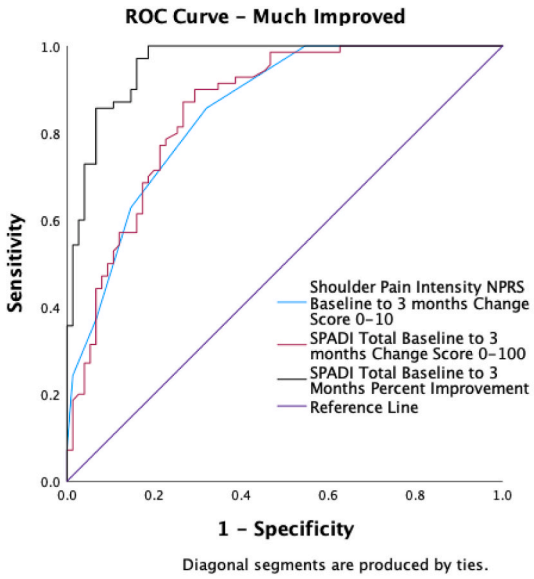
One hundred forty-five patients satisfied the inclusion and exclusion criteria, completed the study, and were included in data analysis. Baseline characteristics are located in Table 1. The mean GROC score for all patients included in the analysis at the 3-month follow-up was +3.8 (SD + 2.6). The mean GROC score for the "stable", "improved" and "much-improved" groups was +0.5 (SD + 1.2), +5.1 (SD + 1.4), and 6.0 (SD + 0.8), respectively. At the 3-month follow-up, 106 (73 %) patients were classified as "improved", 70 (48 %) were "much-improved", 37 (26 %) remained "stable", and 13 (9 %) reported "no change". There was a significant difference ($p < 0.001$) in mean change scores and percentage of change score (NPRS and SPADI) between "stable" vs. "improved" and "stable" vs. "much-improved" groups at the 3-month follow-up (Table 2).

The $ICC_{2,1}$ and MDC_{95} values are reported in Table 3. At the 3-month follow-up, the NPRS (ICC: 0.86, 95 %CI 0.33–0.96) and SPADI (ICC: 0.79, 95 %CI 0.12–0.94) exhibited good reliability. The MDC_{95} was 1.7 points for the NPRS and 20.5 points for the SPADI. The NPRS and the SPADI demonstrated excellent responsiveness (AUC range = 0.85 to 0.98; Table 3). The MCID and the sensitivity/specificity associated with each cutoff score are also located in Table 3. For the NPRS, the MCID was 1.5 points for the "improved" group and 2.5 points for the "much-improved" group. For the SPADI, the MCID was an 18-point or 50 % change and a 25-point or 70 % change for the "improved" and "much-improved" groups, respectively. All PROMs also exhibited adequate construct validity with Pearsons's correlation coefficient (Pearsons *r*: range 0.64–0.80; $p < 0.001$; Table 4).



NPRS=numeric pain rating scale (0-10), SPADI=shoulder pain and disability index (total score 0-100), Improved = Global Rating of change (+3 to +7)

Fig. 1. ROC Curve of Improved Patients
NPRS = numeric pain rating scale (0–10), SPADI = shoulder pain and disability index (total score 0–100), Improved = Global Rating of change (+3 to +7).



NPRS=numeric pain rating scale (0-10), SPADI=shoulder pain and disability index (total score 0-100), Improved = Global Rating of change (+5 to +7)

Fig. 2. ROC Curve of Much Improved Patients
NPRS = numeric pain rating scale (0–10), SPADI = shoulder pain and disability index (total score 0–100), Improved = Global Rating of change (+5 to +7).

Table 1
Baseline characteristics.

	N = 145 Mean (SD)
Gender: Male/Female	71/74
Age: yrs.	47 (±15.7)
Weight: kg	77.3 (±16.8)
Months with Shoulder Pain	3.5 (±5.9)
Numeric Pain Rating Scale	5.3 (±1.5)
Should Pain and Disability Questionnaire -Total	44.1 (±15.4)
Average Number of Treatment Visits	10.1(±2.2)
Medication Intake ≥ 1x/week (%)	67 (46.2)

Table 2
Difference between change scores and percent improvement from baseline to 3-months.

Outcome Measure	Improved GROC (+3 to +7) N = 106 Mean (SD)	Stable GROC (−2 to +2) N = 37 Mean (SD)	Mean Difference (95 % CI)	*P*
NPRS	3.8 (1.5)	0.78 (1.1)	3.02 (2.6; 3.5)	*P<0.001*
SPADI	31.7 (14.5)	11.3 (9.6)	20.4 (15.3; 25.5)	*P<0.001*
SPADI % Change Score	75.2 (19.8)	20.9 (16.1)	54.3 (47.2; 61.4	*P<0.001*
	Much Improved GROC (+5 to +7) N = 70 Mean (SD)	**Stable** GROC (−2 to +2) N = 37 Mean (SD)	Mean Difference (95 % CI)	*P*
NPRS	4.2 (1.6)	0.78 (1.1)	3.4 (2.8; 4.0)	*P<0.001*
SPADI	35.9 (13.7)	11.3 (9.6)	24.6 (19.6; 29.6)	*P<0.001*
SPADI % Change Score	85.3 (11.3)	20.9 (16.1)	64.4 (59.1; 69.7)	*P<0.001*

NPRS = numeric pain rating scale (0–10), SPADI=Shoulder pain and disability index (total score 0–100), GROC = global rating of change (−7 to +7), SD = standard deviation, CI = confidence interval.

4. Discussion

This study examined the clinimetric properties of two commonly used shoulder PROMs in patients clinically diagnosed with SAPS and successfully treated with physical therapy, including extremity mobilization, exercise, spinal/rib manipulation, and electrical dry needling. All outcome measures exhibited good reliability, proper construct validity, and a high level of responsiveness (Tables 2–4). Overall, our analysis suggests that these commonly used PROMs of pain and disability have suitable clinimetric properties in patients with SAPS (Tables 2–4, Figs. 1–2).

4.1. Reliability

The NPRS exhibited good reliability ($ICC_{2,1}$ = 0.86) in the current analysis. (Table 3). Prior studies on patients with NSSP reported somewhat lower reliability, (Michener et al., 2011). These differences could be related to Michener et al. (2011), using the MDC_{90} of Penn Shoulder Score as the criterion measure, instead of the GROC used in the current study. The SPADI also exhibited good reliability ($ICC_{2,1}$ = 0.79) in this cohort of patients with SAPS. Prior studies on RC-tendinopathy have reported wide ranges in ICC values from 0.66 to 0.95 for the SPADI, with large variations in sample sizes (Roy et al., 2009; Spanou et al., 2020).

4.2. Responsiveness & interpretability

The NPRS and SPADI demonstrated a high level of responsiveness (AUC) in both the "improved" (AUC: NPRS = 0.96, SPADI = 0.90) and "much improved" (AUC: NPRS = 0.85, SPADI = 0.86) groups (Table 3; Figs. 1–2). Prior analyses have resulted in somewhat lower responsiveness of the NPRS (AUC from 0.67 to 0.74) (Mintken et al., 2009; Michener et al., 2011) and comparable responsiveness for the SPADI (AUC from 0.81 to 0.87) (Thoomes-de et al., 2017; Paul et al., 2004; Chester et al., 2017) in patients with RC-tendinopathy. Interestingly, the current analysis of the SPADI using percent of change scores the responsiveness (AUC: improved = 0.98, much improved = 0.96) was even stronger with very tight confidence intervals (Table 3). Clinicians and researchers may consider future use of this calculation as well for

Table 3
Clinimetric properties of patient-rated outcome measures for Subacromial Pain Syndrome (3-month).

Outcome	AUC 95 % CI Improved (GROC +3 to +7) N = 106	MCID Sn; Sp Improved (GROC +3 to +7) N = 106	AUC 95 % CI Much Improved (GROC +5 to +7) N = 70	MCID Sn; Sp Much Improved (GROC +5 to +7) N = 70	$ICC_{2,1}$ (95 % CI) No Change (GROC = 0) n = 13	SEM	MDC_{95}
NPRS	0.96 0.92; 0.99	1.5 0.97; 0.79	0.85 0.79; 0.91	2.5 0.86; 0.68	0.86 0.33; 0.96	0.59	1.7
SPADI	0.90 0.84; 0.95	18 0.80; 0.85	0.86 0.79; 0.92	25 0.79; 0.77	0.79 0.12; 0.94	7.4	20.5
SPADI % Change Score	0.98 0.96; 0.99	50 % 0.93; 0.92	0.96 0.93; 0.99	70 % 0.87; 0.89	–	–	–

NPRS = numeric pain rating scale (0–10), SPADI = shoulder pain and disability index (total score 0–100), AUC = area under the curve, GROC = global rating of change (−7 to +7), MCID = minimally clinically important difference, Sn = sensitivity, Sp = specificity, $ICC_{2,1}$ = intraclass correlation coefficient, ICC = intraclass correlation coefficient, SEM = standard error of measure, MDC_{95} = minimal detectable change (95 % confidence interval).

Table 4
Pearson's correlation coefficient (r).

Outcome Measures	SPADI r (95 % CI)	GROC r (95 % CI)
NPRS	0.71 (0.61; 0.78) *P < 0.001*	0.80 (0.74; 0.85) *P < 0.001*
SPADI	–	0.64 (0.53; 0.72) *P < 0.001*

NPRS = numeric pain rating scale (0–10), SPADI = shoulder pain and disability index (total score 0–100), GROC = global rating of change scale (−7 to +7).

setting clinical goals and future research methodology. Overall, the NPRS and SPADI have suitable responsiveness for clinical application in patients with SAPS.

4.3. MDC & MCID

In the current analysis, the MCID of the SPADI for both the "improved" (18 points) and "much improved" groups (25 points) is consistent with findings of reported by Ekeberg et al. (2010) (20 points), and a high-quality clinimetric analysis (SPADI-Dutch) that reported a MCID of 16 and 20 points for patients with RCD who were "slightly improved" and "importantly improved", respectively (Thoomes-de et al., 2017). Additionally, the MCID calculated from the percentage of change score was 50 % change for the "improved" group and 70 % change in the "much-improved" group. These values demonstrated greater sensitivity and specificity than the MCID for point-change scores (Table 3) and can be used in addition or as an alternative to the point-change score. Thoomes-De Graaf et al. (Thoomes-de et al., 2017) also suggested using percent of change scores as a comparison, but no data analysis was done on the responsiveness (AUC) or calculation from a separate ROC curve.

The MDC_{95} of the SPADI in the current analysis was 20.5 points (Table 3), and in the same criterion-based study on the SPADI-Dutch, the MDC_{95} was found to be 22.5 points before extreme data outliers were removed from the analysis, and 19.7 points after removal of the outliers (Thoomes-de et al., 2017). Our results, without manipulation of outliers, resulted in a comparable MDC_{95} to that of the SPADI-Dutch after removal of extreme outliers. (Table 3).

The MDC_{95} for the NPRS was 1.7 points, and the MCID was 1.5 points ("improved") and 2.5 points ("much improved") in the current study. In prior studies, the MCID ranged from 1.1 to 2.2 points in patients with NSSP using the GROC as the criterion measure (Mintken et al., 2009; Michener et al., 2011). Our analysis resulted in an MCID inside this previous range for the "improved" group (1.5 points) and slightly higher in the "much-improved" group (2.5 points). Importantly, these prior studies did not use multiple improvement categories, as in the current analysis. Standardized statistical methodology as well as a more comprehensive analysis seems to be essential to avoid an over/under estimation, and misinterpretation of these clinimetric measures in the clinical setting. Furthermore, as advanced/highly skilled interventions are showing greater efficacy in treating specific musculoskeletal diagnoses, then so should the change scores for the MCID (Draak et al., 2019; Grönkvist et al., 2024; Copay et al., 2018).

4.4. Evaluating measurement error vs. MCID

Interpreting the MDC_{95} and the MCID is worthy of brief discussion. In the current analysis, the MDC_{95} was 1.7 points and 20.5 points for the NPRS and SPADI, respectively. These findings suggest that the point estimate for the MCID of the NPRS and SPADI in the "much improved" group (NPRS = 2.5 points; SPADI = 25 points) exceeded random measurement error, but not in the "improved" group (NPRS = 1.5 points; SPADI = 18 points). In essence, true change may *or* may not have occurred when applying the point estimate of MCID in the "improved" group for either PROM (Grönkvist et al., 2024; Copay et al., 2018). Interestingly, in the present analysis, choosing a more robust anchor/-level of improvement, i.e., -"much-improved" (GROC = +5 to +7), the MCID (NPRS = 2.5; SPADI = 25) may ensure MCID values that exceed and are technically free from random measurement error (Grönkvist et al., 2024; Copay et al., 2018). Similarly findings were noted on the SPADI-Dutch (Thoomes-de et al., 2017). The MCID was 16 points for the "slightly improved" and 20 points for the "importantly improved" groups, whilst the MDC was 22.5 points before, and 19.7 points after removal of extreme outliers. The MCID of 20 points ("importantly improved") lies just outside the measurement error, and was recommended by the authors (Thoomes-de et al., 2017). Without removal of the outliers from the data, the MCID in both groups would have been within/less than measurement error. Intuitively, from a clinical research perspective, choosing the correct threshold for the MCID (greater than the MDC) should be of importance, although this concept is still debated (Draak et al., 2019; Grönkvist et al., 2024/09; Copay et al., 2018).

5. Study limitations

First, although we used the best combination of inclusion/exclusion criteria and special tests to recruit patients with SAPS in the original trial (Dunning et al., 2021), this may not be all inclusive or exclusive to SAPS. Second, we did not adjust the data for outliers as was performed in the Thoomes-de Graaf et al. (Thoomes-de et al., 2017) study. The current authors, who are all practicing clinicians/clinical research scientists, believe random outliers exist in real clinical practice, and therefore should be included in data analysis. Furthermore, it has been suggested that manipulating data outliers may have profound effects on the occurrence of type 1 error (Gress et al., 2018). Third, this analysis reported on the 3-month follow-up, and not the short-term follow-up (1–2 weeks), which has been previously suggested for ICC_{95} analyses (Streiner et al., 2014). Although longer-term clinimetric assessment is needed, mid-term follow-up may have had an effect on recall bias and the results of the $ICC_{2,1}$/MDC_{95}. Lastly, the $ICC_{2,1}$ for both instruments should likely be interpreted with caution, secondary to somewhat wider

confidence intervals.

6. Conclusion

The NPRS and SPADI seem well-suited as short-term self-report outcome measures for up to 3-months follow-up in patients with SAPS. The NPRS and SPADI both exhibited good reliability, strong construct validity, and ideal responsiveness. Clinicians and researchers should expect a 1.5-point change in "improved" patients and 2.5-point change in "much-improved" patients on the NPRS to be considered clinically meaningful. For the SPADI, an 18-point or 50 % change ("improved") and 25-point or 70 % change ("much-improved") should be expected to be considered clinically meaningful. The MCID lies outside measurement error (MDC_{95}) in both PROMs when selecting a more robust level of improvement on the GROC. Notably, shoulder diagnosis, type of intervention, desired level of improvement, and measurement error should all be considered when applying the appropriate MCID.

CRediT authorship contribution statement

Ian Young: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **James Dunning:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **James Escaloni:** Writing – review & editing, Project administration, Methodology. **Filippo Maselli:** Writing – review & editing, Methodology, Formal analysis. **Joshua Prall:** Writing – review & editing, Methodology, Formal analysis. **Firas Mourad:** Writing – review & editing, Methodology. **César Fernández-de-las-Peñas:** Writing – review & editing, Project administration, Methodology, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors noted above have no competing interests to declare.

References

Beaton, D.E., Bombardier, C., Katz, J.N., et al., 2001. Looking for important change/differences in studies of responsiveness. OMERACT MCID working group. Outcome measures in rheumatology. Minimal clinically important difference. J. Rheumatol. 28 (2), 400–405.

Breckenridge, J.D., McAuley, J.H., 2011. Shoulder pain and disability index (SPADI). J. Physiother. 57 (3), 197. https://doi.org/10.1016/S1836-9553(11)70045-5.

Buchbinder, R., Ramiro, S., Huang, H., Gagnier, J.J., Jia, Y., Whittle, S.L., 2020. Measures of adult shoulder function. Arthritis Care Res. 72 (Suppl. 10), 250–293. https://doi.org/10.1002/acr.24230.

Chester, R., Jerosch-Herold, C., Lewis, J., Shepstone, L., 2017. The SPADI and QuickDASH are similarly responsive in patients undergoing physical therapy for shoulder pain. J. Orthop. Sports Phys. Ther. 47 (8), 538–547. https://doi.org/10.2519/jospt.2017.7195.

Copay, A.G., Chung, A.S., Eyberg, B., Olmscheid, N., Chutkan, N., Spangehl, M.J., 2018. Minimum clinically important difference: current trends in the orthopaedic literature, part I: upper extremity: a systematic review. JBJS Rev. 6 (9), e1. https://doi.org/10.2106/jbjs.Rvw.17.00159.

Dabija, D.I., Jain, N.B., 2019. Minimal clinically important difference of shoulder outcome measures and diagnoses: a systematic review. Am. J. Phys. Med. Rehabil. 98 (8), 671–676. https://doi.org/10.1097/phm.0000000000001169.

Dabija, D.I., Pennings, J.S., Archer, K.R., et al., 2019. Which is the best outcome measure for rotator cuff tears? Clin. Orthop. Relat. Res. 477 (8), 1869–1878. https://doi.org/10.1097/corr.0000000000000800.

Desmeules, F., Roy, J.S., Lafrance, S., et al., 2025. Rotator cuff tendinopathy diagnosis, nonsurgical medical care, and rehabilitation: a clinical practice guideline. J. Orthop. Sports Phys. Ther. 55 (4), 235–274. https://doi.org/10.2519/jospt.2025.13182.

Draak, T.H.P., de Greef, B.T.A., Faber, C.G., Merkies, I.S.J., 2019. The minimum clinically important difference: which direction to take. Eur. J. Neurol. 26 (6), 850–855. https://doi.org/10.1111/ene.13941.

Dunning, J., Butts, R., Fernandez-de-Las-Penas, C., et al., 2021. Spinal manipulation and electrical dry needling in patients with subacromial pain syndrome: a multicenter randomized clinical trial. J. Orthop. Sports Phys. Ther. 51 (2), 72–81. https://doi.org/10.2519/jospt.2021.9785.

Ekeberg, O.M., Bautz-Holter, E., Keller, A., Tveitå, E.K., Juel, N.G., Brox, J.I., 2010. A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. J. Clin. Epidemiol. 63 (5), 575–584. https://doi.org/10.1016/j.jclinepi.2009.07.012.

Furtado, R., Bobos, P., Ziebart, C., Vincent, J., MacDermid, J., 2022. Patient-reported outcome measures used for shoulder disorders: an overview of systematic reviews. J. Hand Ther. 35 (2), 174–185. https://doi.org/10.1016/j.jht.2022.03.008.

Gress, T.W., Denvir, J., Shapiro, J.I., 2018. Effect of removing outliers on statistical inference: implications to interpretation of experimental data in medical research. Marshall J Med 4 (2). https://doi.org/10.18590/mjm.2018.vol4.iss2.9.

Grönkvist, R., Vixner, L., Äng, B., Grimby-Ekman, A., 2024. Measurement error, minimal detectable change, and minimal clinically important difference of the short Form-36 health survey, hospital anxiety and depression scale, and pain numeric rating scale in patients with chronic pain. J. Pain 25 (9), 104559. https://doi.org/10.1016/j.jpain.2024.104559.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143 (1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747.

Huang, H., Grant, J.A., Miller, B.S., Mirza, F.M., Gagnier, J.J., 2015. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. Am. J. Sports Med. 43 (10), 2572–2582. https://doi.org/10.1177/0363546514565096.

Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Clinical Trial.

Jensen, M.P., Karoly, P., Braver, S., 1986. The measurement of clinical pain intensity: a comparison of six methods. Pain 27 (1), 117–126. https://doi.org/10.1016/0304-3959(86)90228-9.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15 (2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Lewis, J., 2016. Rotator cuff related shoulder pain: assessment, management and uncertainties. Man. Ther. 23, 57–68. https://doi.org/10.1016/j.math.2016.03.009.

Littlewood, C., Bateman, M., Connor, C., et al., 2019. Physiotherapists' recommendations for examination and treatment of rotator cuff related shoulder pain: a consensus exercise. Physiother. Pract. Res. 40 (2), 87–94. https://doi.org/10.3233/ppr-190129.

McClure, P.W., Michener, L.A., 2015. Staged approach for rehabilitation classification: shoulder disorders (STAR-Shoulder). Phys. Ther. 95 (5), 791–800. https://doi.org/10.2522/ptj.20140156.

Michener, L.A., Snyder, A.R., Leggin, B.G., 2011. Responsiveness of the numeric pain rating scale in patients with shoulder pain and the effect of surgical status. J. Sport Rehabil. 20 (1), 115–128. https://doi.org/10.1123/jsr.20.1.115.

Mintken, P.E., Glynn, P., Cleland, J.A., 2009. Psychometric properties of the shortened disabilities of the arm, shoulder, and hand questionnaire (QuickDASH) and numeric pain rating scale in patients with shoulder pain. J. Shoulder Elb. Surg. 18 (6), 920–926. https://doi.org/10.1016/j.jse.2008.12.015.

Nahm, F.S., 2022. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J. Anesthesiol. 75 (1), 25–36. https://doi.org/10.4097/kja.21209.

Paul, A., Lewis, M., Shadforth, M.F., Croft, P.R., Van Der Windt, D.A., Hay, E.M., 2004. A comparison of four shoulder-specific questionnaires in primary care. Ann. Rheum. Dis. 63 (10), 1293–1299. https://doi.org/10.1136/ard.2003.012088.

Roy, J.S., MacDermid, J.C., Woodhouse, L.J., 2009. Measuring shoulder function: a systematic review of four questionnaires. Arthritis Rheum. 61 (5), 623–632. https://doi.org/10.1002/art.24396.

Schober, P., Boer, C., Schwarte, L.A., 2018. Correlation coefficients: appropriate use and interpretation. Anesth. Analg. 126 (5), 1763–1768. https://doi.org/10.1213/ane.0000000000002864.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86 (2), 420–428. https://doi.org/10.1037//0033-2909.86.2.420.

Spanou, A., Mamais, I., Lamnisos, D., Stasinopoulos, D., 2020. Reliability and validity of the Greek shoulder pain and disability index in patients with shoulder pain. Disabil. Rehabil. 42 (9), 1299–1304. https://doi.org/10.1080/09638288.2018.1519728.

Streiner, D.L., Norman, G.R., Cairney, J., 2014. Health Measurement Scales: a Practical Guide to Their Development and Use. Oxford University Press.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240 (4857), 1285–1293. https://doi.org/10.1126/science.3287615.

Tashjian, R.Z., Deloach, J., Porucznik, C.A., Powell, A.P., 2009. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. J. Shoulder Elb. Surg. 18 (6), 927–932. https://doi.org/10.1016/j.jse.2009.03.021.

Tate, A.R., McClure, P.W., Young, I.A., Salvatori, R., Michener, L.A., 2010. Comprehensive impairment-based exercise and manual therapy intervention for patients with subacromial impingement syndrome: a case series. J. Orthop. Sports Phys. Ther. 40 (8), 474–493. https://doi.org/10.2519/jospt.2010.3223.

Thoomes-de Graaf, M., Scholten-Peeters, W., Duijn, E., et al., 2017. The responsiveness and interpretability of the shoulder pain and disability index. J. Orthop. Sports Phys. Ther. 47 (4), 278–286. https://doi.org/10.2519/jospt.2017.7079.

Williams Jr., J.W., Holleman Jr., D.R., Simel, D.L., 1995. Measuring shoulder function with the shoulder pain and disability index. J. Rheumatol. 22 (4), 727–732.

Young, I.A., Cleland, J.A., Michener, L.A., Brown, C.J., 2010. 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. 89, 831–839.

Young, Ia, Dunning, J., Butts, R., Mourad, F., Cleland, Ja, 2018. Reliability, construct validity, and responsiveness of the neck disability index and numeric pain rating scale in patients with mechanical neck pain without upper extremity symptoms. Physiother. Theory Pract. 1–8. https://doi.org/10.1080/09593985.2018.1471763.

Young, I.A., Dunning, J., Butts, R., Cleland, J.A., Fernandez-de-Las-Penas, C., 2019. Psychometric properties of the numeric pain rating scale and neck disability index in patients with cervicogenic headache. Cephalalgia 39 (1), 44–51. https://doi.org/10.1177/0333102418772584.

Young, I., Dunning, J., Mourad, F., Escaloni, J., Bliton, P., Fernández-de-Las-Peñas, C., 2025a. Clinimetric analysis of the numeric pain rating scale, patient-rated tennis elbow evaluation, and tennis elbow function scale in patients with lateral elbow tendinopathy. Physiother. Theory Pract. 1–9. https://doi.org/10.1080/09593985.2025.2450090.

Young, I., Dunning, J., Mourad, F., Escaloni, J., Bliton, P., Fernández-de-Las-Peñas, C., 2025b. Clinimetric analysis of the visual analogue scale and pain free mouth opening in patients with muscular temporomandibular disorder. Cranio 1–7. https://doi.org/10.1080/08869634.2025.2464227.