PHYSIOTHERAPY THEORY AND PRACTICE
https://doi.org/10.1080/09593985.2018.1471763

Taylor & Francis
Taylor & Francis Group

RESEARCH REPORT

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

Ian A. Young PT, DSc[a], James Dunning PT, DPT[b], Raymond Butts PT, PhD[c], Firas Mourad PT, DPT[d], and Joshua A. Cleland PT, PhD[e]

[a]Cora Physical Therapy, Savannah, GA, USA; [b]Alabama Physical Therapy & Acupuncture, Montgomery, AL, USA; [c]Research Physical Therapy Specialists, Columbia, SC, USA; [d]Università di Roma Tor Vergata, Italy; [e]Department of Physical Therapy, Franklin Pierce University, Manchester, NH, USA

ABSTRACT

Objective: The purpose of this study was to examine the psychometric properties of the neck disability index (NDI) and numeric pain rating scale (NPRS) in patients with neck pain (NP) without concomitant upper extremity (UE) symptoms.

Design: A secondary psychometric analysis of 107 patients with NP without UE symptoms. Test-retest reliability, construct validity, area under the curve (AUC), minimum detectable change (MDC), and minimum clinically important difference (MCID) were calculated.

Results: The NDI exhibited excellent reliability (ICC = 0.88; [0.63 to 0.95]), while the NPRS exhibited moderate reliability (ICC = 0.67; [0.27 to 0.84]). The AUC for both the NDI (0.86; [0.79 to 0.93]) and NPRS (0.81 [0.73 to 0.90]) was acceptable. The MDC for the NDI was 6.9, and the MCID for the NDI was 5.5(Sn = 0.83; Sp = 0.79). For the NPRS, the MDC was 2.6, and the MCID was 1.5(Sn = 0.93; Sp = 0.64).

Conclusion: The threshold for MCID for the NDI and NPRS in patients without UE symptoms is lower (NDI = 5.5; NPRS = 1.5) than that of patients with UE/radicular symptoms (NDI = 8.5 points; NPRS = 2.2). Knowledge of these cut-scores in each presentation of NP is needed for successful research and clinical treatment. Additional outcomes may be warranted for patients with UE symptoms.

ARTICLE HISTORY
Received 21 October 2017
Revised 21 December 2017
Accepted 26 December 2017

KEYWORDS
Disability; neck pain; reliability; validity

Introduction

In order to allow clinicians the ability to determine if an individual patient has experienced a clinically meaningful change with a degree of confidence, the psychometric properties (i.e. the test-retest reliability the construct validity, the minimum detectable change (MDC), and the minimum clinically important difference (MCID)) of a self-report questionnaire must be established in the same patient population as the specific individual in question. For example, the psychometric properties of the Neck Disability Index (NDI) or the Numeric Pain Rating Scale (NPRS) in patients with cervical radiculopathy (CR) cannot be assumed to approximate those in patients with mechanical neck pain (MNP), whiplash-associated disorder (WAD), or mixed non-specific neck pain (Mixed NSNP). The psychometric properties of self-report instruments, such as the NDI or NPRS, should be population specific.

To accurately evaluate the effectiveness of treatment programs for a specific condition, it is necessary that self-report questionnaires are responsive; more specifically, the instrument must be able to distinguish improved from stable patients and recognize change over time. Responsiveness is often reported by the MDC and the MCID. The MDC is the smallest change that must be observed before the change can be considered above the measurement error with a given level of confidence (usually 95% confidence level) (Beaton et al., 2001; Copay et al., 2007; Stratford, 2004), whereas the MCID is the smallest difference which patients perceive as beneficial (Beaton et al., 2001; Copay et al., 2007; Crosby, Kolotkin, and Williams, 2003; Hays and Woolley, 2000). Thus, a valid MCID should be at least as large as the observed MDC (Beaton et al., 2001; Copay et al., 2007).

The psychometric properties of the NDI have been investigated in a variety of populations including: mechanical neck pain (MNP) (Cleland, Childs, and Whitman, 2008; En, Clair, and Edmondston, 2009; Gay, Madson, and Cieslak, 2007; Jorritsma et al., 2012; Shaheen, Omar, and Vernon, 2013; Young et al., 2009); cervical radiculopathy (CR) (Cleland, Fritz,

CONTACT : Ian A. Young, PT, DSc iayoung24@yahoo.com Cora Physical Therapy, 22 West Oglethorpe. Savannah, GA 31401, USA.

Whitman, and Palmer, 2006; Young, Cleland, Michener, and Brown, 2010); and Mixed-NSNP (Ailliet et al., 2013; McCarthy, Grevitt, Silcocks, and Hobbs, 2007; Pool et al., 2007; Riddle and Stratford, 1998; Westaway, Stratford, and Binkley, 1998) with a wide range of symptom durations (i.e. acute, subacute, and chronic). In patients with Mixed-NSNP, the NDI has been found to possess: adequate construct validity when compared with the Mental Component Summary (MCS) score $r = 0.47$ (Riddle and Stratford, 1998) or the Physical Component Summary (PCS) score of the SF-36 $r = 0.53$ (Riddle and Stratford, 1998); and strong/excellent construct validity when compared with the Patient-Specific Functional Scale $r = 0.73$ at admission (Westaway, Stratford, and Binkley, 1998); and $r = 0.81$ at discharge (Westaway, Stratford, and Binkley, 1998); or the Disabilities of the Arm, Shoulder and Hand (DASH) scale $r = 0.75$ (Ailliet et al., 2013).

In patients with MNP, the minimum detectable change (MDC) for the NDI has been reported to be: 19.6 (Cleland, Childs, and Whitman, 2008); 8.4 (Jorritsma et al., 2012); 10.2 (Young et al., 2009) 4.2 (Westaway, Stratford, and Binkley, 1998); and 5.0 (Stratford et al., 1999); whereas the reported MCID values are: 19 (Cleland, Childs, and Whitman, 2008); 3.5 (Jorritsma et al., 2012); 7.5 (Young et al., 2009); and 5.0 (Stratford et al., 1999) percentage points. The NDI has demonstrated fair (ICC = 0.50] (Cleland, Childs, and Whitman, 2008), moderate [ICC = 0.64] (Young et al., 2009), and excellent test-retest reliability [ICC = 0.86] (Jorritsma et al., 2012), $r = 0.89$ (Vernon and Mior, 1991), $r = 0.94$ (Stratford et al., 1999), $r = 0.96$ (Shaheen, Omar, and Vernon, 2013) in patients with MNP. In addition, the NDI has been found to possess strong/excellent construct validity in patients with MNP when compared with: Global Rating of Change $r = 0.52$ (Young et al., 2009) and $r = 0.81$ (Shaheen, Omar, and Vernon, 2013); Neck Bournemouth Questionnaire pretreatment 0.80 (Gay, Madson, and Cieslak, 2007) and post-treatment 0.77; Neck Pain and Disability Scale $r = 0.86$ (En, Clair, and Edmondston, 2009); and Problem Elicitation Technique $r = 0.62$ (En, Clair, and Edmondston, 2009).

Nevertheless, although several studies have investigated the psychometric properties of the NDI in patients with MNP, most of these did not exclude patients with concomitant cervical radiculopathy and/or upper extremity symptoms (Cleland, Childs, and Whitman, 2008; En, Clair, and Edmondston, 2009; Gay, Madson, and Cieslak, 2007; Jorritsma et al., 2012; Shaheen, Omar, and Vernon, 2013; Young et al., 2009). For example, Young et al. (2009) reported on the "Responsiveness of the Neck Disability Index in patients with MNP"; yet, 60% of the patients actually "presented with concomitant upper extremity radicular symptoms."

In regards to the NPRS, very few studies examined its specific psychometric properties in patients specifically with MNP. Cleland, Childs, and Whitman (2008) reported a MDC of 2.1, a MCID of 1.3 and moderate agreement (ICC = 0.76 (95% CI: 0.51–0.87) for the test-retest reliability of the NPRS in patients with "MNP". Pool et al. (2007) reported and MDC of 4.3, an MCID of 4.5, and no reliability data. Additionally, Cleland, Childs, and Whitman (2008) reported acceptable construct validity for the NPRS ($p < 0.001$) between baseline and 2.5 day follow-up scores in "improved" vs. "stable" patients based on the Global Rating of Change (GROC) score. However, Cleland, Childs, and Whitman (2008) included patients with "a primary complaint of neck pain with or without referral to the upper extremity or extremities". Therefore, the psychometric properties of the NPRS in patients with MNP without upper extremity symptoms cannot be assumed to approximate the findings of the broader population. Given the inconsistencies in the literature relative to the psychometric properties, the primary purpose of this secondary analysis was to examine the test-retest reliability, construct validity, and responsiveness of the NDI and NPRS in a large cohort of patients with MNP of any duration, but without concomitant upper extremity symptoms.

Materials and methods

This study is a secondary analysis of a larger, multicenter randomized clinical trial (RCT) (Dunning et al., 2012) that investigated the effects of two different manual physical therapy interventions in 107 consecutive patients with mechanical neck pain without upper extremity symptoms who presented to 1 of 7 outpatient physical therapy clinics in a variety of geographical locations (Arizona, Hawaii, Massachusetts, South Carolina, Texas, and Virginia), over a 20-month period (from August 2009 to March 2011). In the original trial (Dunning et al., 2012), patients were randomized to receive either high-velocity low-amplitude (HVLA) thrust manipulation or non-thrust mobilization to the upper cervical (C1-2) and upper thoracic (T1-2) spine. To be eligible for inclusion, patients had to present with a primary complaint of neck pain defined as pain in the region between the superior nuchal line and first thoracic spinous process of any duration, be between 18 and 70 years of age and have a Neck Disability Index (NDI) score of 20% or greater (i.e. 10 points or greater on a 0-to-50 scale). Patients were excluded if they exhibited any red flags (e.g. tumor, fracture, metabolic diseases, rheumatoid arthritis,

osteoporosis, resting blood pressure greater than 140/90 mmHg, and prolonged history of steroid use), presented with two or more positive neurologic signs consistent with nerve root compression (i.e. muscle weakness involving a major muscle group of the upper extremity, diminished upper extremity deep tendon reflex, or diminished or absent sensation to pinprick in any upper extremity dermatome), presented with a diagnosis of cervical spinal stenosis, exhibited bilateral upper extremity symptoms, had evidence of central nervous system involvement (i.e. hyperreflexia, sensory disturbances in the hand, intrinsic muscle wasting of the hands, unsteadiness during walking, nystagmus, loss of visual acuity, impaired sensation of the face, altered taste, and the presence of pathological reflexes), had a history of whiplash injury within the previous 6 weeks, had prior surgery to the neck or thoracic spine, had received treatment for neck pain from any practitioner within the previous month, or had pending legal action regarding their neck pain.

At baseline, patients completed all outcome measures and then received the intervention. Patients then returned for a 48-hour follow-up, at which the NDI and NPRS were administered by a therapist blinded to group allocation. However, due to the nature of the interventions, it was not possible to blind the patients to group allocation. In addition, at the 48-hour follow-up, patients completed a 15-point Global Rating of Change (GROC) scale (Jaeschke, Singer, and Guyatt, 1989) to rate their own perception of improved function. In order to investigate the psychometric properties of the NDI and NPRS, both the non-thrust mobilization and HVLA thrust manipulation groups in the original RCT (Dunning et al., 2012) were collapsed into a single cohort for this secondary analysis.

To ensure that all examination, outcome assessments, and treatment procedures were standardized, all participating physical therapists were required to study a manual of standard operating procedures, watch a 45-minute instructional DVD and participate in a 4-hour training session with the principal investigator. The study was approved by the Institutional Review Boards of the University of South Carolina, Northeast Hospital Corporation and the Corporate Clinical Research Committee, and all patients provided informed consent before participation. Data from all 107 patients who completed the randomized clinical trial are reported in this secondary analysis.

Outcome measures

Originally developed in 1991, the NDI is the most widely used instrument for assessing self-rated disability in patients with neck pain (MacDermid et al., 2009; Vernon, 2008). The NDI is a self-report questionnaire with 10-items: pain intensity, personal care, lifting, work, headaches, concentration, sleeping, driving, reading, and recreation. The response to each item is rated on a 6-point scale from 0 (no disability) to 5 (complete disability). The numeric responses for each item are summed for a total score ranging between 0 and 50; however, some evaluators have chosen to multiply the raw score by 2 and then report the NDI on a 0–100% scale (MacDermid et al., 2009). Higher scores represent increased levels of disability. The NDI has demonstrated reliability and validity as an outcome measure for patients with NP (MacDermid et al., 2009). The MDC values have been reported between 5 and 10.5 points, while the threshold for MCID have been reported to be between 7.5 and 10.5 points in patients with mechanical NP with potentially concomitant upper extremity symptoms (Cleland, Childs, and Whitman, 2008; MacDermid et al., 2009; Pool et al., 2007; Vernon, 2008; Young et al., 2009).

The NPRS was used to capture the patient's level of 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). The MDC ranges from 2.1 to 4.3, whereas the MCID ranges from 1.3 to 4.5 in patients with NP with or without radiculopathy (Cleland, Childs, and Whitman, 2008; Pool et al., 2007). Reliability (ICC) of the NPRS in patients with neck pain has been reported to be 0.76 (Cleland, Childs, and Whitman, 2008).

Patients completed all outcome measures and then received the intervention. Patients then returned for a 48-hour follow-up, at which the aforementioned outcome measures were collected. At the 48-hour follow-up, patients also completed a 15-point GROC scale, a scale described by Jaeschke, Singer, and Guyatt (1989) 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).

Data analysis

Patient variables for the improved and stable groups were compared at baseline using independent *t*-tests for continuous data and chi-square tests for categorical data. We categorized patients into three non-mutually exclusive groups on the basis of their GROC scores: 1) those scoring from −2 to +2 were considered clinically "stable" (minimal to no change); 2) those scoring ≥ +3 ("moderately better") were considered to have exhibited "clinically meaningful" improvement; 3) those scoring "0" (about the same) were considered "unchanged". Patients could be classified into more than one group, as these different groups were used for one or more analysis of reliability, validity, and responsiveness. Our

main analysis focused on patients who were "stable" and those who demonstrated "clinically meaningful improvement", whereas the "unchanged" group was used for comparative analysis.

Test-retest reliability was examined for the NDI and the NPRS using patients who underwent little to no change during the course of 48 hrs Reliability coefficients were calculated for the two groups of patients who were classified as "unchanged" (n = 13) or "stable" (n = 53) by comparing scores at the initial examination with those at the 48 hr re-evaluation. The ICC was calculated and rated according to procedures described by Shrout and Fleiss (1979). Values < 0.10 indicate no agreement, while values between 0.11–0.40, 0.41–0.60, 0.61–0.80, and > 0.81 denote slight, fair, moderate, and excellent agreement, respectively.

Construct validity of the NDI and NPRS was examined by comparing the change in outcome scores for the "stable" (GROC scores = −2 to +2) and improved (GROC scores ≥ +3) groups using separate, two-way analyses of variance for the repeated measures at baseline and reevaluation. We hypothesized that "stable" patients (n = 53) would have NDI and NPRS values that did not change, whereas patients classified as "improved" (n = 54) would demonstrate a significant change in values. This would be represented by a significant group × time interaction.

Responsiveness, the ability of a measure to recognize change when change has occurred, of the NDI and NPRS was assessed using the "stable" and "improved" patients. Receiver operator characteristic (ROC) curves (Hanley and McNeil, 1982) 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 the NDI and NPRS. Furthermore, for each outcome measure, one ROC curve was constructed with stable *vs.* improved patients. 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. 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). 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, which is considered to be the best cutoff score for distinguishing improved and stable patients (Hanley and McNeil, 1982). Sensitivity and specificity values for the selected cutoff scores were also calculated.

MDC, the amount of change that must be observed before the change can be considered to exceed the measurement error, was calculated by determining the standard error of measurement (SEM) for the NDI and NPRS for the stable group (Beaton et al., 2001). The SEM was estimated using the formula (SD/square root of 2), where SD is the standard deviation of the change scores between the test and retest values. The SEM was multiplied by 1.65 to determine the 90% CI (MDC_{90}) (Swets, 1988). This value was multiplied by the square root of 2 to account for the errors taken with repeated measurements (Swets, 1988).

Results

Of the 266 patients screened for eligibility, 107 (mean age, 42 yrs; SD = 12.8; 68% female) satisfied the inclusion and exclusion criteria, completed the study with all outcomes measures and were included in data analysis. All data met statistical assumptions and had normal distribution. The mean GROC score for all patients included in the analysis was +2.6 (SD = 2.7). The mean GROC score for the improved and stable groups was + 4.7 (SD = 1.3) and +0.4 (SD = 1.9), respectively. Fifty-four (50.5%) patients were classified as improved (GROC scores ≥ +3), and 53 (49.5%) remained stable (GROC scores = −2 to +2). Baseline characteristics are located in Table 1. The mean difference in change scores for the NDI and NPRS are listed in Table 3. There was a significant interaction between mean change-scores of improved versus stable patients in outcomes analyzed (Table 2). The correlation between change scores in all patients for the NDI and GROC (0.66) and the NPRS and GROC (0.67) were significant (p < 0.001).

The ICC values calculated from the unchanged and stable patients are reported in Table 1. The NDI exhibited excellent reliability (ICC = 0.88; [95% CI: 0.63 to 0.95]), while the NPRS exhibited moderate reliability (ICC = 0.67; [95% CI: 0.27 to 0.84]) in the patients considered stable (GROC = −1 to +1). In the unchanged group (GROC = "0"), the reliability was excellent for both the NDI (ICC = 0.99; [95% CI: 0.90 to 0.99]) and the NPRS (ICC = 0.87; [95% CI: 0.56 to 0.96]). The responsiveness (AUC) for all outcome measures is reported in Table 1. The AUC for both the NDI (0.86; [95% CI: 0.79 to 0.93]) and the NPRS (0.81 [95% CI:

Table 1. Baseline variables. difference between stable and improved groups.

Measure	Improved N = 54	Stable N = 53	p
Baseline Neck Disability Index (SD)	22.7(9.2)	20.3 (7.5)	0.15
Baseline Numeric Pain Rating Scale (SD)	5.1 (1.7)	5.5 (2.0)	0.35
Duration of symptoms-days (SD)	328 (526)	375 (425)	0.61
Weight—kg (SD)	73.4 (18.3)	71.8 (20.6)	0.67
Height—cm (SD)	168.8 (9.5)	167.9 (8.1)	0.59
Gender (% female)	68.5	67.9	
Age (SD)	42.0 (12.7)	42.1 (13.2)	0.99

Table 2. Psychometric properties of the neck disability index and numeric pain rating scale.

Measure	Unchanged (n=53)	Stable (n=53)	Improved (n=54)
Neck Disability Index			
Baseline score (SD)	**22.0 (9.7)**	**20.3 (7.5)**	**22.7 (9.2)**
Follow-up score (SD)	**20.5 (9.5)**	**17.4 (7.2)**	**11.6 (8.2)**
Change score (SD [95%CI])	1.46 (1.6 [0.49 to 2.4])	2.9 (4.2 [1.8 to 4.1])	11.1 (6.8[9.2 to 12.9])
ICC 2, 1 (95% CI)	0.99 (0.90 to 0.99)	0.88 (0.63 to 0.95)	–
SEM	1.1	3.0	–
MDC_{90}	2.6	6.9	–
AUC (95% CI)	–	–	0.86 (0.79 to 0.93)
MCID	–	–	5.5 (Sn = 0.83; Sp = 0.79)
Numeric Pain Rating Scale			
Baseline score (SD)	**5.2 (2.2)**	**5.5 (2.0)**	**5.1 (1.7)**
Follow-up score (SD)	**4.5 (2.1)**	**4.3 (1.6)**	**2.2 (1.6)**
Change Score (SD [95%CI])	0.69 (1.4 [−0.14 to 1.5])	1.1(1.6 [0.7 to 1.6])	2.9 (1.2 [2.5 to 3.2])
ICC 2, 1 (95% CI)	0.87 (0.56 to 0.96)	0.67 (0.27 to 0.84)	–
SEM	0.99	1.1	–
MDC_{90}	2.3	2.6	–
AUC (95% CI)	–	–	0.81 (0.73 to 0.90)
MCID (Sn; Sp)	–	–	1.5 (Sn = 0.93; Sp = 0.64)

Table 3. Difference between change scores from baseline to 48-hours on self-report outcomes.

Measure	Improved (SD)	Stable (SD)	Difference in change scores (95% CI)	*p*
Neck Disability Index	11.1(6.9)	2.9 (4.2)	8.2 (6.3 to 10.1)	< 0.0001
Numeric Pain Rating Scale	2.9 (1.2)	1.1 (1.6)	1.8 (1.4 to 2.2)	< 0.0001

SD = standard deviation; CI = confidence interval

0.73 to 0.90]) was acceptable. The MDC, MCID threshold and the sensitivity/specificity associated with the cutoff score are located in Table 1. The MDC for the NDI was 6.9, while the MCID for the NDI was 5.5 (Sn = 0.83; Sp = 0.79). For the NPRS, the MDC was 2.6, and the MCID was 1.5 (Sn = 0.93; Sp = 0.64). The SEM values calculated from the unchanged and stable patients are also reported in Table 1.

Discussion

The importance of obtaining accurate information on reliability, construct validity, and responsiveness of self-report and physical outcome measures is of paramount importance. These properties are used to demonstrate the effectiveness of the clinical trials and provide clinician confidence that true change has occurred over time. Recent research on the psychometric properties of the NDI and NPRS suggested the possibility of self-report error when using the NDI/NPRS across two very different categories of patients with neck pain (with and without radicular/UE symptoms). Young, Cleland, Michener, and Brown (2010) describe the process of centralization, often resulting in a decrease in upper extremity symptoms and an increase in local neck pain in patients treated with symptoms of cervical radiculopathy. The authors suggest this may indirectly alter the self-report response of the NDI/NPRS, and ultimately its psychometric properties. Thus, psychometric analysis should be population specific in order to obtain accurate estimates of reliability, validity, and responsiveness over time. Cleland, Childs, and Whitman (2008) included an estimated 23% of patients with symptoms distal to shoulder and 13 (9.5%) of these with signs of nerve root compression. Young et al. (2009) included an estimated 60% of patients with UE symptoms but 15/25 (60%) were in the "stable" group for analysis of ICC, SEM and MDC. While the Cleland, Childs, and Whitman (2008) study did not attempt to isolate the patients with concomitant radiculopathy/UE symptoms, Young et al. (2009) specifically categorized the patients for a portion of the primary and post-hoc analysis. Importantly, however, the reliability analysis in both studies was not isolated to those specifically without UE symptoms, and both categories were thus "combined" into one cohort. The current study specifically examined the reliability of the NDI and NPRS in a cohort of patients with mechanical NP without radiculopathy/UE symptoms to assess change in this specific presentation of NP.

The results of the present study suggest excellent reliability (ICC = 0.87[95% CI: 0.78 to 0.93]) in a cohort of patients with NP without radicular/UE symptoms. While some original research on the NDI (Stratford et al., 1999; Westaway, Stratford, and Binkley, 1998) yielded similar reliability, other quality studies with similar data analysis (Cleland, Childs, and Whitman, 2008; Young et al., 2009) reported inconsistent SEM and MDC values (Table 4).

Construct validity for outcome measures was examined by comparing the baseline and follow-up scores for both the stable and improved groups. Patients who rated

Table 4. Comparison of reliability, SEM, MDC, AUC, and MCID of the NDI and NPRS across different categories of neck pain.

NDI	*N* Total	*N* Stable	Reliability ICC	SEM	MDC	AUC	MCID
Current study[§]	107	53	0.87	3.0	6.9	0.86	5.5
Young & Walker et al.°	91	25	0.64	4.3	10.2	0.79	7.5
Pool et al.°[¶]	183	87	NR	NR	10.5	NR	10.5
Cleland et al.°	137	89	0.50	8.4	9.8	0.83	9.5
Young & Cleland et al.*	165	43	0.55	5.7	13.4	0.74	8.5
Cleland et al.*	38	17	0.68	4.4	10.2	0.57	7.0
NPRS							
Current study[§]	107	53	0.67	1.1	2.6	0.81	1.5
Pool et al.°[¶]	183	87	NR	NR	4.3	NR	4.5
Cleland et al.°	137	89	0.76	0.91	2.1	0.85	1.3
Young & Cleland et al.*	165	43	0.58	1.8	4.1	0.72	2.2
Cleland et al.*	38	17	0.50	NR	NR	NR	NR

NDI = neck disability index, NPRS = numeric pain rating scale, SEM = standard error of measurement, MDC = minimal detectable change, AUC = area under the curve, MCID = minimal clinically important difference
[§] Data analysis only includes patient with mechanical neck pain without radiculopathy/upper extremity symptoms
° Data analysis includes patient with mechanical neck pain and with radiculopathy/upper extremity symptoms
* Data analysis only includes patients only with cervical radiculopathy/upper extremity symptoms.
[¶] Data Analysis included global perceived effect (GPE) as the measure of anchor based change. All other studies used global rating of change (GROC)
NR = not reported

themselves as improved reported significant changes ($p < 0.001$) on the NDI (disability) and NPRS (pain) (Table 3). These findings are consistent with those previously reported in patients with NP with/without UE symptoms (Cleland, Childs, and Whitman, 2008; Pool et al., 2007; Young et al., 2009).

The AUC is used to determine the probability that the patient exhibiting an improvement can be correctly identified (Hanley and McNeil, 1982). The AUC for a diagnostic test is considered to be satisfactory when it exceeds 0.70 (Swets, 1988). The NDI had an AUC of 0.86, and the NPRS exhibited an AUC of 0.81 (Table 2). These results are similar to prior studies on individuals with NP (Cleland, Childs, and Whitman, 2008; Young et al., 2009), and the results suggest that both outcome measures provide acceptable responsiveness to identify improvement in patient's perceived disability and pain when a true change has occurred.

For purposes of direct comparison, Table 4 includes the most recent (last decade) psychometric studies of the NDI and NPRS in patients with different categories of NP with similar data analyses (i.e. ICC, ROC, and AUC). Table 4 also delineates which studies included a sample with NP only, cervical radiculopathy only and a heterogeneous sample of NP with/without radiculopathy/UE symptoms. Interestingly, the MCID value for the NDI across all studies of patients with NP only (current study) was 5.5. The MCID in studies including a mixed sample of NP categories (with and without radicular and/or UE symptoms) ranges from 7.5 to 10.5, while the MCID for patients specifically with cervical radiculopathy only ranges from 7.0 to 8.5. Table 4. These higher values seem to support prior recommendations that patients with NP and UE symptoms may be generally more refractive from treatment (Vernon, 2008), may have altered levels of responsiveness secondary to centralization/peripheralization and likely have higher thresholds for clinically important change (Young, Cleland, Michener, and Brown, 2010). Hence, the current tools may not be sensitive enough for those patients with UE symptoms, and may be associated with an over estimation of MCID. Further study should incorporate additional upper extremity self-report measures to improve outcomes assessment in this area.

In regards to reliability, Table 4 illustrates similar inconsistent trends with the NDI across different categories of NP. The findings of the current study of patients with NP without UE symptoms (ICC = 0.87) is consistent with original research from Stratford et al. (1999) (0.90) and Westaway, Stratford, and Binkley (1998) (0.80). Notably, the study by Stratford et al. (1999) included only patients with neck pain without UE symptoms, and the study by Westaway, Stratford, and Binkley (1998) reported having only two patients (6%) with UE symptoms. Moreover, the current study in patients with NP without UE symptoms exhibited superior reliability (ICC = 0.87) to that of more recent studies with a mixed sample of NP (ICC = 0.50) (Cleland, Childs, and Whitman, 2008); and 0.64 (Young et al., 2009). These findings are in line with previous theoretical construct (Young, Cleland, Michener, and Brown, 2010) and suggest that test re-test reliability may be affected when specific symptoms (i.e. UE) are not considered in choosing an outcome measure. For example, the GROC does not specifically inquire about neck pain but rather asks about the patient's change in "condition". In contrast, the NDI inquires about disability related to "neck pain" throughout its 10-item construct. Hence, if the patient has changes in UE symptoms > neck

symptoms, they may misinterpret the NDI. This has the potential to affect reliability coefficients in a negative manner. While prior studies on patients with cervical radiculopathy (Cleland, Fritz, Whitman, and Palmer, 2006; Young, Cleland, Michener, and Brown, 2010) have demonstrated similar reliability of the NDI to those with a mixed sample NP (Table 4), some authors recommend also using condition specific tools such as the quick disabilities of arm, shoulder, and hand (QDASH), when assessing outcomes in this patient population (Mehta, MacDermid, Carlesso, and McPhee, 2010).

The MCID for the NPRS in the current study was 1.5. This cut score is consistent with that of Cleland, Childs, and Whitman (2008) who reported a MCID of 1.3 in patients with mixed NP. Pool et al. (2007) reported a much higher MCID of 4.3 for the NPRS in a similar sample of mixed NP patients. However, the global perceived effect (GPE) was used instead of the GROC for the criterion-based measure of true change. Although the GPE has exhibited acceptable reliability and construct validity, it has been challenged for its ability to be used as an acceptable criterion of change when determining minimally important change and responsiveness of other outcomes (Kamper et al., 2010). Thus, comparing MCID values obtained with a different criterion measure should be interpreted with caution.

The reliability of the NPRS in the current study was only moderate (ICC = 0.67). This is similar to prior research in a mixed sample of NP (Cleland, Childs, and Whitman, 2008) but notably larger than studies that specifically include patients with cervical radiculopathy (Cleland, Fritz, Whitman, and Palmer, 2006; Young, Cleland, Michener, and Brown, 2010) (Table 4). The "moderate" reliability of the NPRS vs. the "excellent" reliability of the NDI in this study raises a point for further discussion. Without UE symptoms/radiculopathy, it seems that both outcome measures should have similar test-retest reliability in patients with NP. The difference may exist in the measure used to detect if true change has "not" occurred (GROC). In order to assess reliability, the group defined as stable (GROC; −2 = a little bit worse *to* +2 = a little bit better) was used for analysis and considered to have "minimal to no change". This category is consistent with analysis in prior studies (Cleland, Childs, and Whitman, 2008; Young, Cleland, Michener, and Brown, 2010; Young et al., 2009). However, these cut scores may have included patients who had smaller (non-clinically relevant) levels of "improvement" and/or meaningful "worsening" of symptoms. Interestingly, when a GROC of "0" (about the same) was used for analysis, both the NDI (ICC = 0.99) and the NPRS (ICC = 0.87) exhibited "excellent" agreement. Therefore, the reliability of the NPRS could be stated as "moderate" in patients with minimal (or clinically unimportant) change and "excellent" in patients without change. Although the sample size of "unchanged" patients will likely be small in most effective controlled trials used for psychometric analysis, using a stringent category may give a more accurate measure of reliability. Furthermore, the construct of the NDI includes descriptors for each individual section to be scored, while the NPRS is an 11-point scale without descriptors for each unit. This may lead to greater variability in responses and possibly an underestimation of reliability statistics.

There are several limitations to be noted in this psychometric analysis. First, the results of this this study were based on a 48-hour follow-up only. Secondly, the current data analysis chosen to examine construct validity and responsiveness is not all-inclusive, and may need to be addressed with alternative methods and in further study. Lastly, it is important for clinicians and researchers to consider that the MDC is consistently larger than the MCID in both outcome measures.

Conclusions

The results of this study describes the reliability, construct validity and responsiveness of the NDI and NPRS in a specific cohort of NP patients without radicular/UE symptoms. The NDI exhibited excellent reliability (ICC = 0.88) and the NPRS exhibited moderate reliability (ICC = 0.67) in this patient population. The MCID for the NDI and NPRS were 5.5 and 1.5 respectively. As evidence based medicine progresses with larger quality clinical trials in patients with NP, it seems intuitively necessary to constantly re-evaluate the psychometric properties of the common outcomes used to measure success. Further study of specific categories of neck pain or isolated analysis using separate condition-specific outcomes for UE symptoms may serve to improve our knowledge of appropriate change scores needed for successful research and clinical treatment.

Declaration of Interest

The authors declare no conflict of interest.

References

Ailliet L, Knol DL, Rubinstein SM, De Vet HC, Van Tulder MW, Terwee CB. (2013). Definition of the construct to be

measured is a prerequisite for the assessment of validity. The neck disability index as an example. Journal of Clinical Epidemiology, 66, 775–782.
Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, Strand V, Shea B. (2001). Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome measures in rheumatology. Minimal clinically important difference. Journal of Rheumatology, 28, 400–405.
Cleland JA, Childs JD, Whitman JM. (2008). Psychometric properties of the neck disability index and numeric pain rating scale in patients with mechanical neck pain. Archives of Physical Medicine and Rehabilitation, 89, 69–74.
Cleland JA, Fritz JM, Whitman JM, Palmer JA. (2006). The reliability and construct validity of the neck disability index and patient specific functional scale in patients with cervical radiculopathy. Spine, 31, 598–602.
Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. (2007). Understanding the minimum clinically important difference: A review of concepts and methods. Spine Journal, 7, 541–546.
Crosby RD, Kolotkin RL, Williams GR. (2003). Defining clinically meaningful change in health-related quality of life. Journal of Clinical Epidemiology, 56, 395–407.
Dunning JR, Cleland JA, Waldrop MA, Arnot CF, Young IA, Turner M, Sigurdsson G. (2012). Upper cervical and upper thoracic thrust manipulation versus nonthrust mobilization in patients with mechanical neck pain: A multicenter randomized clinical trial. Journal of Orthopaedic and Sports Physical Therapy, 42, 5–18.
En MC, Clair DA, Edmondston SJ. (2009). Validity of the neck disability index and neck pain and disability scale for measuring disability associated with chronic, non-traumatic neck pain. Manual Therapy, 14, 433–438.
Gay RE, Madson TJ, Cieslak KR. (2007). Comparison of the neck disability index and the neck bournemouth questionnaire in a sample of patients with chronic uncomplicated neck pain. Journal of Manipulative and Physiological Therapeutics, 30, 259–262.
Hanley JA, McNeil BJ. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143, 29–36.
Hays RD, Woolley JM. (2000). The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? PharmacoEconomics, 18, 419–423.
Jaeschke R, Singer J, Guyatt GH. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. Controlled Clinical Trials, 10, 407–415.
Jorritsma W, Dijkstra PU, De Vries GE, Geertzen JH, Reneman MF. (2012). Detecting relevant changes and responsiveness of neck pain and disability scale and neck disability index. European Spine Journal, 21, 2550–2557.
Kamper SJ, Ostelo RW, Knol DL, Maher CG, De Vet HC, Hancock MJ. (2010). Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. Journal of Clinical Epidemiology, 63, 760–766.
MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, Goldsmith CH. (2009). Measurement properties of the neck disability index: A systematic review. Journal of Orthopaedic and Sports Physical Therapy, 39, 400–417.
McCarthy MJ, Grevitt MP, Silcocks P, Hobbs G. (2007). The reliability of the Vernon and Mior neck disability index, and its validity compared with the short form-36 health survey questionnaire. European Spine Journal, 16, 2111–2117.
Mehta S, MacDermid JC, Carlesso LC, McPhee C. (2010). Concurrent validation of the DASH and the QuickDASH in comparison to neck-specific scales in patients with neck pain. Spine, 35, 2150–2156.
Pool JJ, Ostelo RW, Hoving JL, Bouter LM, De Vet HC. (2007). Minimal clinically important change of the neck disability index and the numerical rating scale for patients with neck pain. Spine, 32, 3047–3051.
Riddle DL, Stratford PW. (1998). Use of generic versus region-specific functional status measures on patients with cervical spine disorders. Physical Therapy, 78, 951–963.
Shaheen AA, Omar MT, Vernon H. (2013). Cross-cultural adaptation, reliability, and validity of the arabic version of neck disability index in patients with neck pain. Spine, 38, E609–615.
Shrout PE, Fleiss JL. (1979). Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, 86, 420–428.
Stratford PW. (2004). Getting more from the literature: estimating the standard error of measurment from reliability studies. Physiotherapy Canada, 56, 27–31.
Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B. (1999). Using the neck disability index to make decisions concerning individual patients. Physiotherapy Canada, 51, 107–112.
Swets JA. (1988). Measuring the accuracy of diagnostic systems. Science, 240, 1285–1293.
Vernon H. (2008). The neck disability index: state-of-the-art, 1991-2008. Journal of Manipulative and Physiological Therapeutics, 31, 491–502.
Vernon H, Mior S. (1991). The neck disability index: A study of reliability and validity. Journal of Manipulative and Physiological Therapeutics, 14, 409–415.
Westaway MD, Stratford PW, Binkley JM. (1998). The patient-specific functional scale: validation of its use in persons with neck dysfunction. Journal of Orthopaedic and Sports Physical Therapy, 27, 331–338.
Young BA, Walker MJ, Strunce JB, Boyles RE, Whitman JM, Childs JD. (2009). Responsiveness of the neck disability index in patients with mechanical neck disorders. Spine Journal, 9, 802–808.
Young IA, Cleland JA, Michener LA, Brown C. (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. American Journal of Physical Medicine and Rehabilitation, 89, 831–839.