Original Article

Cephalalgia
An International Journal of Headache

International Headache Society

Cephalalgia
0(0) 1–8

DOI: 10.1177/0333102418772584
journals.sagepub.com/home/cep

SAGE

Psychometric properties of the Numeric Pain Rating Scale and Neck Disability Index in patients with cervicogenic headache

Ian A Young[1,2], James Dunning[2,3], Raymond Butts[2], Joshua A Cleland[4] and César Fernández-de-las-Peñas[3]

Abstract

Background: Self-reported disability and pain intensity are commonly used outcomes in patients with cervicogenic headaches. However, there is a paucity of psychometric evidence to support the use of these self-report outcomes for individuals treated with cervicogenic headaches. Therefore, it is unknown if these measures are reliable, responsive, or result in meaningful clinically important changes in this patient population.

Methods: A secondary analysis of a randomized clinical trial (n = 110) examining the effects of spinal manipulative therapy with and without exercise in patients with cervicogenic headaches. Reliability, construct validity, responsiveness and thresholds for minimal detectable change and clinically important difference values were calculated for the Neck Disability Index and Numeric Pain Rating Scale.

Results: The Neck Disability Index exhibited excellent reliability (ICC = 0.92; [95 % CI: 0.46–0.97]), while the Numeric Pain Rating Scale exhibited moderate reliability (ICC = 0.72; [95 % CI: 0.08–0.90]) in the short term. Both instruments also exhibited adequate responsiveness (area under the curve; range = 0.78–0.93) and construct validity ($p < 0.001$) in this headache population.

Conclusions: Both instruments seem well suited as short-term self-report measures for patients with cervicogenic headaches. Clinicians and researchers should expect at least a 2.5-point reduction on the numeric pain rating scale and a 5.5-point reduction on the neck disability index after 4 weeks of intervention to be considered clinically meaningful.

Keywords

Headache intensity, neck pain, self-report outcomes, minimal clinically important difference, reliability

Date received: 13 December 2017; revised: 12 February 2018; accepted: 22 February 2018

Introduction

Patients with disorders of the cervical spine may present with a multitude of symptoms including neck pain, upper extremity referred/radicular symptoms and or cervicogenic headaches (CeH). Each of these presentations is unique and requires specific evaluation, and treatment procedures. Furthermore, each patient presentation will likely respond differently to the rehabilitation process depending on the clinical presentation. In light of this, it is imperative to have reliable, valid, and responsive condition specific self-report outcomes to guide clinical decision-making and for use in research on each patient population. For example, the threshold for minimum clinically important change (MCID) of the Neck Disability Index (NDI) (1) or the Numeric Pain Rating Scale (NPRS) (2) in patients with symptoms of cervical radiculopathy (NDI = 8.5 points; NPRS = 2.2 points) (3,4) can not be applied to

[1]CORA Physical Therapy, Savannah, GA, USA
[2]American Academy of Manipulative Therapy, Montgomery, AL, USA
[3]Universidad Rey Juan Carlos, Alcorcón, Spain
[4]Franklin Pierce University, Manchester, NH, USA

Corresponding author:
Ian A Young, CORA, 22 West Oglethorpe Ave., Savannah, GA 31401, USA.
Email: iayoug24@yahoo.com

those with neck pain (NDI = 5 points; NPRS = 1.3 points) (3).

The psychometric properties of the NDI and NPRS have been established and are considered reliable, valid and responsive self-report outcomes for patients with neck pain and cervical radiculopathy (3–5). Although self-reported disability and pain intensity are commonly used outcome measures in patients with CeH, there is a paucity of psychometric evidence in this patient population to support the use of these self-report outcomes. Although some form of the NPRS or the visual analogue scale (VAS) for headache intensity is commonly used in clinical trials for individuals with CeH (6–11), inconsistencies exist in the use of a common standardized disability measure across studies. Only three quality clinical trials have assessed disability in adults with CeH, each using a different standardized outcome measure. Hass et al. (6) used the modified Von Korff (12); Jull et al. (9) used the Northwick Park Neck Pain Questionnaire (13) and Dunning et al. (10) used the NDI (1). Regardless, the psychometric properties of these pain and disability outcome measures have not yet been established in patients with CeH. Therefore, it remains unknown if these measures are reliable and responsive, and/or are able to capture clinically important improvement after a treatment intervention in this headache population.

In a recent randomized clinical trial, Dunning et al. (10) used psychometric data on NPRS and NDI developed from patients with neck pain to estimate MCID values in patients with CeH (10). Although these headaches likely arise primarily from a cervical spine derangement (boney, joint, disc, or soft tissue) (14), the characteristics of pain, function and disability before and after treatment may be quite different than those of patients with a primary complaint of neck pain rather than headache. Approximation of psychometric data from patients with neck pain to patients with CeH may pose a threat to the results reported by clinical trials relative to a population with CeH. Determining accurate condition specific psychometric properties of pain and disability is of paramount importance in clinical research and in every day clinical practice. Therefore, the purpose of this study was to examine the psychometric properties (reliability, construct validity and responsiveness) of the NPRS and NDI in a cohort of patients with CeH.

Methods

This study is a secondary analysis of a large multicenter randomized clinical trial (10) that investigated the effects of two different physical therapy interventions in 110 consecutive patients with CeH. In the original trial, patients were randomized to receive both cervical and thoracic manipulation (n = 58) or mobilization and exercise (n = 52) (10). Consecutive patients with CeH were recruited over a 29-month period (from April 2012 to August 2014). For patients to be eligible, they had to present with a diagnosis of CeH according to the revised diagnostic criteria (15) developed by the Cervicogenic Headache International Study Group (CHISG) (15–17). CeH was classified according to the "major criteria" (not including confirmatory evidence by diagnostic anesthetic blockades) and "head pain characteristics" of the CHISG. The inclusion and exclusion criteria are previously described in detail (10). The NDI (1) and NPRS (2) were collected in all patients at baseline, 1 week, 4 weeks, and 3 months. Perceived recovery using the Global Rating of Change Scale (GROC) (18) was collected in all patients at all follow-up points. In order to investigate the psychometric properties of all outcome measures, both the manipulation group and mobilization + exercise group completing the clinical trial were collapsed into a single cohort for this secondary analysis.

Outcome measures

The NDI is the most widely used instrument for assessing self-rated disability in patients with neck pain (5,19). 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 six-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 (19). Higher scores represent increased levels of related disability.

The NPRS was used to capture the patient's level of pain (headache intensity). 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) (2). The minimal detectable change (MDC) ranges from 2.1 to 4.3, whereas the MCID ranges from 1.3 to 4.5 points change in patients with neck pain with or without radiculopathy (3,20) Reliability (ICC) of the NPRS in patients with neck pain has been reported to be 0.76 (3).

Patients also completed a 15-point Global Rating of Change (GROC) scale described by Jaeschke et al. (18) 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 (18).

Scores of +3, +4 and +5 are commonly used to identify improved versus stable patients.

Data analysis

We categorized patients into four non-mutually exclusive groups at each follow-up time point (1 week, 4 weeks, 3 months) on the basis of their GROC scores: (a) those scoring from −2 to +2 were considered clinically "stable" (minimal to no change); (b) those scoring ≥+3 "somewhat better", (3) those scoring ≥+4 "moderately better", and those scoring ≥5, "quite a bit better". Patient variables for the improved (GROC ≥+3 points) and stable (GROC ≥ −2 to ≤ +2) groups were compared at baseline using independent t-tests for continuous data and chi-square tests for categorical data. Patients could be classified into more than one group, as these different groups were used for one or more analyses of reliability, validity, and responsiveness. Our main analysis focused on patients who were "stable" and those who reported being "somewhat better" at the 4-week follow-up, whereas the groups reporting "moderately" or "quite a bit better" at 1 week and 3 months were used for comparative analysis. Although treatment duration varies across clinical trials on CeH, the 4-week follow-up data was used in our main analysis, as this seems to be a more common time frame used in daily clinical practice.

Test-retest reliability was examined for the NDI and NPRS using patients who underwent little to no change. Reliability coefficients for the NDI and NPRS were calculated for the two groups of patients who were classified as "stable" by comparing scores at the initial examination with those at the 1-week, 4-week, and 3-month re-evaluation. The ICC was calculated and rated according to procedures described by Shrout and Fleiss (21). 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", "somewhat better", "moderately better", and "quite a bit better" groups using separate, two-way analyses of variance for the repeated measures at baseline and reevaluation. We hypothesized that "stable" patients in each group would have NDI and NPRS intake values that did not 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.

Responsiveness – the ability of a measure to recognize change when change has occurred – of the NDI and NPRS was assessed using the "stable," and the three individual groups of improved patients (GROC ≥+3, ≥+4, and ≥+5) at each follow-up point. Receiver operator characteristic (ROC) curves (22) 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. 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 (22). 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 (22). 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 have exceeded measurement error, was calculated by determining the standard error of measurement (SEM) for the NDI and NPRS for the stable group (23). 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}) (24). This value was multiplied by the square root of 2 to account for the errors taken with repeated measurements (24).

Results

One hundred and ten patients (35.2 years; SD = 11.5) 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 4-week follow-up was +4.3 (SD = 2.1). The mean GROC score for the improved vs. stable groups was +5.2 (SD = 1.3) and +1.0 (SD = 0.91), respectively. At the 4-week follow-up 86 (78.2%) patients were classified as improved, and 24 (21.8%) remained stable. There was a significant difference ($p < 0.001$) in mean change scores between stable and improved patients, for the NDI and NPRS, at the 1-week, 4-week, and 3-month follow-up (Table 1).

The ICC values SEM and MDC calculated from the stable patients at all follow-up points are reported in Table 2. At the 4-week follow-up, the NDI exhibited excellent reliability (ICC: 0.92, 95%CI: 0.46–0.97), while the NPRS exhibited moderate reliability

Table 1. Difference between change scores from baseline on self-report outcomes over time.

Measure	Improved (GROC ≥ 3)	Stable (GROC; ≥ -2 to $\leq +2$)	Mean difference in change scores (95% CI)	*P*
NDI (SD)				
Baseline				
1 week	6.4 (4.8)	1.8 (3.2)	4.6 (3.0–6.3)	$P < 0.0001$
4 weeks	10.5 (6.4)	3.5 (3.6)	7.0 (4.3–9.6)	$P < 0.0001$
3 months	11.0 (7.1)	2.2 (2.7)	8.8 (5.7–11.7)	$P < 0.0001$
NPRS (SD)				
Baseline				
1 week	2.8 (2.6)	1.2 (1.6)	1.5 (0.6–2.4)	$P < 0.001$
4 weeks	4.0 (2.3)	1.3 (1.5)	2.7 (1.7–3.6)	$P < 0.0001$
3 months	4.1 (2.2)	0.9 (2.0)	3.2 (2.2–4.1)	$P < 0.0001$

NPRS: Numeric Pain Rating Scale, 0–10, lower scores indicate less pain; NDI: Neck Disability Index, 0–50, lower scores indicate greater function; GROC: global rating of change; SD: standard deviation; CI: confidence interval.

Table 2. Reliability of the stable group (GROC ≥ -2 to $\leq +2$) over time, standard error of measurement (SEM) and minimal detectable change (MDC) values.

Measure	ICC (95% CI)	*p*	SEM	MDC
NDI				
1 week (n = 41)	0.94 (0.85–0.97)	$p < 0.001$	2.27	5.3
4 weeks (n = 24)	0.92 (0.46–0.97)	$p < 0.001$	2.55	5.9
3 months (n = 26)	0.95 (0.78–0.98)	$p < 0.001$	1.91	4.5
NPRS				
1 week (n = 41)	0.62 (0.07–0.83)	$p < 0.001$	1.13	2.6
4 weeks (n = 24)	0.72 (0.08–0.90)	$p < 0.001$	1.06	2.4
3 months (n = 26)	0.48 (−0.07–0.76)*	$p = 0.04$	1.42	3.3

NPRS: Numeric Pain Rating Scale, 0–10, lower scores indicate less pain; NDI: Neck Disability Index, 0–50, lower scores indicate greater function; ICC: Intraclass Correlation Coefficient; CI: confidence interval; SEM: Standard Error of Measure; MDC: Minimal Detectable Change.
*no significant interaction between groups at $p < 0.05$.

(ICC: 0.72, 95%CI: 0.08–0.90) in the patients considered stable. At 4 weeks, the MDC for the NDI was 5.9 and the MDC for the NPRS was 2.4.

The responsiveness (AUC) for the NDI and NPRS for all improved categories, and at each follow-up point is reported in Table 3. Regardless of time frame and improvement category, the NDI and NPRS demonstrated acceptable responsiveness (AUC range = 0.78–0.93). The MCID threshold and the sensitivity/specificity associated with each cutoff score are also located in Table 3. At 4 weeks, the MCID for the NDI was 5.5 (“somewhat better” category), 6.5 (“moderately better” category), and 7.5 (“quite a bit better” category), while the MCID for the NPRS was 2.5 (“somewhat better” category), 2.5 (“moderately better” category), and 3.5 (“quite a bit better” category).

Discussion

To date, this is the first study to examine the psychometric properties of commonly used self-report outcome measures in patients with CeH. All outcome measures at all time points exhibited proper construct validity (Table 1). All outcomes analyzed in this cohort of patients with CeH demonstrated acceptable reliability at the 1-week and 4-week follow-up (Table 2); at the 3-month follow-up, the NDI demonstrated excellent

Table 3. Area under the curve (AUC) and cutoff scores for minimal clinically important difference (MCID) of the NDI and NPRS.

	GROC ≥ 3 (somewhat better)			GROC ≥ 4 (moderately better)			GROC ≥ 5 (quite a bit better)		
	AUC (95% CI)	Sn; Sp	MCID	AUC	Sn; Sp	MCID	AUC (95% CI)	Sn; Sp	MCID
NDI-1 wk (n = 69)	0.87* (0.80–0.94)	0.81; 0.79	3.5	0.82* (0.74–0.90)	0.75; 0.80	5.5	0.78* (0.68–0.87)	0.79-0.70	5.5
NDI-4wk (n = 86)	0.87* (0.79–0.95)	0.79; 0.79	5.5	0.83* (0.75–0.91)	0.77; 0.76	6.5	0.83* (0.76–0.91)	0.77; 0.69	7.5
NDI-3 mo (n = 84)	0.93* (0.88–0.98)	0.79; 0.93	5.5	0.87* (0.80–0.94)	0.83; 0.76	5.5	0.87* (0.81–0.94)	0.80; 0.78	7.5
NPRS-1 wk (n = 69)	0.82* (0.74–0.89)	0.63; 0.83	2.5	0.85* (0.78–0.92)	0.75; 0.79	2.5	0.86* (0.78–0.94)	0.86; 0.72	2.5
NPRS-4wk (n = 86)	0.84* (0.76–0.91)	0.72; 0.75	2.5	0.81* (0.73–0.89)	0.78; 0.68	2.5	0.83* (0.75–0.91)	0.72; 0.98	3.5
NPRS-3 mo (n = 84)	0.86* (0.79–0.94)	0.77; 0.75	2.5	0.83* (0.76–0.91)	0.79; 0.63	2.5	0.83* (0.76–0.91)	0.73; 0.83	3.5

NDI: neck disability index; NPRS: numeric pain rating scale; wk: week; mo: month; GROC: global rating of change scale; AUC: area under curve; CI: confidence interval; Sn: sensitivity; Sp: specificity; MCID: meaningful clinically important difference.
*$p < 0.001$.

reliability, while the NPRS had poor and fair reliability, respectively (Table 2). In addition, regardless of time to follow-up and/or improvement category, the NDI and NPRS demonstrated acceptable responsiveness (Table 3). Overall, our analysis suggests that the commonly used self-report outcomes of pain and disability in patients with cervical spine disorders exhibited acceptable psychometric properties in patients with CeH.

The NDI exhibited excellent reliability (ICC) values at all time points (1 week = 0.94, ($p < 0.001$); 4 weeks = 0.92, ($p < 0.001$); 3 months = 0.95, ($p < 0.001$). The MDC was a 5.9 point change at the 4-week follow-up. Notably, based on studies using "stable" patients on the GROC, the test-retest reliability of the NDI in this cohort of patients with CeH is higher than those reported in the majority of studies on patients with neck pain (ICC = 0.64 (25), 0.50 (3)). Nevertheless, Jorritsma et al. (26) used the global perceived effect scale to identify "stable patients" and reported excellent reliability (ICC = 0.86) in patients with non-specific neck pain. Of those studies on neck pain using the more common GROC as the criterion measure for success, both included patients with upper extremity symptoms (3,25). This may have negatively impacted the reliability of the NDI in patients with neck pain. For example, in patients with CeH, the presence of neck pain and headaches is expected, and each of these items is addressed in a specific section on the NDI (1). Conversely, upper extremity symptoms are often expected in a heterogeneous sample of patients with non-specific neck pain, and this is not addressed in any section of the NDI. Similarly, it has been suggested that the varied distribution of symptoms in patients with radiculopathy may have a negative effect on point estimates in psychometric analyses (4). Nevertheless, from the standpoint of test-retest reliability, the NDI seems very well suited for patients with CeH up to a 3-month follow-up (Table 2).

The NPRS for headache intensity exhibited moderate (ICC = 0.62), moderate (ICC = 0.72) and slight (ICC = 0.48) reliability at the 1-week, 4-week, and 3-month follow-up, respectively (Table 2). The ICC values were considered statistically significant at the 1- and 4-week follow-up ($p < 0.001$), but not the 3-month follow-up ($p = 0.04$) (Table 2). Cleland et al. (3) reported similar reliability (ICC = 0.76) to our 4-week data, at 2.5 days follow-up in a cohort of patients with non-specific neck pain. The MDC of the NPRS for headache intensity at 4 weeks (2.4 points) was similar to that of the MCID at 4 weeks (2.5 points) (Tables 2–3). Although the reliability of the NPRS is considered acceptable at the 1- and 4-week follow-up, it seems lower in comparison to the NDI (ICC; 1 week = 0.94, 4 weeks = 0.92) in this cohort of

patients with CeH. This lower (fair to moderate) reliability coefficient for the NPRS is in line with that of patients with neck pain, cervical radiculopathy and low back pain (3,4,27). It appears the reliability of the NDI is superior over time compared to the NPRS; thus, the interpretation of headache intensity values, as measured by the NPRS, should be approached cautiously in this patient population.

It is interesting to note that most studies on patients with CeH often report outcomes on headache intensity, frequency, duration, and medication intake. However, the definition of each of these self-report outcomes and how they were measured is often lacking, or has varied in some manner across all studies. For example, the measurement of headache frequency in the current study and the Jull et al. (9) study (number of headaches in the last week) is different to Haas et al. (6) and Niere et al. (11) (number of headaches per month). Similarly, discrepancies are found when examining pain scales, disability, headache duration, and medication intake across studies. These inconsistencies make it difficult to draw accurate comparisons between multiple trials. Therefore, in this study, we sought only to establish the psychometric properties of the standardized self-report outcomes of pain and disability in patients with CeH for the purpose of their common use in future clinical trials. In essence, this will enable accumulation and analyses of larger data sets to appropriately assess clinical outcomes of selected interventions in future meta-analysis. Further studies are needed to establish the psychometric properties of headache frequency, headache duration and medication intake in patients with CeH.

Finally, measuring responsiveness and deciphering the MCID involves the use of a single cut score (usually 3/7, ''somewhat better'' on the GROC) in order to construct the ROC and AUC (22). In this study, for comparative analysis, we sought to evaluate the responsiveness and MCID using three separate categories of improvement on the GROC (''somewhat better'', ''moderately better'', ''quite a bit better'') at all follow-up time points. In the original trial, we examined the effects of mobilization + exercise versus manipulation alone in 110 patients with CeH (10). In total, 65% (mobilization = 19/52 [37%]; manipulation = 52/58 [90%]) of the patients met the GROC score of $\geq 4/7$ (moderately better) at the 4-week follow-up. Therefore, in the present analysis, we wanted to include data on the changes in the MCID of those patients who were ''moderately'' ($\geq$4) and ''quite a bit better'' ($\geq$5). It seems reasonable that in a highly responsive outcome measure, we should statistically see differences between groups for the MCID that parallel the different categories of clinical improvement. Hence, as level of perceived improvement increases on the GROC, we should see a larger cut-score for the MCID. Of interest clinically is the 4-week follow-up across all improvement categories. Our analysis noted changes in the MCID for the NDI (5.5, 6.5, 7.5) and NPRS (2.5, 2.5, 3.5) for the ''somewhat better'' ($\geq$3), ''moderately better'' ($\geq$4) and ''quite a bit better'' ($\geq$5) categories, respectively (Table 3). Therefore, if a patient with CeH reports being ''somewhat better'' (GROC $\geq 3/7$) at the 4-week follow-up, we should expect to see at least a 5.5-point reduction on the NDI and 2.5-point reduction in the NPRS. Similarly, in a patient who reports their headaches are ''quite a bit better'' (GROC $\geq 5/7$) at the 4-week follow-up, we should expect to see at least a 7.5-point reduction on the NDI and 3.5-point reduction in the NPRS. Young et al. (4) also noted larger MCID values of the NDI and NPRS with greater levels of perceived improvement in patients with cervical radiculopathy (4).

A final point of interest includes the use of CHISG (15–17) instead of the International Classification of Headache Disorders-3rd edition (ICHD-3) (14) diagnostic criteria for CeH. In our original trial (10), our decision to use CHISG criteria was based on the ability to make a clinical diagnosis by the evaluating and treating physical therapist without the use of imaging or anesthetic blockade. Although the reliability and validity of these clinical criteria have been established (15,17,28–34), the ICHD-3 must be acknowledged and considered in all studies involving CeH. More importantly, each set of criteria should be used appropriately within the methodological context of the clinical trial conducted. Therefore, current results should be considered for patients with CeH diagnosed with the CHISG criteria.

Conclusion

This is the first study to examine the psychometric properties of common self-report outcome measures in patients treated for CeH. The NDI and NPRS both exhibited acceptable reliability and strong construct validity at the 1-week and 4-week follow-up. Additionally, the NPRS and NDI exhibited appropriate responsiveness over time. The MCID at 4 weeks for the NDI (5.5, 6.5 and 7.5) and NPRS (2.5, 2.5 and 3.5) was established, based on perceived level of improvement (GROC) scores of ≥ 3, ≥ 4 and ≥ 5, respectively. The NDI and NPRS seem well suited as short-term self-report outcome measures for patients with CeH.

Clinical implications

- The NDI and NPRS are reliable, valid and responsive self-report outcomes that are well suited for use in patients with cervicogenic headache.
- Clinicians and researchers should expect at least a 2.5-point reduction on the NPRS and a 5.5-point reduction on the NDI after 4 weeks of intervention to be considered clinically meaningful.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Ian A Young http://orcid.org/0000-0001-6559-993X

References

1. Vernon H and Mior S. The Neck Disability Index: A study of reliability and validity. *J Manipulative Physiol Ther* 1991; 14: 409–415.
2. Jensen MP, Karoly P, O'Riordan EF, et al. The subjective experience of acute pain. An assessment of the utility of 10 indices. *Clinical J Pain* 1989; 5: 153–159.
3. Cleland JA, Childs JD and 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.
4. Young IA, Cleland JA, Michener LA, et al. 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.
5. Vernon H. The Neck Disability Index: State-of-the-art, 1991–2008. *J Manipulative Physiol Ther* 2008; 31: 491–502.
6. Haas M, Spegman A, Peterson D, et al. Dose response and efficacy of spinal manipulation for chronic cervicogenic headache: A pilot randomized controlled trial. *Spine* 2010; 10: 117–128.
7. Nilsson N. A randomized controlled trial of the effect of spinal manipulation in the treatment of cervicogenic headache. *J Manipulative Physiol Ther* 1995; 18: 435–440.
8. Nilsson N, Christensen HW and Hartvigsen J. The effect of spinal manipulation in the treatment of cervicogenic headache. *J Manipulative Physiol Ther* 1997; 20: 326–330.
9. Jull G, Trott P, Potter H, et al. A randomized controlled trial of exercise and manipulative therapy for cervicogenic headache. *Spine* 2002; 27: 1835–1843, discussion 43.
10. Dunning JR, Butts R, Mourad F, et al. Upper cervical and upper thoracic manipulation versus mobilization and exercise in patients with cervicogenic headache: A multicenter randomized clinical trial. *BMC Musculoskelet Disord* 2016; 17: 64.
11. Niere K and Robinson P. Determination of manipulative physiotherapy treatment outcome in headache patients. *Manual Ther* 1997; 2: 199–205.
12. Underwood MR, Barnett AG and Vickers MR. Evaluation of two time-specific back pain outcome measures. *Spine* 1999; 24: 1104–1112.
13. Leak AM, Cooper J, Dyer S, et al. The Northwick Park Neck Pain Questionnaire, devised to measure neck pain and disability. *Br J Rheumatol* 1994; 33: 469–474.
14. The International Classifcation of Headache Disorders: 3rd Edition. *Cephalalgia* 2013; 33: 629–808.
15. Sjaastad O, Fredriksen TA and Pfaffenrath V. Cervicogenic headache: Diagnostic criteria. The Cervicogenic Headache International Study Group. *Headache* 1998; 38: 442–445.
16. Sjaastad O and Fredriksen TA. Cervicogenic headache: Criteria, classification and epidemiology. *Clin Exp Rheumatol* 2000; 18: S3–S6.
17. Vincent MB and Luna RA. Cervicogenic headache: A comparison with migraine and tension-type headache. *Cephalalgia* 1999; 19: S11–S16.
18. Jaeschke R, Singer J and Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–415.
19. MacDermid JC, Walton DM, Avery S, et al. Measurement properties of the Neck Disability Index: A systematic review. *J Orthop Sports Phys Ther* 2009; 39: 400–417.
20. Pool JJ, Ostelo RW, Hoving JL, et al. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. *Spine* 2007; 32: 3047–3051.
21. Shrout PE and Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420–428.
22. Hanley JA and McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29–36.
23. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome measures in rheumatology. Minimal clinically important difference. *J Rheumatol* 2001; 28: 400–405.
24. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988; 240: 1285–1293.
25. Young BA, Walker MJ, Strunce JB, et al. Responsiveness of the Neck Disability Index in patients with mechanical neck disorders. *Spine* 2009; 9: 802–808.

26. Jorritsma W, de Vries GE, Dijkstra PU, et al. Neck Pain and Disability Scale and Neck Disability Index: validity of Dutch language versions. *Eur Spine J* 2012; 21: 93–100.
27. Cleland JA, Whitman JM, Houser JL, et al. Psychometric properties of selected tests in patients with lumbar spinal stenosis. *Spine* 2012; 12: 921–931.
28. Sjaastad O, Fredriksen TA and Pfaffenrath V. Cervicogenic headache: Diagnostic criteria. *Headache* 1990; 30: 725–726.
29. Antonaci F, Ghirmai S, Bono G, et al. Cervicogenic headache: Evaluation of the original diagnostic criteria. *Cephalalgia* 2001; 21: 573–583.
30. Vincent M. Validation of criteria for cervicogenic headache. *Funct Neurol* 1998; 13: 74–75.
31. van Suijlekom JA, de Vet HC, van den Berg SG, et al. Interobserver reliability of diagnostic criteria for cervicogenic headache. *Cephalalgia* 1999; 19: 817–823.
32. Pfaffenrath V, Dandekar R, Mayer ET, et al. Cervicogenic headache: Results of computer-based measurements of cervical spine mobility in 15 patients. *Cephalalgia* 1988; 8: 45–48.
33. Fredriksen TA, Fougner R, Tangerud A, et al. Cervicogenic headache. Radiological investigations concerning head/neck. *Cephalalgia* 1989; 9: 139–146.
34. Jull G, Amiri M, Bullock-Saxton J, et al. Cervical musculoskeletal impairment in frequent intermittent headache. Part 1: Subjects with single headaches. *Cephalalgia* 2007; 27: 793–802.