

# Neck Pain and Disability Scale and the Neck Disability Index: reproducibility of the Dutch Language Versions

Wim Jorritsma · Grietje E. de Vries ·  
Jan H. B. Geertzen · Pieter U. Dijkstra ·  
Michiel F. Reneman

Received: 28 September 2009/Revised: 15 February 2010/Accepted: 4 April 2010/Published online: 28 April 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** The first aim of this study was to translate the Neck Pain and Disability Scale (NPAD) from English into Dutch producing the NPAD–Dutch Language Version (DLV). The second aim was to analyze test–retest reliability and agreement of the NPAD–DLV and the Neck Disability Index (NDI)–DLV. The NPAD was translated according to established guidelines. Thirty-four patients (mean age 37.5 years, 68% female) with chronic neck pain (CNP), within an outpatient rehabilitation setting, participated in this study. The NPAD–DLV and the NDI–DLV were filled out twice with a mean test–retest interval of 18 days. The intraclass correlation coefficient of the NPAD–DLV was 0.76 (95% confidence interval (CI) 0.57–0.87) and of the NDI–DLV 0.84 (95% CI 0.69–0.92). The limits of agreement of the NPAD–DLV and the NDI–DLV

were, respectively,  $\pm 20.9$  (scale 0–100) and  $\pm 6.5$  (scale 0–50). The reliability of the NPAD–DLV and the NDI–DLV was acceptable for patients with CNP. The variation (‘instability’) in the NPAD–DLV total scores was relatively large and larger than the variation of the NDI–DLV.

**Keywords** Neck pain · Neck Pain and Disability Scale · Neck Disability Index · Reliability · Agreement

## Introduction

Neck pain is a common musculoskeletal complaint in western societies [15]. In the large majority of cases, the pathological basis for the neck pain is unclear and the complaints are labeled as ‘non-specific’ or ‘mechanical’ [7]. Disability, limitations in activities and restrictions in participation in daily living and work, may be the result [30, 32]. The majority of the total costs of neck pain in the Netherlands were costs due to sick leave and disability payment [8].

Self-reported disability in patients with neck pain is often measured by means of region-specific questionnaires [25]. These questionnaires may measure disability or the functional status with greater responsiveness than generic health questionnaires [25]. Questionnaires should have good psychometric qualities, among which is reproducibility [25]. Reproducibility is the extent to which the same results are obtained on repeated tests when no real change in health status has occurred [14, 25]. Reproducibility may be influenced by random measurement errors and within patient variance [14, 27]. Both sources of variance may lead to score instability (natural variation) on repeated tests [14, 27]. If a patient with neck pain fills out the same questionnaire on two occasions, it is relevant to know what score instability can be expected in a predefined retest

---

W. Jorritsma (✉) · G. E. de Vries · J. H. B. Geertzen ·  
P. U. Dijkstra · M. F. Reneman  
Center for Rehabilitation, University Medical Center Groningen,  
University of Groningen, PO Box 30 002, 9750 RA Haren,  
Groningen, The Netherlands  
e-mail: w.jorritsma@cvr.umcg.nl

W. Jorritsma · J. H. B. Geertzen · P. U. Dijkstra ·  
M. F. Reneman  
Graduate School for Health Research,  
University Medical Center Groningen,  
University of Groningen, Groningen, The Netherlands

G. E. de Vries  
Center for Human Movement Sciences,  
University Medical Center Groningen,  
University of Groningen, Groningen, The Netherlands

P. U. Dijkstra  
Department of Oral and Maxillofacial Surgery,  
University Medical Center Groningen,  
University of Groningen, Groningen, The Netherlands

interval or in a waiting period. Reproducibility does have two aspects: reliability and agreement, representing test–retest score stability over time on group level and on individual level, respectively [14]. A measure for reliability is the intra-class correlation coefficient (ICC) [27]. It assesses not only the strength of the correlation between two repeated measures, but also if all measures on each subject are identical and do not differ systematically [27]. To quantify the agreement, the test–retest score stability over time on individual level, the limits of agreement (LOA) can be calculated according to the method of Bland and Altman [3, 6]. The LOA lies two standard deviations ( $SD_{\text{difference}}$ ) above and under the mean total score difference of all patients between the first and second test. This means that, due to score instability, approximately 95% of all differences within patients will lie between these LOA on repeated tests [3, 6]. In an individual patient, the change due to treatment should exceed these LOA before one can state that real change has occurred.

The most used Neck Disability Questionnaires are the Neck Pain and Disability Scale (NPAD) [32] and the neck disability index (NDI) [30], which are both translated into several languages [1, 2, 5, 9, 13, 18–21, 23, 24, 28, 29, 33]. To be able to use the NPAD and NDI in different countries and social environments these questionnaires must not only be translated properly, but also culturally adapted and validated [4, 25]. These translations allow comparison of results of clinical research trials between countries. To investigate which questionnaire is most appropriate, psychometric studies are needed where questionnaires are applied simultaneously to the same sample of patients [25]. The advantage of the NPAD over the NDI may be the simple wording and the unitary structure of the questions; moreover, the questionnaire is easy to complete with its visual analog scale structure [9, 23, 25, 28]. The NPAD has not been formally translated into Dutch and, consequently, psychometric qualities of the Dutch Language Version (NPAD–DLV) are unknown. The reliability of the NDI–DLV has been studied in patients with acute neck pain, but not in patients with chronic neck pain (CNP) other than patients with Whiplash associated disorder (WAD) [17, 31]. The first aim of this study was to translate the NPAD from English into Dutch. The second aim was to analyze test–retest reliability and agreement of the NPAD–DLV and the NDI–DLV in patients with CNP in an outpatient rehabilitation setting.

## Methods

### Translation and cross-cultural adaptation of the NPAD

The NDAP was translated using a forward and backward translation procedure [4]. Two native Dutch speakers

(a clinician, aware of the concepts behind the questionnaire and a staff member of the pain rehabilitation team) independently translated and culturally adapted the original version. The translated versions were critically reviewed reciprocally, compared with one another and with the original English version. Disagreements were discussed and a consensus version was produced. A backward translation (from Dutch into English) of this consensus version was made by a bilingual physiotherapist involved in spine research. The translators examined translation, backward translation and notes about the discussions made during the translation process. A concept version of the NPAD–DLV was developed and pilot-tested on a heterogeneous group of ten patients and employees of the rehabilitation center, who were asked to comment critically on understandability of the questions and instructions, responses, wording and layout. Finally, a final NPAD–DLV was produced.

### Study sample

CNP patients were recruited from referrals by general practitioners or medical specialists for rehabilitation treatment in the Center for Rehabilitation at the University Medical Center Groningen, The Netherlands. Inclusion criteria for this study were: non-specific chronic neck pain (>3 months duration), admitted for rehabilitation, age between 18 and 65 years, less than 2 years out of work due to CNP or still at work with frequent sick leave due to neck pain, and sufficient knowledge of the Dutch language to complete questionnaires. Exclusion criteria were: specific neck pain, status post surgery in the cervical region, cardiovascular or pulmonary diseases significantly diminishing physical capacity, pregnancy, addiction to drugs, extensive psychological or behavioral problems.

### Procedures

Prior to the first visit at the University Medical Center patients filled out a baseline questionnaire assessing demographics and clinical characteristics. During the first visit a review of the medical history and a physical examination was performed. Immediately afterwards patients filled out the NPAD–DLV and the NDI–DLV. A second visit was scheduled, depending on subject availability, 1–5 weeks after the first visit, but prior to the start of the outpatient rehabilitation program. During the second visit the patients filled out the NPAD–DLV and the NDI–DLV for the second time. All patients signed informed consent prior to entering the study.

### Measurements

The NPAD is a questionnaire whose development used the Million Visual Analogue Scale (VAS) as a template [16].

The NPAD consists of 20 items. Each item has a VAS of 100 mm with numeric anchors at 0, 1, 2, 3, 4 and 5 (each 20 mm apart). Item scores range from 0 (no pain or limitation in activities) to 5 (as much pain as possible or maximal limitation). The total NPAD score can vary from 0 to 100 points [16]. Test–retest reliability, expressed as intraclass correlation coefficient (ICC), of different language versions of the NPAD ranged from 0.81 to 0.98 with retest intervals from 1 day to 1–2 weeks. [2, 9, 19, 21, 33].

The NDI is a questionnaire based on the Oswestry low back pain disability questionnaire and consists of 10 items [30]. Each item has six different assertions expressing progressive levels of pain or limitation in activities with a score between 0 (no pain or limitation) and 5 (as much pain as possible or maximal limitation). The total NDI scores can vary from 0 to 50 points [30]. Test–retest reliability, expressed as ICC, of different language versions of the NDI ranged from 0.50 to 0.97 with retest intervals from 1 day to 3 weeks [11, 12, 19, 20, 22, 24, 29, 31, 33]. The ICC of the NDI–DLV was 0.90 in patients with acute neck pain in general practice over a retest interval of 7 days [31]. For patients with WAD the test–retest reliability of the NDI–DLV was  $r = 0.81$  over a retest interval of 3 months [17].

#### Data analyses

Descriptive statistics were calculated for the total scores of the two test sessions for both questionnaires. Reliability of the NDAP–DLV and the NDI–DLV was expressed as ICCs for the total scores. ICCs of 0.75 or higher were interpreted as acceptable reliability [27]. To quantify agreement (the test–retest score stability on individual level) of the NPAD–DLV and the NDI–DLV the limits of agreement were calculated as described by Bland and Altman [3, 6]. Statistical analyses were performed with SPSS 14.0.

#### Results

A total of 181 neck patients were referred to the Center for Rehabilitation between November 2006 and December 2007. From this group 72 (40%) were admitted for rehabilitation. A total of 39 patients were eligible for inclusion in this study. During the waiting period after the first visit, 5 patients decided not to start with the rehabilitation program, 33 completed the NPAD–DLV and 32 the NDI–DLV twice. Characteristics of the study sample are presented in Table 1. In the translation and cross-cultural adaptation of the NPAD, minor changes were made in item 13 and item 18. In item 13 (outlook in life and the future), the given examples ‘depression and hopelessness’ were deleted because the respondents

of the pre-final version found this superfluous. In item 18 (trouble with looking up or down) the text was changed into: bending the head forwards or backwards, because looking up and down can be done with the eyes only and without flexion–extension of the cervical spine. Details were added in the general instruction to emphasize that all items should be answered regarding the intensity of the neck pain or neck pain related disability.

**Table 1** Demographic and clinical characteristics of the study sample ( $n = 34$ )

	Mean (SD) or Median (IQR)	Min–max
Age (years)	37.5 (11.5)	21–57
Duration of chronic pain (months)	12.5 (6.75–26.25) <sup>a</sup>	4–240
Sick leave in the past year (weeks)	21 (18.5)	0–52
	<i>N</i>	%
Female	23	68
Pain radiating to		
Shoulder(s)	30	88
Upper arm(s)	13	38
Forearm(s)	8	24
Hand/fingers	8	24
Between shoulder blades	15	44
Pins and needles below elbow	11	32
Concomitant complaints		
Headache	26	77
Dizziness	13	38
Concentration problems	8	24
Nausea	5	15
Fatigue	22	65
Low back pain	13	38
Self-reported cause of neck pain		
Motor vehicle accident	19	56
Other trauma	1	12
Spontaneously/unknown	1	3
Stress	1	3
Work related	2	6
Other	7	21
Previous treatment for neck pain	31	85
Education		
Low	2	6
Intermediate vocational education	23	68
High	7	21
Work status (self employed/employee)	1/33	3/97
Sick leave	23	67
Involved in litigation	17	50

<sup>a</sup> Median and interquartile range for duration of pain (months)

**Table 2** Total scores of the NPAD–DLV ( $n = 33$ ) and NDI–DLV ( $n = 32$ ) at test and retest, intraclass correlation coefficients (ICC), difference between test and retest and limits of agreement

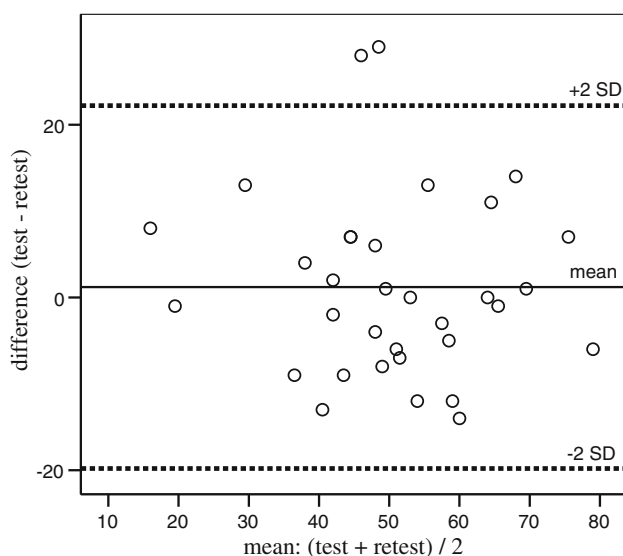
	Test		Retest		ICC (95% CI)	Difference		Limits of agreement
	Mean	SD	Mean	SD		Mean	SD	
NPAD–DLV	50.7	15.7	51.2	14.7	0.76 (0.57–0.87)	1.3	10.4	$\pm 20.9$
NDI–DLV	22.6	5.9	20.6	6.4	0.84 (0.69–0.92)	-1.4	3.2	$\pm 6.5$

NPAD–DLV Neck Pain and Disability Scale Dutch Language Version, NDI–DLV Neck Disability Index Dutch Language Version

### Reliability and agreement

The mean retest interval was 18.2 days (SD 6.2, range 6–34). Test and retest results are presented in Table 2. Item 20 of the NPAD–DLV, concerning pain pills, was left blank by 2 (6%) patients. This item presumes the patient is taking medication. Item 7 of the NDI–DLV concerning driving was left blank by 2 (6%) patients; for these patients the score was adjusted using the mean of the answers on the rest of the questionnaire. The ICC of the NPAD–DLV was 0.76 (95% confidence interval (CI) 0.57–0.87) and of the NDI–DLV 0.84 (95% CI 0.69–0.92). The LOAs of the NPAD–DLV and the NDI–DLV were, respectively,  $\pm 20.9$  (scale 0–100) and  $\pm 6.5$  (scale 0–50) (Table 2).

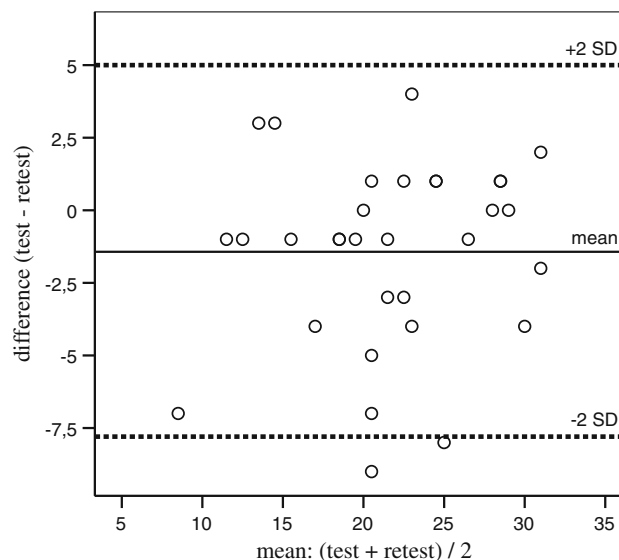
Bland and Altman plots for the NPAD–DLV and the NDI–DLV are presented in Figs. 1 and 2. No visible tendency towards unequal variance of the data appeared present.



**Fig. 1** Mean score of each patient plotted against the difference between test–retest scores of NPAD–DLV. The reference line is the mean of total score difference of all patients. Limits of agreement at 2 SD

### Discussion

The cross-cultural adaptation of the NPAD followed the forward and backward translation procedure. This procedure warranted on the one hand the meaning of the original items and on the other hand capturing of contents and meanings of the questions in the translation into the Dutch language. Other than the production of a more detailed general instruction, only minor modifications were made in items 13 and 18. The adaptation of item 18 is supported by the fact that the developers of the NPAD questionnaire have explicitly related this item to ‘neck problems’ [32] and other authors to ‘neck dysfunction related to activities of the cervical spine’ [2, 9, 13, 23]. The new NPAD–DLV was easy to comprehend. To complete both the NPAD–DLV and the NDI–DLV required maximally 15 min. The number of missing responses was negligible, which was in agreement with other non-English versions of the questionnaires [1, 2, 5, 9, 13, 20, 21, 23, 24, 28, 29, 33].



**Fig. 2** Mean score of each patient plotted against the difference between test–retest scores of NDI–DLV. The reference line is the mean of total score difference of all patients. Limits of agreement at 2 SD

The NPAD–DLV and the NDI–DLV demonstrated acceptable reliability in a sample of patients with CNP. The retest interval depended on the availability of the patient. In our rehabilitation setting it is normal to have a waiting period (1–2 months) between intake and start of the program. Therefore, it is interesting to know the extent of changes in questionnaire outcome occurring in absence of treatment. The sample sizes in our study were similar to other reproducibility studies [2, 9, 12, 19–22, 24, 29, 31, 33] except for four other studies [9, 12, 19, 33] where the sample sizes were 23, 17, 102 and 101, respectively. The female to male ratio in the current study is similar to that in most former reproducibility studies. [9, 12, 19, 20, 22, 31, 33].

In our study, the mean total score of the NPAD–DLV was 50.7 and of the NDI–DLV 22.6. In other studies, the mean total scores of the NPAD ranged from 38.2 to 60.5 and of the NDI from 11.0 to 23.0 [2, 9, 11, 12, 19, 20, 22, 29, 31, 33]. In general, studies carried out in tertiary referral centers have higher total scores than those in primary care settings. In all studies, where both questionnaires were used, including ours, the NPAD scores were approximately 10% higher than the NDI scores when presented in % of a 0–100 scale [19, 21, 24, 33].

The reliability of the NPAD–DLV in our study (ICC = 0.76) was lower than in reliability studies with shorter retest intervals (less than 2 weeks ICC = 0.81–0.98 [2, 9, 19, 21, 33]). The reliability of the NDI–DLV in our study (ICC = 0.84) was somewhat lower than in most former NDI studies with generally shorter retest intervals (less than 2 weeks ICC = 0.50–0.97 [11, 19, 22, 24, 29, 31, 33]; 2 weeks, ICC = 0.88 [20]; 3 weeks ICC = 0.68 [12]). Perceived recovery (change) in the retest interval, to include the ‘stable’ patients in the reliability studies, was assessed in only the half of above mentioned NPAD and NDI studies [2, 29, 31, 33]. Apart from that it seems not to have resulted in differences in the extent of ICCs [2, 9, 20–22, 24, 29, 31, 33]. A trend may be seen that studies with a shorter retest interval do have higher ICCs. When looking at another region-specific questionnaire, the same trend was reported [10]. To test for the bias caused by differences in retest interval duration we assessed a partialled retest correlation; this means that we assessed the test–retest correlation for the NPAD–DLV and NDI–DLV while ‘controlling’ the effect of retest interval duration. The Pearson correlations for the test–retest reliability while ‘controlling’ or ‘not controlling’ for retest interval duration were  $r = 0.70$  and  $r = 0.72$  for the NPAD and  $r = 0.87$  and  $r = 0.87$  for the NDI. These results indicate that the influence of the effect of retest interval duration is minimal or negligible.

The NPAD is claimed to be a questionnaire with four underlying dimensions [32]. Factor analyses in other

language NPADs identified two to four factors on which different items were loading [9, 13, 23, 24, 33]. The factorial structure presented in the original publication was based on a relatively small sample ( $n = 95$ ); therefore, the stability of the observed factor solution may be questioned and too sample specific to be reproducible in different samples. Therefore, comparison of the ICCs for subscales in different (language) NPAD studies is challenging. However, a principal component analysis in a German study with a sample size of 448 indicated a one-factor solution for the NPAD, and it was concluded that the NPAD is a multidimensional assessment instrument measuring different dimensions of one construct neck pain, in a stable manner [28]. Because above mentioned reasons and because factor analysis was not an aim of the current study only total scores were used to analyze the reliability of the NPAD–DLV.

If a patient with neck pain fills out the same questionnaire on two occasions, in a waiting period prior to the start of a rehabilitation program, a (very) short time interval increases the probability of carryover or recall effects due to memory, mood or practice, whereas a larger interval increases the probability the clinical status has changed and that the score of the first session has been forgotten [27]. There are several explanations for possible changes of the clinical status during the waiting period: the effect of the clinic consultation, the anticipation of the patient on the program, the effect of a period of waiting before the real rehabilitation program starts, the chronic neck pain itself with its fluctuations and the questionnaires itself [22].

To quantify the agreement, the test–retest score stability over time on individual level, the ‘limits of agreement’ (LOA) were calculated. No criteria are available for interpretation of the LOA. Smaller LOA means more stability and indicate that the natural variation is smaller. The  $SD_{\text{difference}}$  (the standard deviation of the mean total score difference off all patients between the first and second test) and the LOA of the NPAD–DLV in the present study (10.4 and  $\pm 20.9$ ) were somewhat higher than in one other study (9.0 and  $\pm 17.9$ ) with a retest interval of 1 day [33]. In this French study, a 5 point ordinal transition scale was used to include clinically stable patients. Despite a clear difference in retest intervals, the differences in  $SD_{\text{difference}}$  were small. The  $SD_{\text{difference}}$  and the LOA of the NDI–DLV in the present study (3.2 and  $\pm 6.5$ ) was similar to most other NDI studies with shorter retest intervals (1 day,  $SD_{\text{difference}}$  3.4, LOA  $\pm 6.7$  [33]; 1 week,  $SD_{\text{difference}}$  3.9, LOA  $\pm 7.8$  [31]; 1 week  $SD_{\text{difference}}$  1.5, LOA  $\pm 3.0$  [29]; 1–2 weeks,  $SD_{\text{difference}}$  4.4, LOA  $\pm 8.9$  [22]). In three of these four studies, a transition scale was used to include clinically stable patients [22, 29, 31, 33]. Proportionally, the  $SD_{\text{difference}}$  of the Greek study [29] was similar to the



$SD_{\text{difference}}$  of the present study ( $SD_{\text{difference}}/\text{mean total score}$  was 0.12 and 0.14, respectively). The NDI reliability studies have shown smaller  $SD_{\text{difference}}$  and smaller natural variations compared to the NPAD reliability studies [22, 29, 31, 33]. Larger instability of the NPAD may be explained by differences in operationalizations of ‘neck disability’ between items of the NPAD and the NDI [30, 32]. Post hoc analysis showed that the amount of natural variation of the NPAD–DLV could not be attributed to individual items of the questionnaire. Clinical effects of therapy in an individual patient should exceed the limits of agreement before one can state that real change has occurred. The minimal clinically important difference (MCID) is suggested to be 11 points on the NPAD [9] and 2 to 10 points on the NDI [11, 12, 26, 31]. Based on the variation in the current study, patients have to change at least 21 points on the NPAD–DLV (scale 0–100) and at least 7 points on the NDI–DLV (scale 0–50), will these patients be judged as having ‘really’ changed.

There are limitations to consider in evaluating our research. First, the sample size is relatively small ( $n < 50$ ), therefore, our sample could have misestimated the “true” population ICCs and LOAs. However, the ICCs of the NPAD–DLV and NDI–DLV were in line with two studies with larger samples and with ICCs of, respectively, 0.81 and 0.86 ( $n = 102$ ) [19] and 0.91 and 0.93 ( $n = 101$ ) [33]. Second, the retest interval was longer than in most other NPAD and NDI studies [2, 9, 11, 19–22, 24, 29, 31, 33]. Therefore, the reported ICC and LOA values in the present study probably underestimate the reliability and agreement of the scales. Nevertheless the retest interval duration may be not the only important factor influencing these values, in view of an NDI study with a retest interval of 2.5 days ( $SD \pm 0.95$ ) where the ICC was 0.50 [11]. Other factors such as symptom duration (acute, sub-acute or chronic), patient setting (primary, secondary or tertiary care) and mean disability score on the questionnaires may also influence the values of ICC and LOA. Third, the retest interval was not fixed and perceived recovery (change) in this interval was not controlled for. However,  $SD_{\text{difference}}$  of the NDI–DLV in our sample was similar to studies where change was controlled for.

A strength of this study is that to the authors’ knowledge for the second time a reproducibility study is made for the NPAD with respect to reliability and limits of agreement with a head to head comparison with the NDI. Further study with the NPAD–DLV is necessary to assess the reliability and agreement in other patient groups (e.g. acute, sub-acute, primary care patients), to assess the ICC with a shorter retest interval and to study other measurement properties, such as validity, responsiveness and MCID.

## Conclusion

A reliable DLV of the NPAD was developed. The reliability of the NPAD–DLV and the NDI–DLV was acceptable for patients with CNP within an outpatient rehabilitation setting. The natural variation (‘instability’) in the NPAD–DLV total scores was relatively large and larger than the variation of the NDI–DLV.

**Acknowledgments** The authors thank Marleen Speller for assistance in data collection.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Ackelman BH, Lindgren U (2002) Validity and reliability of a modified version of the neck disability index. *J Rehabil Med* 34(6):284–287
- Agarwal S, Allison GT, Agarwal A, Singer KP (2006) Reliability and validity of the Hindi version of the Neck Pain and Disability Scale in cervical radiculopathy patients. *Disabil Rehabil* 28(22):1405–1411
- Altman DG, Bland JM (1983) Measurement in medicine: the analysis of method comparison studies. *Statistician* 32:307–317
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25(24):3186–3191
- Bicer A, Yazici A, Camdeviren H, Erdogan C (2004) Assessment of pain and disability in patients with chronic neck pain: reliability and construct validity of the Turkish version of the neck pain and disability scale. *Disabil Rehabil* 26(16):959–962
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1(8476):307–310
- Bogduk N (1999) The neck. *Baillieres Clin Rheumatol* 13(2):261–285
- Borghouts JA, Koes BW, Vondeling H, Bouter LM (1999) Cost-of-illness of neck pain in The Netherlands in 1996. *Pain* 80(3):629–636
- Bremerich FH, Grob D, Dvorak J, Mannion AF (2008) The Neck Pain and Disability Scale: cross-cultural adaptation into German and evaluation of its psychometric properties in chronic neck pain and C1-2 fusion patients. *Spine* 33(9):1018–1027
- Brouwer S, Kuijer W, Dijkstra PU et al (2004) Reliability and stability of the Roland Morris Disability Questionnaire: intra class correlation and limits of agreement. *Disabil Rehabil* 26(3):162–165
- 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. *Arch Phys Med Rehabil* 89(1):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 (Phila Pa 1976)* 31(5):598–602
- Cook C, Richardson JK, Braga L et al (2006) Cross-cultural adaptation and validation of the Brazilian Portuguese version of

- the Neck Disability Index and Neck Pain and Disability Scale. *Spine* 31(14):1621–1627
14. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ (2001) Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care* 17(4):479–487
  15. Ferrari R, Russell AS (2003) Regional musculoskeletal conditions: neck pain. *Best Pract Res Clin Rheumatol* 17(1):57–70
  16. Goolkasian P, Wheeler AH, Gretz SS (2002) The neck pain and disability scale: test–retest reliability and construct validity. *Clin J Pain* 18(4):245–250
  17. Heymans WFGJ, Lutke Schipholt HJA, Elvers JWH et al (2002) Neck Disability Index Dutch Version (NDI–DV): investigation of reliability in patients with chronic whiplash. *Ned T Fysiother* 112:94–99
  18. Köke AJA, Heuts PHTG, Vlaeyen JWS et al (1996) Neck Disability Index. Pijn Kennis Centrum Maastricht. Meetinstrumenten chronische pijn, Maastricht
  19. Kose G, Hegguler S, Atamaz F, Oder G (2007) A comparison of four disability scales for Turkish patients with neck pain. *J Rehabil Med* 39(5):358–362
  20. Kovacs FM, Bago J, Royuela A et al (2008) Psychometric characteristics of the Spanish version of instruments to measure neck pain disability. *BMC Musculoskelet Disord* 9:42
  21. Lee H, Nicholson LL, Adams RD et al (2006) Development and psychometric testing of Korean language versions of 4 neck pain and disability questionnaires. *Spine* 31(16):1841–1845
  22. 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. *Eur Spine J* 16(12):2111–2117
  23. Monticone M, Baiardi P, Nido N et al (2008) Development of the Italian version of the Neck Pain and Disability Scale, NPDS-I: cross-cultural adaptation, reliability, and validity. *Spine* 33(13):E429–E434
  24. Mousavi SJ, Parnianpour M, Montazeri A et al (2007) Translation and validation study of the Iranian versions of the Neck Disability Index and the Neck Pain and Disability Scale. *Spine* 32(26):E825–E831
  25. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF (2002) Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. *Spine* 27(5):515–522
  26. 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 (Phila Pa 1976)* 32(26):3047–3051
  27. Portney LG, Watkins MP (2000) Foundations of clinical research, 2nd edn. Prentice-Hall, Upper Saddle River
  28. Scherer M, Blozik E, Himmel W et al (2008) Psychometric properties of a German version of the neck pain and disability scale. *Eur Spine J* 17(7):922–929
  29. Trouli MN, Vernon HT, Kakavelakis KN et al (2008) Translation of the Neck Disability Index and validation of the Greek version in a sample of neck pain patients. *BMC Musculoskelet Disord* 9:106
  30. Vernon H, Mior S (1991) The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther* 14(7):409–415
  31. Vos CJ, Verhagen AP, Koes BW (2006) Reliability and responsiveness of the Dutch version of the Neck Disability Index in patients with acute neck pain in general practice. *Eur Spine J* 15(11):1729–1736
  32. Wheeler AH, Goolkasian P, Baird AC, Darden BV (1999) Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. *Spine* 24(13):1290–1294
  33. Wlodyka-Demaille S, Poiraudau S, Catanzariti JF et al (2002) French translation and validation of 3 functional disability scales for neck pain. *Arch Phys Med Rehabil* 83(3):376–382