RESEARCH ARTICLE

The Responsiveness of Three Persian Outcome Measures Following Physiotherapy Intervention in Patients with Chronic Non-Specific Neck Pain

Seyed Ramin Dabiri, MD; Mohammad Reza Tehrani, PhD; Farzad Omidi-Kashani, MD; Hamed Mami-Pour, PT; Hamed Tabesh, PhD; Ali Moradi, PhD; Salman Nazary-Moghadam, PhD

Research performed at physiotherapy clinic at Ghaem university hospital, Mashhad, Iran

Received: 4 October 2022 *Accepted: 29 January 2023*

Abstract

Objectives: A prospective cohort study to evaluate and compare the responsiveness of the Persian version of the neck disability index (NDI), neck pain & disability scale (NPDS), neck outcome score (NOOS), and to determine the minimal clinically important difference (MCID) and minimal detectable change (MDC). To date, no studies have made a direct comparison between the responsiveness of the Persian version of NPDS, NDI, and NOOS questionnaires.

Methods: At the end of the study, 55 patients with chronic non-specific neck pain completed the NPDS, NDI, and NOOS questionnaires at the beginning and end of three weeks of physiotherapy treatment. Additionally, patients completed the global rating of change scale to differentiate between improved and unimproved patients. Comparison of responsiveness was performed using anchor-based methods (receiver operating characteristic (ROC) curve and correlation analysis). MCID and MDC were assessed to investigate relevant changes for each questionnaire.

Results: ROC curves analysis showed areas under the curves of 0.70, 0.64, and 0.43 to 0.63 for the NPDS, NDI, and NOOS subscales, respectively. The correlation coefficients between the global rating of the change scale and the change scores of the NPDS and NDI were 0.38 (*P<0.01*) and 0.30 (*P<0.05*), respectively. There were no significant correlations between NOOS subscales and global rating of change score (r=0.001- 0.21, *P>0.05*). The MCID for the NPDS, NDI, and NOOS subscales were 28.09 (score 0-100), 7.5 (score 0-50), and 13.75 to 28.64 (score 0-100), respectively. The MDCs were found to be in the following order: 47.1 points for NPDS, 36.1 for NDI, and 23.5 to 39.7 for NOOS subscales.

Conclusion: The Persian NPDS seems more responsive than the NDI and NOOS questionnaires. The level of clinically meaningful change in NDI, NPDS, and NOOS questionnaires is in the range of measurement error.

Level of evidence: IV

Keywords: Disability, Minimal clinically important difference, Neck disability questionnaires, Neck pain, ROC curve

Introduction

Chronic neck pain is a common musculoskeletal problem worldwide, with an estimated incidence of 19%. [1] Chronic non-specific neck pain is defined as pain that lasts more than three months and is provoked by neck movements or sustained neck posture without any specific underlying pathology. [1, 2] chronic neck pain can lead to disability, restriction of daily physical activities, long-term absence from work, and dependence on others. [3] Therefore, it is essential to objectively determine the severity of patients' pain, disability, quality of life, and treatment outcomes. Disability assessment questionnaires are useful tools for evaluating daily activities and social participation. [4] These questionnaires assess outcomes that have a significant impact on personal and social aspects of life, including pain relief, attendance at work, and daily performance. [5]

Corresponding Author: Salman Nazary Moghadam, Department of Physical Therapy, School of Paramedical and Rehabilitation Sciences, Mashhad University of Medical Sciences, Mashhad, Iran

Email: Nazaryms@mums.ac.ir

THE ONLINE VERSION OF THIS ARTICLE ABJS.MUMS.AC.IR

Having an accurate outcome measure that is adequately responsive is crucial for determining the effects of treatment and changes in health status. [6] Poor responsiveness of questionnaires can result in increased false-negative and/or false-positive results. [7] Responsiveness refers to the ability of a tool to assess changes in disease severity over time that have clinical significance. [8] Responsiveness can be evaluated through internal methods, which describe the ability of a measure to detect change over time, or external methods, which define the relationship between a change in a measure and a change in a reference measure. [9] External responsiveness values, such as receiver operating characteristic (ROC) curves and correlation analysis, are more attractive than internal responsiveness. They are frequently used in studies because they provide generalizable results that can be compared between studies. [10]

Assessing the minimal detectable change (MDC) and minimal clinically important difference (MCID) is crucial for understanding changes in questionnaire scores. [11] Evaluating MDC and MCID helps clinicians and researchers define a cut-off to distinguish measurement change from error and identify the change that patients find beneficial. [11]

There are currently three Persian questionnaires available for assessing neck pain: the neck outcome score (NOOS), neck disability index (NDI), and neck pain and disability scale (NPDS). These questionnaires are used in clinics and randomized controlled trials. However, limited data are available on the responsiveness of the Persian versions of NDI and NPDS questionnaires, and the responsiveness of the Persian version of the NOOS has not yet been assessed. Limited data are available for the Persian versions of NDI and NPDS questionnaires. [12] Furthermore, the available studies with NDI and NPDS questionnaires have not been conducted on similar populations, which makes it difficult to compare results. Hence, it is crucial to evaluate the questionnaires in the same sample of patients, at the same time, with similar methods to ensure maximum responsiveness. [4] Therefore, this study aims to evaluate the responsiveness of the three Persian questionnaires and determine which questionnaire is most suitable for measuring chronic non-specific neck pain in a patient sample. Additionally, this study aims to investigate the MCID and MDC in each questionnaire.

Materials and Methods

Study population and sampling

Patients who had been diagnosed with chronic non-specific neck pain by general practitioners or orthopedic surgeons were recruited from the physiotherapy clinic at Ghaem Hospital. Eligible patients were Persian-speaking individuals between the ages of 18 and 65 with a history of neck pain lasting for at least three months. Patients with cognitive disorders, a history of neurological, rheumatologic, cardiovascular, or respiratory problems, a history of heart attack or uncontrolled diabetes, malignancies, illiteracy, tremors, paresthesia, vertigo or imbalance, severe disabling dependency on alcohol and drugs, severe obesity (BMI≥40), a history of trauma or surgery in the cervical spine, or pregnancy were excluded from the study.

Referral data, physical examination, and a system of "red flags" [13, 14] were used to rule out specific neck pain. The sample size was determined based on a preliminary pilot study with ten assigned patients with chronic non-specific neck pain, which showed an improvement rate of 60%. A total sample size of 55 patients was calculated, with a confidence interval of 95%, power of 80%, and standard error of 13%. The study received approval from the local ethical committee (IR.MUMS.MEDICAL.REC.1398.178), and written informed consent was obtained from eligible patients prior to their participation in the study. Patients were free to withdraw from the study at any time.

Procedures

Two research assistants recorded the demographic characteristics and pre-treatment pain severity according to the visual analog scale (VAS). Patients completed the NPDS, NDI, and NOOS questionnaires before and immediately after the rehabilitation program on their own. The research assistants checked the questionnaires and returned incomplete or multiple responses to the patients for correction. To minimize exhaustion's effect on answers, the questionnaires were randomly organized. All patients attended ten physiotherapy sessions over a three-week period. The rehabilitation program included exercises focused on shoulder and neck muscles, scapulothoracic training, dry needling, cryotherapy, massage, and transcutaneous nerve stimulation. All patients received the same treatment from the same physiotherapist and were permitted to take mild analgesics and non-steroidal anti-inflammatory drugs (NSAIDs) during the rehabilitation program. Any disproportionate use of analgesics was monitored. The Global Rating of Change (GRC) was assessed after three weeks of physiotherapy treatment.

Visual Analog Scale (VAS)

A visual analog scale (VAS) was used to assess the patients' pain. Patients were asked to report their pain on a 100-millimeter line based on their perceived pain, with zero indicating the absence of pain and 100 representing the most severe pain that the patient has ever experienced.

Neck pain & disability scale questionnaire (NPDS)

The neck pain and disability scale (NPDS) is a 20-item questionnaire used to measure the severity of pain and disability in patients with neck pain. Each item has a 100 mm VAS line divided into five sections, where zero represents the absence of pain and disability, five represents the worst pain and disability, and the final score ranges between zero and 100. The Persian version of the NPDS has demonstrated acceptable reliability and validity for measuring pain and disability in Persian-speaking patients with neck pain. [3, 12, 15]

Neck disability index questionnaire (NDI)

The neck disability index (NDI) is a 10-item questionnaire used to evaluate the severity of pain and limitation in activities. Each item has six options ranging from zero, indicating the absence of pain and limitation, to five, representing maximum pain and limitation. The total score ranges from zero to 50 points, with a higher score indicating greater pain and disability. The Persian version of the NDI questionnaire has already been validated. [3, 12, 15]

Neck Outcome Score questionnaire (NOOS)

The neck outcome score (NOOS) questionnaire consists of 34 items divided into five independent subscales, which assess mobility and stiffness, symptoms, sleep disturbance, daily activity and pain, and participating in everyday life. Each item is scored based on the severity of pain and limitation on a scale of zero (indicating the absence of a significant problem) to four (representing maximum pain and limitation). The mean score for each subscale is multiplied by 100 and then divided by four. The resulting number is subtracted from 100. A higher score for each subscale indicates less disability. Each NOOS subscale is analyzed and interpreted separately. If any of the subscales are considered invalid, the other subscales can be reported independently. The Persian version of the NOOS questionnaire has already been validated. [16]

Global rating of change scale (GRC)

The Global Rating of Change (GRC) is a self-reported scale used to assess changes over time, particularly after treatment. This method has been frequently utilized in musculoskeletal studies as an external criterion for detecting clinically significant changes. The GRC is a simple and rapid tool for measuring improvement or deterioration, as perceived by the patient. [8, 17-22]

Patients were asked to describe their current status regarding neck pain compared to before starting physical therapy treatment using an 11-point scale ranging from -5 to +5. They were instructed to consider their neck pain before receiving treatment when responding. [17] The -5 score on the 11-point scale indicates a complete worsening of neck pain, while zero indicates no change, and +5 represents complete recovery from pain. The scale has been reported to have weak to moderate validity and moderate to good reliability. [8, 23, 24]

Statistical analysis

Statistical analysis was conducted using SPSS software version 22 (Chicago, IL). The level of statistical significance was set at $P<0.05$. Patients were divided into two groups based on their Global Rating of Change (GRC) scores: improved and unimproved. [8, 17, 19-22] setting a cut-off to distinguish patients into improved and stable groups is a subjective decision, and there is no universal standard for determining this cut-off point. It is generally based on clinical judgment and the research question being addressed. [8, 17] Previous studies have chosen different cut-offs for this purpose. [25-29] A number of studies [30, 31] have defined five groups based on GRC scores: -5 to -4 (marked worsening), -3 to -2 (minimal worsening), -1 to 1 (no change), 2 to 3 (minimal improvement), and 4 to 5 (marked improvement). Concerning prior research, [28,29,32,33] we used a conservative approach and considered the patients with marked improvement (GRC; +4, +5) as an improved group. Patients with a GRC≤3 were classified as unimproved.

To calculate the change score for each questionnaire, the follow-up score was subtracted from the baseline score for both the improved and stable groups. The absolute change score of NOOS subscales was reported since a higher follow-up score indicates an improvement in this questionnaire. Normality analysis indicated that neither group had a normal distribution. Therefore, changes in questionnaire scores were compared between the improved and stable groups using the nonparametric Mann-Whitney U test.

ROC curve analysis was used to assess responsiveness. [4,34-37] The responsiveness of the questionnaires was assessed using the area under the curve (AUC) in the ROC curve. The AUC ranges between 0.5 (no accuracy) and 1 (perfect accuracy), while an AUC≥0.7 was defined as acceptable responsiveness. [38] The gamma correlation coefficient analysis was done as another method to assess responsiveness based on the correlation analysis between the change scores of the NPDS, NDI, and NOOS questionnaires, as well as the raw scores of the GRC scale. [8,20-22] The correlation coefficients were classified based on Munro's classification as follows: 0.00–0.25=little to no correlation, 0.26–0.49=low correlation, 0.50–0.69=moderate correlation, 0.70–0.89=high correlation, and 0.90–1.00=very high correlation. [39]

The minimal clinically important difference (MCID) was calculated based on the nearest point to the upper left corner of the ROC curve, which is the best cut-off with the highest feasible specificity, and sensitivity. [18] According to Froud *et al.* the smallest sum of squares of [1-sensitivity] and [1-specificity] efficiently determines the nearest cut-point to the top-left corner of the ROC curve. [40]

MDC with a confidence level of 95% was evaluated using the following formula:

$$1.95 \times \sqrt{2} \times \text{standard error of measurement (SEM)}$$

Where 1.95 is 95th percentile of the standardized normal distribution, and SEM is the error that is seen after repeated measures in stable group. [33] SEM was calculated in stable patients using the following formula:

$$SD \times \sqrt{1-r}$$

Where SD is the standard deviation of the baseline scores and r is the test-retest reliability that corresponds to the intra-class correlation coefficient (ICC). [8, 20] MDC and MCID were used simultaneously to interpret the changes in questionnaires score.

Results

A total of 84 patients who met the entry criteria were invited to participate in the study and admitted to the physiotherapy clinic at the Ghaem Hospital. Of these, 15 (17.8%) patients declined to participate, and 14 (16.6%) abandoned the study before starting treatment due to personal (8) or economic (6) reasons. The final sample size consisted of 55 consecutive patients who were recruited until the required sample size was achieved. All patients who began treatment completed the rehabilitation program without any serious issues with the procedures, and no excessive use of analgesics was observed. There were no multiple answers or missing parts in the questionnaires. Demographic data and symptom severity are summarized in [Table 1].

According to the Global Rating of Change (GRC) scores, none of the participants reported a score lower than two after three weeks of physiotherapy treatment. Four participants (7.3%) had a score of 2, 16 participants (29.1%) had a score of 3, 25 participants (45.4%) had a score of 4, and 10 participants (18.2%) had a score of 5. Based on these scores, 35 patients (63.6%) were classified as improved,

while 20 (36.4%) were classified as unimproved.

The results of the NDI, NPDS, and NOOS questionnaires before and after the rehabilitation program, as well as the change scores in improved and stable groups, are summarized in [Table 2]. A comparison of the change scores in the questionnaires after three weeks of treatment revealed that only the NPDS questionnaire showed a significant difference between the improved and stable groups (P=0.02).

The NPDS questionnaire demonstrated acceptable responsiveness, with an Area under the Curve (AUC) of 0.70 [Figure 1]. The AUC for NDI was slightly higher at 0.64 compared to the NOOS subscales, which ranged from 0.43 to 0.63. To compare the AUCs of NDI and NOOS, we used DeLong's method 41 which showed no statistically significant difference between the two (P>0.05).

There were low significant correlations between the GRC score and the NPDS (r=0.38, P=0.009) and NDI (r=0.30, P=0.04) scores.

However, no significant correlation was observed between the GRC score and the NOOS subscales (r=0.001-0.21, 0.13<P<0.95).

The Minimum Clinically Important Difference (MCID) with the highest plausible sensitivity and specificity for each questionnaire is reported in [Table 3]. The estimated Minimum Detectable Change at the 95% confidence level (MDC95) for each questionnaire was as follows: 47.1 points (ICC: 0.33, SEM: 17.1) for NPDS, 36.1 points (ICC: 0.32, SEM: 13.1) for NDI, 30.9 points (ICC: 0.70, SEM: 11.2) for mobility and stiffness (NOOS), 23.5 points (ICC: 0.68, SEM: 8.4) for symptom (NOOS), 39.7 points (ICC: 0.64, SEM: 14.4) for sleep disturbance (NOOS), 31 points (ICC: 0.50, SEM: 11.2) for daily activity and pain (NOOS), and 30.6 points (ICC: 0.21, SEM: 11.1) for participating in everyday life (NOOS). Any change in questionnaire score above the values of both MCID and MDC was considered relevant. The psychometric properties of the NPDS, NDI, and NOOS questionnaires are summarized in [Table 3].

Table 1. Demographic characteristics of the patients with chronic nonspecific neck pain (n=55)

Features	Data
Gender N (%)	
Male	8 (14.5)
Female	47 (85.5)
Age (year, Mean ± SD)	43.11±11.72
Body mass index (kg/m²; Mean ± SD)	26.09±4.33
Pain (visual analogs scale, Mean ± SD)	71.29±15.06

Pain is assessed at baseline

Table 2. Base line and follow-up score of NDI, NPDS, and NOOS questionnaires in improved and stable groups

	Baseline mean (SD)	Follow-up mean (SD)	Change mean (SD)
NDI			
Improved (n=35)	20.8 (6.4)	10.4 (5.6)	10.3 (6.7)
Stable (n=20)	22.3 (15.9)	12.9 (5.6)	10.1 (16.4)
NPDS			
Improved (n=35)	56 (16.9)	22.4 (12)	33.6 (19.8)*
Stable (n=20)	53.6 (20.9)	36.1 (15.3)	18.8 (20.3)
Mobility & stiffness (NOOS)			
Improved (n=35)	56.1 (17.2)	73.5 (18.3)	17.4 (20.1)
Stable (n=20)	47.6 (20.5)	70 (17.7)	22.3 (9.8)
Symptoms (NOOS)			
Improved (n=35)	47.7 (16.8)	70 (18.7)	22 (18.2)
Stable (n=20)	52.5 (15)	68.5 (13.7)	16 (11.8)
Sleep disturbance (NOOS)			
Improved (n=35)	53.3 (22.7)	76.7 (20.1)	22.9 (26.8)
Stable (n=20)	51.2 (24)	72.1 (20.5)	20.9 (13)
Daily activity & pain (NOOS)	47.4 (19)	69.7 (14.3)	23.4 (19.5)
Improved (n=35)	40.7 (15.9)	61.4 (14.6)	19 (15.6)
Stable (n=20)			
Participating in everyday life (NOOS)	51.9 (16)	69.9 (17.4)	19.1 (23.1)
Improved (n=35)	53.6 (12.5)	65.9 (8.7)	11.3 (9.7)
Stable (n=20)			

NDI neck disability index, NPDS neck pain & disability scale, NOOS neck outcome score

*According to the Mann-Whitney U test, the NPDS was the only questionnaire that showed a statistically significant difference between the change score of improved and stable patients, P=0.02

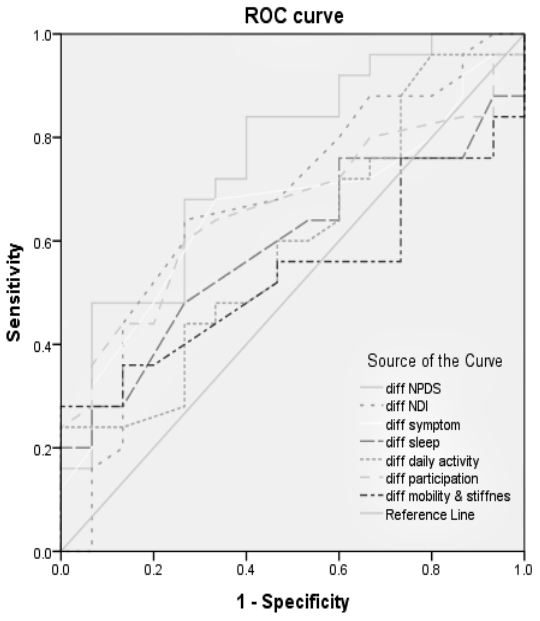

Figure 1. Receiver operating characteristic (ROC) curves of the NPDS, NDI and NOOS subscales

Table 3. Gamma correlation coefficient, area under the ROC curve, minimal detectable change, and minimal clinically important difference based on global rating scale as an external criterion

	Score	AUC (95%CI)	MDC	MCID	Sensitivity	Specificity	Gamma correlation (P value)
NPDS	0-100	0.70 (0.55-0.86)	47.1	28.09	0.65	0.64	0.38 (<0.01)
NDI	0-50	0.64 (0.47-0.81)	36.1	7.5	0.60	0.70	0.30 (<0.05)
Mobility & stiffness (NOOS)	0-100	0.43 (0.28-0.58)	30.9	28.64	0.31	0.80	-0.01 (0.95)
Symptom (NOOS)	0-100	0.63 (0.48-0.77)	23.5	17.5	0.70	0.60	0.21 (0.13)
Sleep disturbance (NOOS)	0-100	0.53 (0.38-0.69)	39.7	28.12	0.41	0.70	0.08 (0.56)
Daily activity & pain (NOOS)	0-100	0.55 (0.39-0.71)	31	21.87	0.52	0.52	0.15 (0.27)
Participating in everyday life (NOOS)	0-100	0.62 (0.46-0.78)	30.6	13.75	0.57	0.70	0.22 (0.15)

AUC area under the curve, MCID minimal clinically important difference, MDC minimal detectable change
NDI neck disability index, NPDS neck pain & disability scale, NOOS neck outcome

Discussion

The aim of this study was to assess the responsiveness properties of three Persian questionnaires through the ROC curve and correlation analysis, using the Global Rating of Change (GRC) scale as an external anchor.

The comparison of the three questionnaires showed that the Neck Pain and Disability Scale (NPDS) exhibited better responsiveness properties in terms of ROC curves. The inadequate sensitivity of the Neck Disability Index (NDI) and Neck Outcome Score (NOOS) to detect changes may be related to some items in the questionnaire that might be unclear or confusing for patients. An investigation into the content validity of the Persian versions of NPDS and NDI revealed that both questionnaires were easy to complete and relevant to patients experiencing neck problems. [3] The study also revealed a high rate of missing values for the driving item in the NDI questionnaire among patients who did not drive regularly. This is because asking about driving issues in the NDI questionnaire assumes that the responder drives regularly. Patients who do not frequently drive might be unsure how to answer this question, which can lead to inaccurate NDI results. [3] We assume that the NPDS questionnaire has addressed this problem by asking specifically about how pain interferes with driving or riding in a car. In contrast, the NOOS uses the phrases "neck problem" or "neck pain" to inquire about patient difficulties in each subscale. However, according to Young et al., using these phrases may be confusing to patients who experience scapular or arm pain instead of primary neck pain. [42] In this case, the patient might report no improvement in their neck pain while the true change has occurred.

As shown in [Table 3], MCID and MDC assessments provide different cut-offs with varying levels of certainty for detecting true changes. It is worth noting that most neck pain studies, including the current study, have reported higher MDC values than MCID values. [8,20,42-45] This finding may be due to patients with longer durations of neck pain having lower expectations and interpreting small changes as improvements. [20, 46] setting a cut-off for improvement as either MDC or MCID depends on the level of confidence that researchers or clinicians require to distinguish true improvement. Using a more conservative MDC cut-off would be reasonable in expensive and challenging treatments, while MCID could be used in primary care centers. [8]

NDI questionnaire

The current study reported an Area under the Curve (AUC) of 0.64 (95% CI: 0.47-0.81), a Minimum Clinically Important Difference (MCID) of 7.5 out of 50 (sensitivity: 0.60, specificity: 0.70), and a correlation coefficient of 0.30 (P=0.04) with the GRC scale for Persian NDI. An earlier assessment of Persian NDI showed an AUC of 0.68 (95% CI: 0.53-0.82), an MCID of 7.5 (sensitivity: 0.70, specificity: 0.75), and a correlation coefficient of 0.33 (P=0.001) with the GRC scale. [22] These findings were in complete agreement with the outcomes of the present study. Investigations into the ROC curves have shown an AUC of 0.57 to 0.96 and MCID of 3.5 to 19 for non-Persian versions of NDI. These findings were in complete agreement with the outcomes of the present study. Investigations into the ROC curves have shown an AUC of 0.57 to 0.96 and MCID of 3.5 to 19 for non-Persian versions of NDI. [8,19-21,42-45,47,48] The correlation coefficients between non-Persian NDI and GRC have also been reported as 0.32 to 0.73.[19-21, 43, 49] Applying different cut-offs in the GRC scale has been identified as a major contributing factor to the divergent interpretability of NDI responsiveness. [8, 43]

In the present study, the reported MDC95 for NDI (36.1 points) was higher than the values reported earlier (5.96 to 13.4 points).[8, 19, 20,43,47,48] MDC is calculated in stable patients, and it is highly dependent on estimated reliability. [20] Thus, variability of baseline scores (that raises the standard deviation), follow-up period, and external criterion are of great importance in MDC assessment. Variation of these factors between different studies can result in variations in MDCs.

NPDS questionnaire

Our study on the Persian version of the NPDS showed an acceptable responsiveness (AUC=0.70) with MCID of 28.09 out of 100 (sensitivity: 0.65, specificity: 0.64), MDC of 47.1, and a low significant correlation (r=0.38 and P<0.01) with the GRC score.

Previous studies examining the responsiveness of Italian and English versions of the NPDS have reported the AUCs of 0.75 and 0.91, MCIDs of 10 (sensitivity: 0.93, specificity: 0.83) and 11.5 (sensitivity: 0.74, specificity: 0.70), and the correlation coefficients of 0.48 and 0.59, respectively. [8, 21] MDC of 31.7 and 10.5 has been reported for English and German NPDS. [8, 50]

Possible causes of these discrepancies are the different external indicators of change, target population, and different formulas to calculate MDC and MCID. The improvement criterion for assessment of English NPDS was set at two or higher on a 7-point scale. [8] A five-level Likert scale with two improvement levels was used to investigate the Italian NPDS. [21] Childs *et al.* also clarified that a shorter treatment period decreases the reminder error and lowers questionnaire responsiveness. [51]

NOOS questionnaire

We demonstrated AUCs of 0.43 to 0.63, MCIDs of 13.75 to 28.64, and MDCs of 23.5 to 39.7 for the Persian NOOS subscales. Unlike the NDI and NPDS questionnaires, the Persian NOOS showed no significant correlation with the GRC scale. Juul *et al.* indicated good internal responsiveness, MCID of 15.0 to 24.1, and MDC of 10.3 to 18 for English NOOS subscales. [52] Considering the present study, differences in responsiveness properties relate to the different approaches and formulas we adopted to measure responsiveness. The Chinese version of the NOOS questionnaire has good sensitivity to change. However, changes in mobility and stiffness showed no significant difference before and after treatment. [53] Our study similarly reported the inability of the mobility and stiffness subscale to dichotomize improved and unimproved patients (AUC=0.43). Although ROC curve analysis showed a moderate power for the NOOS questionnaire in detecting treatment change, it still seems comparable to the NPDS and NDI questionnaires.

There are some limitations regarding the present study. There are some limitations regarding the present study. First, although the sample size of the current study was calculated based on a pilot study with a power of 80%, small sample size is still the main weakness of this study. Regarding this limitation, caution must be applied when concluding that the responsiveness of NDI and NOOS questionnaires is unacceptable. Second, we conducted a study on Persian-speaking patients with no history of surgery. In this respect, our findings may not be attributed to other populations. Third, we used the GRC scale as the best available tool for defining responsiveness. Nevertheless, one of the limitations of GRC is the reminder error of patients' before-treatment status, which can affect the assessment of MCID and MDC in the questionnaires. [54] It should be considered that our study is among those few available studies on Persian versions. Moreover, this study assessed the responsiveness of Persian versions of the NOOS, NDI, and NPDS questionnaires in the same population, which led to an accurate comparison. Assessment of MCID and MDC values for each questionnaire will offer exceptional guidance for researchers to distinguish between improved and unimproved patients in forthcoming studies on patients with neck pain. This study gave an initial insight into the responsiveness of NDI, NPDS, and NOOS questionnaires. The findings of this study will enhance patient management approaches and research methods by evaluating true clinical changes. However, future research with a larger sample size and more power to investigate the responsiveness of self-report questionnaires will give a brighter idea of how to apply these questionnaires in clinical or research settings.

Conclusion

According to the findings of the current study, the Persian NPDS questionnaire seems more responsive than the NDI and NOOS questionnaires. The level of clinically meaningful change in NDI, NPDS, and NOOS questionnaires is in the range of measurement error.

Acknowledgement

The authors thank Mohammad Mahdi Shadman for assistance in data collection.

Conflict of interest: The authors do NOT have any potential conflicts of interest for this manuscript.

Funding: The present study was financially supported by the Mashhad University of Medical Science, Mashhad, Iran (Grant Number: 971497). None of the mentioned authors, their immediate family, and any research foundation with which they are affiliated did not receive any financial payments or other benefits from any commercial entity related to the subject of this article.

Seyed Ramin Dabiri MD [1]
Mohammad Reza Tehrani PhD [2]
Farzad Omidi-Kashani MD [1]
Hamed Mami-Pour PT [2]
Hamed Tabesh PhD [3]
Ali Moradi PhD[4]
Salman Nazary-Moghadam PhD [2,4]

1 Orthopedic Department, Imam Reza Hospital, Mashhad University of Medical Sciences, Mashhad, Iran
2 Department of Physical Therapy, School of Paramedical and Rehabilitation Sciences, Mashhad University of Medical Sciences, Mashhad, Iran
3 Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
4 Orthopedic Research Centre, Ghaem Hospital, Mashhad University of Medical Sciences, Mashhad, Iran

REFERENCES

1. Huisstede BM, Wijnhoven HA, Bierma-Zeinstra SM, Koes BW, Verhaar JA, Picavet S. Prevalence and characteristics of complaints of the arm, neck, and/or shoulder (CANS) in the open population. Clin J Pain. 2008; 24(3):253-259. doi: 10.1097/AJP.0b013e318160a8b4.
2. Ferrari R, Russell AS. Regional musculoskeletal conditions: neck pain. Best Pract Res Clin Rheumatol. 2003; 17(1):57-70. doi: 10.1016/s1521-6942(02)00097-9.
3. Mousavi SJ, Parnianpour M, Montazeri A, et al. Translation and validation study of the Iranian versions of the Neck Disability Index and the Neck Pain and Disability Scale. Spine (Phila Pa 1976). 2007; 32(26):E825-831. doi: 10.1097/BRS.0b013e31815ce6dd.
4. Deyo R. Measuring the functional status of patients with low back pain. Arch Phys Med Rehabil. 1988; 69(12):1044-1053.
5. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine (Phila Pa 1976). 1998; 23(18):2003-2013. doi: 10.1097/00007632-199809150-00018.
6. Shafiee E, Farzad M, Karbalaei M. A Systematic Review of Self-Reported Outcome Measures Assessing Disability Following Hand and Upper Extremity Conditions in Persian Population. Arch Bone Jt Surg. 2021; 9(2):141-151. doi: 10.22038/abjs.2020.48859.2423.
7. Pengel LH, Refshauge KM, Maher CG. Responsiveness of pain, disability, and physical impairment outcomes in patients with low back pain. Spine (Phila Pa 1976). 2004;29(8):879-883. doi: 10.1097/00007632-200404150-00011.
8. Jorritsma W, Dijkstra PU, de Vries GE, Geertzen JH, Reneman MF. Detecting relevant changes and responsiveness of Neck Pain and Disability Scale and Neck Disability Index. Eur Spine J. 2012; 21(12):2550-2557. doi: 10.1007/s00586-012-2407-8.
9. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol. 1989; 42(5):403-408. doi: 10.1016/0895-4356(89)90128-5.
10. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000; 53(5):459-468. doi: 10.1016/s0895-4356(99)00206-1.
11. Lehman LA, Velozo CA. Ability to detect change in patient function: responsiveness designs and methods of calculation. J Hand Ther. 2010; 23(4):361-370; quiz 371. doi: 10.1016/j.jht.2010.05.003.
12. Schellingerhout JM, Heymans MW, Verhagen AP, de Vet HC, Koes BW, Terwee CB. Measurement properties of translated versions of neck-specific questionnaires: a systematic review. BMC Med Res Methodol 2011; 11(1):1-14. doi: 10.1186/1471-2288-11-87.
13. Nordin M, Carragee EJ, Hogg-Johnson S, et al. Assessment of neck pain and its associated disorders: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. J Manipulative Physiol Ther. 2009; 32(2 Suppl):S117-140. doi: 10.1016/j.jmpt.2008.11.016.
14. Rubinstein SM, van Tulder M. A best-evidence review of diagnostic procedures for neck and low-back pain. Best Pract Res Clin Rheumatol. 2008; 22(3):471-482. doi: 10.1016/j.berh.2007.12.003.
15. Alexopoulos EC, Stathi I-C, Charizani F. Prevalence of musculoskeletal disorders in dentists. BMC Musculoskelet Disord .2004; 5(1):1-8. doi: 10.1186/1471-2474-5-16.
16. Mazaheri Tehrani MR, Nazary-Moghadam S, Dabiri SR. Translation, cross-cultural adaptation, reliability, and validity of the Neck OutcOme Score in Persian speaking participants with non-specific neck pain. Available at: http://www.koos.nu/noospersian.pdf. Accessed November, 2021.
17. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009;17(3):163-170. doi: 10.1179/jmt.2009.17.3.163.
18. de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res. 2007; 16(1):131-142. doi: 10.1007/s11136-006-9109-9.
19. Guzy G, Vernon H, Polczyk R, Szpitalak M. Psychometric validation of the authorized Polish version of the Neck

Disability Index. Disabil Rehabil. 2013; 35(25):2132-2137. doi: 10.3109/09638288.2013.771706.
20. Pereira M, Cruz EB, Domingues L, Duarte S, Carnide F, Fernandes R. Responsiveness and Interpretability of the Portuguese Version of the Neck Disability Index in Patients With Chronic Neck Pain Undergoing Physiotherapy. Spine (Phila Pa 1976). 2015;40(22):E1180-1186. doi: 10.1097/BRS.0000000000001034.
21. Monticone M, Ambrosini E, Vernon H, et al. Responsiveness and minimal important changes for the Neck Disability Index and the Neck Pain Disability Scale in Italian subjects with chronic neck pain. Eur Spine J. 2015; 24(12):2821-2827. doi: 10.1007/s00586-015-3785-5.
22. Salehi R, Negahban H, Saghayezhian N, Saadat M. The Responsiveness of the Persian Version of Neck Disability Index and Functional Rating Index Following Physiotherapy Intervention in People with Chronic Neck Pain. Iran J Med Sci. 2019; 44(5):390-396. doi: 10.30476/ijms.2019.44963.
23. Ngo T, Stupar M, Côté P, Boyle E, Shearer H. A study of the test–retest reliability of the self-perceived general recovery and self-perceived change in neck pain questions in patients with recent whiplash-associated disorders. European Spine Journal. 2010; 19(6):957-962. doi: 10.1007/s00586-010-1289-x.
24. Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H. Capturing the patient's view of change as a clinical outcome measure. Jama. 1999; 282(12):1157-1162. doi: 10.1001/jama.282.12.1157.
25. Schreiber S, Parent EC, Hill DL, Hedden DM, Moreau MJ, Southon SC. Patients with adolescent idiopathic scoliosis perceive positive improvements regardless of change in the Cobb angle - Results from a randomized controlled trial comparing a 6-month Schroth intervention added to standard care and standard care alone. SOSORT 2018 Award winner. BMC Musculoskelet Disord. 2019; 20(1):319. doi: 10.1186/s12891-019-2695-9.
26. Nordqvist SF, Boesen VB, Rasmussen Å K, et al. Determining minimal important change for the thyroid-related quality of life questionnaire ThyPRO. Endocr Connect. 2021; 10(3):316-324. doi: 10.1530/EC-21-0026. doi: 10.1530/EC-21-0026.
27. Maughan EF, Lewis JS. Outcome measures in chronic low back pain. Eur Spine J. 2010;19(9):1484-1494. doi: 10.1007/s00586-010-1353-6.
28. Freitas P, Pires D, Nunes C, Cruz EB. Cross-cultural adaptation and psychometric properties of the European Portuguese version of the Global Perceived Effect Scale in patients with chronic low back pain. Disabil Rehabil. 2021; 43(7):1008-1014. doi: 10.1080/09638288.2019.1648568.
29. Ferreira ML, Ferreira PH, Latimer J, et al. Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. Pain. 2007; 131(1-2):31-37. doi: 10.1016/j.pain.2006.12.008.
30. Tashjian RZ, Deloach J, Green A, Porucznik CA, Powell AP. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am. 2010; 92(2):296-303. doi: 10.2106/JBJS.H.01296.
31. Sorensen AA, Howard D, Tan WH, Ketchersid J, Calfee RP. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg Am. 2013; 38(4):641-649. doi: 10.1016/j.jhsa.2012.12.032.
32. Costa LO, Maher CG, Latimer J, et al. Clinimetric testing of three self-report outcome measures for low back pain patients in Brazil: which one is the best? Spine (Phila Pa 1976). 2008; 33(22):2459-2463. doi: 10.1097/BRS.0b013e3181849dbe.
33. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. 2006; 4:54. doi: 10.1186/1477-7525-4-54.
34. Terwee CB, Schellingerhout JM, Verhagen AP, Koes BW, de Vet HC. Methodological quality of studies on the measurement properties of neck pain and disability questionnaires: a systematic review. Journal of Manipulative and Physiological Therapeutics. 2011; 34(4):261-272. doi: 10.1016/j.jmpt.2011.04.003.
35. Holtermann A, Hansen JV, Burr H, Søgaard K. Prognostic factors for long-term sickness absence among employees with neck-shoulder and low-back pain. Scand J Work Environ Health. 2010:34-41. doi: 10.5271/sjweh.2883.
36. Harrison JJ, Yorgey MK, Csiernik AJ, Vogler JH, Games KE. Clinician-friendly physical performance tests for the knee. J Athl Train. 2017; 52(11):1068-1069. doi: 10.4085/1062-6050-52.11.19.
37. Yeung TS, Wessel J, Stratford P, MacDermid J. Reliability, validity, and responsiveness of the lower extremity functional scale for inpatients of an orthopaedic rehabilitation ward. J Orthop Sports Phys Ther. 2009; 39(6):468-477. doi: 10.2519/jospt.2009.2971.
38. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther. 2002; 82(1):8-24. doi: 10.1093/ptj/82.1.8.
39. Zeinalzadeh A, Nazary-Moghadam S, Sayyed Hoseinian SH, Ebrahimzadeh MH, Imani E, Karimpour S. Intra- and Inter-Session Reliability of Methods for Measuring Reaction Time in Participants with and without Patellofemoral Pain Syndrome. Arch Bone Jt Surg. 2021; 9(1):102-109. doi: 10.22038/abjs.2020.46213.2270.
40. Froud R, Abel G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: the forgotten lesson of pythagoras. Theoretical considerations and an example application of change in health status. PLoS One. 2014; 9(12):e114468. doi: 10.1371/journal.pone.0114468.
41. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988; 44(3):837-845.
42. Young IA, Cleland JA, Michener LA, Brown C. Reliability, construct validity, and responsiveness of the neck disability index, patient-specific functional scale, and numeric pain rating scale in patients with cervical radiculopathy. Am J Phys Med Rehabil. 2010; 89(10):831-839. doi: 10.1097/PHM.0b013e3181ec98e6.
43. Young BA, Walker MJ, Strunce JB, Boyles RE, Whitman JM, Childs JD. Responsiveness of the Neck Disability Index in

patients with mechanical neck disorders. Spine J. 2009;9(10):802-808. doi: 10.1016/j.spinee.2009.06.002.
44. Cleland JA, Childs JD, Whitman JM. Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. Arch Phys Med Rehabil. 2008; 89(1):69-74. doi: 10.1016/j.apmr.2007.08.126.
45. Cleland JA, Fritz JM, Whitman JM, Palmer JA. The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. Spine (Phila Pa 1976). 2006; 31(5):598-602. doi: 10.1097/01.brs.0000201241.90914.22.
46. Demoulin C, Ostelo R, Knottnerus JA, Smeets RJ. Quebec Back Pain Disability Scale was responsive and showed reasonable interpretability after a multidisciplinary treatment. J Clin Epidemiol. 2010; 63(11):1249-1255. doi: 10.1016/j.jclinepi.2009.08.029.
47. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine (Phila Pa 1976). 2007; 32(26):3047-3051. doi: 10.1097/BRS.0b013e31815cf75b.
48. Johansen JB, Roe C, Bakke E, Mengshoel AM, Andelic N. Reliability and responsiveness of the Norwegian version of the Neck Disability Index. Scand J Pain. 2014; 5(1):28-33. doi: 10.1016/j.sjpain.2013.10.001.
49. Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M. Responsiveness of pain and disability measures for chronic whiplash. Spine (Phila Pa 1976). 2007;32(5):580-585. doi: 10.1097/01.brs.0000256380.71056.6d.
50. Bremerich FH, Grob D, Dvorak J, Mannion AF. 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 (Phila Pa 1976). 2008; 33(9):1018-1027. doi: 10.1097/BRS.0b013e31816c9107.
51. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine. 2005; 30(11):1331-1334. doi: 10.1097/01.brs.0000164099.92112.29.
52. Juul T, Søgaard K, Davis AM, Roos EM. Psychometric properties of the Neck OutcOme Score, Neck Disability Index, and Short Form–36 were evaluated in patients with neck pain. Journal of clinical epidemiology. 2016; 79:31-40. doi: 10.1016/j.jclinepi.2016.03.015.
53. Li P, Ning N, Liu H, et al. Responsiveness of Chinese version of neck outcome score in subjects with neck pain. Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi. 2018; 32(5):554-557. doi: 10.7507/1002-1892.201801021.
54. Cook CE. Clinimetrics corner: the minimal clinically important change score (MCID): a necessary pretense. J Man Manip Ther. 2008; 16(4):82E-83E. doi: 10.1179/jmt.2008.16.4.82E