Minimal clinically important change of the neck disability index and the numerical rating scale for patients with neck pain

Pool, J.J.M.; Ostelo, R.W.J.G.; Hoving, J.L.; Bouter, L.M.; de Vet, H.C.W.

published in
Spine
2007

DOI (link to publisher)
10.1097/BRS.0b013e31815cf75b

document version
Publisher's PDF, also known as Version of record

SPINE Volume 32, Number 26, pp 3047–3051

Minimal Clinically Important Change of the Neck Disability Index and the Numerical Rating Scale for Patients With Neck Pain

Jan J. M. Pool, PhD, PT, MT,* Raymond W. J. G. Ostelo, PhD,*† Jan L. Hoving, PhD,*‡ Lex M. Bouter, PhD,* and Henrica C. W. de Vet, PhD*

Study Design. Prospective, single-cohort study.

Objective. To assess the minimal clinically important change (MCIC) on the Neck Disability Index (NDI) and the Numerical Rating Scale (NRS) for pain in patients with neck pain.

Summary of Background Data. Both measurement instruments are frequently used in research and clinical practice, but which changes are clinically relevant is still unknown.

Methods. The MCIC was estimated with 2 different methods, both integrating an anchor-based and distribution-based approach: the minimal detectable change (MDC) and the optimal cutoff point of the ROC curve. The study population consisted of 183 patients with nonspecific neck pain.

Results. The results show an MDC of 10.5 points for the NDI (scale range, 0–50) and 4.3 points for the NRS (scale range, 0–10), and optimal cutoff points of the ROC curve of 3.5 for the NDI and 2.5 for the NRS.

Conclusion. The estimated MCIC should be used as an indication for relevant changes in clinical practice. Using the optimal cutoff point of the ROC curve, false positives and false negatives are equally weighted; and if there are no objections doing so, the optimal cutoff point of the ROC curve may be a good choice.

Key words: minimal clinically important change, neck pain, disability, pain. **Spine 2007;32:3047–3051**

Neck pain is a common musculoskeletal disorder, and its point prevalence in the general population of the Netherlands varies between 9% and 22%.[1,2] Approximately one third of all adults will experience neck pain during the course of 1 year.[3] Although neck pain is often self-limiting within a few weeks, 40% of the patients contact their general practitioner. Of these, 30% are referred for further diagnosis to a medical specialist and 32% to physiotherapy, manual therapy, or some other type of conservative therapy.[1,2] To evaluate the effect of treatment for neck disorders, it is necessary to assess relevant outcome measures, such as pain and functional disability.

The Neck Pain Disability Index (NDI) is a questionnaire that is commonly used in clinical trials to measure the functional status of patients with neck pain.[4,5] The NDI was originally developed for assessing the functional status of patients with disabling neck pain, particularly whiplash-associated disorders.[6] The psychometric properties of the NDI, in terms of validity and reproducibility, is still a topic of research,[7–10] which also counts for how to interpret change scores.[11]

Vernon and Mior[10] assessed face validity through peer-review and patient feedback sessions and concurrent validity of the NDI on the Visual Analogue Scale (n = 10 and a correlation of 0.60) and the McGill Pain Questionnaire (n = 30 and a correlation of 0.69). Furthermore, a test-retest reliability was calculated and found a correlation of 0.89. Hoving *et al*[8] assessed the construct and content validity of the NDI using 71 patients with whiplash-associated disorders, comparing the NDI with a patient preference questionnaire, the correlation was 0.57 with the remark that the patient preference questionnaire identified more disabilities.

In the review of Pietrobon *et al*,[9] the NDI was found to be 1-dimensional, the validity was established by concurrent criterion validity and showed a correlation coefficient of 0.6 with the VAS and 0.7 with the McGill Pain Questionnaire and was reported to be the scale, which was most widely validated among different patient populations, the responsiveness was not reported. The Numerical Rating Scale (NRS) is frequently used to measure pain intensity.[7] Patients are asked to rate their pain on a 0 to10 point rating scale. Bolton and Wilkinson[7] compared the responsiveness of 3 pain scales, Visual Analogue Scale, the Verbal Rating Scale, and the NRS on patients, n = 79, and using effects sizes. The NRS showed to be the most responsive (effect size 0.86).

For the interpretation of treatment effects, it is not only important to know whether results are statistically significant, but also whether they are relevant for patients or clinicians. Consequently, insight into the clinically important difference or change is needed. A well-accepted definition of minimal clinically important difference has been proposed by Jaeschke *et al*[12] as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would

From the *EMGO Institute, VU University Medical Centre, Amsterdam, the Netherlands; †Department of 'Allied Health Care Research, Amsterdam School for Allied Health Care Education, Amsterdam, the Netherlands; and ‡Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia.
Acknowledgment date: December 8, 2005. First revision date: June 19, 2006. Second revision date: February 20, 2007. Acceptance date: June 1, 2007.
The manuscript submitted does not contain information about medical device(s)/drug(s).
Foundation funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.
Address correspondence and reprint requests to Jan J. M. Pool, PhD, PT, MT, EMGO institute, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands; E-mail: j.pool@vumc.nl

mandate, in the absence of troublesome side-effects and excessive cost, a change in patient's management." We prefer to use the term minimal clinically important change (MCIC) for the change in health status *within* patients and the term minimal clinically important difference to indicate differences *between* patients. The aim of this study is to assess the MCIC of both the NDI and the NRS for pain in patients with neck pain. A number of different methods have been proposed to determine the MCIC.[13] Crosby *et al* distinguish an anchor-based and a distribution based method.[14] Anchor-based approaches use an external criterion to operationalize clinically important change, and distribution-based approaches are based on statistical characteristics of the sample, for example effect sizes, relating observed change to the sample variation.[14] For neck pain, effect sizes and standard response mean have been used.[15,16] The estimate of the effect size and of the standard response mean are parameters without any dimension which makes it difficult to interpret them for clinicians. Therefore, we used 2 methods, both integrating an anchor-based and a distribution-based approach: the minimal detectable change (MDC) and the optimal cutoff point of the receiver operator characteristic curve (ROC),[17] because these are expressed in scale points which improves the interpretability of change scores. For both methods the global perceived effect (GPE) is used as an external criterion for change.[4]

■ Methods

Study Population. The study population consisted of participants included in a randomized controlled trial to evaluate the effectiveness of 3 conservative treatment options for neck pain.[4] General practitioners (n = 42) referred patients with neck pain to 1 of the 4 research centers for study selection. The eligibility criteria were: age between 18 and 70 years, pain and/or stiffness in the neck for at least 2 weeks, neck symptoms reproducible during physical examination, willingness to adhere to treatment and measurement regimens, and no physical therapy or manual therapy for neck pain during the previous 6 months. The participants were randomly allocated to either physiotherapy, manual therapy, or continued care provided by a general practitioner. Data were collected at the research centers at baseline, after 7 weeks and after 52 weeks of follow-up, for the present analysis only measurements at baseline and after 7 weeks of follow-up were used. Approval was obtained from the medical ethics committee of the VU University Medical Center, Amsterdam.

Measurement Instruments. The NDI consists of 10 items addressing functional activities, such as personal care, lifting, reading, work, driving or cycling, sleeping and recreational activities, and a number of symptoms such as pain intensity, concentration, and headache.[10] For each item, answering options range from 0 = no disability to 5 = total disability, resulting in a total range of scores from 0 to 50 points.

The NRS is an 11-point rating scale for pain in which 0 = no pain and 10 = worst pain imaginable. Patients were asked to rate their average pain in the previous week.

To assess the GPE, the patients rated this on a 6-point Likert scale ranging from 1 = completely recovered to 6 = much worse.[4] We trichotomized this scale: patients who indicated that they were "much worse" were labeled as "importantly deteriorated"; patients who indicated that they were "slightly improved," "no change," or " slightly worse," were labeled as "not importantly changed"; and were consequently considered not to have experienced an important or clinically relevant improvement or deterioration; patients who indicated that they were "completely recovered" or "much improved" were labeled as "importantly improved." The distributions of these subgroups (labels) were used to estimate the MCIC, thereby integrating anchor based and distribution based methods.

Data Analysis. We defined the MDC as the smallest difference in a score that can be detected, considering the variation in changes on the NDI and the NRS observed in persons who were not importantly changed on the external criterion.[17,18] To determine the MDC, first the standard error of measurement (SEM) was assessed. The SEM indicates the precision of outcome measure and was estimated by taking the square root of the within-subject variance of patients categorized as "not importantly changed" on the GPE. To be 95% confident that observed change is real change and not caused by measurements error, the MDC was calculated as $1.96 * \sqrt{2} * \text{SEM}$. Observed change is a result of 2 measurements, baseline and follow-up and therefore occur twice, hence $\sqrt{2}$. Changes greater than the MDC are consequently considered to indicate real change[17–19] because only "not importantly improved" patients were assessed.

The optimal cutoff point of the ROC curve considers the NDI and the NRS as a diagnostic test for discriminating between "importantly improved" and "not importantly improved" patients. The external anchor (GPE) functions as the gold standard and distinguishes those patients who showed a clinically important change from those who did not. The diagnostic accuracy of a measurement instrument can thus be expressed in terms of sensitivity and specificity for clinically important change, and can be depicted in a ROC curve. The ROC is a graph of the percentage of true-positive values (sensitivity) *versus* the percentage of false-positive values (1 − specificity) for each possible cutoff change score of the NDI and the NRS. The optimal cutoff point was chosen in such a way that the overall misclassification, *i.e.*, the sum of the percentages of false-positive and false-negative outcomes, was minimized. False-positive outcomes are persons who are "not importantly changed" according to the GPE but show a change that is greater than the cutoff value on the measurement instrument. False negatives are persons who are "importantly improved" on the GPE but show less change than the cutoff value on the measurement instrument. For all statistics, SPSS 12 for Windows was used.

■ Results

Patient Characteristics

During a period of 22 months, a total of 183 patients with nonspecific neck pain were included, of whom completed the 7-week follow-up. The mean age of the patients was 45.8 years and 60.8% was female. Table 1 shows the characteristics of the participants at baseline.

Table 2 shows the mean scores and standard deviations at baseline and after 7 weeks of follow-up for sub-

Table 1. Characteristics of Patients (N = 183)

Characteristic	Value
Age (yr) (mean ± SD)	45.8 ± 11.6
Female (%)	60.8
Previous episodes of neck complaints (%)	64.8
Duration of current episode (%)	
2–6 wk	48.0
7–12 wk	26.1
≥13 wk	25.9
Pain score (mean ± SD)*	6.0 ± 1.9
NDI score (mean ± SD)†	14.5 ± 7.0
Work status employed (%)	73.8

*Numeric Rating Scale ranging from 0 (no pain) to 10 (worst pain).
†Neck Disability Index: 10 items ranging from 0 to 5 points; maximal disability = 50 points.

Table 3. Minimal Detectable Change and Several Possible Cutoff Scores of the ROC Curve for the NDI and the NRS

Questionnaire	Range	MDC	ROC Cutoff*	Sensitivity	Specificity
NDI	0–50	10.5	10.5	0.3	0.9
			3.5	**0.9**	**0.7**
			1.5	0.9	0.5
NRS	0–10	4.3	4.5	0.4	0.9
			2.5	**0.8**	**0.8**
			1.5	0.9	0.6

*Optimal cutoff point of the ROC curve for clinically important change with a sensitivity and a specificity of the Neck Disability Index and Numeric Rating Scale for pain.
MDC indicates the minimal detectable change.

jects in each of the 6 categories of the GPE and for the combined categories, as used in the analysis. After 7 weeks, 94 patients were labeled as "importantly improved," *i.e.*, completely recovered or were much improved. Only 2 patients were deteriorated, so due to the small numbers, they were excluded from the analysis; 87 patients were labeled as "not importantly changed," *i.e.*, unchanged, slightly improved, or slightly worse.

Table 3 presents the MDC and the optimal cutoff point for the stable subjects. The MDC for the NDI is 10.5, and the optimal cutoff point of the ROC curve is 3.5. The MDC for the NRS is 4.3, and the optimal cutoff point of the ROC curve is 2.5. The optimal cutoff point of the ROC curve for the NDI corresponds to a sensitivity of 0.9 and a specificity of 0.7. For the NRS for pain the sensitivity and specificity both were 0.8.

■ Discussion

In research and also in clinical practice, the NDI and the NRS are often used as questionnaires to evaluate the effects of interventions on functional status and pain perception in patients with neck pain. Hence, it is important to know what the smallest change in score is on both questionnaires, which patients and clinicians label as clinically important. This study demonstrates quite a difference between the 2 methods used to estimate the MCIC for the NDI as well as the NRS. Using the optimal cutoff point of the ROC curve as a method, both improved and unchanged patients are included. The optimal cutoff point of the ROC curve is chosen in such a way that the percentages of false-positive and false-negative outcomes are minimized. So, if one wants to weight false-positive and false-negative misclassifications equally, the optimal cutoff point of the ROC curve is preferred. If one hesitates to classify patients as "improved," of whom the change scores fall within the measurement error of the unchanged patients, one may prefer the more conservative MDC method. The choice between the 2 methods may depend on the type of intervention or the clinical consequences of being "false positive" or being "false negative."

The MDC for the NDI is 10.5 points on a scale of 50 points. This MDC can be considered as quite large, since this magnitude is greater than the change score of patients who consider themselves as "much improved" (mean NDI score = 8.82). So in other words, the MDC considers nearly all patients as being within the measurement error of the questionnaire. As a consequence, if applying this MDC as a cutoff point in clinical practice and the change score is more than the MDC, one knows

Table 2. Mean Scores and Standard Deviations of the Neck Disability Index (NDI) and the Numeric Rating Scale for Pain (NRS) at Baseline (T0) and After 7 Weeks (T7) for Categories of Global Perceived Effect (GPE)

	NDI (n = 183)			NRS (n = 182)		
Categories of GPE	T0	T7	Mean Change	T0	T7	Mean Change
Completely recovered (n = 19)	10.74 (5.32)	0.68 (1.64)	10.05 (5.58	5.37 (1.89)	0.16 (0.37)	5.21 (2.10)
Much improved (n = 75)	14.67 (7.57)	6.16 (4.69)	8.51 (6.40)	5.92 (1.77)	2.11 (1.48)	3.81 (1.94)
Slightly improved (n = 48) (NRS, n = 47)	15.71 (6.96)	12.00 (7.20)	3.71 (4.48)	6.21 (2.13)	4.57 (1.86)	1.64 (2.08)
No change (n = 29)	14.52 (6.60)	14.24 (6.65)	0.26 (3.79)	5.97 (1.96)	5.59 (1.96)	0.38 (1.52)
Slightly worse (n = 10)	13.50 (5.34)	14.10 (7.98)	−0.60 (7.50)	6.10 (1.37)	5.70 (2.00)	0.40 (1.71)
Much worse (n = 2)	19.00 (5.57)	26.50 (13.44)	−7.50 (7.78)	7.00 (0.00)	8.00 (1.41)	−1.00 (1.41)
Total	14.49 (7.00)	9.06 (7.46)		5.97 (1.88)	3.36 (2.46)	
Importantly changed* (n = 94)	13.87 (7.32)	5.05 (4.78)	8.82 (6.25)	5.81 (1.80)	1.71 (1.54)	4.10 (2.04)
Not importantly changed† (n = 87) (NRS, n = 86)	15.06 (6.65)	12.99 (7.11)	2.07 (4.99)	6.12 (1.98)	5.05 (1.96)	1.07 (1.95)

*Categories "completely recovered" and "much improved" were considered to indicate importantly changed.
†Categories "slightly improved," "unchanged," and "slightly worse" were considered to indicate "not importantly changed."

almost for sure (with a uncertainty of 5%) that a patient really is changed. A possible explanation for this large MDC for the NDI can be that "slightly improved" patients are included in the "unchanged" group and not in the "improved" group. However, the inclusion of this group into the "unchanged" group is sensible and has been demonstrated before.[17] Another remark has to be made. Based on these analyses, it is unclear whether for deterioration a similar value applies. Despite the small number (n = 2), we could not make a estimation of the "minimal important deterioration." However, based on the Farrar *et al* study,[20] there is some evidence that patients interpret deterioration quite differently to improvement.

However, the disadvantage of the MDC method is that the false-negative rate is not taken into account; in other words, if there is no reason for weighting false negative different from false positive, we recommend the use of the optimal cutoff point of the ROC curve.

When we used the method of the optimal cutoff point of the ROC curve the MCIC of the NDI appeared to be smaller: 3.5 points. This score makes more clinical sense because this change is quit similar to the change in score for those patients who consider themselves as "slightly improved." Since both "improved" and "unchanged" patients are included in the analysis of the cutoff point of the ROC curve, this method anticipates to a more clinical perspective on the change of the questionnaire. However, their still is a possibility of false-positive outcomes.

The MDC for the NRS for pain is also quite large: 4.3 points. Again, this magnitude equals the mean change score of the patients who consider themselves as much improved. The optimal cutoff point of the ROC curve is 2.5 points. This MCIC for pain is in line with the findings of other studies[17,20] in which the ROC curve was used to define clinically important change and in which an average reduction of 2 points was found to be clinically important.

In summary, we consider the optimal cutoff point of the ROC curve as the most optimal method, since false positive and false negative can be weighted equally.

Although the GPE is often applied as an external anchor, the use of this scale has been criticized by Norman *et al*.[21] They question the validity of a single-item design compared with a multi-item scale. Another disadvantage of the GPE is that it may be difficult for patients to recall their initial health status and to compare it with their current status in order to assess any changes, and this may introduce bias. Fritz and Irrgang found that a global rating of change could be used to differentiate unchanged patients from improved patients in the dimension of physical impairment.[22] In line with previous studies,[17,18,20,23] we used a GPE scale to cover the whole range from severely deteriorated, slightly deteriorated, no change, slightly improved, much improved, and completely recovered. To calculate the MCIC, the cutoff point for clinically important change was set at "much improved." Therefore, the category "slightly improved" was labeled as "not importantly changed." We had several reasons for this. First, in our opinion, this more accurately reflects the concept of clinically important change. Setting the cutoff point for improvement at slightly improved may reflect more accurately the smallest detectable change, and not the minimal clinically important change. Second, we think that patients are likely to give "slightly improved" as a socially desirable answer, even if they did not perceive a relevant improvement. Finally, in previous studies,[17,18,20] it was found that the difference between the categories "no change" and "slightly improved" was small whereas the difference between "slightly improved" and "much improved" was greater. These results were confirmed in the present study.

■ Conclusion

The estimated MCIC should be used as an indication for relevant changes in clinical practice. Using the optimal cutoff point of the ROC curve, false positives and false negatives are equally weighted; and if there are no objections doing so, the optimal cutoff point of the ROC curve may be a good choice. However, if there are objections against classifying as improved those patients whose results fall within the measurement error of the "unchanged" patients, the more conservative MDC method would be more appropriate.

■ Key Points

- The NDI and the NRS for pain are frequently used measurement scales assessing neck pain patients.
- MCIC can be used to estimated clinically important change, using 2 different methods. Using the optimal cutoff point of the ROC curve, a change of score of 3.5 points on the NDI and of 2.5 points on the NRS best distinguished those patients who are clinically improved from those who are not.
- If there are any objections against classifying as improved those patients whose results fall within the measurement error of the "unchanged" patients, the more conservative MDC method would be more appropriate.
- Using the optimal cutoff point of the ROC curve, false positives and false negatives are equally weighted; and if there are no objections doing so, the optimal cutoff point of the ROC curve may be a good choice.

References

1. Borghouts JA, Koes BW, Vondeling H, et al. Cost-of-illness of neck pain in The Netherlands in 1996. *Pain* 1999;80:629–36.
2. Picavet HSJ, Schouten JSAG. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC3-study. *Pain* 2003; 102:167–78.
3. Croft PR, Lewis M, Papageorgiou AC, et al. Risk factors for neck pain: a longitudinal study in the general population. *Pain* 2001;93:317–25.

4. Hoving JL, Koes BW, de Vet HC, et al. Manual therapy, physical therapy, or continued care by a general practitioner for patients with neck pain: a randomized, controlled trial. *Ann Intern Med* 2002;136:713–22.
5. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, et al. French translation and validation of 3 functional disability scales for neck pain. *Arch Phys Med* Rehabil 2002;83:376–82.
6. Vernon H. The Neck Disability Index: patient Assessment and Outcome Monitoring in Whiplash. *J Musculoskeletal Pain* 1996;4:95–104.
7. Bolton JE, Wilkinson RC. Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. *J Manipulative Physiol Ther* 1998;21:1–7.
8. Hoving JL, O'Leary EF, Niere KR, et al. Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. *Pain* 2003;102:273–81.
9. Pietrobon R, Coeytaux RR, Carey TS, et al. Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. *Spine* 2002;27:515–22.
10. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther* 1991;14:409–15.
11. Cepeda MS, Africano JM, Polo R, et al. What decline in pain intensity is meaningful to patients with acute pain? *Pain* 2003;105:151–7.
12. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10: 407–15.
13. Terwee CB, Dekker FW, Wiersinga WM, et al. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003;12:349–62.
14. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56:395–407.
15. Bolton JE. Sensitivity and specificity of outcome measures in patients with neck pain: detecting clinically significant improvement. *Spine* 2004;29: 2410–7.
16. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, et al. The ability to change of three questionnaires for neck pain. *Joint Bone Spine* 2004;71:317–26.
17. van der Roer N, Ostelo RW, Bekkering GE, et al. Minimal clinically important change for pain intensity, functional status and general health status in patients with non-specific low back pain. *Spine* 2006;31:578–82.
18. Ostelo RW, de Vet HC, Knol DL, et al. 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *J Clin Epidemiol* 2004;57:268–76.
19. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002;14:109–14.
20. Farrar JT, Young J, LaMoreaux L, et al. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001;94:149–58.
21. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997;50:869–79.
22. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther* 2001;81:776–88.
23. Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 1996;65:71–6.