Responsiveness and minimal important changes for the Neck Disability Index and the Neck Pain Disability Scale in Italian subjects with chronic neck pain

Marco Monticone · Emilia Ambrosini · Howard Vernon · Roberto Brunati · Barbara Rocca · Calogero Foti · Simona Ferrante

Received: 12 May 2014 / Revised: 29 January 2015 / Accepted: 29 January 2015

Introduction

Patient-reported outcomes (PRO) on health and function have increasingly become useful means of assessing outcomes in cervical disorders by adding valuable information to physiological measurements such as articular range of motion or muscle strength [1]. One area in which PRO instruments make an important contribution to patient management and research is in measuring clinical change. Responsiveness is the ability of an instrument to detect changes in the construct to be measured over time, while the minimal important change (MIC) is the smallest change in score of the construct to be measured that patients perceive to be important. This is important for researchers studying the effectiveness of various treatments in clinical trials, power calculations, sample size estimates, and cost

M. Monticone (✉) · E. Ambrosini · R. Brunati · B. Rocca
Physical Medicine and Rehabilitation Unit, Scientific Institute of Lissone, Institute of Care and Research (IRCCS), Salvatore Maugeri Foundation, Via Monsignor Bernasconi, 16, 20035 Lissone, Milan, Italy
e-mail: marco.monticone@fsm.it

E. Ambrosini · S. Ferrante
Neuroengineering and Medical Robotics Laboratory, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

H. Vernon
Canadian Memorial Chiropractic College, Toronto, ON, Canada

C. Foti
Chair of Physical and Rehabilitation Medicine, Tor Vergata University of Rome, Rome, Italy

evaluations as well as for studies of prognosis over the natural history of the condition. Clinicians, as well, require such instruments to evaluate the effectiveness of their treatments and to guide treatment decisions [2, 3].

In the area of neck pain, the Neck Disability Index (NDI) and the Neck Pain Disability Scale (NPDS) are two commonly used condition-specific health status measures for the assessment of disability related to neck pain (NP)[4–6]. Both questionnaires were originally developed in English and have been culturally adapted in various languages, with both showing satisfactory psychometric properties (internal consistency, reproducibility and validity) in a wide variety of situations [4–7].

The Italian versions of the NDI and NPDS have been psychometrically analysed and found to have similar properties to those of the original versions [8, 9], but their responsiveness and MICs have not been determined yet, limiting their use to clinical and research purposes.

The primary aim of this study was to determine the responsiveness and MICs of the NDI and NPDS in Italian subjects with chronic NP undergoing rehabilitation using both distribution-based and anchor-based methods mainly suggested in the current literature and based on the "COnsensus-based Standards for the selection of health status Measurement INstruments" (COSMIN) [3, 10, 11]; influences of different baseline scores on MICs were also assessed. Our secondary aim was to investigate which questionnaire was the most responsive in the population under investigation.

Methods

This research was part of an observational study approved by the Institutional Review Board of the Salvatore Maugeri Foundation's Scientific Institute in Lissone. Patients gave their written consent to participate.

Subjects

Outpatients admitted to our rehabilitation unit and to an affiliated centre were enrolled between January 2013 and December 2013. The inclusion criteria were: a diagnosis of chronic non-specific NP (i.e. a documented history of pain lasting for more than 12 weeks), a good under-standing of Italian, and an adult age. The exclusion cri-teria were acute (lasting up to 4 weeks) and subacute non-specific NP (lasting up to 12 weeks), specific causes of NP (e.g. disc herniation, canal stenosis, spinal deformity, fracture, spondylolisthesis, or infections), and central or peripheral neurological signs. Patients with systemic illness, cognitive impairment, recent myocardial infarctions, cerebrovascular events, or chronic lung or renal diseases were excluded on the basis of their case histories. Patients who participated in previous conservative treatments for their NP were also excluded. The patients' socio-demographic and clinical characteristics were investigated us-ing a specific form.

Procedures and outcome measures

All of the participants were provided written information concerning the questionnaires and procedures by two research assistants. Those satisfying the entry criteria underwent an 8-week outpatient rehabilitation programme that included exercises aimed at improving postural con-trol, strengthening and stabilising the neck muscles, and stretching; patients also received cognitive-behavioural therapy and education in ergonomic principles. Mild analgesics and non-steroidal anti-inflammatory drugs (NSAIDs) were permitted during the study and an excessive use of medicines for pain control was checked.

This rehabilitation programme was the same for all of the subjects and was already tested for its efficacy [12].

The NDI and NPDS were administered to all of the patients as part of the pre- and post-rehabilitation assessment. The NDI is a ten-item self-administered questionnaire modelled on the Oswestry Disability Index by Vernon and Mior in 1991 [4]; each question is scored on a 6-point scale ranging from 0 (no disability) to 5 (full disability), and these are added together providing a percentage score as a means of dealing with unanswered questions, ranging from 0 (no disability) to 100 (complete disability) [5]. The NPDS is a 20-item self-administered questionnaire developed by Wheeler et al. in 1999 using the Million Visual Analogue Scale as a model [6], where each item ranges from 0 (meaning normal function) to 5 (meaning the worst possible situation your problem has caused you), and patients respond to each of them by marking along a 10-cm visual analogue scale, in which the total score ranges from 0 (no disability) to 100 (complete disability). We used the adapted Italian versions of the NDI and NPDS [8, 9].

At the end of treatment we also evaluated the global perceived effect (GPE) using the question: "Overall, how much did the treatment you received help your neck problem?"; the GPE was determined using a five-level Likert scale with two improvement levels (helped a lot = 1, helped = 2), one no change level (helped only a little = 3), and two worsening levels (did not help = 4, made things worse = 5) [13].

At both assessments, the questionnaires were administered by secretarial staff who checked them and returned any incomplete part to the patients for completion to minimise the rate of missing/multiple responses.

Statistics

Responsiveness was determined using distribution and anchor-based methods [1, 10]. The distribution methods included the effect size (ES), also using Guyatt's approach, and the standardised response mean (SRM). The ES is a standardised measure of change over time calculated on the whole sample by dividing the difference between the pre- and post-test scores by the pre-test standard deviation (SD); in the case of Guyatt's approach, the change computed on the whole sample is divided by the pre-test SD calculated only for stable subjects whose clinical status remained unchanged (GPE = 3). The ES therefore represents individual change in terms of the number of pre-test SDs, with values of 0.20, 0.50, and 0.80, respectively, representing small, moderate, and large changes. The SRM (also referred to as the responsiveness-treatment coefficient or efficacy index) is the ratio between individual change and the SD of that change. It has been suggested that SRM values of 0.20, 0.50, and 0.80, respectively, represent small, moderate, and large changes.

As an anchor-based method, receiver operating characteristic (ROC) curves were selected, which are useful indicators of the relationship between a measure and an external indicator of change, such as the GPE. Patients were dichotomised into two groups based on GPE scores. Patients were considered improved when the GPE score was equal to 1 or 2, and stable when the GPE score was equal to 3. Responsiveness is described in terms of sensitivity (the probability that the measure correctly classifies patients who demonstrate change when an external criterion of clinical change is used) and specificity (the probability that the measure correctly classifies patients who do not demonstrate change when the external criterion is used). The sensitivity and specificity of each value of change in the measure are calculated and used to plot a ROC curve. The sensitivity values and false-positive rates (1-specificity) are plotted on the y and the x axis of the curve, and the area under the curve (AUC) represents the probability a measure correctly classifies patients as improved or unchanged. This area theoretically ranges from 0.5 (no discriminating accuracy) to 1.0 (perfect accuracy), and an AUC of at least 0.70 is considered to be acceptable [2]. The optimal cutoff point was computed using the Youden index and taken as the MIC, which indicates the change score associated with the least misclassification [14].

ROC curves adjusted using the baseline scores as covariate were also computed to investigate the impact of the baseline scores on the responsiveness analysis. If a significant impact was found, the patients were divided into two subgroups based on the baseline scores and MIC values for the two subgroups were computed. The median NDI/NPDS score was used to divide the population to maximise the group size and thus to optimise the statistical power [15].

External responsiveness was also investigated by means of correlation analyses with external criteria (GPE) [10]. We tested the correlations between the pre–post treatment change scores in the NDI and NPDS and the GPE scores by estimating Spearman's rank order correlation coefficients.

Results

Subjects

260 patients were invited to participate, of whom 35 (13.5 %) refused. Of the 225 selected subjects, 25 dropped out before starting the rehabilitation sessions because of logistic problems (18), economic difficulties (2), or per-sonal problems (5), and so the final study population con-sisted of 200 subjects (120 females, 60 %, and 80 males, 40 %) with a mean age of 52.5 ± 15.8 years and a mean pain duration of 10.8 ± 11.9 months. The body mass index was 24.4 ± 3.7 kg/m^2. Table 1 shows the patients' clinical and socio-demographic characteristics.

Mean values (standard deviation) for NDI at pre- and post-treatment were 33.7 (19.2) and 21.0 (15.4) out of 100, respectively; mean values (standard deviation) for NPDS at pre- and post-treatment were 48.2 (20.7) and 33.0 (18.7) out of 100, respectively (see Table 2). Pain was also measured using a numerical rating scale and was 5.4 (1.9) and 3.0 (1.5) at pre- and post-treatment assessment, respectively.

Procedures

The study procedures were well accepted by all of the patients, who did not raise any specific questions during the instruction phase or the administration of the questionnaires; no missing or multiple answers were found for both NDI and NPDS. None of the procedures led to any problems and all of the patients completed the rehabilitation program. No specific issues were raised by the patients or the physiotherapists. There was no excessive use of medicines to control pain.

Psychometric Properties

Dichotomisation of the GPE showed that 124 subjects (62 %) improved and 64 subjects (32 %) were stable; 12 patients (6 %) were not included in the ROC curves analysis, since they had worsened clinical condition. As there were more than 50 subjects per subgroup, these estimates assured an adequate sample size for calculating responsiveness [15]. Baseline and post-treatment scores of NDI

Table 1 Socio-demographic characteristics of the study population ($n = 200$)

Variable	No.	%
Marital status		
Unmarried	40	20
Married	160	80
Employment		
Employee	94	47
Self-employed	37	18.5
Housewife	44	22
Pensioner	25	12.5
Education		
Elementary school	18	9
Middle school	54	27
High school	112	56
University	16	8
Smoking		
Yes	28	14
No	172	86
Use of drugs		
Antidepressants	10	5
Analgesics	68	34
Muscle relaxants	19	9.5
NSAIDs	35	17.5
None	68	34
Comorbidities (principal)		
Hypertension/heart disease	55	27.5
NIDDM	15	7.5
Gastro-enteric disease	20	10.0
Liver disease	7	3.5
None	103	51.5

Table 2 Mean values (standard deviation) of the NDI and NPDS for the total, improved, and stable group of patients

	Pre-treatment	Post-treatment
NDI		
Total ($n = 200$)	33.7 (19.2)	21.0 (15.4)
Improved ($n = 124$)	36.7 (20.1)	17.9 (14.5)
Stable ($n = 64$)	28.8 (18.2)	25.6 (16.9)
NPDS		
Total ($n = 200$)	48.2 (20.7)	33.0 (18.7)
Improved ($n = 124$)	51.4 (20.6)	30.9 (19.8)
Stable ($n = 64$)	41.6 (20.7)	36.3 (17.0)

and NPDS for improved and stable patients are reported in Table 2.

The results of the distribution-based and anchor-based methods to determine responsiveness are reported in Table 3.

The ES of the rehabilitation programme, as measured by the NDI and NPDS, was 0.66 and 0.73, respectively, indicating a moderate magnitude of the change scores, and was similar when Guyatt's approach was used (0.70 and 0.73, respectively); the SRM of all of the subscales was large (1.09 and 1.26, respectively).

The ROC analyses of the NDI and NPDS revealed AUCs of 0.96 and 0.91, respectively, thus showing a good capacity to discriminate between improved and stable subjects (see Figs. 1, 2); the best cutoff points (i.e. MICs) were 7 (sensitivity: 98 %; specificity: 81 %) for the NDI, and 10 (sensitivity: 93 %; specificity: 83 %) for the NPDS, respectively. This means that a pre–post treatment change of >7 and >10 for NDI and NPDS, respectively, would have been considered a clinically important change. For both the NDI and NPDS, none of the patients was characterised by a baseline score preventing any achievement of a clinically important change. The estimated AUC was slightly higher for the NDI than for the NPDS. Thus, we tested the equality of the two AUC values and could not reject the hypothesis that the two tests have equal AUC at the 0.01 level. This means that both the NDI and NPDS showed equivalent responsiveness properties in terms of ROC curves.

When computing the ROC analysis adjusted for baseline scores, different results were obtained for the NDI and NPDS. Concerning the latter, the covariate did not affect the ROC analysis, which otherwise affected the former. Thus, a subgroup analysis was carried out on the NDI. The median NDI score at baseline was 27. Of the 92 patients with a baseline score of <27, 56 were improved and 36 were stable, while, of patients with a baseline score of ≥ 27 ($n = 96$), 68 were improved and 28 were stable. The MIC [AUC; sensitivity; specificity] was 17 [0.97; 81; 100] and 6 [1.00; 98; 100] for patients with baseline scores above and below 27, respectively.

The correlation between change scores of the NDI and NPDS and GPE were, respectively, high (Spearman correlation of 0.71, $p < 0.01$) and moderate (Spearman correlation of 0.59, $p < 0.01$).

Discussion

This paper describes the estimation of responsiveness and the MICs of the NDI and NPDS in a population of Italian subjects with chronic NP undergoing rehabilitation. Analysing the responsiveness and MIC of an outcome measure is a continuous process that is strongly recommended to strengthen its properties and expand its applicability [11, 16]. Different approaches have been used to calculate responsiveness, but there is still no consensus as to which method is the best [1]. Thus, in this study we used both

Table 3 Results of the distribution-based and anchor-based methods for determining the responsiveness and minimal important changes of the NDI and NPDS outcome measures

Methods	NDI (0–100)	NPDS (0–100)
Effect size; $n = 200$	0.66	0.73
Effect size (Guyatt); $n = 200$	0.70	0.73
Standardised response mean (SRM); $n = 200$	1.09	1.26
MIC (optimal cutoff point) [AUC; sensitivity; specificity]; $n = 188$	7 [0.96; 98; 81]	10 [0.91; 93; 83]

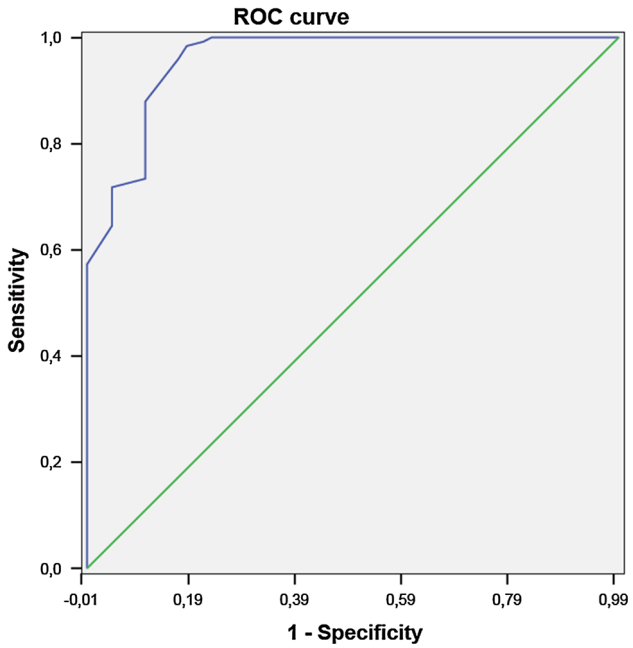

Fig. 1 ROC characteristics of the NDI

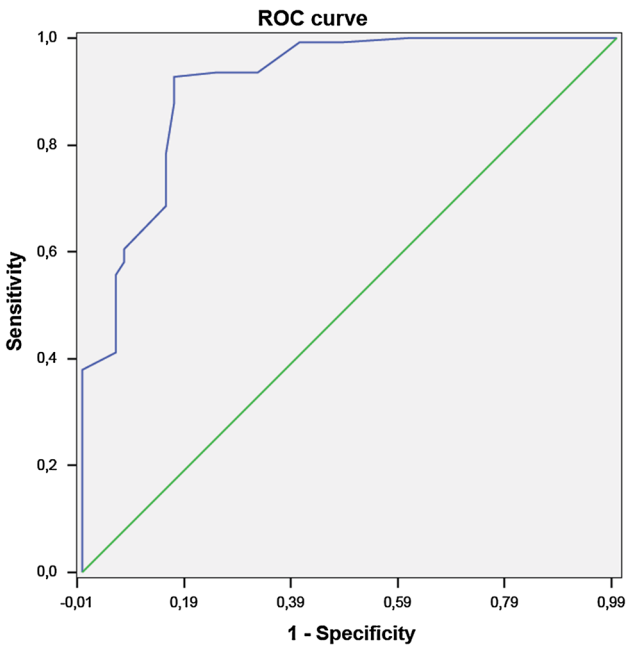

Fig. 2 ROC characteristics of the NPDS

distribution-based (ES, Guyatt's ES and SRM) and anchor-based methods (ROC analysis).

Distribution-based methods showed a moderate to large responsiveness to the rehabilitation programme of both NDI and NPDS. The published ES estimates on the NDI for subjects with chronic NP undergoing exercises varied from 0.13 to 1.17 ([17–22] in [5]). Excepting one study [22] which showed a small effect, all of the others showed a moderate to large effect (0.81–1.17), which is in line with our findings. Concerning NPDS, larger estimates in ES (1.44) and SRM (1.37) were found in Korean subjects [23], while smaller estimates were shown in the Turkish population with chronic NP undergoing exercises (SRM = 0.92) [24].

However, it has been recently recommended that distribution methods should be used cautiously because they measure the magnitude of change scores rather than their validity [16]. When a general measure of change in PRO such as the GPE is available and can be dichotomised into two subgroups representative of improved and stable subjects, an anchor-based method such as ROC analysis is preferred because the AUC measures the ability of an instrument to discriminate improved and stable subjects [11]. The findings of this study showed an AUC that was always >0.90, thus assuring the satisfactory discriminatory ability of both NDI and NPDS in the enrolled population. The optimal cutoff point estimated on the basis of ROC analysis was about 7 for the NDI, in line with those previously found by other authors in chronic NP subjects (7-19, considering a NDI score ranging from 0 to 100) [25–29]. The optimal cutoff point for the NPDS was about 10, again in line with a previous study that reported an MIC of 11.5 in subjects with chronic NP [29].

Recently, it was demonstrated that MICs for the NDI differed for subgroups of patients with higher and lower baseline scores [15]. This was confirmed by our results that showed higher MICs for patients with worst disability levels. Patients with moderate to severe disability at baseline need a larger improvement to perceive the treatment as helpful, and this may explain the higher estimate of the MIC achieved for this subgroup of patients [15].

The external responsiveness was also investigated by means of correlation analyses with GPE, which reflect the extent to which changes in a PRO measure over a specific time relate to corresponding changes in an external standard, defined as an accepted indication of change in the condition of a patient [10]. We found that the pre–post treatment changes in the NDI and NPDS were moderately to highly correlated to the change in perceived effect, which was chosen as the external standard. This was in line with previous findings (NDI:0.49; NPDS:0.48 [24]) and confirmed that the NDI and NDPS were responsive to the GPE score, being able to predict changes in perceived treatment effect.

This study has some limitations. First of all, GPE was assessed using a five-point Likert scale, and clinically important changes would probably have been more discriminating if a seven-point scale had been used. Secondly, the NDI and NPDS might not have been responsive to worsening outcomes as the patients who were a "little worse" or "worse" were excluded from the analyses. Thirdly, responsiveness and MICs were calculated for subjects with chronic NP undergoing rehabilitation, and further investigations are needed to calculate the estimates pre- and post-surgery, as well as after longer periods of follow-up. Fourthly, the applicability of this study is limited to an Italian population and similar studies are recommended in other countries. Lastly, the sample size may have been relatively small for the ROC subgroup analyses of the NDI.

In conclusion, the findings of this study show that the NDI and NPDS are both responsive measures in Italian subjects with chronic NP undergoing multi-modal rehabilitative therapy. It is recommended taking these MIC estimates into account when assessing improvement or planning clinical studies on a similar sample. Baseline scores did not affect MIC estimates for the NPDS. In contrast, when the NDI is used, MIC values should be selected carefully paying attention to the disability level of the population under investigation.

Acknowledgments The authors would like to thank all the physiotherapists who took part in the study.

Conflict of interest None.

Ethical standard Our institutional review board approved the research and the study was conducted in conformity with ethical and humane principles of research.

References

1. Revicki D, Hays RD, Cella D, Sloan J (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 61:102–109
2. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient reported outcomes. J Clin Epidemiol 63(7):737–745
3. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM (2006) Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes 4:54
4. Vernon H, Mior S (1991) The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 14:409–415
5. Vernon H (2008) The Neck Disability Index: state-of-the-art, 1991–2008. J Manipulative Physiol Ther 31:491–502
6. Wheeler AH, Goolkasian P, Baird AC, Darden BV 2nd (1999) Development of the Neck Pain and Disability Scale: item analysis, face, and criterion-related validity. Spine 24:1290–1294
7. Goolkasian P, Wheeler AH, Gretz SS (2002) The neck pain and disability scale: test-retest reliability and construct validity. Clin J Pain 18:245–250
8. Monticone M, Ferrante S, Vernon H, Rocca B, Dal Farra F, Foti C (2012) Development of the Italian version of the Neck Disability Index. Cross-cultural adaptation, factor analysis, reliability, validity and sensitivity to change. Spine (Phila Pa 1976) 37(17):E1038–E1044
9. Monticone M, Baiardi P, Nido N, Righini C, Tomba A, Giovanazzi E (2008) Development of the Italian version of the Neck Pain and Disability Scale, NPDS-I. Spine (Phila Pa 1976) 33:E429–E434
10. Husted JA, Cook RJ, Farewell VT, Gladman DD (2000) Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 53:459–468
11. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC (2010) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 18(10):22
12. Monticone M, Baiardi P, Vanti C, Ferrari S, Nava T, Montironi C, Rocca B, Foti C, Teli M (2012) Chronic neck pain and treatment of cognitive and behavioural factors. Results of a randomised controlled clinical trial. Eur Spine J 21(8):1558–1566
13. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ (2012) Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol 63:760–766
14. Zweig M, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 39:561–577
15. Schuller W, Ostelo RW, Janssen R, de Vet HC (2014) The influence of study population and definition of improvement on the smallest detectable change and the minimal important change of the neck disability index. Health Qual Life Outcomes 12(1):53
16. de Vet HCW, Terwee CB, Mokkink LB, Knol DL (2011) Measurement in medicine. A practical guide. Cambridge University Press, Cambridge
17. Bronfort G, Evans R, Nelson B, Aker PD, Goldsmith CH, Vernon H (2001) A randomized clinical trial of exercise and spinal manipulation for patients with chronic neck pain. Spine 26:788–797
18. Evans R, Bronfort G, Nelson B, Goldsmith CH (2002) Two-year follow-up of a randomized clinical trial of spinal manipulation and two types of exercise for patients with chronic neck pain. Spine 27:2383–2389
19. Hoving JL, Koes BW, de Vet HC, van der Windt DA, Assendelft WJ, van Mameren H, Devillé WL, Pool JJ, Scholten RJ, Bouter LM (2002) Manual therapy, physical therapy, or continued care

by a general practitioner for patients with neck pain—a randomized, controlled trial. Ann Intern Med 136:713–722
20. Kjellman G, Oberg B (2002) A randomized clinical trial comparing general exercise, McKenzie treatment and a control group in patients with neck pain. J Rehabil Med 34:183–190
21. Hoving JL, De Vet HCW, Koes BW, van Mameren H, Deville WLJM, van der Windt DAWM, Assendelft WJJ, Pool JJM, Scholten RJPM, Korthals–De Bos IBC, Bouter LM (2006) Manual therapy, physical therapy, or continued care by the general practitioner for patients with neck pain—long-term results from a pragmatic randomized clinical trial. Clin J Pain 22:370–377
22. Kietrys DM, Galper JS, Verno V (2007) Effects of at-work exercises on computer operators. Work 28:67–75
23. Lee H, Nicholson LL, Adams RD, Maher CG, Halaki M, Bae SS (2006) Development and psychometric testing of Korean language versions of 4 neck pain and disability questionnaires. Spine 31:1841–1845
24. Kose G, Hepguler S, Atamaz F, Oder G (2007) A comparison of four disability scales for Turkish patients with neck pain. J Rehabil Med 39:358–362
25. Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B (1999) Using the Neck Disability Index to make decisions concerning individual patients. Physiother Can 51:107–112
26. Pool JJ, Ostelo RW, Hoving JL, de Bouter LM, Vet HC (2007) Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine 32:3047–3051
27. 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
28. Young BA, Walker MJ, Strunce JB, Boyles RE, Whitman JM, Childs JD (2009) Responsiveness of the Neck Disability Index in patients with mechanical neck disorders. Spine J 9(10):802–808
29. Jorritsma W, Dijkstra PU, de Vries GE, Geertzen JH, Reneman MF (2012) Detecting relevant changes and responsiveness of Neck Pain and Disability Scale and Neck Disability Index. Eur Spine J 21(12):2550–2557.24