

The reliability of the Vernon and Mior neck disability index, and its validity compared with the short form-36 health survey questionnaire

M. J. H. McCarthy · M. P. Grevitt ·
P. Silcocks · G. Hobbs

Received: 9 March 2007 / Revised: 10 August 2007 / Accepted: 9 September 2007 / Published online: 6 October 2007
© Springer-Verlag 2007

Abstract Prospective single cohort study. To evaluate the NDI by comparison with the SF36 health Survey Questionnaire. The NDI is a simple ten-item questionnaire used to assess patients with neck pain. The SF36 measures functional ability, well being and the overall health of patients. It is used as a gold standard in health economics to assess the health utility, gain and economic impact of medical interventions. One hundred and sixty patients with neck pain attending the spinal clinic completed self-assessment questionnaires. A second questionnaire was completed in 34 patients after a period of 1–2 weeks. The internal consistency of the NDI and SF36 was calculated using Cronbach's alpha. The test–retest reliability was assessed using the Bland and Altman method. The concurrent validity of the NDI with respect to the SF-36 was assessed using Pearson correlations. Both questionnaires showed robust internal consistency: Cronbach's alpha for the NDI scale was acceptable (0.864, 95% confidence limits 0.825–0.894) though slightly smaller than that of the SF36. The correlations between each item of the NDI scores and the total NDI score ranged from 0.447 to 0.659, (all with $P < 0.001$). The test–retest reliability of the NDI was high (intra-class correlation 0.93, 95% confidence limits 0.86–0.97) and comparable with the best values

found for SF36. The correlations between NDI and SF36 domains ranged from -0.45 to -0.74 (all with $P < 0.001$). We have shown that the NDI has good reliability and validity and that it compares well with the SF36 in the spinal surgery out patient setting.

Keywords Neck disability index (NDI) · Short form 36 health survey questionnaire (SF36) · Reliability · Validity

Introduction

Neck pain is a common and important problem. In one study, 34% from a sample of 10,000 adults in the population experienced neck pain in the previous year [4]. In another, 43% of the normal population reported neck pain and in one third of these the duration was greater than 6 months [10]. It has been postulated that neck pain may cause as many days lost from work as low back pain [24]. In the Netherlands, the total estimated costs of neck pain are £437million per year [17].

Self-assessment questionnaires can be powerful instruments in assessing the outcome of medical management and interventions. To do this, amongst other things, they need to be both valid and reliable. Validity indicates that the questionnaire primarily measures what it is intended to measure. Reliability indicates that it is measuring something in a consistent and reproducible way [19, 20].

The Neck disability index (NDI) is a condition specific disability measure. It was devised in an outpatient physiotherapy department by Vernon and Mior in 1991 and is based on the Oswestry disability index (the reader is referred to Hains et al. and Vernon and Mior for examples of the questionnaire) [6, 7, 11, 22]. The questionnaire was

M. J. H. McCarthy · M. P. Grevitt (✉) · G. Hobbs
Department of Spinal Studies and Surgery,
Queens Medical Centre, Derby Road,
NG7 2UH Nottingham, UK
e-mail: michael.grevitt@qmc.nhs.uk

M. J. H. McCarthy
e-mail: mikemccarthy@doctors.org.uk

P. Silcocks
Trent RSDU, University of Nottingham, Nottingham, UK

devised and validated in English. It consists of ten questions each with six answers (scoring 0–5 points). The sum of the scores obtained is doubled to give a percentage score out of 100 (0–20 normal, 21–40 mild disability, 41–60 moderate, 61–80 severe and 80+ complete/exaggerated). It is simple and takes around 5 min to complete and score. Internal consistency was measured using the Cronbach's alpha statistic on 52 patients with neck pain [22]. Test–retest reliability was assessed using correlations on 17 patients with whiplash who repeated the questionnaire after a 2-day interval [22]. The construct validity was gauged using frequency histograms for the scores and the concurrent validity was assessed by correlating the NDI with the McGill pain questionnaire in 30 patients. This was done in a similar way to the technique used by Fairbank et al. [6, 7, 22].

There have been several studies which have looked at the reliability of the NDI [1, 5, 11–13, 16, 20–23]. However, the majority of these were performed either in physiotherapy departments or on patients with whiplash-associated disorders. To our knowledge the NDI has not been validated in the spinal surgery out-patient setting where patients may exhibit a different spectrum and severity of disease.

The primary aim of this study was to validate the NDI in the spinal surgery out-patient setting (to include all causes of neck pain), in particular by comparison with the Short Form 36 Health Survey Questionnaire (SF36). The SF36 is considered the gold standard generic health assessment tool being both valid and reliable and having been vigorously tested [8, 9, 14, 15]. However, unlike the NDI the SF36 covers emotional and social functioning which is an important aspect of neck disease. A secondary aim was to evaluate a visual analogue score (VAS) in the same patient population. The VAS is a simple and common tool used in assessing pain, but lacks formal evaluation.

Methods

All patients attending the Spinal Out-Patients Department at Queen's Medical Centre (Nottingham, UK) complete our standard neck pain assessment form (in English). The Spinal Surgery Department at the Queen's Medical Centre is a tertiary referral centre and serves a population of 2 million. The diagnoses of patients attending the Out Patient Department with neck pain is 85% degenerate (of which 40% axial neck pain, 50% associated radiculopathy and 10% whiplash associated disorders), 10% post-trauma and 5% others (tumours, infection and inflammatory).

Over a 4-month period, the following data was analysed: age, sex, VAS, NDI and SF36. Only responses from patients with fully completed forms were used. During the study period the nursing staff in clinic ensured all forms were completed. Forty patients were asked to complete and

return a second postal questionnaire within 1–2 weeks [1, 18, 23]. These were posted to them after their initial clinic appointment. This number was chosen for pragmatic reasons, being a one-month period out of the four-month study. A postal questionnaire was used on grounds of cost and convenience.

The study was registered with the hospital audit department. Ethics committee approval was not required as there was no variation from normal practice.

The internal consistency of each questionnaire was assessed using Cronbach's alpha. We also assessed the correlations of the individual item scores to the total score excluding that of item under consideration so as not to artificially inflate the correlation. We do not believe previous studies carried out this correction [1, 11, 13, 22]. Acceptable levels of internal consistency are indicated by a Cronbach's alpha above 0.7 [3].

Construct validity was gauged by comparing mean scores for each NDI item with results presented in other studies. Concurrent validity between the questionnaires was assessed using Pearson correlation between NDI score, VAS score and SF36 domains.

The test–retest reliability of each questionnaire was assessed by Bland and Altman Plots [2]. A Bland and Altman Plot illustrates the spread of the difference in scores between the test and retest examination for each individual. One should expect 95% of the differences to be less than two standard deviations (this is the definition of a repeatability coefficient adopted by the British Standards Institution). Intra-class correlations were also estimated on the patients who repeated the questionnaire in order to roughly compare our results to other studies (exact methodology between studies varies).

Results

One hundred and sixty questionnaires were completed in full. There were ten incomplete questionnaires (the patients could not answer question 8—see [Discussion](#)). The mean age of respondents was 51.2 years (range 14–93 years). There were 64 males (40%) and 96 females (60%). Thirty-four of the forty patients (80%) completed and returned the second questionnaire within 2 weeks. The average NDI in our study was 46%. [Figure 1](#) shows the distribution of the NDI scores. The distribution of scores in our sample was approximately normal with mean 45.8 and standard deviation 18.43. [Table 1](#) compares the NDI mean item scores with those quoted in other studies. There was no statistically significant difference between the sexes for the mean NDI, SF36 and VAS scores ([Table 2](#)). Age did not influence the NDI and VAS scores. Age was negatively associated with the Physical Functioning, Role Physical,

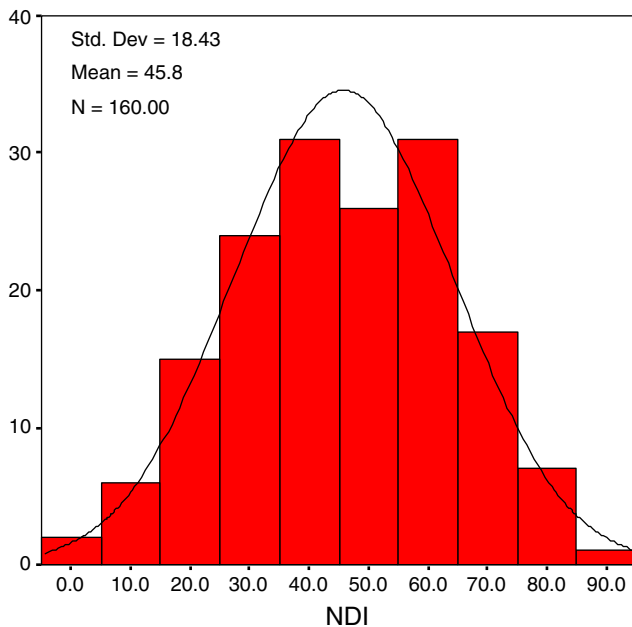


Fig. 1 Distribution of categorised NDI scores

Energy, General Health and Social domains of the SF36 ($P < 0.05$, Table 3).

Cronbach's alpha for both the SF36 and NDI scales showed good internal consistency, respective values being 0.878 (se = 0.014, 95% CI = 0.843–0.906) and 0.864 (se = 0.017, 95% CI = 0.825–0.894). Pearson correlations between the individual NDI item scores and the total ranged from 0.45 to 0.66, all with $P < 0.001$ (see Table 4). No item dominated with an especially high correlation and no item appeared redundant by virtue of a negligible correlation.

For concurrent validity of the NDI with respect to the SF36, Table 5 shows that all domains of the SF36 were at least moderately correlated with the NDI. The eight domains of the SF36 can be summarized into physical and mental component scores. This is done by averaging the scores in the relevant four domains. These were then correlated with the NDI to permit comparison with previous studies (Table 5). Correlations between SF36 domains and the VAS are shown in Table 6, and these were consistently smaller in magnitude than those for the NDI, being at most only moderate.

Figure 2a–d display Bland–Altman plots. The middle reference line is the mean difference in NDI scores between the two occasions. The outer reference lines are the 95% reference limits of agreement (equal to two standard deviations around the total score difference). There is a small, negative trend in the mean difference—the higher the initial score the higher the second test score seems to be—but this result was not statistically significant and the spread of values also appears constant. Both the point estimates and lower 95% confidence limits for the test

retest intra-class correlation coefficients are high, also indicating good reproducibility (Table 7).

Discussion

Our study has shown that the NDI is valid and reliable in spinal surgery patients and is comparable with the SF36. The male to female ratio in our study is similar to that in other studies [1, 10–13, 22]. The average age of patients in previous studies ranges from 32 to 46 years and the average NDI from 30 to 41% [1, 5, 11–13, 21, 22]. These studies were carried out in physiotherapy and chiropractic clinics. In our study the average age was 51 years and the average NDI was 46%. It would seem that our population is older and reflects one with a more severe spectrum of neck disease. As one might expect, an increasing age correlated with worsening scores in several domains of the SF36 (as shown in Table 3).

Table 1 shows a comparison of published data for the mean item scores [11, 13, 22]. Of note, our study demonstrates a higher mean item score across all domains except headaches. This seems to score higher, and perhaps be more common, in chiropractic, physiotherapy and whiplash associated disorder groups.

The justification for assessing Cronbach's alpha for the SF-36 in our sample, rather than using “known published” values, is that conceivably the SF36 might not perform so well in our patients. By re-estimating the SF-36 alpha, we have reassured ourselves that the SF-36 is behaving itself in our sample, and thus calibrates the NDI result as it were. Vernon and Mior [22] quote Cronbach's alpha for the NDI of 0.8 and Hains et al. [11] found one of 0.92 (in physiotherapy and chiropractic out patients respectively). In the spinal surgery patient population our study also shows that the NDI has good internal consistency (alpha 0.864). The correlations of the individual item scores to the total are similar to, albeit slightly lower than, those published by Hains et al. [11] (Table 4). Possible reasons are sampling error, the inclusion of the item in the total score by Hains et al. or a different spectrum of disease in the two study populations.

All eight SF-36 domains were at least moderately correlated with the NDI—not just the physical and mental components. Our values were slightly lower than in the other two studies possibly reflecting the different patient population (spinal surgery versus physiotherapy or chiropractic out patients) [12, 21] (Table 5).

Our response rate for the test–retest analysis was good (80%), although it would have been preferable to choose a random sample of patients. Previous studies have shown that there is little difference in the test–retest interval of 2 days versus 2 weeks [18]. It is becoming accepted that test–retest reliability is best described by the Bland and Altman method [2, 19]. The standard deviation (SD) of the difference

Table 1 A comparison of the mean item scores for the NDI (SD where available) [11, 13, 22]

Study NDI item	Present study (<i>n</i> = 160)	Vernon and Mior *(<i>n</i> = 44)	Hoving et al.**(<i>n</i> = 71)	Hains et al. ***(<i>n</i> = 237)
1. Pain intensity	2.2 (1.0)	1.7	2.7 (1.0)	1.8 (1.2)
2. Personal care	1.4 (1.2)	0.75	0.7 (0.8)	0.6 (0.9)
3. Lifting	2.9 (1.5)	2.2	2.5 (1.4)	2.1 (1.6)
4. Reading	2.3 (1.2)	2.1	2.3 (1.3)	2.0 (1.3)
5. Headaches	2.0 (1.7)	2.6	2.1 (1.5)	2.2 (1.5)
6. Concentration	1.3 (1.2)	1.1	1.4 (1.1)	1.3 (1.2)
7. Work	2.6 (1.4)	1.5	2.2 (1.3)	1.6 (1.6)
8. Driving	2.6 (1.6)	2.0	2.0 (1.3)	2.0 (1.3)
9. Sleeping	2.6 (1.4)	1.4	2.0 (1.5)	1.9 (1.4)
10. Recreation	3.0 (1.5)	2.2	2.5 (1.4)	2.1 (1.5)

*Physiotherapy Out Patient Department, **patients with whiplash associated disorders, ***Chiropractic Out Patient Department

Table 2 NDI, SF36 and VAS scores between the sexes

Sex	Male		Female	
	Mean	95% CI	Mean	95% CI
Score				
NDI	44.9	40.4 to 49.5	47.3	43.5–51.1
VAS	5.8	5.2 to 6.5	5.4	4.9–6.0
SF36 Physical	47.1	40.5 to 53.7	42.3	35.8–48.8
SF36 Role Physical	18.8	11.2 to 26.3	17.7	10.6–24.9
SF36 Pain	33.4	27.5 to 39.2	34.5	29.2–39.8
SF36 General Health	46.2	40.0 to 52.3	45.9	40.7–51.2
SF36 Energy	33.0	27.0 to 39.0	34.8	30.1–39.5
SF36 Social	46.3	38.7 to 53.8	48.7	42.3–55.1
SF36 Role Mental	52.8	41.0 to 64.6	46.1	36.3–55.9
SF36 Mental Health	58.2	52.4 to 64.0	61.0	56.9–65.1

Table 3 Influence of age on NDI, SF36 and VAS scores

Score	Regression coefficient (B)	95% CI (B)	<i>P</i> value
NDI	0.065	–0.077 to 0.207	0.365
VAS	–0.570	–1.65 to 0.513	0.300
SF36 Physical	–0.238	–0.322 to –0.154	<0.001
SF36 Role Physical	–0.146	–0.224 to –0.067	<0.001
SF36 Pain	–0.061	–0.172 to 0.051	0.284
SF36 General Health	–0.148	–0.256 to –0.040	0.008
SF36 Energy	–0.130	–0.246 to –0.015	0.027
SF36 Social	–0.125	–0.213 to –0.038	0.005
SF36 Role Mental	–0.040	–0.098 to 0.017	0.166
SF36 Mental Health	0.053	–0.075 to 0.180	0.419

between the first and second measurements for the NDI was 8.87 and this was deemed small compared with the baseline value (the mean of repeated NDI scores). The minimum clinically important difference (MCID) is the change in score that represents a true clinical change. This has recently been estimated to be around ten points for the NDI¹². For the Oswestry disability index it is also generally agreed to be

around ten points with studies quoting a significant change being of between 5 and 16 points [19]. The standard deviation (SD) of the difference between the first and second measurements for the NDI in our study was 8.87. Hence, the MCID corresponds to $10/8.87 = 1.13$ standard deviations. Plus or minus 1.13 standard deviations would cover 74% of a normally-distributed population. Therefore, 26% will lie outside this—i.e. will have experienced a clinically significant change between measurements. While our test–retest reliability would seem reasonable on the basis of the SD of the change being relatively small compared to the baseline (and a high intraclass correlation), nevertheless 26% of the patients may have experienced a clinically significant difference between testing. There are several possible explanations for this—(1) the disease and its evolution; (2) our methodology—(a) the timing of the questionnaire and the effects of the clinic consultation, (b) the attention paid by the patient when answering the questionnaire (at home vs. in the out patients clinic), (c) the mixture of patients (different diagnoses), (d) the number of patients studied; and (3) the questionnaire itself—is it an adequate instrument to assess neck conditions?

Table 4 Pearson correlation values of individual item scores to the total NDI [11]

Study NDI item	Present study			Hains et al.*
	Item total correlations (<i>r</i>) (<i>n</i> = 160)	95% CI	<i>P</i> value	Item total correlations (<i>r</i>) (<i>n</i> = 237)
1. Pain intensity	0.53	0.41–0.63	<0.001	0.70
2. Personal care	0.60	0.49–0.69	<0.001	0.68
3. Lifting	0.49	0.36–0.60	<0.001	0.75
4. Reading	0.64	0.54–0.72	<0.001	0.77
5. Headaches	0.45	0.31–0.56	<0.001	0.64
6. Concentration	0.53	0.41–0.63	<0.001	0.76
7. Work	0.59	0.47–0.68	<0.001	0.87
8. Driving	0.61	0.50–0.70	<0.001	0.87
9. Sleeping	0.61	0.51–0.70	<0.001	0.76
10. Recreation	0.66	0.56– 0.74	<0.001	0.83

*Chiropractic Out Patient Department

Table 5 Correlation values between the NDI and domains of the SF36 [12, 21]

NDI (<i>r</i> value) SF 36 domain	Present study * (<i>n</i> = 160)	95% CI	Present study (<i>n</i> = 160)	Riddle and Stratford (<i>n</i> = 146)	Herman et al. (<i>n</i> = 80)*
Physical	-0.57	-0.67 to -0.46	-0.74	-0.53	-0.67
Role Physical	-0.53	-0.63 to -0.41			
Pain	-0.74	-0.80 to -0.66			
General Health	-0.47	-0.58 to -0.34			
Energy	-0.63	-0.72 to -0.53	-0.68	-0.47	-0.39
Social	-0.66	-0.74 to -0.56			
Role Mental	-0.45	-0.57 to -0.32			
Mental Health	-0.51	-0.62 to -0.39			

Individual domains and then the physical and mental components are shown

* All *P* < 0.001

** The present study and Riddle and Stratford used Pearson Correlation whereas Herman et al. used Spearman Rank Correlation

Table 6 Correlation values between the VAS and the NDI and SF36

	<i>r</i> value	95% CI	<i>P</i> value
NDI	0.50	0.37 to 0.61	0.000
SF36 Physical	-0.11	-0.26 to 0.04	0.18 NS
SF36 Role Physical	-0.28	-0.41 to -0.13	0.001
SF36 Pain	-0.51	-0.61 to -0.38	0.000
SF36 General Health	-0.20	-0.34 to -0.05	0.016
SF36 Energy	-0.33	-0.47 to -0.19	0.000
SF36 Social	-0.33	-0.46 to -0.18	0.000
SF36 Role Mental	-0.25	-0.39 to -0.09	0.003
SF36 Mental Health	-0.31	-0.45 to -0.17	0.000

Recently, Vos et al. [23] studied the reliability and validity of the NDI in Dutch patients with acute neck pain in the general practice setting. They performed Bland and Altman plots which showed limits of agreement ranging

from -7.4 to +7.92. However, they scored the NDI out of 50 (not 100) and therefore our values are roughly the same (if doubled). Their mean NDI score was 28% and this may well reflect the different populations being studied. They found the intra-class correlation coefficients (ICC) in 42 patients after a 1-week interval to be 0.9 (95% CI 0.82–0.95). This is similar to our finding (see Table 7). Cleland et al. [5] found the ICC to be 0.68 (95% CI 0.3–0.9), however, they only included 17 patients.

Question 8 in the NDI focuses on the ability to drive. In patients who cannot drive (i.e. have never been able to do so) this question becomes redundant. A previous study computed average values for this question based on the other nine question responses [11]. Scores for incomplete Oswestry disability index questionnaires in which there are missing question responses have been tabulated and are acceptable [7]. For simplicity, our study only included complete questionnaires.

Fig. 2 a–d Bland and Altman plots illustrating the test retest reproducibility of the NDI, VAS and SF36 (each point represents a patient)

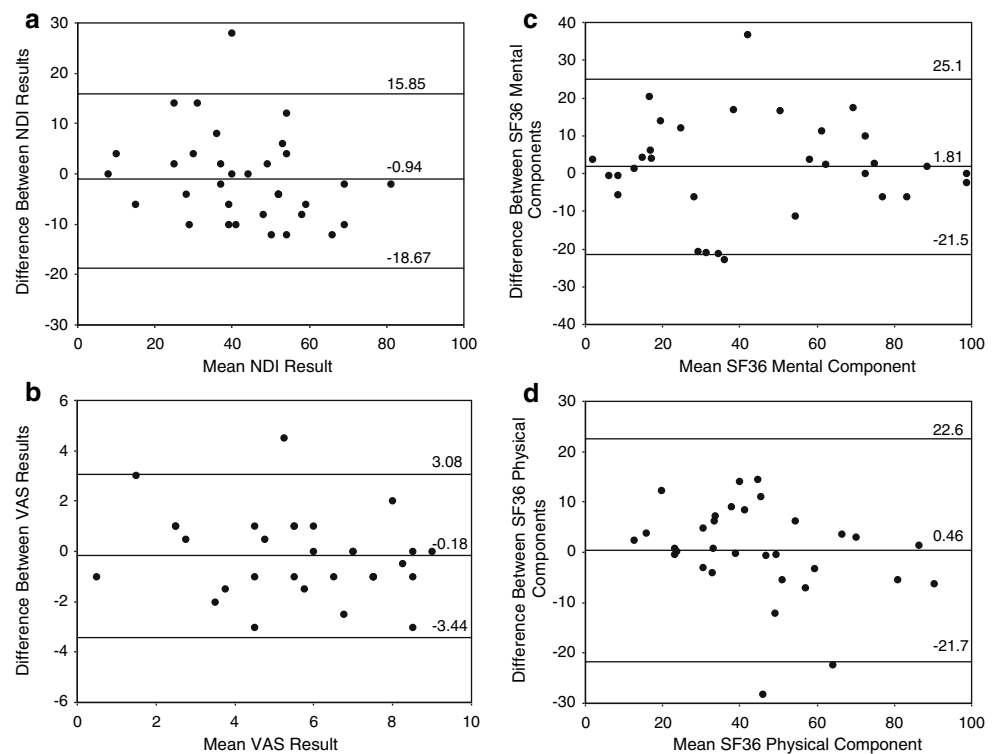


Table 7 Intra-class correlation coefficients for reliability

Questionnaire domain	Intra-class correlation coefficient	95%CI
NDI	0.93	0.86–0.97
VAS	0.87	0.72–0.94
SF36 Physical	0.94	0.89–0.97
SF36 Role Physical	0.83	0.66–0.92
SF36 Pain	0.91	0.81–0.96
SF36 General Health	0.93	0.86–0.97
SF36 Energy	0.91	0.81–0.96
SF36 Social	0.91	0.82–0.96
SF36 Role Mental	0.81	0.60–0.91
SF36 Mental Health	0.90	0.80–0.95

When we designed this study we specifically set out to include all patients attending the spinal out-patients clinic with neck pain. We did not want to investigate specific causes of neck pain (e.g. whiplash associated disorders, trauma, tumour, etc.) because we needed to ascertain whether the questionnaire would be useful for all of our patients. In our hospital, on arrival to clinic and before seeing a doctor, the patient completes the questionnaire; it is scored and then used as an adjunct in the clinical assessment. It is conceivable that the NDI may not be a satisfactory tool in some neck pathologies. In addition we have not looked at the responsiveness to clinical change of the NDI (mentioned above) and any floor and ceiling effects of the NDI.

Conclusions

This study was performed to benchmark the NDI versus the SF36 in the spinal surgery out-patient setting. We have shown that the NDI has good reliability and validity and that it stands up well to the SF36. In agreement with a previous study, there is no need to do both [21]. The NDI is shorter, quicker to answer and easier to score.

References

- Ackelman BH, Lindgren U (2002) Validity and reliability of a modified version of the neck disability index. *J Rehabil Med* 34:284–287
- Bland JM, Altman DG (1996) Measurement error. *Br Med J* 313:744–746
- Bland JM, Altman DG (1997) Statistics notes: Cronbach's alpha. *BMJ* 314:572
- Bovim G, Schrader H, Sand T (1994) Neck pain in the general population. *Spine* 19:1307–1309
- Cleland JA, Fritz JM, Whitman JM, Palmer JA (2006) The reliability and construct validity of the neck disability index and patient specific functional scale in patients with cervical radiculopathy. *Spine* 31:598–602
- Fairbank JCT, Pynsent PB (2000) The Oswestry disability index. *Spine* 25:2940–2953
- Fairbank JCT, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66:271–273
- Garratt AM, Ruter DA, Abdalla MI, Buckingham JK, Russell IT (1993) The SF36 health survey questionnaire: an outcome

- measure suitable for routine use within the NHS. *Br Med J* 306:1440–3
9. Grevitt M, Khazim R, Webb JK, Mulholland R, Shepperd (1997) The short form-36 health survey questionnaire in spine surgery. *Br J Bone Joint Surg* 79:48–52
 10. Gruez M, Hildingsson C, Nilsson M, Toolanen G (2002) The prevalence of neck pain—a population-based study from northern Sweden. *Acta Orthop Scand* 73:455–759
 11. Hains F, Waalen J, Mior S (1998) Psychometric properties of the neck disability index. *J Manip Physiol Thera* 21:75–80
 12. Hermann KM, Reese CS (2001) Relationships among selected measures of impairment, functional limitation, and disability in patients with cervical spine disorders. *Phys Ther* 81:903–914
 13. Hoving JL, O’Leary EF, Niere KR, Green S, Buchbinder R (2003) Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. *Pain* 102:273–281
 14. Jenkinson C, Coulter A, Wright L (1993) Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *Br Med J* 306:1437–1440
 15. Jenkinson C, Layte R, Wright L, Coulter A (2006) The UK SF-36: An analysis and interpretation manual. Health Services Research Unit, University of Oxford March 2006
 16. Jette DU, Jette AM (1996) Physical therapy and health outcomes in patients with spinal impairments. *Phys Ther* 76:930–941
 17. Korthals-de Bos IB, Hoving JL, van Tulder MW, Rutten-van Molken MP, Ader HJ, de Vet HC, Koes BW, Vondeling H, Bouter LM (2003) Cost effectiveness of physiotherapy, manual therapy, and general practitioner care for neck pain: economic evaluation alongside a randomised controlled trial. *BMJ* 326:911
 18. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF (2003) A comparison of two time intervals for test–retest reliability of health status instruments. *J Clin Epidemiol* 56:730–735
 19. Muller U, Duetz MS, Roeder C, Greenough CG (2003) Condition-specific outcome measures for low back pain. *Eur Spine J* 13:301–324
 20. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF (2002) Standard scales for measurement of functional outcome for cervical pain or dysfunction. *Spine* 27:515–522
 21. Riddle DL, Stratford PW (1998) Use of generic versus region specific functional status measures on patients with cervical disorders. *Phys Ther* 78:951–963
 22. Vernon H, Mior S (1991) The neck disability index: a study of reliability and validity. *J Manip Physiol Ther* 14:409–415
 23. Vos CJ, Verhagen AP, Koes BW (2006) Reliability and responsiveness of the Dutch version of the Neck Disability Index in patients with acute neck pain in general practice. *Eur Spine J* 15:1729–1736
 24. White AR, Ernst E (1999) A systematic review of randomised controlled trials of acupuncture for neck pain. *Rheumatology* 38:143–147