ORIGINAL ARTICLE

Danish version of the Oswestry Disability Index for patients with low back pain. Part 1: Cross-cultural adaptation, reliability and validity in two different populations

Henrik Hein Lauridsen · Jan Hartvigsen · Claus Manniche · Lars Korsholm · Niels Grunnet-Nilsson

Abstract Over the past 10 years, a plethora of back-specific patient-orientated outcome measures have appeared in the literature. Standardisation has been advocated by an expert panel of researchers proposing a core set of instruments. Of the condition-specific questionnaires the Oswestry Disability Index (ODI) is recommended for use with low back pain (LBP) patients. To date, no Danish version of the ODI exists which has been cross-culturally adapted, validated and published in the peer-reviewed literature. A cross-cultural adaptation and validation of the ODI for the Danish language was carried out according to established guidelines: 233 patients [half of the patients were seen in the primary sector (PrS) and half in the secondary sector (SeS) of the Danish health care system] with LBP and/or leg pain completed a questionnaire booklet at baseline, 1 day or 1 week and 8 weeks follow-up. The booklet contained the Danish version of the ODI, along with the Roland Morris Disability Questionnaire, the LBP Rating Scale, the SF36 (physical function and bodily pain scales) and a global pain rating. For the ODI test–retest analysis (93 stable patients) resulted in an intraclass correlation coefficient of 0.91, a mean difference of 0.8 and 95% limits of agreements of – 11.5 to + 13. Thus, a worsening greater than 12 points and improvement greater than 13 points can be considered a "real" change above the measurement error. A substantial floor effect was found in PrS patients (14.1%). The ODI showed satisfactory cross-sectional discriminant validity when compared to the external measures. Concurrent validity of the ODI revealed: (a) a 10% and 21% lower ODI score compared to the disability and pain measures, respectively, (b) a poorer differentiation of patient disabilities and (c) an acceptable individual ODI score level compared to the external measures. Longitudinal external construct validity showed moderate correlations (range 0.56–0.78). We conclude that the Danish version of the ODI is both a valid and reliable outcome instrument in two LBP patient populations. The ODI is probably most appropriate for use in SeS patients.

Part 2 of this article is available at: http://dx.doi.org/10.1007/s00586-006-0128-6

H. H. Lauridsen (✉) · J. Hartvigsen · N. Grunnet-Nilsson
Institute of Sports Science and Clinical Biomechanics,
University of Southern Denmark, Odense, Denmark
e-mail: hlauridsen@health.sdu.dk

C. Manniche · H. H. Lauridsen · J. Hartvigsen ·
N. Grunnet-Nilsson
Clinical Locomotion Science, Odense, Denmark

J. Hartvigsen
Nordic Institute for Chiropractic and Clinical Biomechanics,
Odense, Denmark

C. Manniche
Backcenter Funen, Ringe, Denmark

L. Korsholm
Department of Statistics, University of Southern Denmark,
Odense, Denmark

Introduction

Patient views in measuring functional health status is important in order to understand and document both

the impact of pain and symptoms and the effect of treatment in low back pain (LBP) patients [1, 2].

Patient-based outcome measures are usually classified as generic or disease-specific [3]. The generic measures are designed to measure the domains of general health, overall disability and quality of life and are important for broad comparisons across conditions. This is often at the expense of the responsiveness to clinically relevant change in specific diseases [4]. Therefore, disease-specific instruments measuring attributes of symptoms and functional status relevant to a particular disease or condition were developed and they are often found to be more responsive to the target condition when compared to generic measures [2, 5–7].

A plethora of back-specific instruments have been developed over the last decade, and in a recent review a total of 36 back-specific questionnaires attempting to address patient perceptions of their back trouble have been identified [8]. Choosing the ''ideal'' outcome measure for a clinical trial is virtually impossible since most instruments offer advantages and disadvantages depending for example on the type of study or patient population. As a result, Deyo and colleagues [9] proposed a standardised core set of instruments measuring five domains: pain symptoms, back related function, generic well-being, disability and satisfaction with care. These recommendations were updated by an expert panel in 2000 [4]. In the domain of back related function, the recommended and most widely used measures are the Oswestry Disability Index (ODI) and the Roland Morris Disability Questionnaire (RDQ). Consequently, a MEDLINE search revealed more than 300 citations in which the ODI had been used to assess disability in LBP and it has been found to be reliable, valid and responsive in particular in patients with a higher level of disability [10–13]. The ODI exists in four versions, and to facilitate a comparison of results among studies, version 2.1 is recommended [8, 10, 11].

The ODI is a self-administered questionnaire initially developed by John O'Brien in 1976 and version 1.0 was published and validated in 1980 [14]. Version 1.0 has been adapted by the American Academy of Orthopaedic Surgeons (AAOS) omitting sections 1,8 and 9 and changing the score of each item from 0–5 to 1–6 [9, 11]. Another revision of version 1.0 was carried out by the Medical Research Council and was published as version 2.0 in 1989 [15]. This is not to be confused with the revised (modified) ODI also published in 1989 by a chiropractic study group [16]. In 2000 Fairbank and Pynsent [11] published a thorough review of the ODI with reprints of the four versions. However, in section 10 (travelling) of version 2.0 as published by Fairbank and Pynsent [11] the third and fourth response options have subtle mistakes. This was corrected in a subsequent publication by Roland and Fairbank [10] and is now referred to as version 2.1 [17].

Most questionnaires are developed in English-speaking countries and a direct translation for use in a different language may be problematic. Published guidelines for standardised translation and cross-cultural adaptation exist [18, 19]. The ODI has been cited in nine languages, some of which have followed a rigid cross-cultural adaptation process, and published in the literature. Several publications refer to a Danish version of the ODI [20–23], however, a systematic search of the literature revealed no published translation, cross-cultural adaptation or validation into the Danish language.

The objectives of this two article series are twofold: (1) to translate and cross-culturally adapt the ODI version 2.1 into the Danish language and (2) to investigate the psychometric properties of the Danish ODI in a large population of back pain patients seen in the primary (PrS) and secondary sectors (SeS) of the Danish health care system. In paper 1 of this series, the translation and cross-cultural adaptation process, test–retest stability, scale width and construct validity are examined in two distinct back pain populations, and in paper 2 we examine the sensitivity, specificity and clinically significant improvement in the same two LBP populations.

Materials and methods

Translation and cross-cultural adaptation

The translation and cross-cultural adaptation process followed the five stages outlined in the recent guidelines [18, 19]. Written documentation was produced for each stage of the process serving as a memory aid for the expert committee review. Version 2.1 of the ODI was translated from English to Danish by two different and independent translators whose mother tongue was Danish. Translator 1 (T-1) was a professional translator with a secretarial job and, thus, naïve to the purpose and health concepts of the questionnaire. Being naïve to the purpose and concepts was useful in eliciting unexpected meanings from the original instrument. The other translator (T-2), a professor in clinical biomechanics, was aware of the purpose and the concepts involved in the instrument. This would improve the reliability of the ODI by allowing for a better idiomatic and conceptual rather than literal equivalence between the two versions of the questionnaire.

Both Danish translations (T-1 and T-2) were compared with one another to produce a preliminary translated version (T-12p). To evaluate the quality of the translation process, two independent raters judged the T-12p version of the ODI before retranslation into English. The quality was rated according to clarity of translation, common language use and conceptual equivalence on a scale ranging from 0 (not at all perfect) to 100 (perfect) [24]. A panel consisting of the forward translators, the independent raters and the main author evaluated the comments from the two raters to produce the final translated version (T-12).

The final T-12 version was then retranslated into English by two independent translators with English (British English and Australian English) as their mother tongue and Danish as their secondary language. Back-translator 1 (BT-1) was a teacher of English and back-translator 2 (BT-2) was a professor at the local University. Both had been living in Denmark for several years and were blinded to the original version of the ODI. In preparation for the expert committee review, the two retranslations were compared with the original version of the instrument resulting in a checklist highlighting major discrepancies in content between the two versions.

A bilingual expert committee including the forward and back-translators, a language specialist, a methodologist, a clinician and a recorder/coordinator was assembled to review all the versions of the forward and back translations. The purpose of the expert committee was to resolve major discrepancies detected in the translation and retranslation process, detect errors of interpretation and missed nuances, and assess the necessity of performing a cultural adaptation for use among Danish back pain patients. All issues raised during the expert committee review process were resolved by consensus and documented in a written report.

During the final part of the adaptation process the pre-final version was tested for content (face validity), wording, ease of understanding and missing items. Forty patients participated in the pre-testing of the questionnaire; 20 patients seen in the PrS of the Danish health care system (a chiropractic clinic) and 20 in the SeS (an out-patient hospital back pain clinic). Each patient completed the questionnaire followed by a questionnaire developed for the purpose of detecting comprehension. At completion, they were briefly interviewed to explore any problem areas in-depth. The findings were discussed among the translators resulting in only minor changes to the pre-final version. Further psychometric testing of the final version of the Danish ODI was carried out in a validation study.

Validation study

The study was reported and accepted by The Danish Data Protection Agency.

Patients and setting

Back pain patients' initial entry point into the Danish health care system is the primary health care sector comprising general practitioners, chiropractors and physical therapists (via the general practitioner). Patients who do not respond to the initial treatment may get referred to a hospital-based multidisciplinary spinal unit in the SeS for further evaluation and management. Thus, sociodemographic and illness profiles of the patients in the two sectors are very different [25] and we recruited participants in both sectors. The PrS patients were recruited from seven chiropractic practices, whereas the SeS patients were enrolled from a multidisciplinary spinal unit (Backcenter Funen, Ringe). A total of 301 consecutive patients were recruited: 168 from the PrS and 133 SeS patients from the Danish health care system (Fig. 1). Sixty-eight patients refused to participate resulting in a baseline study population of 233.

Questionnaires

A questionnaire booklet was constructed for the validation study which included the final version of the Danish ODI, the 23-item RDQ [26, 27], the two subscales of the LBP Rating Scale—Pain ($LBPRS_{pain}$) and disability ($LBPRS_{disability}$) [23]—and the two subscales of the SF36—physical function SF36 (pf) and bodily pain SF36 (bp) scales [24, 28, 29]. Furthermore a global 0–10 numeric rating scale (NRS_{pain}) measuring back and/or leg pain intensity "today" was included.

The questionnaire booklet used for test–retest reliability contained the Danish version of the ODI with the questions rearranged at follow-up and a single question asking whether the patient had experienced any change since the last time completing the questionnaire.

The patients' global retrospective assessment of treatment effect (transition question) was measured using a 7-point Likert scale ranging from "much better" to "much worse" [30]. All patients were told their baseline global rating of pain severity (NRS_{pain}) before answering the transition question [31, 32]. In addition, they were asked to rate the importance of any changes in their back/leg pain since baseline using a 0–10 NRS.

Fig. 1 Flow of participants and dropouts in the validation study. Groups A and B received a 7-point and 15-point transition question, respectively
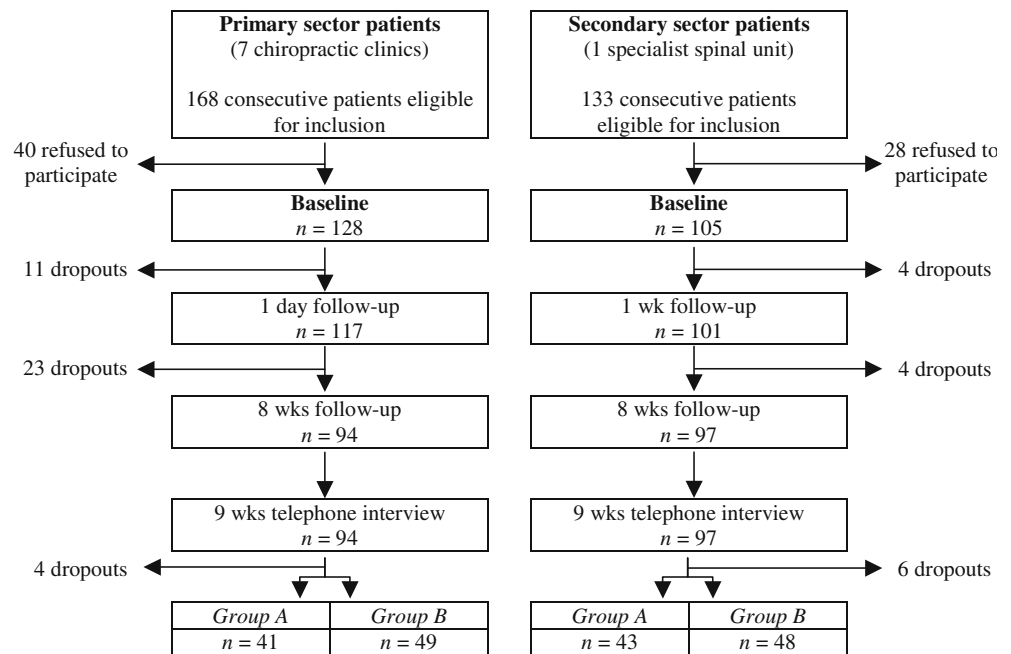
Data collection

Patients eligible to participate in the study had to fulfil certain criteria: (1) age above 18, (2) presence of LBP and/or leg pain and (3) able to read and understand Danish. Exclusion criteria were: (1) suspected pathological disorder of the spine (fractures, spinal infections or malignancy, ankylosing spondylitis, rheumatoid arthritis, or other inflammatory diseases) and (2) patients with a known psychiatric disorder.

Twenty minutes before the initial consultation, the purpose of the study was explained and oral consent was obtained. The patients filled in the baseline questionnaire booklet. A test–retest questionnaire booklet was completed 1 day after for the PrS patients and 1 week after for the SeS patients. The shorter interval for the PrS test–retest patients was selected as these patients are likely to demonstrate true change due to the natural history of back pain and a possible treatment effect [33, 34]. This is more unlikely in SeS patients as the duration of LBP is longer. Only patients reporting to be stable were included in the test–retest analysis. At 8 weeks the patients received the final questionnaire booklet. All patients who completed the questionnaire at 8 weeks participated in a telephone interview carried out by a professional interviewer from the Danish National Institute of Social Research. Information on the patient's retrospective assessment of the treatment was obtained, and to reduce dependence between the transition question and the questionnaires the interview was conducted 3–5 days after the 8 weeks follow-up [35].

Analysis

Data transformation. The two subscales of both the SF36 and the LBPRS, the RDQ and the NRS_{pain} were transformed to cover an interval ranging from 0% to 100%, with a high score representing higher disability or pain [36].

The raw change score for each outcome measure was obtained by subtracting the 8 weeks follow-up score from the baseline score. For the last part of the concurrent validity calculations, the raw change scores were converted into standard scores with a mean of 0 and a standard deviation of 1, thus, allowing for between-scale comparisons [37].

Reliability

Psychometricians have for years used *reliability* as a generic term to indicate both homogeneity (internal consistency) of a scale and reproducibility of scores [38, 39].

Homogeneity (internal consistency) assesses to which extent the items in a scale are interrelated and taps different aspects of the same attribute (unidimensionality). We used item-total correlations and Cronbach's alpha coefficient to assess internal consistency. Item-total correlation is the correlation of the individual item with the scale total omitting that item. Cronbach's alpha was calculated from the baseline values and homogeneity is considered acceptable when Cronbach's alpha exceeds 0.7 although it is often recommended that values should not be above 0.9 as this

suggests item redundancy [37, 40]. To further evaluate each items contribution to the total score, we graphed the item score against the five score categories as described by Fairbank et al. [14]. If the item correlates well with the latent variable (pain related function) an increase in the line is expected as the total ODI score increases. On the other hand, a more horizontally oriented line may represent an item which belongs to a different latent variable [27].

Reproducibility was measured using the intraclass correlation coefficient (ICC) for repeated trials [39] and using the limits of agreement (LOA) as outlined by Bland and Altman [41, 42]. The Bland and Altman method has several advantages when compared to all correlation coefficients. First, correlation coefficients depend on the range and distribution of the variables and, hence, the way in which the sample of subjects was chosen. Lastly, correlation coefficients may be high despite a poor agreement between the repeated measurements [43].

Scale width

The lowest and highest possible scores of a scale are known as the ''floor'' and ''ceiling''. If a high proportion of patients score at or very close to the floor or ceiling, no further improvement or deterioration can be detected resulting in biased results [44].

Scale width is defined as the region of the score range of an instrument with the capacity to allow detection of change in scores over time and is an extension of the ''floor'' and ''ceiling'' concepts [45]. In addition to reporting floor and ceiling effects. We used the LOA interval at each end of the scale to be 95% confident that a change greater than instrument measurement error can take place, in addition to reporting floor and ceiling effects [45].

Validity

Cross-sectional discriminant validity assesses whether the scales under investigation can differentiate among groups of patients with different levels of a chosen factor (e.g. symptom location). We chose to assess the following baseline factors from the medical history at two levels: (1) location of symptoms (LBP only vs. leg pain ± LBP) [46, 47], (2) pain duration of the current episode (≤ 30 days vs. > 30 days) [48] and (3) frequency of taking medication during the last week (less than a couple of times during the last week vs. more than a couple of times during the last week) [46].

Concurrent validity analysis was carried out at baseline and 8 weeks follow-up. We tested the ODI

and the external instruments for within- and between-scale systematic differences in patient grading by calculating the difference in the mean score of the instruments for the two patient populations. Between-scale systematic differences were tested using an interaction term in the regression model. Second, the ability of an instrument to distinguish between different degrees of patient disability can be expressed as how well the patients are spread out on the response scale (0–100%). Using a variance comparison test, we compared the spread of the ODI scores to the other instruments. Lastly, we examined whether the individual patient score level on the ODI scale was comparable to the external instruments. Bland–Altman LOA plots of standardised scores were used for this analysis [41, 43].

Longitudinal external construct validity examines whether or not a scale measuring a certain domain over time correlates appreciably well with other scales that theory suggests should be related to it [49, 50]. Longitudinal external construct validity was assessed by comparing the change score of the ODI with that of the external measures using Pearson's correlation coefficient (r).

All statistical calculations were carried out using the statistical package STATA® v. 8.2 SE (StataCorp). Robust variance estimation was applied whenever possible in order to reduce the dependency on normality assumption and statistical significance was accepted at the $P < 0.05$ level.

Results

Translation and cross-cultural adaptation

During the translation process, several noteworthy issues arose. First, in section 1 there was disagreement among the expert committee members as to how to scale the severity of pain in Danish. Many words exist describing for example ''mild pain'' or ''moderate pain''. Consensus was reached by close scrutiny of (a) common language and (b) conceptual equivalence. Second, as noted in the German translation of the ODI [51], it seems illogical to have ''*very* painful'' in answer category 2 *and* 3 of section 2 (personal care) as category 2 reflects less disability compared to category 3. Thus, we omitted the word ''very'' from answer category 2 in this section of the Danish ODI. Third, the expert committee discussed how to translate ''travelling'' as the equivalent Danish word ''rejse'' is conceptually slightly different. However, in lack of a more precise word, the committee agreed on using this word.

The quality of the translation process showed an overall difficulty rating (average of clarity, common language and conceptual equivalence) well above 90 for all sections of the questionnaire (data not shown). Item 1 (pain) and 6 (standing) showed the poorest difficulty ratings (91 and 94) corresponding to a high number of comments but only minor wording changes. The Danish version of the ODI is available from the official ODI website [17] or from the authors on request.

Validation study

Participants and missing data

Three hundred and one consecutive patients (PrS: $n = 168$; SeS: $n = 133$) were eligible for inclusion into the study (Fig. 1). The baseline response rate was 77% leaving 233 included patients at baseline (PrS: $n = 128$; SeS: $n = 105$). At 8 weeks the follow-up response rate was 82% of the baseline entry; thus, 191 patients (PrS: $n = 94$; SeS: $n = 97$) were available for analysis at 8 weeks follow-up. An additional ten patients dropped out at the 9 weeks telephone interview mostly from the SeS.

Table 1 Baseline descriptive data for the two study groups

Characteristic	Study population	
	Primary sector patients ($n = 128$)	Secondary sector patients ($n = 105$)
Age (mean; range)	42 (18–76)	46 (20–85)
Sex (n; %)		
Male	65 (51)	50 (47)
Female	63 (49)	55 (52)
Pain location (n; %)		
Low back only[a]	96 (75)	30 (29)
Pain in the lower extremities only	1 (1)	6 (6)
Low back pain and pain in the lower extremities	31 (24)	69 (66)
Duration of current problem (n; %)		
≤ 30 days	93 (73)	12 (11)
> 30 days	32 (25)	85 (81)
Missing	3 (2)	8 (8)
Use of pain medication[b] (n; %)		
≤ a couple of times a week	85 (66)	44 (42)
> a couple of times a week	41 (32)	61 (58)
Missing	2 (2)	–

[a] Low back is defined as the area between the lower costal margin and the gluteal folds

[b] Any pain-killing medication taken for back and/or leg pain over the past week

The baseline demographics of the two study populations are shown in Table 1. Age distribution and the ratio of male/female were similar in the two groups whereas all the other characteristics were distinctly different. Patients from the PrS had mostly LBP only, shorter duration of the current LBP episode and used less medication compared to SeS patients.

A dropout analysis showed a lower mean age for the dropouts (8 years lower) in both PrS and SeS patients and dropouts from the SeS were more likely to be males with longstanding problems but lower medication use.

At baseline 25 patients (11%) failed to answer item 8 (sex life) and 15 patients (6%) failed to answer item 10 (travelling) and this was equally distributed between PrS and SeS patients.

Reliability and stability

Homogeneity was assessed using Cronbach's alpha and item-total correlations at baseline ($n = 233$). For the whole group alpha was 0.88. For PrS and SeS patients we found an alpha of 0.89 and 0.85, respectively. Item-total correlations ranged from 0.54 (item 7, sleeping) to 0.73 (item 10, travelling) in the whole group.

The influence of each item on the total ODI score is depicted in Fig. 2. In general, all item scores increase with an increasing total ODI score. Thus, each item contributes to the total score and belongs to the same latent variable (pain related function). Items 8 and 10 (sex life and travelling) seem to respond better at higher ODI scores; however, caution should be taken as to the validity of this since the number of patients is low ($n = 5$).

Repeatability was carried out on 93 stable patients (PrS: $n = 36$; SeS: $n = 57$). The mean (SD) time interval for completion of the two questionnaires was 9.1 (10.6) days for all patients, 4.4 (9.8) days for PrS patients and 12.0 (10.1) days for SeS patients. The ODI showed excellent test–retest reliability, as evidenced by the ICC and LOA. ICC was 0.91 among all patients, 0.93 in PrS patients and 0.89 in SeS patients. The mean difference and 95% LOA for all patients were 0.8 (–11.5 to + 13.0) with no noteworthy difference between PrS and SeS patients [2.2 (–9.2 to + 13.6) and –0.1 (–12.7 to + 12.4), respectively]. Thus, no systematic bias was found between the test and retest and the spread of the dots was uniform (Fig. 3). All normal plots of the differences were acceptable.

Scale width

Only one patient obtained the lowest possible score (floor effect) whereas no patients reached the ceiling of
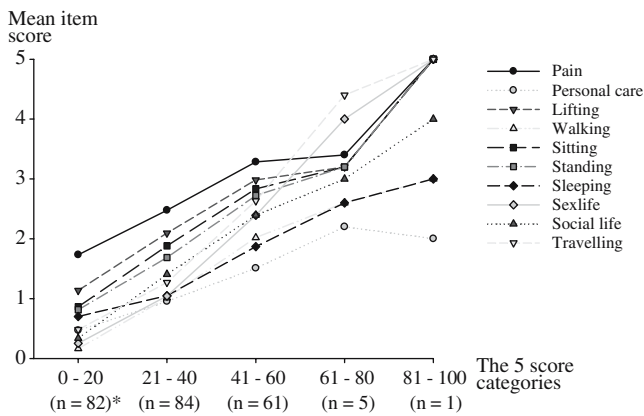
Mean item
score

Fig. 2 Mean score for each ODI item in relation to the five total score categories. The average score of each item is depicted as a function of the total baseline entry score (divided into five score categories). An increase in the average item score with an increasing total baseline entry score signifies good correlation with the latent variable of the instrument. *Number of patients in each category

the scale at baseline. However, the proportion of patients scoring outside the scale width (as indicated by the LOA) showed a different picture. A total of 25 patients (10.7%) scored within the lower score range (0–11.5%) with 18 (14.1%) being PrS patients and 7 (6.7%) being SeS patients. No patients scored within the upper score range (87–100%).

Validity

Cross-sectional discriminant validity. Table 2 provides a summary of the findings for the cross-sectional discriminant validity analysis. The results show a small monotonic decrease in the ODI score with more proximal symptoms ($P < 0.001$), shorter pain duration
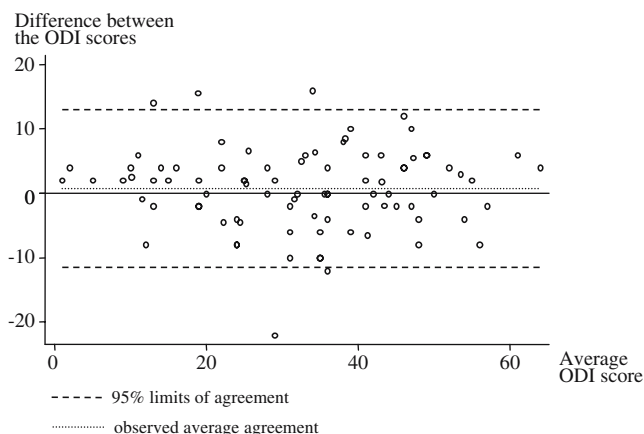
Difference between
the ODI scores

Fig. 3 Limits (95%) of agreement plot for repeated measures of the ODI in stable patients. The plot shows the difference in ODI score against the average ODI score

of the current episode ($P < 0.05$) and a larger increase in ODI score with more medication usage ($P < 0.001$). No differences were observed between the PrS and SeS patient groups.

Concurrent validity. We looked at three different aspects of concurrent validity. First, the ODI was tested for systematic differences when compared to the other instruments. At baseline the ODI measured $\approx 10\%$ ($P < 0.01$) lower compared to the external disability measures [RMQ, LBPRS$_{disability}$ and SF36 (pf)] and $\approx 21\%$ ($P < 0.01$) lower compared to the external pain measures (LBPRS$_{pain}$, SF36 (bp) and NRS$_{pain}$). The same trend was noted at 8 weeks follow-up and between PrS and SeS patients. We also looked at the within- and between-scale systematic differences at baseline between the two study populations to evaluate if any differences existed. The within-scale mean difference between the PrS and SeS patients for the ODI was 5 points. A similar result was found for the RMQ (5 points); however, LBPRS$_{disability}$ and SF36 (pf) showed a somewhat higher mean difference of 10 and 11 points, respectively (data not shown). Between-scale mean differences are shown in Table 3. No statistically significant differences were found between the mean score of the ODI and the external instruments except for the two subscales of the SF36. The results from the 8 weeks between-group comparison are not included as the data in the PrS patients were biased due to a floor effect.

Second, we compared the spread of the ODI scores to the disability and pain measures at baseline and 8 weeks follow-up. At baseline, the ODI scores are spread over a narrower window (SD ± 15.85) when compared to the external measures (SD range 17.40–25.38). This was statistically significant ($P < 0.01$) for all comparisons except the RMQ (SD ± 17.40; $P = 0.16$). When comparing ODI and RMQ for the PrS

Table 2 The ability of the ODI to distinguish between clinically important subgroups (cross-sectional discriminant validity) at baseline

Medical history variables	Oswestry Disability Index		
	n	Mean (± SD)	*P* Values*
Location of symptoms			
Low back pain only	126	27 (16)	< 0.001
Leg pain ± low back pain	100	34 (14)	
Pain duration			
≤ 30 days	105	29 (17)	< 0.05
> 30 days	117	32 (15)	
Medication			
≤ a couple of times a week	129	25 (15)	< 0.001
> a couple of times a week	102	38 (14)	

*P Values based on the Wilcoxon rank sum test

Table 3 Baseline comparisons of systematic differences between mean scores of the ODI and the external instruments in PrS and SeS patients

Instrument	Mean scores		Difference between mean score and mean ODI score[b]	
	PrS ($n = 128$)	SeS ($n = 105$)	PrS	SeS
ODI	28.0	33.1		
RMQ	39.7	44.2	11.7	11.1
LBPRS$_{disability}$	33.8	44.2	5.8	11.1
SF36 (pf)	33.8	44.8	7.3	11.7[a]
LBPRS$_{pain}$	38.4	47.4	10.4	14.3
SF36 (bp)	65.9	63.3	37.9	30.2[a]
NRS$_{pain}$	47.0	48.3	19.0	15.2

[a] Statistically significant difference between PrS and SeS patients (*P* values based on the regression analysis with an interaction term)

[b] Difference calculated as the mean score of the external instrument minus the mean ODI score

and SeS patients at baseline, no significant difference in the score spread was seen. The same trend was observed at 8 weeks follow-up in both patient populations.

Finally, the individual patient score level was examined by Bland–Altman LOA plots of standardised scores (Fig. 4). ODI score level at baseline is within ± 1.3 SD when compared to the other disability measures and within ± 1.7 SD in comparison to the pain measures. Furthermore, the ODI score level is comparable in PrS and SeS patients. The same pattern was seen at 8 weeks follow-up (data not shown).

Longitudinal external construct validity. Correlations between the change score of the ODI and the external measures were calculated using Pearson's *r*. The results showed correlation coefficients of 0.78 (RDQ), 0.69 (LBPRS$_{disability}$), 0.75 (SF36 (pf)), 0.56 (LBPRS$_{pain}$), 0.65 (SF36 (bp)) and 0.61 (NRS$_{pain}$). As expected, the ODI correlated less strongly to the pain measures compared to the disability measures. All correlations were statistically significant ($P < 0.01$), indicating acceptable external longitudinal construct validity of the ODI change score.

Discussion

This paper reports on the Danish cross-cultural adaptation of the frequently used back-specific ODI, and presents results of the first part of the psychometric testing. The validation procedures were carried out in two different back pain populations for several reasons. First, few studies have cross-culturally adapted and validated functional scales in patients with LBP of differing severity [47]. Second, we specifically wanted to psychometrically test the ODI in a broad range of LBP patients since a cross-culturally adapted outcome measure should be tested in target populations relevant for clinical research and clinical practice.

We included consecutive patients in the study to get a true representation of LBP patients in the two patient populations. The dropout analysis did show some differences between the participants and dropouts; however, we consider these differences minor.

Translation of the ODI

During the translation and cross-cultural adaptation procedures we followed the recommendations described by Guillemin et al. and Beaton et al. [18, 19].

Fig. 4 Limits of agreement between the mean standardised ODI score and the standardised RDQ, LBPRS$_{disability}$, LBPRS$_{pain}$, SF36 (pf), SF36 (bp) and NRS$_{pain}$ scores at baseline
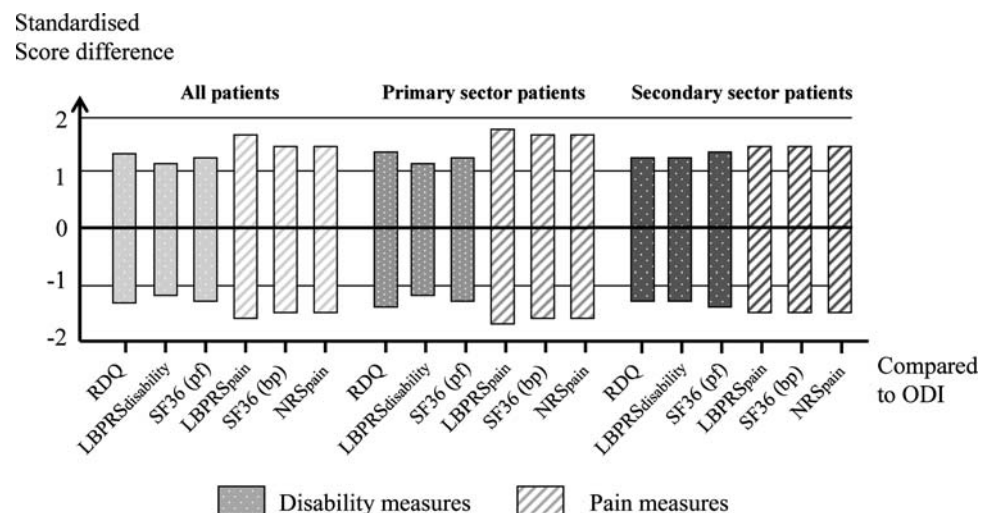
The problems encountered during the process were minor and documented at all stages, and we conclude that our attempt to translate the ODI into Danish is both reliable and conceptually valid.

Reliability

Homogeneity (internal consistency), as measured by Cronbach's alpha, was found to be 0.88 for the whole study population (PrS 0.89; SeS 0.85) which falls well within the recommended interval of 0.7–0.9 for group comparisons [37]. Our ODI alpha is in the top end when compared to previously reported coefficients ranging from 0.76 to 0.94 [46, 52–55]. Item-total correlations ranged from 0.54 to 0.73 for all patients and were generally higher for the PrS patients.

We used the ICC and LOA as a measure of repeatability. The study showed that the ODI had an excellent ICC of 0.91 which compared well with the literature [15, 45, 56]. We found a mean difference of 0.8 and a 95% LOA of –11.5 to + 13.0 with no noteworthy difference in the two patient populations. This indicates that the ODI showed negligible systematic bias on the repeated measurements. The 95% LOA signifies change greater than the measurement error and is therefore conceptually equivalent to the minimum detectable change (MDC) as reported by Stratford and Binkley [57]. Thus, a worsening greater than 12 points and improvement greater than 13 can be considered a ''real change'' at the very stringent 95% confidence level. At the less stringent 90% confidence level the LOA was found to be (–9.6 to 11.0). To the author's knowledge, this is first time LOAs for the ODI have been reported in the literature [13]. In several studies values for the MDC for the ODI have been reported; however, the comparability is questionable as the ODI version and level of confidence differ. Hägg et al. [58] reported an $MDC_{95\%}$ of 10 points for ODI version 1.0, Frits et al. and Grotle et al. found an $MDC_{95\%}$ of 13 and 11 points, respectively, for the modified (revised) ODI and Mannion et al. [51] found an $MDC_{95\%}$ of nine points for ODI version 2.1. Furthermore, the $MDC_{90\%}$ was reported to be 10.5 points for the modified ODI [45]. Thus, our LOA of 13 points is in the high end in comparison to reported values. Apart from ODI version and confidence level, we ascribe this to differences in the patient population and test–retest time interval.

The mean time spans between completions of the two questionnaires were 4.4 and 12.0 days for the PrS and SeS patients, respectively. The shorter test–retest interval in PrS patients was carefully chosen balancing the risk of not finding stable patients and introducing bias from patients memorising their previous answer. To reduce the memory effect, the sequence for ten items of the ODI were changed at the retest. When examining the LOAs for the two patient populations no differences were found.

Scale width

Traditionally floor and ceiling effects describe the percentage of subjects scoring maximal or minimal points. As a benchmark McHorney and Tarlov [44] suggested that questionnaires with more than 15% of the respondents scoring at the floor or ceiling initially should not be used. We did not find any floor or ceiling effect of the Danish ODI using this criterion as only one patient reached the floor of the scale. However, using the more sensible scale width approach, the Danish ODI showed a fairly pronounced floor effect in the PrS patients (14.1%) compared to the SeS patients (6.7%). Similar results were found by Patrick et al. [26] in a non-surgical patient group and it is thus questionable how useful the Danish ODI is as a primary outcome measure in a PrS patient population.

Validity

We examined several aspects of criterion and construct validity of the Danish ODI. The results of the cross-sectional discriminant validity analysis showed that the ODI can discriminate between groups of subjects that are expected to differ in their level of disability for all the chosen variables (symptom location, pain duration and medication usage). Interestingly, the group score difference was the largest for medication usage (13 points) in comparison to symptom location (7 points) and pain duration (3 points) indicating that this variable is important for discriminating among LBP patients when using the ODI.

In the concurrent validity analysis we looked at the differences between the ODI and external disability and pain measures at baseline and 8 weeks follow-up. Three aspects were analysed: systematic differences among the instruments, patient spread on the response scale and specific response scale scores for the different instruments. In comparison to the disability and pain measures the mean score of the ODI was 10 and 21% lower, respectively. This confirms previous findings that the ODI may be more appropriate for patients with a greater degree of disability [10], particularly so when the pain level is high. Comparing PrS and SeS patients, the results showed similar systematic differences for the ODI and RMQ (5 points) but higher for the $LBPRS_{disability}$ and SF36 (pf) (10–11 points). This is

important when comparing results of similar patient populations in clinical trials. Further comparisons of the two patient populations showed that the difference between the mean scores of the external instruments compared to the mean ODI score was negligible except for the two subscales of the SF36. We suspect this to be due to the generic nature of the SF36 and the finding supports the validity of disease-specific instruments such as the ODI.

The second analysis evaluated the ability of the ODI to distinguish between different patient disabilities (patient spread). Several interesting points were noted. First, of all the external pain and disability scales the ODI showed the narrowest window indicating a poorer spread of the PrS and SeS patients on a scale ranging from 0 to 100%. Second, the ODI and RMQ seem to be almost equally good at differentiating patient disabilities in both study populations except at lower disabilities where the ODI has a tendency to reach the floor of the scale (data not shown). Third, the pain scales showed a superior ability at differentiating patients (in particular NRS_{pain}) in comparison to the disability scales highlighting the importance of including both pain and disability measures in clinical trials. Lastly, the global scale of SF36-pf showed a better differentiating ability compared to the disease-specific scales (ODI and RMQ) proving that disease-specific scales are not necessarily the best scales for the cross-sectional differentiation of LBP patients.

In the last analysis we compared the ODI score level to the external pain and disability scales using standardised LOAs. Agreement on the individual score level ranges from ± 1.3 SD for the disability measures and ± 1.7 SD for the pain scales reflecting that pain and disability are two related but different dimensions. We consider the agreement between the ODI score level as compared to the external measures acceptable.

Kirshner and Guyatt [49] recommended evaluative measures be tested for longitudinal external construct validity. In lack of a "golden standard" we examined the correlation of the ODI change scores against well-validated instruments purporting to measure the domains of pain and disability. The moderate to strong correlation coefficients ranging from 0.69 to 0.78 for the disability measures and from 0.56 to 0.65 for the pain measures supported a good longitudinal external construct validity of the ODI.

Finally, our SeS population contains chronic LBP patients ranging from the moderately disabled patient to the surgical patient. Thus, the mean pain and disability scores are lower compared to a purely surgical population such as those reported by Fairbank et al. [59] and Fritzell et al. [60]. In other words, our

estimates apply to the majority of the chronic LBP patients but specific values may vary between subgroups.

Conclusion

The Danish ODI version 2.1 was translated, culturally adapted and psychometrically tested in two different LBP populations relevant for future clinical research. The ODI is a reliable and valid tool to assess pain related function when compared to well-established pain and disability scales. It is probably a more appropriate outcome measure in patients seen in the SeS due to a negligible floor effect and its ability to assess patients with a greater degree of disability and pain.

References

1. Guyatt GH, Feeny DH, Patrick DL (1993) Measuring health-related quality of life. Ann Intern Med 118:622–629
2. Kopec JA (2000) Measuring functional outcomes in persons with back pain: a review of back-specific questionnaires. Spine 25:3110–3114
3. Patrick DL, Deyo RA (1989) Generic and disease-specific measures in assessing health-status and quality of life. Med Care 27:S217–S232
4. Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine 25:3100–3103
5. Stucki G, Liang MH, Fossel AH, Katz JN (1995) Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. J Clin Epidemiol 48:1369–1378
6. Lurie J (2000) A review of generic health status measures in patients with low back pain. Spine 25:3125–3129
7. Suarez-Almazor ME, Kendall C, Johnson JA, Skeith K, Vincent D (2000) Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. Rheumatology (Oxford) 39:783–790
8. Grotle M, Brox JI, Vollestad NK (2005) Functional status and disability questionnaires: what do they assess?: a systematic review of back-specific outcome questionnaires. Spine 30:130–140
9. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B et al (1998) Outcome measures for low back pain research. A proposal for standardized use. Spine 23:2003–2013

10. Roland M, Fairbank J (2000) The Roland-Morris disability questionnaire and the Oswestry disability questionnaire. Spine 25:3115–3124
11. Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. Spine 25:2940–2952
12. Muller U, Roeder C, Dubs L, Duetz MS, Greenough CG (2004) Condition-specific outcome measures for low back pain. Part II: Scale construction. Eur Spine J 13:314–324
13. Muller U, Duetz MS, Roeder C, Greenough CG (2004) Condition-specific outcome measures for low back pain. Part I: Validation. Eur Spine J 13:301–313
14. Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain disability questionnaire. Physiotherapy 66:271–273
15. Baker D., Pynsent PB, Fairbank JC (1989) The Oswestry Disability Index revisited: its reliability, repeatability and validity, and a comparison with the St Thomas's Disability Index. In: Roland M, Jenner J (eds) Back pain: new approaches to rehabilitation and education. Manchester University Press, Manchester, pp 174–186
16. Hudson-Cook N, Tomes-Nicholson K, Breen A (1989) A revised Oswestry disability questionnaire. In: Roland M, Jenner J (eds) Back pain: new approaches to rehabilitation and education. Manchester University Press, Manchester, pp 187–204
17. Oswestry Disability Index homepage. http://www.ortho-surg.org.uk/odi/ (accessed on: 20-1-2006)
18. Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 46:1417–1432
19. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25:3186–3191
20. Malmros B, Mortensen L, Jensen MB, Charles P (1998) Positive effects of physiotherapy on chronic pain and performance in osteoporosis. Osteoporos Int 8:215–221
21. Malmros B, Jensen MB, Charles P, Mortensen LS (1999) Effect of specific physiotherapy on chronic pain, functional level and quality of life in osteoporosis. A prospective randomized single-blind placebo-controlled study (in Danish). Ugeskr Laeger 161:4636–4641
22. Christensen TH, Bliddal H, Hansen SE, Jensen EM, Jensen H, Jensen R et al (1993) Severe low-back pain. I: Clinical assessment of two weeks conservative therapy. Scand J Rheumatol 22:25–29
23. Manniche C, Asmussen K, Lauritsen B, Vinterberg H, Kreiner S, Jordan A (1994) Low back pain rating scale: validation of a tool for assessment of low back pain. Pain 57:317–326
24. Bjorner JB, Thunedborg K, Kristensen TS, Modvig J, Bech P (1998) The Danish SF-36 Health Survey: translation and preliminary validity studies. J Clin Epidemiol 51:991–999
25. Lonnberg F (1997) The management of back problems among the population. I. Contact patterns and therapeutic routines (in Danish). Ugeskr Laeger 159:2207–2214
26. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB (1995) Assessing health-related quality of life in patients with sciatica. Spine 20:1899–1908
27. Albert HB, Jensen AM, Dahl D, Rasmussen MN (2003) Criteria validation of the Roland Morris questionnaire. A Danish translation of the international scale for the assessment of functional level in patients with low back pain and sciatica (in Danish). Ugeskr Laeger 165:1875–1880
28. Bjorner JB, Kreiner S, Ware JE, Damsgaard MT, Bech P (1998) Differential item functioning in the Danish translation of the SF-36. J Clin Epidemiol 51:1189–1202
29. Bjorner JB, Damsgaard MT, Watt T, Groenvold M (1998) Tests of data quality, scaling assumptions, and reliability of the Danish SF-36. J Clin Epidemiol 51:1001–1011
30. Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H (1999) Capturing the patient's view of change as a clinical outcome measure. JAMA 282:1157–1162
31. Guyatt GH, Berman LB, Townsend M, Taylor DW (1985) Should study subjects see their previous responses. J Chronic Dis 38:1003–1007
32. Guyatt GH, Townsend M, Keller JL, Singer J (1989) Should study subjects see their previous responses—data from a randomized control trial. J Clin Epidemiol 42:913–920
33. Coste J, Delecoeuillerie G, Cohen dL, Le Parc JM, Paolaggi JB (1994) Clinical course and prognostic factors in acute low back pain: an inception cohort study in primary care practice. BMJ 308:577–580
34. Roland MO, Morrell DC, Morris RW (1983) Can general practitioners predict the outcome of episodes of back pain? Br Med J (Clin Res Ed) 286:523–525
35. Norman GR, Stratford P, Regehr G (1997) Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 50:869–879
36. Bjorner JB, Damsgaard MT, Watt T, Bech P, Rasmussen NK, Modvig J et al (1997) Danish Manual to the SF36. LIF, Lægemiddelindutriforeningen
37. Streiner DL, Norman GR (2003) Health measurment scales. A practical guide to their development and use. Oxford Medical Publications, Oxford
38. Scientific Advisory Committee (1995) Instrument Review Criteria. Med Outcomes Trust Bull 3:I–IV
39. Deyo RA, Diehr P, Patrick DL (1991) Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials 12:142S–158S
40. Fayers PM, Machin D (2000) Mulit-item scales. In: Fayers PM, Machin D (eds) Quality of life. Assessment, analysis and interpretation. Wiley, New York, pp 72–90
41. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310
42. Bland JM, Altman DG (2003) Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol 22:85–93
43. Bland JM, Altman DG (1995) Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 346:1085–1087
44. McHorney CA, Tarlov AR (1995) Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 4:293–307
45. Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther 82:8–24
46. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL et al (1995) The Quebec back pain disability scale. Measurement properties. Spine 20:341–352
47. Leclaire R, Blier F, Fortin L, Proulx R (1997) A cross-sectional study comparing the Oswestry and Roland-Morris functional disability scales in two populations of patients with low back pain of different levels of severity. Spine 22:68–71

48. Stratford PW, Binkley JM (2000) A comparison study of the back pain functional scale and Roland Morris questionnaire. North American Orthopaedic Rehabilitation Research Network. J Rheumatol 27:1928–1936

49. Kirshner B, Guyatt G (1985) A methodological framework for assessing health indices. J Chronic Dis 38:27–36

50. de Vet HC, Terwee CB, Bouter LM (2003) Current challenges in clinimetrics. J Clin Epidemiol 56:1137–1141

51. Mannion AF, Junge A, Fairbank JC, Dvorak J, Grob D (2005) Development of a German version of the Oswestry Disability Index. Part 1: cross-cultural adaptation, reliability, and validity. Eur Spine J 15(1):55–65

52. Fisher K, Johnston M (1997) Validation of the Oswestry low back pain disability questionnaire, its sensitivity as a measure of change following treatment and its relationship with other aspects of the chronic pain experience. Physiother Theory Prac 13:67–80

53. Hsieh CY, Phillips RB, Adams AH, Pope MH (1992) Functional outcomes of low back pain: comparison of four treatment groups in a randomized controlled trial. J Manipulative Physiol Ther 15:4–9

54. Boscainos PJ, Sapkas G, Stilianessi E, Prouskas K, Papadakis SA (2003) Greek versions of the Oswestry and Roland-Morris disability questionnaires. Clin Orthop 411:40–53

55. Grotle M, Brox JI, Vollestad NK (2003) Cross-cultural adaptation of the Norwegian versions of the Roland-Morris disability questionnaire and the Oswestry disability index. J Rehabil Med 35:241–247

56. Fritz JM, Irrgang JJ (2001) A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec back pain disability scale. Phys Ther 81:776–788

57. Stratford PW, Binkley JM (1999) Applying the results of self-report measures to individual patients: an example using the Roland-Morris questionnaire. J Orthop Sports Phys Ther 29:232–239

58. Hagg O, Fritzell P, Nordwall A (2003) The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J 12:12–20

59. Fairbank J, Frost H, Wilson-MacDonald J, Yu LM, Barker K, Collins R (2005) Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. BMJ 330:1233

60. Fritzell P, Hagg O, Wessberg P, Nordwall A (2002) Chronic low back pain and fusion: a comparison of three surgical techniques: a prospective multicenter randomized study from the Swedish lumbar spine study group. Spine 27:1131–1141