

Urs Müller
Christoph Röder
Charles G. Greenough

Back related outcome assessment instruments

Received: 11 October 2005
Accepted: 27 October 2005
Published online: 16 November 2005
© Springer-Verlag 2005

U. Müller (✉) · C. Röder
Institute for Evaluative Research in
Orthopaedic Surgery, University of Bern,
Stauffacherstrasse 78, 3014 Bern,
Switzerland
E-mail: urs.mueller@MEMcenter.unibe.ch
Tel.: +41-31-6315942
Fax: +41-31-6315931

C. G. Greenough
Middlesbrough General Hospital,
Middlesbrough, UK

Abstract A literature review of the most widely used condition specific, self administered assessment questionnaires for low back pain had been undertaken. General and historic aspects, reliability, responsiveness and minimum clinically important difference, external validity, floor and ceiling effects, and available languages were analysed. These criteria, however, are only part of the consideration. Of similar importance are the content, wording of questions and answers in each of the six questionnaires and an analysis of the different score results. The

issue of score bias is discussed and suggestions are given in order to increase the construct validity in the practical use of the individual questionnaires.

Keywords Spine · Outcome · Assessment · Review · Low back pain

Introduction

The evaluation of therapies for low back pain requires consideration of a number of variables. It has been recommended that for a full evaluation a condition specific disability measure, a general health measure (e.g. EQ-5D [20], WHODAS II [61], SF36 [62]), a pain measure (e.g. VAS [32]), a satisfaction measure, and a measure of employment should be considered [4]. For further technical details the reader is referred to two recent articles on the same subject [43, 44]. The present article aims to provide a more practical approach on the five most frequently used back related specific outcome scales.

The application of a widely used tool allows comparisons between the study group and other populations. All currently available measures have flaws or restrictions regarding their construction, validation or application. In the absence of an ideal instrument [5], the

choice of a commonly used measurement tool may be considered reasonable.

The present article aims to provide a more practical approach of the five most frequent used back related specific outcome scales. Further technical details are available in the previous publications [43, 44].

Methods

Eighty two back related questionnaires were identified reviewing the major medical databases. The ten most frequently used questionnaires were analysed based on various criteria such as general characteristics, external validity, internal consistency, responsiveness to changes, floor and ceiling effect, question focus, offered answers, item masking score bias, item weighting score bias and cross contamination score bias [43, 44]. This study comprises a short and more practical version of the

mentioned review summarising the five most widely used LBP assessment instruments. General measures such as the SF36 [62] as well as isolated pain measures were not included in the investigation.

History and technical validation concepts:

A number of specific areas were examined, defined below. More specific information is available in the literature [4, 48, 58].

General Characteristics: The population from which the score was developed, number of items, item scoring, information about subscales, and a brief description of the domains are provided.

Reliability: There are a variety of ways of examining the reproducibility of a measure administered on different occasions. Test-retest reliability is the most important. It is measured best by using tests of agreement such as Kappa [3, 42, 58]. The Pearson correlation coefficient [3] is a measure of correlation and although commonly used is a less precise measure. Pearson correlation values should exceed 0.8 and kappa values should exceed 0.5. Another measure is the Bland Altman plot [3]. This describes the spread of the score values within the same individuals between the test and the retest examination and provides a 95% confidence interval.

Internal Consistency: Measures of internal consistency are based on a single administration of the outcome measure. If the outcome measure has a relatively large number of items addressing the same dimension such as measures of physical function, it is reasonable to expect that the scores on each item would be correlated with the scores on all other items. Thus, if the internal consistency is low, the different items should not be summed, because they measure different domains. Internal consistency is predominantly measured by Cronbach's alpha correlations [10]. Values above 0.8 are acceptable.

Responsiveness to changes: The minimum clinically important difference (MCID) is the value of the change in the score which equates to the smallest change in the condition of interest the patient perceives as beneficial. Responsiveness can also be evaluated using the receiver operating characteristic (ROC) curve which is constructed by calculating the sensitivity (true positive rate) and specificity (true negative rate) of the cut-off point for each of the possible score values [58]. An index of the "goodness" of the questionnaire is the area under this curve (AUC), which is usually abbreviated as D' . A poorly discriminating questionnaire has an area of 0.5 and a perfect test has an AUC of 1.0 [58].

External validation: Comparison of a new score with the existing scores allows the assessment of its performance against known measures, particularly in the

selection of measurement domains, responsiveness and floor and ceiling effects.

Floor and Ceiling effect: Floor and ceiling effects describe the percentage of clients which have maximal or minimal points in the score [4, 58]. Here the measure is inefficient in discriminating between subjects. A similar problem occurs when the results are skewed in a certain region. Floor and ceiling effects may be observed if a measure developed in one population, e.g. severely disabled subjects in a pain clinic, is used in a very different population, e.g. attendees in primary care.

Apart from the technical aspects of the validation process, the softer aspects of scale construction are of major importance and are mainly underestimated. The following definitions were employed

Question focus: Each question should have one single target (e.g. "Do you have pain in the groin?") and they should be easy to understand and unambiguous [48, 58]. This aspect was recognised in most of the questionnaires. Questions should follow a logical structure and potential sources of inaccuracy should be specifically examined. For example, gender differences may exist in questions such as "Do you have back pain when doing chores?" as in many households chores are not evenly distributed between men and women.

In the ICF classification [26], three dimensions are described: (1) impairments sub-divided into functions and structures) (2) activity and participation limitations and (3) environmental factors. Impairment is divided into (1a) impairments of body functions (e.g. sleep as a mental impairment, pain as a sensory impairment [50], blood pressure as a cardiovascular impairment, etc.) and (1b) impairments of body structures (e.g. cartilage damage, impairments of spinal cord or peripheral nerves, etc.). Activities and participation include activities (such as learning, walking, doing housework etc.) and participation includes e.g. relationships, religion etc. Environmental factors are health care system, food, climate, etc. All these four dimensions are part of an outcome assessment and are valuable indicators for the quality of life. The four dimensions can be asked within one questionnaire like in the low back outcome score (LBOS) [25] or different questionnaires are used to fulfil the ICF criteria as is the case in the NASS LSO questionnaire [11]. Asking all three dimensions in one questionnaire is usually preferred because the type of questions and answer scaling as well as the question flow can be harmonised and similar questions can be united in one. The importance of assessing all the four ICF dimensions is stressed but summing the answer results in questions that belong to different ICF classes or subclasses in order to get a final sum score is not recommended (see below).

Offered answers: Answers should be clear and the scale has to be comprehensive and disjunctive [48] (answers do not overlap and focus on one single issue). It is

helpful for the patient to use one single answer type throughout the questionnaire. Technically this issue is important for questions that will be grouped in order to get a sum score (see item weighting below).

Score: All answers of a questionnaire are given a certain numeric value. Adding up the various answer values results in a score. Sum scores facilitate comparison between patients or patient groups or between pre-operative and postoperative conditions. Summing the scores of individual questions can result in a number of sources of error.

Item masking score bias: This may be present if unrelated questions are summed in a single score [18, 29]. For example, if a person suffers from severe back pain (ICF group 1a) and marks 9 out of a maximum of 10 points but has almost unrestricted walking (ICF group 2) with 3 out of 10 points. This results in a sum score of 12 points. During follow-up the same person suffers from moderate back pain only and marks 4 out of 10 points. Meanwhile, walking is severely restricted (8 out of 10 points). The follow-up sum score remains unchanged at 12 points, which indicates a seemingly unchanged outcome. The sum score has masked a significant change. Masking is more likely to be present if a sum score is composed of items focusing on different domains. The effect has also been called score bias [17, 19]. In a questionnaire comprising a number of domains the use of sub-scores focused on only one dimension should be considered.

Item weighting score bias: Some measures allocate different weights to questions. For example, weighting the ability to do work with 10 points and walking with 3 points places much emphasis on the work item. If the ability to work changes slightly, the sum score is affected significantly. If the walking capacity is altered, the change will hardly influence the sum score. Often no rationale is presented for these different weights, which may lead to under or overestimation of certain outcome parameters. Further research on the importance that patients place on various activities may improve this weighting. Weighting problems can also arise if the sum score contains questions that relate to abilities that are of no relevance for certain patients such as doing chores for certain male individuals or sex life for some elderly people. By giving the answers of such questions a certain amount of points, we put weight on questions that are meaningless for some individuals.

Cross contamination score bias: This is present if answers can be influenced by other diseases. If the question is not properly phrased in relation to the symptoms of the addressed disease, co-lesions or comorbidities can alter the outcome. The question “do you have pain in your leg?” for example might produce a positive answer in a patient with severe radiating low back pain but degenerative hip disease might also produce a positive response.

Suggestions: suggestions are made for avoiding bias when using each of the five questionnaires. These suggestions should be carefully considered. Changes only in the calculation of the main and sub-scores will not change the content of the questionnaire. However, changes to the content of questions or answers will necessitate a new validation process before the changed questionnaire can be used.

None of the mentioned scale construction problems will be detected by internal or external validation and the same is the case with questions that do not offer disjunctive answers. These issues are structural problems that should be eliminated before validating a questionnaire. A proper outcome validation is hardly possible if one of the mentioned effects or a score bias is present. Note that the overview analysis of each questionnaire is always based on the latest available English version.

The Oswestry disability index (ODI)

The ODI was initiated in 1976 and version 2.0 [21] is recommended [1, 52] for use. The administration is easy. A slightly modified ODI is used in the NASS [11] questionnaire (see below). The ODI and the Roland Morris questionnaire are the most thoroughly validated questionnaires [23, 24, 28, 31, 36] and have a good reliability and internal consistency. The ODI responsiveness seems to be acceptable but not as good as that of the Roland Morris score [12, 56]. Nevertheless, it can also be used for cervical problems [64]. However, Taylor et al [59] found that the ODI is more sensitive to patients who had improved and is less sensitive for patients whose condition remained unchanged. This fact is closely related to its floor effect [47]. External validation to different questionnaires shows neither an advantage nor a disadvantage of the ODI [12, 24, 25, 31, 35, 36, 39, 64] compared to other assessment tools. Slight item weighting, cross contamination and item masking bias are expected. The ODI is validated in English [22], German [40, 41], French [16], Finnish [27] and Greek [6]. Translations in several other languages do not appear to be validated.

Conclusion: The ODI is a simple, well analysed questionnaire widely used in comparative studies. The ODI is not recommended to use for the assessment of preventive measures because of its floor effect. If used in seriously diseased collectives the ODI can be recommended for in-depth scientific research studies if combined with the NASS (see below).

The Roland Morris disability questionnaire (RMDQ)

The RMDQ was derived from the Sickness Impact Profile, 24 out of 136 items were selected and published

1983 [53]. The RMDQ is short and simple to administer and widely used. Despite several published modification proposals [15, 47, 57, 60], the original version of the RMDQ is favoured by an international expert group [14]. A good reliability [13, 33, 53] and internal consistency [31, 39, 51, 57] is reported. The RMDQ seems to detect changes over time slightly better than the Oswestry scale [12, 39] especially in patients having mild disabilities [1], provided that the initial score is in the range between 4 and 20 [55]. Nevertheless, the RMDQ lacks a ceiling effect [55] in severely affected individuals. The RMDQ is validated in English [53], French [9], German [63], Greek [7], Portuguese [45], Spanish [37], Swedish [33], Turkish [38] and is available in several other invalidated language versions. A significant advantage is that the RMDQ questions are straightforward and consistently related to the back, offering simple Yes/No answers. Thus the RMDQ does not display uncontrolled item weighting or cross contamination.

Conclusion: The RMDQ is the low back assessment tool of choice if combined with a general health assessment and used in a mild to moderately affected low back pain collective.

The low back outcome score (LBOS)

The questionnaire was first published in 1992. The answering possibilities of each item are scaled. For pain an eleven point scale ranging from “no pain” to “maximum pain possible” is used. The other items use a four point scaled text. The total score gives different weights to different questions. Although this weighting has to be conducted with care, the administration is easy. All aspects of the ICF classification [26] are considered. Slight extensions to the questionnaire have been reported [30]. The LBOS is reliable [30] and provides a good internal consistency [30]. The MCID is reported to be 7.5 points [34]. The LBOS correlates well with the ODI ($r=0.87$) [59]. A ceiling effect was reported in a non compensated low back pain population [25]. The English version is validated and non validated German and a Spanish version are available. The questionnaire considers the different dimensions of the ICF classification, thus summing to provide a total sum score may lead to item masking. Nevertheless, the consideration of the different ICF dimensions provides a multidimensional assessment of the patient and is a strength of the LBOS. There is a possibility of item weighting bias and cross contamination bias. It is suggested that the sum score should not be used but that sub-scores for each of the dimensions (pain functional pain, and ability items) are to be used instead.

Conclusion: If the above suggestions are considered, the LBOS is useful because it is short, covers the important aspects of the treatment outcome and clearly discriminates between pain and disability. Thus the

LBOS can be suggested if only a short assessment (such as for registries or for the private use of one surgeon’s patient collective) is envisioned. Unfortunately only the English version is validated.

The Quebec back pain disability scale (QBPDS)

The QBPDS was validated on a back pain population and published in 1995 [36]. The questions were designed using a conceptual model. Item selection was done using factor analysis for 46 disability items. Twenty items were selected and tested for reliability. The QBPDS measures only functional disability (self care, walking, sitting, standing, lifting, sport, stairs and housework) and sleep. Pain has to be evaluated with other tools. Items about social life, sex life and the need for help are not included. Nevertheless, the items give a comprehensive view of the patient’s disabilities because easy as well as more difficult functional abilities are asked. The administration is easy. The MCID lies between 14 and 19 points [12, 35] and the questionnaire seems to be as reliable as the ODI or the RMDQ. Floor and ceiling effects are not reported and the QBPDS is validated in English, French [36] and Dutch [54]. The simple, on disability focused questions and the consistently offered five answers are the strength of this questionnaire. Item masking is not present but uncontrolled item weighting and cross contamination might be present.

Conclusion: The questionnaire is well focused on disabilities and offers consistent answers that makes it an excellent disability assessment tool. If the QBPDS is combined with an independent pain assessment tool, it can be recommended for low back pain assessment. Unfortunately this questionnaire misses validated translations to other languages and is not as often used as the RMDQ or the ODI.

The NASS lumbar spine outcome (NASS LSO) assessment instrument

The NASS was first published by Daltroy et al. [11] and is based on a consensus of the North American Spine Society. It considers all aspects of the ICF classification, e.g. demographic data (age, sex, race, education and insurance information), medical history (co-morbidities, past surgeries etc.), body functions (pain, neurogenic symptoms etc.) and employment history. In the questionnaire construct the SF36, a modified ODI, and a modified employment assessment published by Bigos et al. [2] are included. For the follow-up assessment the questionnaire is slightly modified. It takes 20–25 min to fill in the form and the scoring is complex. As all the other mentioned questionnaires, the NASS is reliable and shows a good internal consistency [11]. Data on the

MCID as well as research on floor and ceiling effects are not available in the literature. The NASS LSO is validated in English [11], German [49] and Italian [46]. The NASS LSO baseline questionnaire is by far the longest of all analysed instruments with its 62 main questions. Pain is a very dominant factor assessed with different methods. The pain questions themselves (ODI), however, do not explicitly relate pain and disability to the back. Instead, a pain locator (picture where the client can mark the location of pain) is offered where the various painful body regions can be clearly indicated. The pain locator is a useful instrument for drawing a precise pattern of pain distribution that reduces undetected cross contamination effects. Location unspecific questions can be set in relation to the location marked in the pain locator. This increases the value of the ODI significantly. Because of the use of different questionnaires, no homogenous way of posing questions exists. As a consequence, certain questions are asked twice. Item masking is present if all dimensions of the NASS are summed but this does not make sense since sub-scores are calculated due to the original questionnaires.

Conclusion: The NASS enables an extensive outcome assessment where pain is very dominant in the low back specific questions. In contrast to the ODI, pain can be indicated precisely (pain drawing), so that changes related to the pain origin become visible. Time consumption, administration and statistics limit the use of the NASS to scientists who can take advantage from a scientific environment (statistician, study nurse or equivalent, etc.).

Discussion

No gold standard exists for outcome assessment in low back therapies. Each of the studied questionnaires has its advantages or disadvantages. However, all of the questionnaires have been validated and have proved to be reliable and consistent.

Scientists who stress the importance of comparing data with other collectives should use the RMDQ or the ODI where the latter seems to be more discriminative in patients having severe low back pain while the RMDQ seems to be more sensitive in the less severely affected patients. Those two questionnaires have to be combined with a general health assessment tool like the EuroQuol (no fee) [8, 20], the WHODAS II (no fee) [61] or SF36/SF12 (fee to pay) [62].

Spine specialists who want to assess their own patient collectives should favour an easy to administer questionnaire which gives them a quick and good overview. In this case the LBOS can be suggested. Ideally LBOS is combined with the five questions of the EuroQuol [8, 20]. Unfortunately the LBOS is not validated in languages other than English.

Scientists working together with a research department and not discouraged by the needed administrative effort should take advantage of the NASS LSO assessment instrument. Nevertheless, this instrument is available only in a few languages.

References

1. Baker D, Pynsent P, Fairbank J (1989) The Oswestry index revisited. In: Roland M, Jenner J (eds) *Back pain: new approaches to rehabilitation and education*. Manchester University Press, Manchester pp 174–186
2. Bigos SJ, Battie MC, Spengler DM, et al. (1991) A prospective study of work perceptions and psychosocial factors affecting the report of back injury. *Spine* 16(1):1–6
3. Bland J, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 8:307–310
4. Bombardier C (2000) Outcome assessment in the evaluation of treatment of spinal disorders. Summary and general recommendations. *Spine* 25:3100–3103
5. Bombardier C (2000) Spine focus issue introduction. *Spine* 25:3097–309
6. Boscainos P, Sapkas G, Stilianesi E (1999) Clinical relevance of specific parameters isolated within the Oswestry and Roland Morris functional disability index. *J Bone Joints Surg [Br]* 81(Suppl):239
7. Boscainos PJ, Sapkas G, Stilianessi E, et al. (2003) Greek versions of the Oswestry and Roland-Morris disability questionnaires. *Clin Orthop* 411:40–53
8. Brooks R (1996) EuroQol: the current state of play. *Health Policy* 37(1):53–72
9. Coste J, Le Parc JM, Berge E, et al. (1993) [French validation of a disability rating scale for the evaluation of low back pain (EIFEL questionnaire)]. *Rev Rhum Ed Fr* 60(5):335–341
10. Cronbach L (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334
11. Daltroy LH, Cats-Baril WL, Katz JN, et al (1996) The North American spine society lumbar spine outcome assessment instrument: reliability and validity tests. *Spine* 21(6):741–749
12. Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 82(1):8–24

13. Deyo R, Centor R (1986) Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 39:897–906
14. Deyo RA, Battie M, Beurskens AJ, et al. (1998) Outcome measures for low back pain research A proposal for standardized use. *Spine* 23(18):2003–2013
15. Dionne CE, Von Korff M, Koepsell TD, et al. (1999) A comparison of pain, functional limitations, and work status indices as outcome measures in back pain research. *Spine* 24(22):2339–2345
16. Dropy R, Marty M (1994) Quality-of-life indexes for assessment of low-back-pain. *Rev Rhum* 61:S44–S48
17. Dubs L (1999) Everything was done—patient still not satisfied: introduction to evidence-based surgery. *Swiss Surg* 5(4):160–166
18. Dubs L (2000) Der score bias. *Schweiz. Aertzeitung* 81(6):292–295
19. Dubs L (2000) The patient as expert—introduction to evidence-based orthopedics. *Z Orthop* 138(4):289–294
20. EuroquoI-Group (1990) EuroquoI—a new facility for the measurement of health-related quality of life. *Health Policy* 16:199–208
21. Fairbank J (1995) Use of Oswestry disability index (ODI). *Spine* 20(13):1535–1537
22. Fairbank JC, Couper J, Davies JB, et al. (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66(8):271–273
23. Fisher K, Johnson M (1997) Validation of the Oswestry low back pain questionnaire, its sensitivity as a measure of change following treatment and its relationship with other aspects of the chronic pain experience. *Physiother Theory Pract* 22:61–80
24. 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(2):776–778
25. Greenough CG, Fraser RD (1992) Assessment of outcome in patients with low-back pain. *Spine* 17(1):36–41
26. Grimby G, Smedby B (2001) ICF approved as the successor of ICIDH. *J Rehabil Med* 33–5:193–194
27. Gronblad M, Hupli M, Wennerstrand P, et al. (1993) Intercorrelation and test-retest reliability of the pain disability index (PDI) and the Oswestry disability questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain* 9(3):189–195
28. Gronblad M, Jarvinen E, Hurri H, et al. (1994) Relationship of the pain disability index (PDI) and the Oswestry disability questionnaire (ODQ) with three dynamic physical tests in a group of patients with chronic low-back and leg pain. *Clin J Pain* 10(3):197–203
29. Hefti F (1996) Numerische “Scores” zur Evaluation von Behandlungsergebnissen am Bewegungsapparat. *Schweiz. Aertzeitung* 77:118–25
30. Holt AE, Shaw NJ, Shetty A, et al. (2002) The reliability of the low back outcome score for back pain. *Spine* 27(2):206–210
31. Hsieh CY, Phillips RB, Adams AH, et al. (1992) Functional outcomes of low back pain: comparison of four treatment groups in a randomized controlled trial. *J Manipulative Physiol Ther* 15(1):4–9
32. Huskisson EC (1974) Measurement of pain. *Lancet* 2(7889):1127–1131
33. Johansson E, Lindberg P (1998) Subacute and chronic low back pain. Reliability and validity of a Swedish version of the Roland and Morris disability questionnaire. *Scand J Rehabil Med* 30(3):139–143
34. Kahtri H, Greenough CG (2002) Minimum clinically important difference in the low back outcome score. *Society of Back Pain Research, Leeds*
35. Kopec JA, Esdaile JM (1995) Functional disability scales for back pain. *Spine* 20(17):1943–1949
36. Kopec JA, Esdaile JM, Abrahamowicz M, et al. (1995) The Quebec back pain disability scale. Measurement properties. *Spine* 20(3):341–352
37. Kovacs FM, Llobera J, Gil Del Real MT, et al. (2002) Validation of the spanish version of the Roland–Morris questionnaire. *Spine* 27(5):538–542
38. Kucukdeveci AA, Tennant A, Elhan AH, et al. (2001) Validation of the Turkish version of the roland–morris disability questionnaire for use in low back pain. *Spine* 26(24):2738–2743
39. Leclaire R, Blier F, Fortin L et al (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(1):68–71
40. Mannion A, Junge A, Fairbank J, et al. (2005) Development of a German version of the Oswestry disability index. Part 1: Cross-cultural adaptation, reliability, and validity. (in press)
41. Mannion A, Junge A, Grob D, et al. (2005) Development of a German version of the Oswestry disability index. Part 2: Sensitivity. *Eur Spine* (in press)
42. McDowell I, Newell C (1996) The theoretical and technical foundations of health measurement. *Measuring health. A guide to rating scales and questionnaires*. Oxford University Press, Oxford pp 10–42
43. Mueller U, Duetz M, Roeder C, et al. (2004) Condition specific outcome measures for low back pain. Part I: Validation. *Eur Spine J* 13:301–313
44. Mueller U, Roeder C, Dubs L, et al (2004) Condition Specific Outcome Measures for low back pain Part II: Scale construction. *Eur Spine J* 13:314–324
45. Nusbaum L, Natour J, Ferraz MB, et al. (2001) Translation, adaptation and validation of the Roland–Morris questionnaire—Brazil Roland–Morris. *Brazilian Journal of Medical and Biological Research = Revista Brasileira de Pesquisas Medicas E Biologicas/Sociedade Brasileira de Biofisica* 34(2):203–210
46. Padua R, Padua L, Ceccarelli E, et al. (2001) Cross-cultural adaptation of the lumbar North American Spine Society questionnaire for Italian-speaking patients with lumbar spinal disease. *Spine* 26(15):E344–E347
47. Patrick DL, Deyo RA, Atlas SJ et al. (1995) Assessing health-related quality of life in patients with sciatica. *Spine* 20(17):1899–908-discussion 909
48. Porst R (2000) Question wording—Zur Formulierung von Fragebogen-Fragen. *ZUMA, Mannheim* 2:1–11
49. Pose B, Sangha O, Peters A, et al. (1999) Validation of the North American Spine Society Instrument for assessment of health status in patients with chronic backache. *Zeitschrift Fur Orthopadie Und Ihre Grenzgebiete* 137(5):437–441
50. Raspe HH (1997) Minimum requirements in medical expert assessment of occupational disability of patients with chronic nonspecific pain. *Versicherungsmedizin* 49(4):118–125
51. Roes I, Kohlmann T, Raspe H (1996) Measuring functional capacity in backache patients in rehabilitation: a comparison of standardized questionnaires 35(2):103–108
52. Roland M, Fairbank J (2000) The Roland–Morris disability questionnaire and the Oswestry disability index. *Spine* 25:3115–3124
53. Roland M, Morris R (1983) A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8(2):141–144

-
54. Schoppink LE, van Tulder MW, Koes BW, et al. (1996) Reliability and validity of the Dutch adaptation of the Quebec Back pain disability scale. *Phys Ther* 76(3):268–275
 55. Stratford PW, Binkley J, Solomon P, et al. (1996) Defining the minimum level of detectable change for the Roland–Morris questionnaire. *Physi Ther* 76(4):359–365 discussion 66–8
 56. Stratford PW, Binkley J, Solomon P, et al. (1994) Assessing change over time in patients with low back pain. *Physi Ther* 74(6):528–33
 57. Stratford PW, Binkley JM (1997) Measurement properties of the RM-18. A modified version of the Roland–Morris disability scale. *Spine* 22(20):2416–2421
 58. Streiner D, Normann G (1995) Health measurement scales. A Practical Guide to their Development and Use. Oxford Medical Publications 2 edn. ISBN 0-19-262670-1
 59. Taylor SJ, Taylor AE, Foy MA, et al. (1999) Responsiveness of common outcome measures for patients with low back pain. *Spine* 24(17):1805–1812
 60. Underwood MR, Barnett AG, Vickers MR (1999) Evaluation of two time-specific back pain outcome measures. *Spine* 24(11):1104–1112
 61. Van Tubergen A, Landewe R, Heuft-Dorenbosch L, et al. (2003) Assessment of disability with the World Health Organisation Disability Assessment Schedule II in patients with ankylosing spondylitis. *Ann Rheum Dis* 62(2):140–145
 62. Ware JEJ, Sherbourne CD. (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30(6):473–483
 63. Wiesinger GF, Nuhr M, Quittan M, et al. (1999) Cross-cultural adaptation of the Roland-Morris questionnaire for German-speaking patients with low back pain. *Spine* 24(11):1099–1103
 64. Zoega B, Karrholm J, Lind B (2000) Outcome scores in degenerative cervical disc surgery. *Eur Spine J* 9(2):137–143