40 Outcome Assessment in Spinal Surgery

Mathias Haefeli, Norbert Boos

Core Messages

- ✔ The evaluation of treatment modalities for spinal disorders by self-administered questionnaires has entered into clinical practice
- ✔ Functional and psychosocial aspects often exhibit a closer correlation with fair or poor outcome after spinal surgery than organ-specific symptoms and morphological alterations and must therefore be evaluated in outcome research
- ✔ The main subjects addressed by outcome tools are pain, disability, health-related quality of life and work status
- ✔ For more thorough investigations, psychosocial aspects, work-related parameters and fear avoidance behavior should additionally be assessed
- ✔ There are several standardized and validated questionnaires available
- ✔ Current research is trying to facilitate data assessment by developing short but reliable instruments

General Concepts of Outcome Assessment

The evaluation of treatment modalities in spinal orders by self-administered assessment tools has become standard in most institutions. In many fields of medicine and particularly in spinal surgery, it has become evident that treatment outcome is influenced by a large variety of non-morphological factors [100]. Psychosocial aspects and work-related factors often exhibit a higher predictive value than pathomorphological and surgical aspects [47]. Therefore, it has become apparent that a meaningful outcome assessment should consider most of these confounding variables, which, however, is not always possible to achieve in a busy clinical practice. The **minimal data set** that should be collected consists of:

- pain
- disability
- quality of life
- work status

Several criteria should be considered when data assessment is performed by **self-rating questionnaires:**

- comparability
- validity
- availability
- scale characteristics

When a comparison between treatment groups is chosen in a study, the criteria of comparability of a questionnaire must be defined. If the results are to be com-

pared with a control group out of the literature, an identical questionnaire must be used.

Validity [2] is the degree to which an instrument measures what it is intended to measure. It is the most important quality of a questionnaire and there are different types of validity. A **questionnaire ideally should fulfill:**

- **content validity**, i.e. the extent to which the instruments include the domain of the target phenomenon
- **criterion validity**, i.e. extent of agreement when comparing with a "gold standard"
- **construct validity**, i.e. extent to which the instrument corresponds to theoretical concepts of the target phenomenon

Most of the questionnaires are developed for the English language. If these tools are used in non-English speaking countries, these versions should ideally be translated and validated first for the used language (availability). Several rules should be considered in this process of **cross-cultural adaptation** [13]. According to this, such a process should start with at least two forward translations into the target language. In a second step a synthesis of the two translations should be done before performing at least two back translations in the next step. After a consolidation of all versions of the instruments resulting from the first three

Table 1. Outcome tools in spinal surgery

Topic	Tool	Available languages (validated versions only)
Pain	**VAS/GRS/NRS/VRS**	
Disability	**RMDQ**	**English [131]** French [38] German [156] Greek [24] Portuguese [115] Spanish [88] Swedish [82] Turkish [90]
	ODI	English [50] Finnish [63] French [157] German [11, 101, 102] Greek [24]
	NASS-Q	English [39] German [123] Italian [119]
	FAQH	German [86]
	NDI	English [145] French [157] Swedish [3]
	NPDI	English [154] French [157] Turkish [20]
Quality of life	**WHOQOL-100/-Bref**	**www.who.int**
	SF-36/-12/-8	www.sf36.com
	EQ-5D	www.euroqol.org
	SRS-22/-30	English: www.srs.org Spanish [10]
Fear avoidance behavior	**FABQ**	**English [149]** German [121, 138]
Core item tools	**Low back pain**	**English [41]** German [99]
	Neck pain	English [155]

steps by an expert committee, a testing of the instrument and further refinements have to be done.

Since there are many aspects influencing outcome of spinal surgery, a well designed questionnaire will include different standardized and validated tools to cover these different fields (scale characteristics).

A questionnaire should be comparable, valid and comprehensive

A broad range of outcome tools are available (Table 1), of which only a limited number are frequently used. In the following, the most important questionnaires in the field of spinal surgery are briefly discussed including pain assessment, disability, quality of life and work assessment. Presented in regard to their strengths and weaknesses and their best feasible clinical setting, this survey should enable the best possible decision when searching for a self-administered assessment tool in spinal surgery.

Pain

General Aspects

Back pain is one of the most frequent reasons for spinal surgery and therefore pain relief is the major aim in the vast majority of cases. Pre- and postoperative assessment of pain and pain relief serves to evaluate the effectiveness of a specific therapy [68]. However, some important findings of the past two decades of research have to be kept in mind when the gathering and interpreting of such data is intended. As perception of pain may differ within a time period, recent studies have mentioned that it is more valuable to ask patients to rate their "usual" pain on average over a past short period of time, e.g. 1 week, than to ask for "current" pain at the specific time of completion of the questionnaire [21, 22, 147]. Posing such questions relies on the assumption that patients are able to accurately recall their pain levels in a past period of time. Whether or not this is reliable is controversial. Whereas some studies find it unreliable to assess pain retrospectively [40, 94–96], others report acceptable levels of validity up to a 3 months recall period [21, 139, 146]. It has been found that pain is usually overestimated when the actual intensity of pain is higher and underestimated when it is lower [30, 45, 94–96]. Moreover, Haas et al. [66] found that pain and disability recall became more and more influenced by present pain and disability during a period of 1 year while the influence of actual relief and pain and disability reporting at the initial consultation decreased. On the other hand, Von Korff et al. [146] stated that recall of chronic pain in terms of its average intensity, interference with activities (disability due to pain), number of days with pain and number of days with activity limitation, leads to acceptable validity levels.

The objective assessment of pain for outcome research remains controversial

When assessing pain in the context of a spinal intervention, it is necessary to use some kind of pain recall when not using "current pain" as the test parameter as discussed above. Based on the literature, it is justifiable to use short time periods of pain and disability recall for comparison of patients' pain status. The interpretation of whether or not a statistically significant change in pain corresponds to a significant clinical change remains challenging and requires further research [12]. Similarly, the definition of a threshold for a significant clinical change needs to be explored.

Short time periods of pain recall are superior to current pain assessment

Pain Duration

There are different definitions of chronic back pain. Nachemson et al. [112] defined it in 1984 as a period of at least 3 months with persistent pain. Von Korff et al. [147] defined it in 1996 as back pain which has to be present on at least half of the days during 1 year. Raspe et al. [127] investigated 40 epidemiological/ther-

apeutic studies between 1998 and 2000 with regard to the definitions of chronic back pain that were used. Finding periods between 4 weeks and more than 1 year of persistent pain, he showed that there is no consensus about this definition.

Pain Affect

The experience of pain is subjective, complicating an objective assessment

Pain can be described in terms of the intensity but also in terms of its effect on the individual. Pain intensity describes **how much** a patient is in pain, whereas **pain affect** describes the "degree of emotional arousal or changes in action readiness caused by the sensory experience of pain" [146]. It has been shown that pain intensity may quite easily be described by most patients and that different methods of measuring pain intensity showed high intercorrelation [80, 81]. Contrary to these findings, alternative methods of pain affect assessment did not intercorrelate as highly as those of pain intensity, making the utilization of this part of pain characterization more complicated [109, 110].

Instruments

Visual Analogue Scale (VAS)/Graphic Rating Scale (GRS)

A visual analogue scale (VAS) consists of a straight line with endpoints

The VAS consists of a straight line with the endpoints defining extreme limits such as "no pain at all" and "pain as bad as it could be" (Fig. 1) [2]. The patient is asked to mark his or her pain level on the line between the two endpoints, the distance between "no pain at all" and the mark defining the subject's pain. This tool was first used in psychology by Freyd in 1923 [56].

A graphic rating scale (GRS) adds descriptive terms or a numerical scale

A GRS additionally uses descriptive terms such as "mild", "moderate", "severe" or a numerical scale (Fig. 2) [2]. A line length of 10 or 15 cm showed the smallest measurement error compared to 5 and 20 cm versions and seems to be most convenient for respondents [135].

Scott and Huskisson demonstrated that the configuration of a graphic rating scale may influence the distribution pattern of the answers [134]. Moreover, they showed that the experience of patients with this tool influenced the outcome. While patients who had no experience with a graphic rating scale with numbers of 1 – 20 underneath the line showed a preference for the numbers 10 and 15, sub-

Figure 1. Visual analogue scale (VAS)

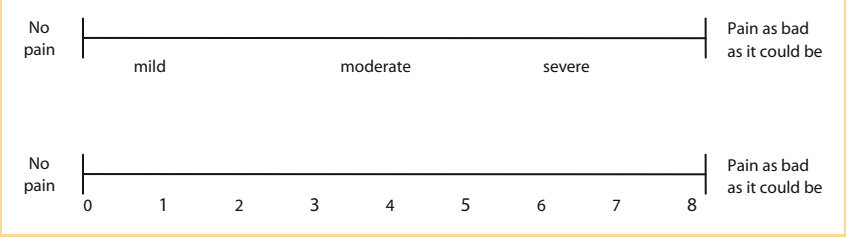

Figure 2. Examples of graphic rating scales (GRS)

jects who were experienced in the use ignored the numbered scale and showed no preferences and, therefore, a nearly uniform distribution of the answers. Analogue observations were made with descriptive terms. In several studies, VAS and GRS have demonstrated to be sensitive to treatment effects [80, 83, 89, 135]. They were found to correlate positively with other self-reporting measures of pain intensity [80, 89]. In addition, differences in pain intensity measured at two different points of time by VAS represent the real difference in magnitude of pain, which seems to be the major advantage of this tool compared to the others [125, 126].

VAS indicate real differences between measurements at two points of time

As the distance between "no pain" and the patient-made mark has to be measured, scoring is more time consuming and susceptible to measurement errors than a rating scale for example. Hence, a mechanical VAS has been developed where subjects position a slider on a linear pain-scale instead of marking a cross on a drawn line. Several studies have shown this system to be strongly associated with the original VAS [36, 62]. Moreover, it has been shown that a mechanical VAS exhibits a good test-retest reliability and appears to have ratio qualities [146].

Mechanical visual analogue scales are easy to handle

Besides the disadvantage mentioned above, the VAS seems to be more difficult to understand than other measurement methods and, hence, more susceptible to misinterpretations or "zero values". This is particularly true in elderly patients [37, 80, 89]. In conclusion, the VAS, mechanical VAS and GRS are valuable instruments for assessment of pain intensity and changes due to therapy when respondents are given good instructions and one bears in mind the limitations [37, 134].

Numerical Rating Scale (NRS)

When using an NRS, patients are asked to circle the number between 0 – 10, 0 – 20 or 0 – 100 that best fits their pain intensity [2]. Zero usually represents "no pain at all" whereas the upper limit represents "the worst pain ever possible". In contrast to the VAS/GRS, only the numbers are valuable answers, meaning that there are only 11 possible answers in a 0 – 10, 21 in a 0 – 21 and 101 in a 0 – 100 point NRS. The NRS allows a less subtle distinction of pain levels compared to VAS/GRS, where there is theoretically an unlimited number of possible answers.

The NRS allows less subtle distinction of pain levels compared to VAS and GRS

The NRS has shown high correlations with other pain assessment tools in several studies [80, 89]. The feasibility of its use and good compliance have also been proven [37, 52]. As it is easily possible to administer NRS verbally, it can be used in **telephone interviews** [146]. On the other hand, results cannot necessarily be treated as ratio data as is possible in VAS/GRS [124].

Verbal Rating Scale (VRS)

In a verbal rating scale, **adjectives** are used to **describe** different levels of **pain** [2]. The respondent is asked to mark the adjective which fits best to the pain intensity. Also in the VAS two endpoints such as "no pain at all" and "extremely intense pain" should be defined. Between these extremes different adjectives are placed which describe different pain intensity levels. Mostly, 4- to 6-point VRS are used in clinical trials. A different form of VRS is the behavioral rating scale, where different pain levels are described by sentences including behavioral parameters [32].

As well as VAS, VRS have been shown to strongly correlate with other pain assessment tools [80, 89, 118]. Compared to other instruments, respondent's compliance is often as good or even better even though subjects must be familiar with reading the entire list before answering [37, 80]. However, due to the limited number of possible response categories some patients may have problems defin-

Verbal rating scales are less suited to assessing changes in pain intensity and interindividual comparisons

ing which answer fits best to their pain situation. Moreover, the intervals between different adjectives describing pain may not be equal or may be interpreted differently by respondents. Thus, interpretation of a VRS does not allow conclusions to be drawn on the magnitude of a change in pain intensity between two assessments, for example, pre- and postoperatively, and interrespondent comparison is problematic.

Disability

General Aspects

Back and neck problems often lead to disability in daily activities due to pain or deformity. Several tools have been developed in respect of this aspect of spinal disorders. In the field of low back pain the most commonly used questionnaires are the **Roland & Morris Disability Questionnaire** (RMDQ) and the **Oswestry Disability Index** (ODI). Both are available in several languages and have proven good internal consistency and test-retest reliability [76, 130, 141]. The North American Spine Society Lumbar Spine Outcome Assessment Instrument (NASS LSO) and the Hannover Functional Ability Questionnaire (HFAQ) are two other disability questionnaires, the latter only existing for the German language. In the field of neck pain the **Neck Disability Index** (NDI) [145] and the **Neck Pain and Disability Index** (NPDI) [154] are the most commonly used tools.

Instruments

Roland & Morris Disability Questionnaire (RMDQ)

This tool was developed by Roland and Morris in 1983 [131]. It is frequently used and has been validated for the English, French [38], Swedish [82], German [49, 156], Turkish [90], Spanish [88], Portuguese [115], Japanese [142], Norwegian [64] and Greek [24] languages. Twenty-four questions from the Sickness Impact Profile (SIP) [17] were selected and added with the phrase "because of my back", leaving it open whether an impairment is due to pain or disability. The answering possibilities are **dichotomous** (yes/no) and, therefore, filling in the questionnaire requires little time and is easy to do. On the other hand, this might leave subtle changes in the abilities unrecognized. In contrast to the ODI, sex life is not included, and similar to the ODI neurological leg deficits are not addressed.

The RMDQ is more sensitive than the ODI in detecting changes over time

Compared to the ODI, the RMDQ is regarded as being more sensitive in detecting changes over time [19, 76, 140]. This is especially true in patients with minor disabilities. For patients with severe disabilities the RMDQ seems to perform worse than the ODI [19, 130]. Internal consistency has been shown to be equal [91, 129] or slightly superior to the ODI [76, 87].

Oswestry Disability Index (ODI)

This tool was developed by Fairbank et al. [50] in 1980. It is used frequently and has been validated in English, German [11, 101, 102], Danish [98], Finnish [63], Norwegian [64], French [43], and Greek [24]. It contains ten items about pain level and interference with physical activities, sleeping, self-care, sex life, social life and traveling. Each question offers six answers, which allows the assessment of subtle differences of disability.

The ODI performs better in patients with severe back-related disability than the RMDQ

In contrast to the RMDQ, respondents are only given an introduction, which points out that the questionnaire is about back pain, instead of being reminded in every question about the main topic. This might lead to misunderstanding if

patients are suffering from pain of different origin. Other differences between the ODI and the RMDQ are described above.

NASS Questionnaire

This questionnaire was designed by the North American Spine Society in the early 1990s [39]. Validated German [123] and Italian [119] versions are available. It is based on the ODI, from which a selection of items was adopted and adapted. Questions from the SF-36 and the Health Survey Questionnaire were added to allow the assessment of a broad patient profile.

The NASS is based on the ODI, the SF-36 and the Health Survey Questionnaire

Hannover Functional Ability Questionnaire (HFAQ)

The back pain version of the HFAQ belongs to a series of self-administered questionnaires about **functional limitations in the daily life** of patients suffering from musculoskeletal disorders [86]. It consists of 12 questions about abilities in daily activities such as lifting a heavy item. Each ability must be graded by "yes", "yes, but with trouble" or "no, or only with help". The HFAQ has been frequently used in German-speaking areas.

The HFAQ has been compared with different other disability questionnaires. Roese et al. [129] found it to be as feasible, practicable, valid and reliable as the RMDQ. Haase et al. [67] compared it with the physical functioning domain of the MOS SF-36 in a rehabilitation collective. In 4.3 % of all respondents, they found confusion with positive and negative ratings in the SF-36 subscale, while no similar problems could be detected in the HFAQ, and it was argued that the SF-36 seems to be more valuable for use in the ambulant medical sectors than in a rehabilitation setting. Finally, Schochat et al. [133] compared it with the NASS questionnaire in a rehabilitation collective and found high correlations indicating high concurrent validity. However, both questionnaires were not able to detect changes in the "impairment" domains after a 3-week period, again indicating that these instruments might be more suitable in short-term outcome research than in the field of rehabilitation.

The HFAQ is more applicable for short-term outcome research

Neck Disability Index (NDI)

The NDI is a ten-item questionnaire derived from the ODI [145]. It is designed to assess **neck pain and disability** and consists of ten six-point Likert scales covering the following ten sections: Pain intensity, Personal care (washing, dressing, etc.), Lifting, Reading, Headaches, Concentration, Work, Driving, Sleeping, Recreation. Each question is rated from zero to five points, allowing a maximum of 50 points. The score achieved by the patient is divided by the maximum possible and multiplied by 100 to get a percentage score of the possible total. If one section is missed, the maximum score of 50 points is reduced by 5 points.

The NDI has been used in different populations and has been validated against multiple measures of function and pain [122]. Besides the original English version, a validated form for the French [157] and Swedish [3] languages is available.

The NDI assesses neck pain and related disability by ten six-point Likert scales

Neck Pain and Disability Index (NPDI)

The NPDI was introduced in 1999 and consists of 20 VAS items assessing **neck pain and linked disability** [154]. Each VAS ranges from zero (normal function) to five (worst possible situation). It is divided into four sections: Neck problems, Pain intensity, Effect of neck pain on emotional and cognitive status, Interference of neck pain with daily activities.

The NPDI responds well to changes in neck pain and disability

It was found to show high internal consistency [154] and proved to have high test-retest reliability and a good response to changes in pain perception following treatment [61]. Besides the original validated English version, validated Turkish [20] and French [157] forms are available.

Quality of Life

General Aspects

The assessment of quality of life is related to health

The Constitution of the World Health Organization (WHO) defines quality of life as:

> "individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person's physical health, psychological state, level of independence, social relationships, personal beliefs and their relationship to salient features of their environment".

Consequently, not only the WHOQOL questionnaires but also the MOS SF-36/-12/-8 and the EuroQol questionnaires cover these general aspects, usually integrating them into a physical and mental health score without addressing disease specific parameters. In the field of spinal surgery, these tools are mainly used in combination with disease-specific pain and disability questionnaires.

Julious et al. [84] and Roset et al. [132] stated that sample sizes should always be calculated to allow the opportunity to detect changes at a pre-set level of statistical significance when planning a trial with health related quality of life (HRQL) instruments. However, only a small amount data is to be found in the literature on this topic. They published guidelines for calculating sample sizes for the use with the SF-36 [84] and for the use with the EQ-5D [132], respectively.

The Psychological General Well Being Index (PGWBI) focuses on psychological and psychosocial aspects and therefore may not be considered to be an all-embracing tool to assess quality of life. However, as psychological aspects comprise an important part of the quality of life, it will be described in this section.

Instruments

WHOQOL-100/WHOQOL-Bref

The WHO Quality of Life instruments have been developed with the intention of creating questionnaires that allow **quality of life** to be assessed as outlined above. Moreover, the aim was to evolve an international tool in several culturally diverse settings to simplify cross-cultural comparisons. To achieve this, 15 so-called Field Centers all over the world were involved in every stage of instrument development and further centers participated in the field testing [65].

The WHOQOL instruments assess health-related quality of life

The WHOQOL-100 consists of 100 questions referring to six domains [65]: Physical domain, Psychological domain, Level of independence, Social relationships, Environment, Spirituality/religion.

Each question has a **five-point answering scale**. For each domain a separate score is computed and transformed to a scale with a maximum of 100 points. It is obvious that such an extensive questionnaire is not practicable in a clinical setting where quality of life is only one part beside the more disease specific ones to be assessed. The evaluation of the data gathered with the WHOQOL-100 showed that the six domains may be grouped into four domains: Physical domain, Psychological domain, Social relationships, Environment.

Consequently, a core questionnaire consisting of 24 items was built and field tested in 17 centers with approximately 300 respondents each [1]. It was con-

cluded that this WHOQOL-Bref questionnaire showed validity and reliability and, thus, would be interesting for use in clinical trials. Meanwhile, the WHOQOL-Bref has been translated into and validated for further languages [53, 72, 79, 108, 114, 158]. It has been used in several recent studies in different fields of medicine: psychiatric disease [6, 42, 75, 85, 93, 113, 144, 150], geriatrics [60, 79], cancer [77, 159], liver disease [116] and HIV infection [35, 51]. In the field of musculoskeletal disorders it has been used in three studies [25, 69, 111]. The extensive validation procedures and translation into nearly 20 languages make the WHOQOL-Bref an interesting instrument for the future. Further detailed information is available from www.who.int.

MOS SF-36/SF-12/SF-8

The SF-36 is widely used for the assessment of health-related quality of life

The SF-36 was developed in 1992 by Stewart and Ware as a short form of the questionnaires used in the Medical Outcomes Study (MOS) [152]. It consists of 36 items, most of which have their roots in established instruments such as the General Psychological Well-Being Index (PGWBI) [44], the Health Perceptions Questionnaire [153] and other tools which have proved to be useful during the Health Insurance Experiment (HIE) [27]. **Eight scales** are built to **describe quality of life**: Physical functioning, Physical role (problems with work or other daily activities due to physical health), Bodily pain, General health, Vitality, Social functioning, Emotional role (problems with work or other daily activities due to emotional problems), Mental health.

The results of these scales are then grouped into two **summary measures**:

- **Physical health** (scales 1 – 4)
- **Mental health** (scales 5 – 8)

The SF-36 sensitively detects changes over time

The SF-36 is the most commonly used self-assessed generic quality of life instrument [59]. The mean internal consistency and test-retest validity of the first version has been shown to exceed 0.80 in several studies [71, 105, 106]. In 1996, the second version, SF-36v2, was introduced offering several improvements based on experience with the first version: Instructions and questionnaire items were shortened and simplified. The layout was adapted to reduce missing responses. Some dichotomous response choices were replaced by five-point scales whereas others were shortened from six- to five-point scales as well. These adaptations led to a decrease in standard deviation and percentage of ceiling and floor scoring. Today the SF-36 is available in a 4-week (standard) and a 1-week (acute) recall version. Compared to other generic health status instruments, it has shown several advantages [48, 97]. It was found to be most sensitive to detecting changes over time and showed the highest levels of internal consistency.

Peto et al. [120] compared the mental health subscale with the PGWB questionnaire in a sample of patients with amyotrophic lateral sclerosis and found good internal reliability and high correlations for both the PGWB and the SF-36 subscale. They stated that the mental health subscale provided comparable psychometric performance and, thus, may be used to measure and compare mental health in defined groups.

The SF-12 and SF-8 are short forms of the SF-36 with good validity

In 1994 the development of a 12-item questionnaire began which led to the SF-12, a subset of the SF-36, that is now available in the second version [151]. Though improving efficiency and practicability in the clinical setting, one has to accept some restrictions leading to less information about health status compared to the SF-36. Finally, an 8-item subset of the SF-36 has been developed. The SF-8 assesses every domain described in the SF-36 by only one item each. Besides a 24-h recall version there is a 4-week and a 1-week recall version available. It has been translated and validated for more than 30 countries [99].

In conclusion, the SF questionnaires represent **valuable tools for the assessment of general quality of life**. Their widespread use in clinical trials leads to broad possible comparisons. It is recommended to use these instruments in combination with disease-specific questionnaires to obtain an all-embracing picture of the respondents. Extensive information about the use, validity and norm-based scoring and interpretation is available on the SF internet homepage (www.sf36.com) and in the SF manuals.

EuroQol 5D

This tool was developed by the EuroQol Group, which started in 1987 with the intention of constructing an instrument for the assessment of standardized **non-disease-specific health-related quality of life**. It was thought to complement other tools such as the SF-36. The EuroQol Group is a multi-country, multi-center and multi-disciplinary group and, thus, the developed instrument should more easily allow cross-cultural comparisons to be performed.

The EuroQol exhibits validity comparable to the SF-36

The EQ-5D is a self-completion tool consisting of four components [28]. The first two parts address HRQL whereas the latter parts address further background information such as occupation, activity, age, sex, education and so on. In the first part HRQL is assessed by five statements about **mobility, self-care, usual activities, pain/discomfort and anxiety/depression**, which are divided into three degrees of severity. The respondents are asked to sign the one statement fitting best to their situation. This leads to a score of one to three in each statement. The second part consists of a Graphic Rating Scale ranging from zero to 100 in which respondents are asked to indicate their actual state of health today. Several studies were made to compare the EQ-5D with other quality of life tools, for example the SF-36. Generally, it was found to be a valuable instrument, simple to use by the patients and showing clinically relevant correlations with other condition-specific tools [26, 78]. Nevertheless, Brazier et al. [26] found it to be less sensitive and more susceptible to ceiling effects than the SF-36, preferring the latter for detecting changes over time. Further, detailed information is available on www.euroqol.org.

Psychological General Well-Being Index (PGWBI)

This questionnaire was developed by Dupuy in 1969 and first published after modification in 1984 [44]. It consists of 22 questions on the following **six** domains: **Anxiety, Depression, Well-being, Self-control, and Health vitality**.

Each domain consists of three to five questions which have to be rated on a six-point Likert scale. Every answer is validated by zero to five points. This results in a maximum score of 110. Revicki et al. [128] developed the PGWB into a version suitable for use in telephone interviews and successfully validated it for an American population.

The PGWB is a reliable tool with which to assess psychological distress

The PGWB has been extensively validated and has been used in many clinical studies, for example in the field of chronic pain, often in combination with other general health state questionnaires such as the SF-36 [14–16, 143].

Scoliosis Research Society Questionnaires: SRS-22/-24/-30

The **Scoliosis Research Society (SRS)** developed instruments to evaluate and monitor patients with **idiopathic scoliosis**. In 1999, the initial 24-item SRS-24 questionnaire was developed based on several previously validated questionnaires [70]. It is divided into seven equally weighted domains: Pain, General self-image, Post-operative self image, General function, Overall level of activity, Post-operative function and satisfaction.

This initial version was found to be reliable for postoperative outcome in scoliosis surgery as well as for dynamic monitoring in patients as they become adults. Nevertheless, some concerns about low internal consistency for some domains and some questions led to the creation of the current SRS-22.

This questionnaire is divided into five domains: Pain, Function/activity, Self-image/appearance, Mental health, Satisfaction about previous treatment.

The SRS-22/-30 questionnaires are specifically designed for scoliosis patients

As the SRS-22 no longer integrates specific questions about the postoperative status of the patients, the SRS-30 was developed. This version includes all questions of the 22-item tool and the postoperative questions of the 24-item tool. While the SRS-22 is validated for the English and Spanish [10] languages, the SRS-30 has not been validated so far. The SRS-22 was shown to be reliable with internal consistency and reproducibility comparable to the SF-36 [8, 18]. Moreover, it was found to be responsive to changes postoperatively [9] and to discriminate well between patients with no, moderate and severe scoliosis [7]. In one study it was even found to be useful in choosing non-surgical treatment in borderline cases [7]. The questionnaires and more information on scoring are available on the Scoliosis Research Society website (www.srs.org).

Psychosocial Aspects, Work Situation and Fear Avoidance Beliefs

General Aspects

In the past two decades, psychosocial and work-related aspects as well as the potential influence of behavior patterns have attracted interest in research on the development and course of chronic back pain [4, 33, 55, 57, 73, 149]. In this context, some instruments have been developed to assess these important aspects.

Instruments

Assessment of Occupational Status

As a minimum data set the extent of work incapacity should be assessed preoperatively and at follow-up as it is easy to assess and of great societal relevance [5]. Bombardier [23] proposed a categorization including the following:

- employed at usual job
- on light duty or some restricted work assignment
- paid leave/sick leave
- unpaid leave
- unemployed because of health problems
- unemployed because of other reasons
- student, keeping house/homemaker
- retired
- disability

Besides the occupational status, sickness absence is quite easily accessible too and is also of economic relevance. Hensing et al. [74] proposed five measures for sick leave assessment. Nevertheless, it has become apparent that age, gender, cultural factors, economic and health policy factors, job satisfaction, psychosocial job factors and factors not related to work at all influence work status and sickness absence [46]. Therefore, **multivariate methods** must be used to control these confounding parameters when work status is analyzed [148], and additional measures of work-related outcome such as work ability, **job-related resignation** and **job satisfaction** should be used.

Occupational status and sickness absence should be assessed preoperatively and at follow-up

Job Satisfaction and Job-Related Resignation

General job satisfaction and job-related resignation can be assessed by four 5-point Likert scales each. The items for the two scales are derived from a larger set of items developed by Oegerli [117] on the basis of the concept of "different forms of job satisfaction" by Bruggemann [29] (English description [34]). The two scales have been found to be reliable in several investigations.

Fear-Avoidance Beliefs Questionnaire (FABQ)

The FABQ predicts treatment outcome in subacute and chronic low back pain

Lethem and Slade [92, 136] first mentioned in 1983 that an avoidance behavior may result in an exaggerated pain perception and in 1993 Waddell et al. [149] introduced the FABQ, which consists of 16 items and is designed as a self-reporting tool. The questions are pain-specific and divided into one part assessing **fear-avoidance beliefs about work** and another part assessing **fear-avoidance beliefs about physical activities**. It has been shown to be a valid and reliable questionnaire and several studies have found it to be useful in predicting treatment outcome in subacute and chronic low back pain [31, 54, 58, 138].

Validated German and Swiss-German versions are available [121, 138]. McCracken et al. [103] compared the FABQ with three other validated instruments for the assessment of anxiety and fear in chronic pain patients: (1) the Spielberger Trait Anxiety Inventory (STAI) with more general response tendencies [137]; (2) the Fear of Pain Questionnaire (FPQ) [107] with more general response tendencies in addition; and (3) the Pain Anxiety Symptoms Scale (PASS) with more pain-specific response tendencies [104]. The FABQ and the PASS as more pain-specific questionnaires were found to be better predictors than the less pain-specific ones. However, it was recommended to use these tools in combination with general emotional distress measures in a clinical setting to achieve valuable information about the influence of pain avoidance beliefs and other psychosocial stressors on the course of chronic pain situations.

Clinical Feasibility and Practicability

Data completeness is mandatory for valid and reliable outcome assessment

As in most questionnaires a total score or several subscores are computed with the data from a small number of questions, and it is mandatory that questionnaires are filled in completely. Often, lacking the answer from only one or two questions makes analysis of the score impossible.

It is therefore important to inform patients about the importance of thorough questionnaire completion. Possible consequences of the planned investigation on future treatment modalities should be explained to the participants to increase their understanding. The patients' health and social condition have a significant impact on the willingness to participate in a study.

Short, valid reliable and easy to handle questionnaires are needed to increase questionnaire response and participation

It is desirable to use simple and short questionnaires in a clinical setting. This would not only minimize the patients' effort but also analysis of data by the health professionals. Therefore different groups are endeavoring to develop short, valuable, standardized outcome assessment tools. Deyo et al. [41] proposed a six-item core set of questions measuring several dimensions of outcome, each with a single item which has been studied and validated elsewhere. This short set of questions covering the core dimensions pain, function, well-being, disability (work), disability (social) and satisfaction post-treatment could be used as a basic battery for checking treatment outcome or developing quality improvements. A more detailed data assessment, for example within the scope of clinical trials with specific problems addressed, could easily be achieved by add-

ing further items in one of the core dimensions without necessarily expanding the whole questionnaire and therefore increasing the effort for respondents and analysts.

Mannion et al. [99] evaluated a modified German version of the standardized short core-measure tool proposed by Deyo and found it to be simple, practical, reliable and valid. Cronbach's alpha (internal consistency) for each core measure was between 0.41 and 0.78. Composing an index from all the core measures, Cronbach's alpha increased to 0.85. Test-retest reliability was moderate to excellent. There were floor and ceiling effects notable in the function domain whereas the disability dimension showed floor effects at follow-up. The correlations between the single items and their corresponding reference questionnaire were 0.60 – 0.79. The Sensitivity to Change was a little bit lower than in the reference questionnaires. Recently, White et al. [155] adapted the Deyo core questions to the neck pain setting and tested them on 104 patients. This first evaluation demonstrated a good repeatability and validity with absent floor or ceiling effects. These promising findings provide motivation for further research because the standardized use of such an instrument in future clinical trials would improve outcome assessment. It would improve the comparability between clinical studies and therefore build a better basis for treatment improvements in spinal surgery.

Recapitulation

For the evaluation of spinal interventions **self-administered assessment tools** are widely used. An instrument must be comparable, translated into and validated for the corresponding language and must embrace at least **pain**, **disability**, **health-related quality of life** and **work status**. For more thorough investigations, psychosocial aspects, **work-related parameters** and **fear avoidance behavior** should additionally be assessed. For these purposes an array of well validated standardized questionnaires are available.

Pain. As the predominant complaint in patients with spinal disorders, the evaluation of pain is one of the pillars of outcome assessment. Pain assessment seems to be most reliable when asking for an average pain level during a short recall period of time from 1 week to 4 weeks. Pain experience is very individual, complicating an interindividual comparison. In well informed patients **visual analogue** (VAS) and **graphic rating scales** (GRS) are valuable instruments for assessment of pain intensity and changes due to therapy. Some restrictions have to be taken into account when using these tools in an elderly population as they may be misunderstood and misinterpreted. NRS and VRS are other methods in pain assessment. Although well understandable and easy to handle (also in telephone interviews), they are not as appropriate for detecting changes over time as are VAS and GRS.

Disability. Neck- or back-related disability is another predominant complaint. The **Roland and Morris Disability Questionnaire** and **Oswestry Disability Index** are by far the most used instruments for assessment of disability in back patients. While the former seems to be more sensitive in detecting changes over time, the latter seems to be more useful in patients with severe disability. The North American Spine Society Questionnaire and the Hannover Functional Ability Questionnaire are also valuable tools though less frequently used.

Quality of life. Besides disease-specific tools, questionnaires on health-related quality of life are widely used in medicine. Several instruments have been developed and broadly tested in terms of reliability and validity. The most commonly used questionnaire is the **SF-36**, but also the WHO has edited a valuable tool (**WHOQOL-Bref**). The third well explored and frequently used instrument is the Euro-**Qol EQ-5D**. The **PGWB** concentrates on psychological general well-being as an important part of quality of life and is a valuable questionnaire in more thorough investigations. For the special setting in scoliosis patients, the **Scoliosis Research Society** introduced the SRS-22 and SRS-30 questionnaires. They include pain, disability, quality of life and satisfaction with treatment and allow a pre- and postoperative evaluation of these patients.

Recapitulation

Psychosocial aspects. It has been realized that psychosocial aspects and work situation are related to back pain. They may figure as risk factors or even predictors in subacute and chronic back pain. One aspect in this context is **fear avoidance behavior**, which can negatively influence outcome in spinal surgery. The most frequently used questionnaire in this field is the FABQ.

Work situation. As a minimum the work situation should be assessed by occupational status measures and sick absence measures. Because of the shortcomings of these simple methods additional instruments on job satisfaction and job-related resignation should be used for a more comprehensive assessment.

Feasibility/practicability. As in most questionnaires a total score or several subscores are computed with the data from a small number of questions, it is mandatory that questionnaires are filled in completely. Nevertheless, the patient's compliance is often insufficient for various reasons. Recent research is thus attempting to develop short and easily understandable tools which allow the gathering of enough data for meaningful conclusions.

Key Articles

Bombardier C (ed) (2000) Spine Focus Issue: Outcome assessments in the evaluation of treatment of spinal disorders. Spine 25:3097 – 3199

Boos N (ed) (2006) Outcome assessment and documentation. Eur Spine J 15 Suppl 1: S1 – 123

These two special journal issues summarize the state of the art in outcome assessment, research, and documentation in the treatment of spinal disorders and are a source for further reading.

References

1. (1998) Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 28:551 – 558
2. (2000) Glossary. Spine 25:3200 – 3202
3. Ackelman BH, Lindgren U (2002) Validity and reliability of a modified version of the neck disability index. J Rehabil Med 34:284 – 287
4. Al-Obaidi SM, Nelson RM, Al-Awadhi S, Al-Shuwaie N (2000) The role of anticipation and fear of pain in the persistence of avoidance behavior in patients with chronic low back pain. Spine 25:1126 – 1131
5. Amick BC, 3rd, Lerner D, Rogers WH, Rooney T, Katz JN (2000) A review of health-related work outcome measures and their uses, and recommended measures. Spine 25:3152 – 3160
6. Amir M, Lev-Wiesel R (2003) Time does not heal all wounds: quality of life and psychological distress of people who survived the holocaust as children 55 years later. J Trauma Stress 16:295 – 299
7. Asher M, Min Lai S, Burton D, Manna B (2003) Discrimination validity of the scoliosis research society-22 patient questionnaire: relationship to idiopathic scoliosis curve pattern and curve size. Spine 28:74 – 78
8. Asher M, Min Lai S, Burton D, Manna B (2003) The reliability and concurrent validity of the scoliosis research society-22 patient questionnaire for idiopathic scoliosis. Spine 28:63 – 69
9. Asher M, Min Lai S, Burton D, Manna B (2003) Scoliosis research society-22 patient questionnaire: responsiveness to change associated with surgical treatment. Spine 28:70 – 73
10. Bago J, Climent JM, Ey A, Perez-Grueso FJ, Izquierdo E (2004) The Spanish version of the SRS-22 patient questionnaire for idiopathic scoliosis: transcultural adaptation and reliability analysis. Spine 29:1676 – 1680
11. Basler HD, Jakle C, Kroner-Herwig B (1997) Incorporation of cognitive-behavioral treatment into the medical care of chronic low back patients: a controlled randomized study in German pain treatment centers. Patient Educ Couns 31:113 – 124
12. Beaton DE (2000) Understanding the relevance of measured change through studies of responsiveness. Spine 25:3192 – 3199

13. 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
14. Becker N, Bondegaard Thomsen A, Olsen AK, Sjogren P, Bech P, Eriksen J (1997) Pain epidemiology and health related quality of life in chronic non-malignant pain patients referred to a Danish multidisciplinary pain center. Pain 73:393 – 400
15. Becker N, Hojsted J, Sjogren P, Eriksen J (1998) Sociodemographic predictors of treatment outcome in chronic non-malignant pain patients. Do patients receiving or applying for disability pension benefit from multidisciplinary pain treatment? Pain 77:279 – 287
16. Becker N, Sjogren P, Bech P, Olsen AK, Eriksen J (2000) Treatment outcome of chronic non-malignant pain patients managed in a Danish multidisciplinary pain centre compared to general practice: a randomised controlled trial. Pain 84:203 – 211
17. Bergner M, Bobbitt RA, Carter WB, Gilson BS (1981) The Sickness Impact Profile: development and final revision of a health status measure. Med Care 19:787 – 805
18. Berven S, Deviren V, Demir-Deviren S, Hu SS, Bradford DS (2003) Studies in the modified scoliosis research society outcomes instrument in adults: validation, reliability, and discriminatory capacity. Spine 28:2164 – 2169; discussion 2169
19. Beurskens AJ, de Vet HC, Koke AJ (1996) Responsiveness of functional status in low back pain: a comparison of different instruments. Pain 65:71 – 76
20. Bicer A, Yazici A, Camdeviren H, Erdogan C (2004) Assessment of pain and disability in patients with chronic neck pain: reliability and construct validity of the Turkish version of the neck pain and disability scale. Disabil Rehabil 26:959 – 962
21. Bolton JE (1999) Accuracy of recall of usual pain intensity in back pain patients. Pain 83: 533 – 539
22. Bolton JE, Wilkinson RC (1998) Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. J Manipulative Physiol Ther 21:1 – 7
23. Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine 25:3100 – 3103
24. Boscainos PJ, Sapkas G, Stilianessi E, Prouskas K, Papadakis SA (2003) Greek versions of the Oswestry and Roland-Morris Disability Questionnaires. Clin Orthop:40 – 53
25. Bowman SJ, Booth DA, Platts RG (2004) Measurement of fatigue and discomfort in primary Sjögren's syndrome using a new questionnaire tool. Rheumatology 43:758 – 764
26. Brazier J, Jones N, Kind P (1993) Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Qual Life Res 2:169 – 180
27. Brook R, Ware J, Davies-Avery A, Stewart A, Donald C, Rogers W, Williams K, Johnston S (1979) Overview of adult health status measures fielded in RAND's Health Insurance Study. Med Care 17:1 – 131
28. Brooks R (1996) EuroQol: the current state of play. Health Policy 37:53 – 72
29. Bruggemann A (1974) Zur Unterscheidung verschiedener Formen von 'Arbeitszufriedenheit'. Arbeit und Leistung 28:281 – 284
30. Bryant RA (1993) Memory for pain and affect in chronic pain patients. Pain 54:347 – 351
31. Buchbinder R, Jolley D, Wyatt M (2001) 2001 Volvo Award Winner in Clinical Studies: Effects of a media campaign on back pain beliefs and its potential influence on management of low back pain in general practice. Spine 26:2535 – 2542
32. Budzynski TH, Stoyva JM, Adler CS, Mullaney DJ (1973) EMG biofeedback and tension headache: a controlled outcome study. Psychosom Med 35:484 – 496
33. Buer N, Linton SJ (2002) Fear-avoidance beliefs and catastrophizing: occurrence and risk factor in back pain and ADL in the general population. Pain 99:485 – 491
34. Buessing A (1992) A dynamic view of job satisfaction in psychiatric nurses in Germany. Work Stress 6:239 – 259
35. Chandra PS, Deepthivarma S, Jairam KR, Thomas T (2003) Relationship of psychological morbidity and quality of life to illness-related disclosure among HIV-infected persons. J Psychosom Res 54:199 – 203
36. Choiniere M, Amsel R (1996) A visual analogue thermometer for measuring pain intensity. J Pain Symptom Manage 11:299 – 311
37. Closs SJ, Barr B, Briggs M, Cash K, Seers K (2004) A comparison of five pain assessment scales for nursing home residents with varying degrees of cognitive impairment. J Pain Symptom Manage 27:196 – 205
38. Coste J, Le Parc JM, Berge E, Delecoeuillerie G, Paolaggi JB (1993) [French validation of a disability rating scale for the evaluation of low back pain (EIFEL questionnaire)]. Rev Rhum Ed Fr 60:335 – 341
39. Daltroy LH, Cats-Baril WL, Katz JN, Fossel AH, Liang MH (1996) The North American Spine Society Lumbar Spine Outcome Assessment Instrument: reliability and validity tests. Spine 21:741 – 749
40. Dawson EG, Kanim LE, Sra P, Dorey FJ, Goldstein TB, Delamarter RB, Sandhu HS (2002) Low back pain recollection versus concurrent accounts: outcomes analysis. Spine 27:984 – 993; discussion 994
41. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M,

Von Korff M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. Spine 23:2003 – 2013
42. Dogan S, Dogan O, Tel H, Coker F, Polatoz O, Dogan FB (2004) Psychosocial approaches in outpatients with schizophrenia. Psychiatr Rehabil J 27:279 – 282
43. Dropsy R, Marty M (1994) [Indices of quality of life for evaluating lumbago]. Rev Rhum Ed Fr 61:44S–48S
44. Dupuy H (1984) The Psychological General Well-Being (PGWB) Index. Assessment of quality of life in clinical trials of cardiovascular therapies. New York: Le Jacq:170 – 183
45. Eich E, Reeves JL, Jaeger B, Graff-Radford SB (1985) Memory for pain: relation between past and present pain intensity. Pain 23:375 – 380
46. Elfering A (2006) Work-related outcome assessment instruments. Eur Spine J 15 Suppl 1: S32 – 43
47. Elfering A, Semmer NK, Schade V, Grund S, Boos N (2002) Supportive colleague, unsupportive supervisor: the role of provider-specific constellations of social support at work in the development of low back pain. J Occup Health Psychol 7:130 – 140
48. Essink-Bot ML, Krabbe PF, Bonsel GJ, Aaronson NK (1997) An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. Med Care 35:522 – 537
49. Exner V, Keel P (2000) [Measuring disability of patients with low-back pain – validation of a German version of the Roland & Morris disability questionnaire]. Schmerz 14:392 – 400
50. Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain disability questionnaire. Physiotherapy 66:271 – 273
51. Fang CT, Hsiung PC, Yu CF, Chen MY, Wang JD (2002) Validation of the World Health Organization quality of life instrument in patients with HIV infection. Qual Life Res 11:753 – 762
52. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM (2001) Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 94:149 – 158
53. Fleck MP, Louzada S, Xavier M, Chachamovich E, Vieira G, Santos L, Pinzon V (2000) [Application of the Portuguese version of the abbreviated instrument of quality life WHOQOL-bref]. Rev Saude Publica 34:178 – 183
54. Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, Butler B, Garber M, Allison S (2002) A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine 27:2835 – 2843
55. Fordyce WE, Shelton JL, Dundore DE (1982) The modification of avoidance learning pain behaviors. J Behav Med 5:405 – 414
56. Freyd M (1923) The graphic rating scale. J Educ Psychol 43:83 – 102
57. Fritz JM, George SZ (2002) Identifying psychosocial variables in patients with acute work-related low back pain: the importance of fear-avoidance beliefs. Phys Ther 82:973 – 983
58. Fritz JM, George SZ, Delitto A (2001) The role of fear-avoidance beliefs in acute low back pain: relationships with current and future disability and work status. Pain 94:7 – 15
59. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R (2002) Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ 324:1417
60. Golimbet V, Trubnikov V (2001) Evaluation of the dementia carers situation in Russia. Int J Geriatr Psychiatry 16:94 – 99
61. Goolkasian P, Wheeler AH, Gretz SS (2002) The neck pain and disability scale: test-retest reliability and construct validity. Clin J Pain 18:245 – 250
62. Gracely RH, McGrath P, Dubner R (1978) Validity and sensitivity of ratio scales of sensory and affective verbal pain descriptors: manipulation of affect by diazepam. Pain 5:19 – 29
63. Gronblad M, Hupli M, Wennerstrand P, Jarvinen E, Lukinmaa A, Kouri JP, Karaharju EO (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 9:189 – 195
64. 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
65. Group W (1994) Development of the WHOQOL: Rationale and Current Status. Int J Ment Health 23:24 – 56
66. Haas M, Nyiendo J, Aickin M (2002) One-year trend in pain and disability relief recall in acute and chronic ambulatory low back pain patients. Pain 95:83 – 91
67. Haase I, Schwarz A, Burger A, Kladny B (2001) [Comparison of Hannover Functional Ability Questionnaire (FFbH) and the SF-36 subscale "Physical Functioning"]. Rehabilitation (Stuttg) 40:40 – 42
68. Haefeli M, Elfering A (2006) Pain assessment. Eur Spine J 15 Suppl 1:S17 – 24
69. Haefeli M, Elfering A, Kilian R, Min K, Boos N (2006) Nonoperative treatment for adolescent idiopathic scoliosis: a 10- to 60-year follow-up with special reference to health-related quality of life. Spine 31:355 – 366; discussion 367

70. Haher TR, Gorup JM, Shin TM, Homel P, Merola AA, Grogan DP, Pugh L, Lowe TG, Murray M (1999) Results of the Scoliosis Research Society instrument for evaluation of surgical outcome in adolescent idiopathic scoliosis. A multicenter study of 244 patients. Spine 24:1435 – 1440
71. Haley SM, McHorney CA, Ware JE, Jr. (1994) Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. J Clin Epidemiol 47:671 – 684
72. Hasanah CI, Naing L, Rahman AR (2003) World Health Organization Quality of Life Assessment: brief version in Bahasa Malaysia. Med J Malaysia 58:79 – 88
73. Hasenbring M, Hallner D, Klasen B (2001) [Psychological mechanisms in the transition from acute to chronic pain: over- or underrated?]. Schmerz 15:442 – 447
74. Hensing G, Alexanderson K, Allebeck P, Bjurulf P (1998) How to measure sickness absence? Literature review and suggestion of five basic measures. Scand J Soc Med 26:133 – 144
75. Herrman H, Hawthorne G, Thomas R (2002) Quality of life assessment in people living with psychosis. Soc Psychiatry Psychiatr Epidemiol 37:510 – 518
76. 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
77. Hsu C, Wang JD, Hwang JS, Tien HF, Chang SM, Cheng AL, Chen YC, Tang JL (2003) Survival-weighted health profile for long-term survivors of acute myelogenous leukemia. Qual Life Res 12:503 – 517
78. Hurst NP, Jobanputra P, Hunter M, Lambert M, Lochhead A, Brown H (1994) Validity of Euroqol – a generic health status instrument – in patients with rheumatoid arthritis. Economic and Health Outcomes Research Group. Br J Rheumatol 33:655 – 662
79. Hwang HF, Liang WM, Chiu YN, Lin MR (2003) Suitability of the WHOQOL-BREF for community-dwelling older people in Taiwan. Age Ageing 32:593 – 600
80. Jensen MP, Karoly P, Braver S (1986) The measurement of clinical pain intensity: a comparison of six methods. Pain 27:117 – 126
81. Jensen MP, Karoly P, O'Riordan EF, Bland F, Jr., Burns RS (1989) The subjective experience of acute pain. An assessment of the utility of 10 indices. Clin J Pain 5:153 – 159
82. 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 30:139 – 143
83. Joyce CR, Zutshi DW, Hrubes V, Mason RM (1975) Comparison of fixed interval and visual analogue scales for rating chronic pain. Eur J Clin Pharmacol 8:415 – 420
84. Julious SA, George S, Campbell MJ (1995) Sample sizes for studies using the short form 36 (SF-36). J Epidemiol Community Health 49:642 – 644
85. Kilian R, Matschinger H, Loeffler W, Roick C, Angermeyer MC (2002) A comparison of methods to handle skew distributed cost variables in the analysis of the resource consumption in schizophrenia treatment. J Ment Health Policy Econ 5:21 – 31
86. Kohlmann T, Raspe H (1996) [Hannover Functional Questionnaire in ambulatory diagnosis of functional disability caused by backache]. Rehabilitation (Stuttg) 35:I–VIII
87. Kovacs FM, Abraira V, Zamora J, Teresa Gil del Real M, Llobera J, Fernandez C, Bauza JR, Bauza K, Coll J, Cuadri M, Duro E, Gili J, Gestoso M, Gomez M, Gonzalez J, Ibanez P, Jover A, Lazaro P, Llinas M, Mateu C, Mufraggi N, Muriel A, Nicolau C, Olivera MA, Pascual P, Perello L, Pozo F, Revuelta T, Reyes V, Ribot S, Ripoll J, Rodriguez E (2004) Correlation between pain, disability, and quality of life in patients with common low back pain. Spine 29:206 – 210
88. Kovacs FM, Llobera J, Gil Del Real MT, Abraira V, Gestoso M, Fernandez C, Primaria Group KA (2002) Validation of the Spanish version of the Roland-Morris questionnaire. Spine 27:538 – 542
89. Kremer E, Atkinson JH, Ignelzi RJ (1981) Measurement of pain: patient preference does not confound pain measurement. Pain 10:241 – 248
90. Kucukdeveci AA, Tennant A, Elhan AH, Niyazoglu H (2001) Validation of the Turkish version of the Roland-Morris Disability Questionnaire for use in low back pain. Spine 26:2738 – 2743
91. 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
92. Lethem J, Slade PD, Troup JD, Bentley G (1983) Outline of a Fear-Avoidance Model of exaggerated pain perception – I. Behav Res Ther 21:401 – 408
93. Lin MR, Huang W, Huang C, Hwang HF, Tsai LW, Chiu YN (2002) The impact of the Chi-Chi earthquake on quality of life among elderly survivors in Taiwan – a before and after study. Qual Life Res 11:379 – 388
94. Linton SJ (1991) Memory for chronic pain intensity: correlates of accuracy. Percept Mot Skills 72:1091 – 1095
95. Linton SJ, Gotestam KG (1983) A clinical comparison of two pain scales: correlation, remembering chronic pain, and a measure of compliance. Pain 17:57 – 65

96. Linton SJ, Melin L (1982) The accuracy of remembering chronic pain. Pain 13:281 – 285
97. Lurie J (2000) A review of generic health status measures in patients with low back pain. Spine 25:3125 – 3129
98. 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
99. Mannion AF, Elfering A Outcome assessment in low back pain: how low can you go? Eur Spine J 14(10):1014 – 1026
100. Mannion AF, Elfering A (2006) Predictors of surgical outcome and their assessment. Eur Spine J 15 Suppl 1:S93-S108
101. Mannion AF, Junge A, Fairbank JC, Dvorak J, Grob D (2006) Development of a German version of the Oswestry Disability Index. Part 1: cross-cultural adaptation, reliability, and validity. Eur Spine J 15:55 – 65
102. Mannion AF, Junge A, Grob D, Dvorak J, Fairbank JC (2006) Development of a German version of the Oswestry Disability Index. Part 2: sensitivity to change after spinal surgery. Eur Spine J 15:66 – 73
103. McCracken LM, Gross RT, Aikens J, Carnrike CL, Jr. (1996) The assessment of anxiety and fear in persons with chronic pain: a comparison of instruments. Behav Res Ther 34: 927 – 933
104. McCracken LM, Zayfert C, Gross RT (1992) The Pain Anxiety Symptoms Scale: development and validation of a scale to measure fear of pain. Pain 50:67 – 73
105. McHorney CA, Ware JE, Jr., Lu JF, Sherbourne CD (1994) The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 32:40 – 66
106. McHorney CA, Ware JE, Jr., Raczek AE (1993) The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 31:247 – 263
107. McNeil D, Rainwater A, Al-jazireh L (1986) Development of a methodology to measure fear of pain. Paper presented at the annual meeting of the Association for Advancement of Behavior Therapy, Chicago
108. Min SK, Kim KI, Lee CI, Jung YC, Suh SY, Kim DK (2002) Development of the Korean versions of WHO Quality of Life scale and WHOQOL-BREF. Qual Life Res 11:593 – 600
109. Morley S (1989) The dimensionality of verbal descriptors in Tursky's pain perception profile. Pain 37:41 – 49
110. Morley S, Pallin V (1995) Scaling the affective domain of pain: a study of the dimensionality of verbal descriptors. Pain 62:39 – 49
111. Muller K, Schwesig R, Leuchte S, Riede D (2001) [Coordinative treatment and quality of life – a randomised trial of nurses with back pain]. Gesundheitswesen 63:609 – 618
112. Nachemson A, Bigos SJ (1984) The low back. In: Cruess J, Rennie WRJ (eds) Adult orthopedics. New York: Churchill-Livingstone, pp 843 – 937
113. Nasermoaddeli A, Sekine M, Hamanishi S, Kagamimori S (2003) Associations between sense of coherence and psychological work characteristics with changes in quality of life in Japanese civil servants: a 1-year follow-up study. Ind Health 41:236 – 241
114. Norholm V, Bech P (2001) The WHO Quality of Life (WHOQOL) Questionnaire: Danish validation study. Nord J Psychiatry 55:229 – 235
115. Nusbaum L, Natour J, Ferraz MB, Goldenberg J (2001) Translation, adaptation and validation of the Roland-Morris questionnaire – Brazil Roland-Morris. Braz J Med Biol Res 34:203 – 210
116. O'Carroll RE, Smith K, Couston M, Cossar JA, Hayes PC (2000) A comparison of the WHOQOL-100 and the WHOQOL-BREF in detecting change in quality of life following liver transplantation. Qual Life Res 9:121 – 124
117. Oegerli K (1984) Arbeitszufriedenheit. Versuch einer quantitativen Bestimmung. Paul Buetiger AG, Biberist
118. Ohnhaus EE, Adler R (1975) Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale. Pain 1:379 – 384
119. Padua R, Padua L, Ceccarelli E, Romanini E, Bondi R, Zanoli G, Campi A (2001) Cross-cultural adaptation of the lumbar North American Spine Society questionnaire for Italian-speaking patients with lumbar spinal disease. Spine 26:E344 – 347
120. Peto V, Jenkinson C, Fitzpatrick R, Swash M (2001) Measuring mental health in amyotrophic lateral sclerosis (ALS): a comparison of the SF-36 Mental Health Index with the Psychological General Well-Being Index. Amyotroph Lateral Scler Other Motor Neuron Disord 2:197 – 201
121. Pfingsten M, Kroner-Herwig B, Leibing E, Kronshage U, Hildebrandt J (2000) Validation of the German version of the Fear-Avoidance Beliefs Questionnaire (FABQ). Eur J Pain 4: 259 – 266
122. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF (2002) Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. Spine 27:515 – 522

123. Pose B, Sangha O, Peters A, Wildner M (1999) [Validation of the North American Spine Society Instrument for assessment of health status in patients with chronic backache]. Z Orthop Ihre Grenzgeb 137:437 – 441
124. Price DD, Bush FM, Long S, Harkins SW (1994) A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain 56:217 – 226
125. Price DD, Harkins SW, Baker C (1987) Sensory-affective relationships among different types of clinical and experimental pain. Pain 28:297 – 307
126. Price DD, McGrath PA, Rafii A, Buckingham B (1983) The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 17:45 – 56
127. Raspe H, Huppe A, Matthis C (2003) [Theories and models of chronicity: on the way to a broader definition of chronic back pain]. Schmerz 17:359 – 366
128. Revicki DA, Leidy NK, Howland L (1996) Evaluating the psychometric characteristics of the Psychological General Well-Being Index with a new response scale. Qual Life Res 5:419 – 425
129. Roese I, Kohlmann T, Raspe H (1996) [Measuring functional capacity in backache patients in rehabilitation: a comparison of standardized questionnaires]. Rehabilitation (Stuttg) 35:103 – 108
130. Roland M, Fairbank J (2000) The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine 25:3115 – 3124
131. 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 8:141 – 144
132. Roset M, Badia X, Mayo NE (1999) Sample size calculations in studies using the EuroQol 5D. Qual Life Res 8:539 – 549
133. Schochat T, Rehberg W, von Kempis J, Stucki G, Jackel WH (2000) [The North American Spine Society Lumbar Spine Outcome Assessment Instrument: translation and psychometric analysis of the German version in rehabilitation patients with chronic back pain]. Z Rheumatol 59:303 – 313
134. Scott J, Huskisson EC (1976) Graphic representation of pain. Pain 2:175 – 184
135. Seymour RA, Simpson JM, Charlton JE, Phillips ME (1985) An evaluation of length and end-phrase of visual analogue scales in dental pain. Pain 21:177 – 185
136. Slade PD, Troup JD, Lethem J, Bentley G (1983) The Fear-Avoidance Model of exaggerated pain perception – II. Behav Res Ther 21:409 – 416
137. Spielberger C, Gorsuch R, Lushene P, Vagg P, Jacobs G (1983) Manual for the State-Trait Anxiety Inventory (Form Y). Palo Alto, CA: Consulting Psychologists Press
138. Staerkle R, Mannion AF, Elfering A, Junge A, Semmer NK, Jacobshagen N, Grob D, Dvorak J, Boos N (2004) Longitudinal validation of the Fear-Avoidance Beliefs Questionnaire (FABQ) in a Swiss-German sample of low back pain patients. Eur Spine J 13:332 – 340
139. Stewart WF, Lipton RB, Simon D, Liberman J, Von Korff M (1999) Validity of an illness severity measure for headache in a population sample of migraine sufferers. Pain 79:291 – 301
140. Stratford PW, Binkley J, Solomon P, Gill C, Finch E (1994) Assessing change over time in patients with low back pain. Phys Ther 74:528 – 533
141. Stratford PW, Binkley JM, Riddle DL (2000) Development and initial validation of the back pain functional scale. Spine 25:2095 – 2102
142. Suzukamo Y, Fukuhara S, Kikuchi S, Konno S, Roland M, Iwamoto Y, Nakamura T (2003) Validation of the Japanese version of the Roland-Morris Disability Questionnaire. J Orthop Sci 8:543 – 548
143. Thomsen AB, Sorensen J, Sjogren P, Eriksen J (2002) Chronic non-malignant pain patients and health economic consequences. Eur J Pain 6:341 – 352
144. Umansky R, Amir M, Fridmann M, Zidon E, Chen D, Nemetz B (2003) Was it a good move? Improvement in quality of life among chronic mental patients moving from a mental hospital to a hostel in the community. Isr J Psychiatry Relat Sci 40:248 – 257
145. Vernon H, Mior S (1991) The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 14:409 – 415
146. Von Korff M, Jensen MP, Karoly P (2000) Assessing global pain severity by self-report in clinical and health services research. Spine 25:3140 – 3151
147. Von Korff M, Saunders K (1996) The course of back pain in primary care. Spine 21:2833 – 2837; discussion 2838 – 2839
148. Waddell G, Burton AK, Main CJ (2003) Screening to identify people at risk of long-term incapacity for work. A conceptual and scientific review. Royal Society of Medicine Press, London
149. Waddell G, Newton M, Henderson I, Somerville D, Main CJ (1993) A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 52:157 – 168
150. Wang X, Gao L, Zhang H, Zhao C, Shen Y, Shinfuku N (2000) Post-earthquake quality of life and psychological well-being: longitudinal evaluation in a rural community sample in northern China. Psychiatry Clin Neurosci 54:427 – 433

151. Ware J, Jr., Kosinski M, Keller SD (1996) A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 34:220 – 233
152. Ware J, Sherbourne C (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473 – 483
153. Ware JE, Jr. (1976) Scales for measuring general health perceptions. Health Serv Res 11:396 – 415
154. Wheeler AH, Goolkasian P, Baird AC, Darden BV, 2nd (1999) Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. Spine 24: 1290 – 1294
155. White P, Lewith G, Prescott P (2004) The core outcomes for neck pain: validation of a new outcome measure. Spine 29:1923 – 1930
156. Wiesinger GF, Nuhr M, Quittan M, Ebenbichler G, Wolfl G, Fialka-Moser V (1999) Cross-cultural adaptation of the Roland-Morris questionnaire for German-speaking patients with low back pain. Spine 24:1099 – 1103
157. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Revel M (2002) French translation and validation of 3 functional disability scales for neck pain. Arch Phys Med Rehabil 83:376 – 382
158. Yao G, Chung CW, Yu CF, Wang JD (2002) Development and verification of validity and reliability of the WHOQOL-BREF Taiwan version. J Formos Med Assoc 101:342 – 351
159. Yu CL, Fielding R, Chan CL, Tse VK, Choi PH, Lau WH, Choy DT, O SK, Lee AW, Sham JS (2000) Measuring quality of life of Chinese cancer patients: A validation of the Chinese version of the Functional Assessment of Cancer Therapy-General (FACT-G) scale. Cancer 88:1715 – 1727