

U. Müller  
C. Roeder  
L. Dubs  
M. S. Duetz  
C. G. Greenough

# Condition-specific outcome measures for low back pain

## Part II: Scale construction

Received: 1 December 2003  
Accepted: 5 December 2003  
Published online: 17 March 2004  
© Springer-Verlag 2004

U. Müller (✉) · C. Roeder  
Institute for Evaluative Research  
in Orthopaedic Surgery, University of Bern,  
Murtenstrasse 35, P.O. Box 8354,  
3001 Bern, Switzerland  
e-mail: umueller@memced.org

L. Dubs  
WIG, Winterthur, Switzerland

M. S. Duetz  
Institute for Social  
and Preventive Medicine,  
University of Bern, Bern, Switzerland

C. G. Greenough  
The James Cook University Hospital,  
Middlesbrough, UK

**Abstract** A literature review of the most widely used, condition-specific, self-administered assessment questionnaires for low back pain has been undertaken. In part I, technical issues such as validity, reliability, availability and comparability were analyzed for the nine most widely used outcome tools. This second part focuses on the content and wording of questions and answers in each of the nine questionnaires, and an analysis of the different score results is performed. 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 · Score bias · Review · Low back pain

### Introduction

In part I of the study, the nine most commonly used questionnaires specifically for low back pain were subjected to a complete validation process. The validation procedure examined reliability, responsiveness and minimum clinically important difference, external validity and floor and ceiling effects. This article will study a number of different aspects of questionnaire composition, content, and understandability. The first of these aspects is the relationship between the questionnaire and the modern concepts of disability, activity limitation, handicap and restrictions in society.

The 1980 WHO International Classification of Impairments, Disabilities and Handicaps (ICIDH) [27] considered the manifestations of a disease in three domains: (1) impairment which is related to loss or abnormality of

body structures or function (e.g. liver damage, knee damage, sensory impairment, pain [17]), (2) disability related to the individual (e.g. disabilities in activities of daily living, domestic tasks, communication) and (3) handicap related to society (e.g. employment, social integration, sport).

Pathologies express their manifestations in all three ICIDH dimensions but (depending on the individuality of the patient) the values of the three dimensions are expressed individually. The values of each of the three dimensions are correlated (low or high correlation) but there is no causality between them. Thus some patients with slight impairments of body structures (e.g. minimal contusion in the low back) can present major disability and handicap after one year, and others with major impairment (e.g. severe lumbar fracture) can present only minor disability or handicap after three months. That is why each of the three dimensions must be assessed individually.

In 2001, the WHO introduced the International Classification of Functioning, Disability and Health (ICF) [26]. The three domains (body, individual and society) are now ordered into three basic lists: (1a) body functions and (1b) structures, (2) activities and participation, and (3) personal and environmental factors. Each of these four lists (1a, 1b, 2 and 3) contains a domain where a disease can manifest its presence. As in the ICIDH, pathologies express their manifestations in all four ICF dimensions but (depending on the individuality of the patient) the values of the four dimensions are expressed individually. The values of each of the four dimensions are correlated (low or high correlation) but there is no causality between them [28].

Other important features are the design of the questions (understandable? unambiguous?) and the answer options available. How do the questions and answers interact and fit together? If the questionnaire produces a summed score, are the questions addressing the same domain or is there a possibility of item masking? Item weighting is used in some questionnaires but the basis for this has often not been validated. Cross-contamination between back pain and other pathologies, e.g. hip arthrosis, will be examined.

It is hoped that the results of this analysis (when combined with the conclusions reached in part I) will provide the reader with the necessary information enabling him or her to choose the best option when searching for a self-administered assessment tool in the field of low back problems.

## Methods

The nine scores examined in part I were analysed under a number of different headings. The questionnaires were also examined according to the concepts promulgated by the ICF classification [10, 22, 26]. Some basic concepts are set out below. For more specific information, readers are referred to the literature [1, 16, 21].

### 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 [16, 21]. 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.

According to the ICF classification [26], three main types of questions can be defined: questions regarding impairments, questions regarding activity limitations, and questions regarding environmental factors. Impairment is divided into impairments of body functions (e.g. sleep as a mental impairment, pain as a sensory impairment [17], blood pressure as a cardiovascular impairment etc.) and impairments of body structures (e.g. cartilage damage, impairments of spinal cord or peripheral nerves etc.). Activities are divided into activities (e.g. learning, walking, doing housework etc.) and participation (e.g. relationships, religion etc.). Environmental factors are the health care system, food, climate etc.. All these four dimensions are part of an outcome and are valuable indicators for quality of life.

The four dimensions can be asked about within one questionnaire, like in the Low Back Outcome Score [11], or different questionnaires are used to fulfil the ICF criteria as is the case in the NASS LSO questionnaire [3]. Asking about all three dimensions in one questionnaire is usually favoured because the type of questions and answer scaling as well as the question flow and similar questions can be harmonised. The importance of assessing all four ICF dimensions is stressed, but summing up 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 [16] (answers do not overlap and target 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 preoperative and postoperative conditions. Some authors use statistically defined weights for the different answers, resulting in a greater or smaller influence of the concerned item on the final score. None of the nine examined condition-specific spine scores uses such statistically defined weights. Summing the scores of individual questions can result in a number of sources of error.

### Item-masking score bias [5]

Item-masking score bias may be present if unrelated questions are summed up in a single score [5, 12]. For example, if a person suffers from severe back pain (ICF group 1a) and receives 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. If, during follow-up, the same person suffers from moderate back pain only and receives 4 out of 10 points but 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, therefore, 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 [6, 7]. If a questionnaire comprises a number of domains, then the use of sub-scores, focussed on only one dimension, should be considered.

### Item-weighting score bias

Some measures allocate different weights to questions. For example, if a question asking about ability to do work offers a 10-point scale and one on walking, a 3-point scale, this puts great 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 to certain patients, such as do-

ing household 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 weightless for some individuals.

Cross-contamination score bias

Cross-contamination score bias is present if answers can be influenced by other diseases. If the question is not properly phrased in relation to the symptoms of the disease addressed, then co-lesions or co-morbidities can alter the outcome. For example: the question “do you have pain in your leg?” 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 are made for avoiding bias when using each of the nine questionnaires. These suggestions should be used with care. Changes only in the calculation of 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 different above-mentioned scale construction problems can 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 following analysis of each questionnaire is always based on the latest available English version.

Analysis

The Oswestry Disability Index (ODI) [9]

Questions

The ten questions of the ODI scale are straightforward and simple. Pain is the central issue in all the questions. In the introduction of the questionnaire, the patient is in-

**Table 1** Content and question-and-answer characteristics in the chosen evaluation tools

Characteristic	ODI	RMDQ	LBOS	QBPDS	MVAS	ALBDS	NASS LSO	LBPRS	WDI
Assessment of Pain	1a	1a	1a	–	1a	1a	1a	1a	–
Sleep	1a	1a	1a	1a	1a	1a	1a	1a	1a
Self-care	2	2	–	2	–	2	–	–	–
Walking	2	2	2	2	2	2	2	2	2
Sitting	2	2	2	2	2	2	2	2	2
Standing	2	2	–	2	2	2	2	–	2
Lifting	2	2	–	2	–	–	2	2	2
Sex life	2	–	2	–	–	–	2	–	2
Travelling	2	–	2	–	–	–	2	–	2
Social life	3	–	–	–	3	–	3	3	3
Work	–	2	2	–	2	–	–	2	–
Dressing	–	2	2	–	–	–	2	2	2
Sport	–	–	2	2	–	2	–	–	–
Stairs	–	2	–	2	–	–	–	2	–
Housework	–	2	2	2	–	2	–	2	–
Resting	–	2	2	–	–	2	–	–	–
Appetite	–	1a	–	–	–	–	–	–	–
Need of help	–	2	–	–	–	–	–	–	–
Psychological factors	–	–	–	–	–	–	–	1a	–
Need of treatment	–	–	X.9	–	–	–	–	–	–
Need of medication	–	–	X.8	–	X.8	X.8	–	–	–
Car driving	–	–	–	2	–	–	–	2	–
Throwing	–	–	–	2	–	–	–	–	–
Stiffness	–	–	–	–	1a	–	–	–	–
Twisting	–	–	–	–	1a	–	–	–	–
Bending	–	–	–	–	–	1a	–	–	–
Loss of feeling	–	–	–	–	–	1b	1b	–	–
Leg weakness	–	–	–	–	–	1b	1b	–	–
Special features							SF36 and health survey item 18–28 are included		
Scoring points	0–50	0–24	0–75	100	150	0–100	0–102	0–90	0–9

**Table 2** Number of questions in the different ICF lists, and answer possibilities in each of the different questionnaires

Score	ICF 1a	ICF 1b	ICF 2	ICF 3	ICF Add-on	Answer possibilities	Remarks
ODI	2	0	8	0	0	Scaled text	Functional pain assessment
RMDQ	5	0	19	0	0	Yes/no	Disability and functional pain assessment
LBOS	2	0	8	1	1	Scaled text, 11-point VAS for pain	Disability, functional pain and general assessment
QBPDS	1	0	19	0	0	6-point Likert scale	Disability assessment
MVAS	5	0	10	0	0	100 mm VAS	Functional pain assessment
ALBDS	8	0	10	0	1	Multiple choice and scaled text	Disability and functional pain assessment
NASS LSO <sup>a</sup>	10	1	8	0	0	Pain locator and 6-point Likert scale	Pain and pain intensity locator, functional pain assessment, combined with other instruments
LBPRS	7	0	13	0	1	11-point VAS and 3-point Likert scale	Disability and functional pain assessment, specific pain and pain intensity locator
WDI	1	0	8	0	0	Yes/no	Disability assessment

<sup>a</sup>NASS LSO includes all four ICF lists, including add-on questions. In the table above, only the low back pain-specific questions like the pain locator, the specific, quantitative pain and neurological deficiency questions and the ODI are analysed.

**Table 3** Problems with question-and-answer inconsistencies, and the presence of score bias in each of the different questionnaires

Score	Question inconsistencies	Answer inconsistencies	Score bias I	Score bias II	Score bias III	Overall value
ODI	(+)	+(+)	++	(+)	+	+
RMDQ	+	-	+	(+)	-	+++
LBOS	-	+	++	+	(+)	++
QBPDS	+	-	(+)	+	+	+++
MVAS	++	-	++	+	+	+
ALBDS	+	++	++	++	-	-
NASS LSO <sup>a</sup>	-	+	++	+	+	++
LBPRS	+	-	++	+	-	+++
WDI	+	-	+	+	+	+

<sup>a</sup>NASS LSO includes all four ICF lists, including add-on questions. In the table above, only the low-back-pain-specific questions like the pain locator, the specific, quantitative pain and neurological deficiency questions and the ODI are analysed.

formed that the questionnaire deals with back problems. The questions themselves, however, do not explicitly mention pain and disability in the context of the lower back. By consequence, answers in the questionnaire can be contaminated by pain issues not related to low back pain. For example, a patient with severe pain in the groin due to hip arthritis will note this pain somehow in the questionnaire even if it is not related to low back disorders.

#### *Offered answers*

Six answer possibilities are consistently offered. In 52 out of 60 answers, pain is an issue. This means that the ODI is an almost pure functional pain scale. In some questions, the answers are inconsistent. E.g. in a question on walking, four answer possibilities mention pain in relation to different walking distances. The fifth deals with walking

and the use of crutches, and the sixth concerns lying in bed and the need of crawling to the toilet. Answers five and six do not mention pain. This means that the answers to this question are not disjunctive. Questions 2, 3, 4 and 5 do not offer disjunctive answers.

#### *ICF classification*

The three main foci of the ICF classification are addressed in the ODI: two questions relate to impairments of body functions (sleep = mental dysfunction, pain = sensory dysfunction), one question relates to environmental factors, and the other questions relate to activity limitations (Table 2 in Part I). All these foci are somehow, but not consistently related to pain (Tables 1, 2, and 3).

*Item masking*

A sum score is suggested for the ODI. Summing up answer results of questions in different ICF classes creates possible item masking.

*Item weighting*

All questions offer the same answer scale. One question asks about sex life, but some patients do not have a sex life. This question puts weight on an issue that is unimportant to some people, creating a minor problem with item weighting.

*Cross-contamination*

The introduction to the questionnaire (version 2.0) refers to pain in the back, but also to pain in the leg. Thus, a low-back status free of pain can become obscured by a developing painful hip arthritis. Therefore, possible cross-contamination is present.

*Suggestion for an unbiased use*

Results might be of reduced value because of non-disjunctive answers (questions 2–5). Item-weighting problems can be eliminated if question 8 is excluded from the sum score. If the use of the ODI is limited to functional pain assessment, the influence of item masking is minimal.

*Conclusion*

The ODI is suitable for the evaluation of functional low back pain but there might be some concern about non-disjunctive answer options. A questionnaire to assess activities of daily living should be added to the functional pain assessment of the ODI.

The Roland-Morris Disability Questionnaire (RMDQ) [18, 19]

*Questions*

The 24 questions in the RMDQ are straightforward and consistently related to the back. Question 18 (“I sleep less well on my back”) is an exception. A more consistent question would have been “I sleep less well because of my back”. A second problem with the original wording is the fact that it concerns a sleeping habit. Many people prefer to sleep either on their side or on their chest, which is not related to back troubles.

*Offered answers*

The answers are dichotomous and simple (Yes or No).

*ICF classification*

Two questions relate to impairments of body functions (sleep, pain), 21 questions relate to activity limitations and one question relates to environmental factors.

*Item masking*

Questions with different foci (pain, abilities etc.) are present. This may lead to item masking if a sum score is calculated. If the sum score is used solely for a quantification of functional disability, then this bias is minimal.

*Item weighting*

Consistent dichotomous answer types avoid this score bias.

*Cross-contamination*

Cross-contamination is not present as questions are consistently related to the back (except question 18).

*Suggestion for an unbiased use*

Calculating a total sum score is not recommended. Sub-scores adding up answers that belong to the same ICF group should be used as outcome measures instead. Question 18 should be evaluated independently from any sub-score. Nevertheless, it would be worth the effort to change the content of question 18 as stated above.

*Conclusion*

This questionnaire is very useful for a functional assessment. Consideration should be given to combining the questionnaire with a specific pain assessment.

The Low Back Outcome Score (LBOS) [11]

*Questions*

The 13 questions are easy to understand and have a single and clear focus. Pain is assessed independently. This issue allows discrimination between the influences of pain and the influences of disability on the patient’s health status.

This discrimination is important because pain and dysfunction do not necessarily relate to each other.

#### *Offered answers*

The pain question is answered with an 11-point visual analogue scale. The question on work offers six different answer options. All the other questions provide four different answer possibilities. For scoring, the 11 answer possibilities on pain as well as the six answer options on work will be reduced to four categories with distribution of 9, 6, 3 and 0 points. Using this method of reduction, all questions of the questionnaire offer four answer categories.

The answers regarding work are not disjunctive because “full time at your usual job”, “full time at a lighter job”, “part time”, “not working/unemployed” and “disability benefit” do not follow a continuous reduction of inability to work. E.g. “working part time” may be someone’s own choice and independent of a possible disease, while “full time at a lighter job” might be caused by a low back disease.

#### *ICF classification*

According to the ICF classification, the questionnaire provides two questions considering impairments of body functions (pain and sleep) and two questions on the use of health care (doctors visits and pain medication). In addition, the questionnaire considers pain (question 1) and working ability. The other questions concern activity limitations.

#### *Item masking*

The questionnaire considers different levels of the ICF classification. Summing to provide a total sum score may lead to item masking.

#### *Item weighting*

The scoring consistently uses four answer categories per question. However, because different questions are differently weighted (e.g. “sex life” up to 6 points, “sleeping” up to 3 points) there is an item-weighting bias.

#### *Cross-contamination*

Some questions do not explicitly refer to the back, e.g. “How often do you have to take pain killers for your pain?” The medication intake could be influenced by other types of pain like headache.

#### *Suggestion for an unbiased use*

The questionnaire is short, comprehensive and includes all three ICF groups – which is its main advantage. On the other hand, the questionnaire should not be summed because of its score bias. Instead, individual sub-scores for pain, functional pain, and ability items should be preferred. In addition, it is advisable to include only the questions that have the same weight (e.g. all questions that allow a maximum of 3 points) in a sum score. Results of the other items should be presented independently. The introduction should clearly state that the questions relate to back problems. The use of such an introductory sentence would render the focus of the questionnaire more precisely on the back.

#### *Conclusion*

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

#### The Quebec Back Pain Disability Scale (QBPDS) [13]

##### *Questions*

The 20 questions in the QBPDS scale are straightforward and simple. In the introduction, the patient is informed that the questionnaire deals with back problems. However, the questions themselves do not explicitly mention disability in the context of the back. This could lead to misunderstandings because patients might forget the introductory sentence and note shoulder troubles when they are asked about throwing a ball. The questions do not distinguish between functional pain and disabilities, causing an uncontrolled interaction between these two entities. Thus a reduction in pain might not be recorded if a dysfunction were to increase in the meantime.

##### *Offered answers*

Five answer options are consistently offered.

##### *ICF classification*

One question relates to impairments of body functions (sleep) and the other questions relate to activity limitations. Pain is not referred to in the questionnaire.

*Item masking*

The questionnaire solely addresses disabilities when performing activities. In that respect, a score bias is not present.

*Item weighting*

If the patient is not asked about pain, he includes his pain perception in the answers. Functional pain may therefore get a hidden and uncontrolled weight in the answers. In that respect a theoretical bias might occur. For non-drivers, the question about driving a car puts weight on a question that is weightless for the patient.

*Cross-contamination*

Questions like “throw a ball”, “reach up to high shelves” etc. might be influenced by other pathologies unrelated to the back. For example, patients with severe shoulder or other troubles might mark their complaints in these answers because of the lack of other alternatives.

*Suggestion for an unbiased use*

If a questionnaire does not specifically refer to pain, patients may include this aspect of the disease in the answers. The addition of a questionnaire about pain and functional pain should be considered.

*Conclusion*

The questionnaire is well focussed on disabilities and offers consistent answers. The questionnaire should be combined with a pain assessment.

The Million Visual Analogue Scale (MVAS) [15]

*Questions*

The 15 questions in the MVAS scale are mainly straightforward and simple. Nevertheless, some questions like “Is there anything that you do or are there any circumstances...?” are complex. The word “anything” makes the question very unspecific as it may relate to various aspects of daily living. Question 4, addressing pain medication, does not specifically refer to back pain, raising the possibility of cross-contamination with pain from other areas. The same problem exists with other questions concerning pain and disabilities, which do not specify the body region of interest.

*Offered answers*

A 100 mm visual analogue scale is consistently used. Scaling problems are not present.

*ICF classification*

Four out of 12 questions refer to impairments of body functions (sleep, pain, stiffness, twisting), and nine questions refer to activity limitations. One question addresses pain medication and one question is about possible modifications at work in order to facilitate job performance.

*Item masking*

The questionnaire considers different levels of the ICF classification – which leads to a strong bias if a sum score is calculated.

*Item weighting*

A consistent answer scale is used. The question “Do you get relief from pain killers?” is important, but may create a score bias if the patient does not take pain medication.

*Cross-contamination*

Pain is not always clearly referred to the back. Consequently, other pain origins can distort the success of treatment for low back pain. A question like “Do you get relief from pain killers?” is strongly dependent on the type of medication.

*Suggestion for an unbiased use*

It is recommended to use only sub-scores containing questions that are within the same ICF class. Question 3 should not be integrated into sub-scores.

*Conclusion*

Despite the interesting questions and answers, this questionnaire should be used with caution as some questions can lead to imprecise answers.

The Aberdeen Low Back Disability Scale (ALBDS) [20]

*Questions*

In the introduction of the ALBDS (19 items), the patient is informed that the questionnaire deals with back prob-

lems. Two questions address the topic of specific pain locations, two questions are on pain intensity and appearance, one on loss of feeling and two on its possible location. Twelve questions do not explicitly mention pain and disability in the context of the back. Like other questionnaires, these questions can lead to cross-contamination from other pain locations.

#### *Offered answers*

There are six multiple choice questions offering inconsistent answers. The single choice questions offer different numbers of answer options ranging from three to six answers. Questions 11 (sleep and pain) and 14 (walking and pain) are not disjunctive.

#### *ICF classification*

Six questions relate to impairments of body functions (sleep, pain, bending, loss of feeling, weakness), and 12 questions relate to activity limitations. One question is about pain medication.

#### *Item masking*

Bias is present because neurological findings, pain, sleep and abilities of daily living are mixed. A sum score is difficult to interpret because of the multiple choice answers. Choosing two out of four answers leaves six possibilities to choose further answers. Each possibility results in the same score but shows a completely different pattern of pain distribution for probably different reasons.

#### *Item weighting*

A heavy bias is also present because of the use of inconsistent answer options.

#### *Cross-contamination*

Pain medication is not related to the back. This means it can be influenced by different factors.

#### *Suggestion for an unbiased use*

The questionnaire is difficult to use in an unbiased fashion. Suggestions will not solve the present bias problems.

#### *Conclusion*

This questionnaire is of limited value because of structural problems.

The NASS Lumbar Spine Outcome Assessment Instrument (NASS LSO) [4]

#### *Questions*

The NASS LSO baseline questionnaire is by far the longest of all analysed instruments with its 62 main questions. It is a combination of existing questionnaires like the SF36 and ODI, and of additional questions about education, culture and other topics. This analysis will focus on the low-back-pain-specific questions and exclude all the other general 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. 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. This may be helpful for assessing internal consistency, but it is an irritating fact for the patients and might influence their compliance.

#### *Offered answers*

Six answers are consistently offered in the low-back-specific questions (Q. 38–54). Answers in questions 38–45 use a six-point disjunctive Likert scale. The ODI (Q. 46–54) has already been reviewed.

#### *ICF classification*

The entire questionnaire implements all aspects of the ICF classification.

#### *Item masking*

All dimensions of the ICF classification are introduced, thus a total sum score is biased. However, sub-scores are calculated due to the original questionnaires. Questions 38–45 offer such a sub-score. This sub-score is biased because questions about loss of feeling and pain are mixed

despite the knowledge that these two symptoms are not related to each another.

#### *Item weighting*

Item weighting is not present in questions 38–45 (refer to the ODI chapter above).

#### *Cross-contamination*

Cross-contamination is not present in questions 38–45 (refer to the ODI chapter above).

#### *Suggestion for an unbiased use*

The sub-score bias of questions 38–45 can be avoided if only the pain questions 38, 39, 42 and 43 and the neurological questions 40, 41, 44 and 45 are summed together. For the ODI, see chapter above.

#### *Conclusion*

Pain is very dominant in the low-back-specific questions. In contrast to the Oswestry score, pain can be indicated precisely, so that changes related to the pain origin become visible. All the other, earlier-stated problems of the ODI remain unchanged (see above). Use only sub-scores as stated.

### The Low Back Pain Rating Scale (LBPRS) [14]

#### *Questions*

The 21 questions are straightforward and simple. Low back pain explicitly is the central issue in 16 out of 21 questions. Because they consistently refer to low back pain (except in the last question), pain location is always specified. Some questions like question 14 (“If it was of present interest, do you think that there are certain jobs which you would not be able to manage because of your back pain”) are complex and refer to uncertain factors that might influence the answer.

#### *Offered answers*

Pain intensity is evaluated using an 11-point VAS scale. All other questions are answered with a consistent three-point Likert scale.

#### *ICF classification*

Seven questions refer to impairments of body function, 13 questions refer to activity limitations and one question refers to environmental factors.

#### *Item masking*

Item masking is present because three different dimensions are asked about: pain, functional pain and abilities of daily living.

#### *Item weighting*

The sum score is influenced by a weighting bias since three answer options exist for ability and functional pain assessment compared to 11 options to indicate pain.

#### *Cross-contamination*

Cross-contamination is not present.

#### *Suggestion for an unbiased use*

It is recommended not to use the total sum score. Sub-scores provide valuable information instead. Suggested grouping: a sub-score for back and leg pain questions, a sub-score for the ten functional pain questions and a sub-score for the five questions on abilities of daily living.

#### *Conclusion*

The sub-scores provide helpful information and the questionnaire is recommended.

### The Waddell Disability Index (WDI) [24]

#### *Questions*

With its nine questions, the WDI is the shortest questionnaire discussed in the present paper. Pain is indirectly addressed in one question (Q. 9) that relates to back pain. All the other questions relate to activities. Some questions use the word “or” (e.g. “help required or avoid heavy lifting...”) and therefore target two different foci.

#### *Offered answers*

Answer options are simple yes/no statements. Scaling problems are not present.

### ICF classification

One question refers to impairments of body function (sleep), one question refers to environmental factors, one question refers to need of help and the remaining six questions refer to activity limitations.

### Item masking

Item masking is present because three ICF domains are included.

### Item weighting

Item weighting is present because restrictions of sex and social life are measured, and therefore put weight on issues that might be of minor importance to some individuals.

### Cross-contamination

Cross-contamination is present, e.g. question 9 (“Help regularly required with footwear...”) can depend on hip arthritis as well as on low back pain.

### Suggestion for an unbiased use

It is recommended to exclude the questions about social life restrictions, sex life and sleep from the sum score. The

questionnaire could increase its value if the questions using the word “or” were redesigned.

### Conclusion

The WDI is a simple and short questionnaire with slight structural problems. It can be helpful in certain circumstances.

### Conclusions

Most of the questionnaires suffer from different score biases. This issue can often be solved by adapting the scoring algorithm. Changes in this algorithm do not interfere with the validity of the questionnaire but they do reduce score bias and increase the data value. Questionnaires with structural problems within questions and answers would need to be reworded. Changes in this aspect would need a re-validation process.

The RMDQ or the QBPDS are recommended for disability assessment, the LBPRS is recommended for functional pain evaluation, and the LBOS is recommended as a short general assessment for backache, pain medication, ability to work, leisure activities etc.. These questionnaires fulfil the validation process discussed in part I, as well as the considerations set out here. These condition-specific questionnaires should be combined with general health assessment tools like the EuroQuol [2, 8], the WHO DAS II [23] or the SF36 [25] for quality assessment in low back patients.

### References

- Bombardier C (2000) Outcome assessment in the evaluation of treatment of spinal disorders. Summary and general recommendations. *Spine* 25:3100–3103
- Brooks R (1996) EuroQuol: the current state of play. *Health Pol* 37:53–72
- 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
- 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
- Dubs L (2000) Der Score Bias. *Schweiz. Aertzezeitung* 81(6):292–295
- Dubs L (1999) [Everything was done – patient still not satisfied: introduction to evidence-based surgery]. *Swiss Surg* 5(4):160–166
- Dubs L (2000) [The patient as expert – introduction to evidence-based orthopedics]. *Z Orthop Ihre Grenzgeb* 138: 289–294
- Euroquol-group (1990) Euroquol – a new facility for the measurement of health-related quality of life. *Health Pol* 16:199–208
- Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66:271–273
- Gray DB, Hendershot GE (2000) The ICDH-2: developments for a new era of outcomes research. *Arch Phys Med Rehabil* 81:S10–4
- Greenough CG, Fraser RD (1992) Assessment of outcome in patients with low-back pain. *Spine* 17:36–41
- Hefti F (1996) Numerische “Scores” zur Evaluation von Behandlungsergebnissen am Bewegungsapparat. *Schweiz. Aertzezeitung* 12:492–496
- Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood Dauphinee S, Lamping DL, Williams JI (1995) The Quebec Back Pain Disability Scale. Measurement properties. *Spine* 20: 341–352
- 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
- Million R, Hall W, Nilsen KH, Baker RD, Jayson MI (1981) Assessment of the progress of the back-pain patient 1981 Volvo Award in Clinical Science. *Spine* 7:204–212
- Forst R (2000) Question wording – Zur Formulierung von Fragebogen-Fragen. *ZUMA, Mannheim* 2:1–11

- 
17. Raspe HH (1997) [Minimum requirements in medical expert assessment of occupational disability of patients with chronic nonspecific pain]. *Versicherungsmedizin* 49:118–125
  18. Roland M, Fairbank J (2000) The Roland-Morris Disability Questionnaire and the Oswestry Disability Index. *Spine* 25:3115–3124
  19. 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
  20. Ruta DA, Garratt AM, Wardlaw D, Russell IT (1994) Developing a valid and reliable measure of health outcome for patients with low back pain. *Spine* 19:1887–1896
  21. Steiner D, Normann G (1995) Health measurement scales. A practical guide to their development and use. Oxford Medical, 2nd edn.
  22. Stucki G, Cieza A, Ewert T, Kostanjsek N, Chatterji S, Ustun TB (2002) Application of the International Classification of Functioning, Disability and Health (ICF) in clinical practice. *Disabil Rehabil* 24:281–282
  23. Tubergen A van, Landewe R, Heuft-Dorenbosch L, Spoorenberg A, van der Heijde D, van der Tempel H, van der Linden S (2003) Assessment of disability with the World Health Organisation Disability Assessment Schedule II in patients with ankylosing spondylitis. *Ann Rheum Dis* 62:140–145
  24. Waddell G, Main CJ (1984) Assessment of severity in low-back disorders. *Spine* 9:204–208
  25. Ware JEJ, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30:473–483
  26. WHO (2001) The International Classification of Functioning, Disability and Health – ICF. WHO, Geneva
  27. Wood PL (1980) Introduction to the International Classification of Impairments, Disabilities and Handicaps. (ed), WHO, Geneva
  28. Zycha H (1998) Theorie und Erfahrung in der Medizin: Orthodoxe Wissenschaft und ganzheitliche Kybernetik. *Forsch Komplementarmed* 5 [Suppl] S1: 52–59