

V. Exner · P. Keel

Psychiatrische Universitätspoliklinik, Kantonsspital, Basel

Erfassung der Behinderung bei Patienten mit chronischen Rückenschmerzen

Validierung einer deutschen Version des „Roland & Morris disability questionnaire“ sowie verschiedener numerischer Ratingskalen

Zusammenfassung

Fragestellung. Ziel dieser Arbeit war es, 3 Instrumente zur Erfassung der momentanen subjektiv erlebten Behinderung chronischer Rückenschmerzpatienten vorzustellen und über erste Evaluationen zu berichten: 1. Behinderungsfragebogen (RM) – eine deutschsprachige Version des Roland & Morris disability questionnaire (RDQ), 2. numerische Ratingskala zur Erfassung der gesamthaften Behinderung (QL1) und 3. 8 numerische Ratingskalen zur Erfassung der bereichsspezifischen Behinderung (QL3).

Methode. Die psychometrische Prüfung der Instrumente basiert auf einer Stichprobe von 345 Rückenpatienten. Für den Behinderungsfragebogen (RM) liegen zusätzlich Daten zur Konstruktvalidität aus einer weiteren Studie mit 41 Rückenpatienten vor.

Ergebnisse. Alle 3 Instrumente erwiesen sich als reliabel und änderungssensitiv. Bei der Überprüfung der Validität bestätigte sich, dass Patienten in der Wahrnehmung ihrer Behinderung nur wenig von ihrer momentanen Leistungsfähigkeit beeinflusst sind, dafür aber stärker von ihrer momentanen psychischen Befindlichkeit.

Schlussfolgerungen. Mit dem *Behinderungsfragebogen RM* liegt eine deutsche

Version des im englischsprachigen Raum weit verbreiteten Roland & Morris disability questionnaire vor, welche die Erfassung der subjektiv erlebten Behinderung von Rückenpatienten ohne großen Zeitaufwand für Patient und Auswerter ermöglicht. *Numerische Ratingskalen QL1 und QL3:* Die 8 numerischen Ratingskalen zur Erfassung der bereichsspezifischen Behinderung (QL3) haben alle Vorteile einer einzelnen numerischen Ratingskala (Einfachheit der Instruktion, hohe Plausibilität für den Laien sowie schnelle und wenig aufwendige Durchführbarkeit), zeigen aber ein differenzierteres Bild von der Behinderung des Patienten, was einen Vergleich zu unterschiedlichen Messzeitpunkten möglich macht. Sie sind deshalb der einzelnen numerischen Ratingskala zur Erfassung der gesamthaften Behinderung (QL1) vorzuziehen. Numerische Ratingskalen sind für Patienten mit einer geringen Abstraktionsfähigkeit nicht geeignet. Aufgrund der besseren Testqualitäten ist QL3 RM vorzuziehen.

Schlüsselwörter

Rückenschmerzen · Behinderung · Roland & Morris disability questionnaire · Numerische Ratingskalen (NRS)

Die Folgen einer Erkrankung oder Verletzung lassen sich allgemein mit den Begriffen

- *Schädigung* (impairment),
- *Behinderung* (disability) und
- *Benachteiligung* (handicap)

beschreiben [18]. Während der Begriff „*impairment*“ die physikalische Einschränkung einer Funktion des Körpers meint und als Abweichung von der vorgegebenen biomedizinischen Bezugsnorm definiert ist, bedeutet „*disability*“ eine fehlende oder eingeschränkte Fähigkeit in der Ausübung von bestimmten Verhaltensweisen und Aktivitäten, die funktionell als normal anzusehen sind. Der Begriff „*handicap*“ meint die Einschränkungen bei der Ausübung von Rollenaktivitäten einer Person, die Folge einer Schädigung oder Behinderung sein können [7, 16].

Gerade im Bereich der chronischen Rückenschmerzkrankungen gibt es aber keine kausale Beziehung zwischen

V. Exner
Psychiatrische Universitätspoliklinik,
Kantonsspital, Petersgraben 4, 4031 Basel,
Schweiz

V. Exner · P. Keel

Measuring disability of patients with low-back pain – validation of a German version of the Roland & Morris disability questionnaire

Abstract

Introduction. In this study three instruments measuring disability of patients with low-back pain are presented and evaluated: (1) the Behinderungsfragebogen (RM) – a German version of the Roland & Morris disability questionnaire (RDQ) (2) a numerical rating scale measuring disability in general and (2) eight numerical rating scales measuring specific dimensions of disability (standing, sitting, walking, driving a car, carrying light loads, carrying heavy loads, sleeping, and sexual intercourse).

Methods. The psychometric evaluation, including the item analysis, test reliability, test validity, and responsiveness of the instruments, is based on two samples. Sample A comprises 345 patients with low-back pain: 282 of these patients took part in the Swiss multicentre intervention study testing the effectiveness of in-patient rehabilitation of sub-chronic and chronic low-back pain under an integrative group treatment program. The instruments were administered at different times in the therapeutic process (t1: at hospital admission; t4: follow-up after one year). 63 patients were hospitalized (orthopedic or rheumatological units) for medical examinations (myelography or infiltration of facets) or rehabilitation of low-back pain.

The instruments were administered twice within 24 h to measure test-retest correlation. In order to determine the psychometric parameters as accurately as possible, the two samples were examined jointly. Sample B is composed of 41 patients with low-back pain participating in the study „Prädiktoren des Erfolgs bei stabilisierenden Wirbelsäuleneingriffen“ (Success predictors of effectiveness of surgical interventions for spinal stabilization).

Results. All instruments proved to be generally reliable and valid (high or medium correlations with each other and with a German version of the Oswestry Disability Questionnaire) as well as responsive tools for measuring the momentary disability of patients with back pain. The psychometric examination of the test validity showed that patients' perceptions of their disability were influenced by their psychological well-being. The correlation between the 3 instruments

and physical tests was low. The RM is not a homogeneous instrument. Factor analysis (principal component analysis, rotation Varimax) indicated 6 factors. Because of the small number of items for each factor it is not appropriate to treat RM in terms of dimensions of disability.

Conclusions. The RM is an instrument measuring patients' perception of their disability that offers simple, fast practicability for patients and tester. The 2 rating scales: The 8 numerical rating scales measuring specific dimensions of disability (QL3) offer all the advantages of the numerical rating scale measuring disability in general (QL1) (simple instruction, high plausibility for the patients, and simple, fast practicability), but they provide more information about the patient's disability, which allows comparisons of disability at different times in the therapeutic process. Numerical rating scales are not suitable for patients with poor ability to abstract. For these patients it is necessary to use a questionnaire which asks concretely about what the patient can or cannot do (e.g. RM). Because of its better psychometric properties, the QL3 should be favored over the RM.

Keywords

Low-back pain · Pain-related disability · Roland & Morris disability questionnaire · Numerical rating scales (NRS)

Schädigung, Schmerz und Behinderung: Entweder fehlen medizinisch objektive Befunde, die auf eine Schädigung hindeuten, völlig (trotz Anwendung moderner und kostenintensiver Techniken wie Computertomographie und Kernspintomographie) oder die Befunde reichen nicht aus, um die vom Patienten erlebten Schmerzen und Behinderungen zu erklären. Andererseits gibt es immer wieder Patienten, die trotz erheblicher somatischer Schädigung nur wenig über Schmerzen klagen und sich in ihren Alltagsaktivitäten nur wenig eingeschränkt fühlen. Diese Diskrepanz zwischen somatischer Schädigung einerseits und erlebter Schmerzintensität und Behinderung andererseits weist auf die Bedeutung psychischer Faktoren bei der Wahrnehmung und Bewertung somatischer Schädigungen hin.

Eine umfassende Abklärung des Patienten sollte daher nicht nur den somatischen Status des Patienten erfassen, sondern auch sein subjektives Krankheitserleben mit einbeziehen: Neben der Schmerzintensität gibt auch das Ausmaß der Behinderung Auskunft über den Schweregrad der Erkrankung. Eine zusätzliche Erfassung der Behinderung ist daher angebracht [16].

Konzeptionell und methodisch lassen sich verschiedene Möglichkeiten der Behinderungserfassung unterscheiden [7, 10, 14]: Neben *Fremdeinschätzungsverfahren* (beispielsweise durch das Klinikpersonal) gewinnen zunehmend auch Verfahren der *Selbsteinschätzung* (durch den Patienten) an Bedeutung. Hierbei wird der vom Patienten wahrgenommene Grad der Behinderung erfasst. Des Weiteren lassen sich *syndromübergreifende* (d. h. von den besonderen Gegebenheiten der zugrunde liegenden Erkrankung abstrahierende) von *syndromspezifischen* (d. h. auf die Probleme der speziellen Erkrankung zugeschnittenen) Erhebungsinstrumenten unterscheiden. Ein 3. Unterscheidungsmerkmal besteht darin, ob lediglich einzelne *Bewegungsfunktionen* mit unterschiedlicher Relevanz im Alltag geprüft werden oder ob der Patient direkt nach seinen Einschränkungen bei *Alltagstätigkeiten* befragt wird. Hinsichtlich der Befragung des Patienten nach seinen Schwierigkeiten im Alltag lassen sich grob *Fähigkeitsfragen* (capacity based questions) von *Fragen nach dem aktuellen Tun* (performance based questions) unterscheiden. Bei Fähigkeitsfragen werden die Patienten nach der persönlichen Einschätzung gefragt („Sind sie in der Lage, eine bestimmte Bewegung bzw. Tätigkeit auszuführen?“). Bei Fragen nach dem aktuellen Tun lautet die Frage: „Tun sie das?“ Das Problem bei Fähigkeitsfragen ist, dass Patienten ihre Fähigkeiten häufig über- oder unterschätzen. Deshalb ist es besser, danach zu fragen, was ein Patient wirklich tut, statt was ein Patient glaubt, tun zu können [2].

Messinstrumente

Ziel dieser Arbeit ist es, die 3 folgenden Instrumente zur Erfassung rücken-schmerzbezogener Behinderungen vorzustellen und über Untersuchungen zu deren Testqualitäten zu berichten:

Wenn Ihnen Ihr Rücken weh tut, kann es für Sie schwierig sein, gewisse alltägliche Tätigkeiten auszuführen. Die folgende Liste enthält Aussagen von Leuten, die unter Rückenschmerzen gelitten haben. Wenn Sie die Liste durchgehen, stossen Sie möglicherweise auf Sätze, die für Sie am **heutigen Tag** Geltung haben. Denken Sie beim Durchlesen an Ihre Situation **heute**. Wenn Sie eine Aussage lesen, die für Sie am heutigen Tag zutrifft, kreisen Sie die Zahl 2 hinter der Aussage ein. Trifft die Aussage jedoch nicht zu, so kreisen Sie die Zahl 1 ein.

Datum		trifft nicht zu	trifft zu
RM1	Wegen meinem Rücken bleibe ich die meiste Zeit zu Hause.	1	2
RM2	Ich ändere meine Körperhaltung häufig, um so für meinen Rücken eine bequeme Haltung zu finden.	1	2
RM3	Ich gehe langsamer als gewohnt wegen meinem Rücken.	1	2
RM4	Wegen meinem Rücken kann ich meine gewohnten Tätigkeiten zu Hause nicht verrichten.	1	2
RM5	Ich benütze beim Treppensteigen den Handlauf (Treppengeländer) wegen meinem Rücken.	1	2
RM6	Wegen meinem Rücken lege ich mich vermehrt hin, um auszurufen.	1	2
RM7	Wegen meinem Rücken muss ich mich beim Aufstehen aus einem Sessel an etwas festhalten.	1	2
RM8	Wegen meinem Rücken versuche ich, andere Leute dazu zu bringen, Dinge für mich zu tun.	1	2
RM9	Wegen meinem Rücken benötige ich mehr Zeit zum Ankleiden als sonst.	1	2
RM10	Ich stehe jeweils nur für kurze Zeit auf wegen meinem Rücken.	1	2
RM11	Wegen meinem Rücken vermeide ich wenn möglich Bücken und Niederknien.	1	2
RM12	Wegen meinem Rücken habe ich Mühe, mich von einem Stuhl zu erheben.	1	2
RM13	Mein Rücken tut fast immer weh.	1	2
RM14	Ich habe Mühe, mich im Bett zu drehen wegen meinem Rücken.	1	2
RM15	Mein Appetit ist nicht sehr gut wegen meinem Rücken.	1	2
RM16	Wegen meinem Rücken habe ich Mühe, die Socken (oder Strümpfe) anzuziehen.	1	2
RM17	Ich gehe nur kurze Strecken wegen meinem Rücken.	1	2
RM18	Ich schlafe weniger gut wegen meinem Rücken.	1	2
RM19	Wegen meinem Rücken muss mir jemand beim Ankleiden behilflich sein.	1	2
RM20	Tagsüber sitze ich die meiste Zeit wegen meinem Rücken.	1	2
RM21	Ich vermeide schwerere Arbeiten zu Hause wegen meinem Rücken.	1	2
RM22	Wegen meinem Rücken bin ich im Umgang mit anderen Leuten schlecht gelaunt und mehr gereizt als sonst.	1	2
RM23	Wegen meinem Rücken bin ich langsamer als sonst beim Treppensteigen.	1	2
RM24	Ich bleibe die meiste Zeit im Bett wegen meinem Rücken.	1	2

Abb. 1 ▲ Behinderungsfragenbogen (RM)

- Behinderungsfragebogen (RM)
- 11-stufige numerische Ratingskala zur Erfassung der gesamthaften Behinderung (QL1)
- 8 11-stufige numerische Ratingskalen zur Erfassung der bereichsspezifischen Behinderung (QL3)

Behinderungsfragebogen

Der *Behinderungsfragebogen (RM)* [13] (Abb. 1) ist eine Übersetzung und geringe Modifikation des Roland disability questionnaire (RDQ) [17]. Dazu wurde das Original nach der Forward-backward-translation-Methode aus dem Englischen übersetzt.

Der RDQ geht auf eine Auswahl von Aussagen des Sickness impact profile (SIP) zurück. Das SIP ist ein Fragebogen zum allgemeinen Gesundheitsstatus und wird bei einer Vielzahl von chronischen Erkrankungen eingesetzt [1, 2]. 24

Aussagen wurden so ausgewählt, dass sie die wichtigsten Aspekte des täglichen Lebens abdecken. Die Phrase „because of my back“ kam zu jeder Aussage hinzu, um deutlich zu machen, dass es hier nur um die Behinderung geht, die durch den Rücken verursacht wird.

Der Fragebogen ist so konstruiert, dass er vom Patienten selbst ohne Hilfe innerhalb von 5 min ausgefüllt werden kann. Trifft eine Aussage am Tag der Untersuchung (today) zu, ist der Patient gebeten, diese anzukreuzen. Für jedes angekreuzte Item wird 1 Punkt vergeben. Ein individueller Patientenscore variiert also von 0 Punkten (keine Behinderung) bis 24 Punkten (schwerste Behinderung) [17]. Der RDQ hat sich als ein reliables, valides und änderungssensitives Instrument erwiesen, das auch zwischen Patienten mit unterschiedlich starken Rückenschmerzen differenzieren kann [2, 17].

Die deutsche Version unterscheidet sich vom Original [17] und auch von der deutsch-österreichischen Version von Wiesinger et al. [19] insofern, dass auch bei nichtzutreffenden Items dies vom Patienten vermerkt werden muss (2-stufige Ratingskala: „trifft nicht zu“ vs. „trifft zu“). Dadurch ist es erst möglich, fehlende Angaben von nichtzutreffenden Items zu unterscheiden. Mit Ausnahme von Item 4 werden die Patienten hinsichtlich ihres aktuellen Tuns befragt.

Numerische Ratingskala

Die *numerischen Ratingskalen* sind Unterskalen des Fragebogens „Beeinträchtigung der Lebensqualität (QL)“ (Deutschmann, pers. Mitteilung) – eines mehrdimensionalen Fragebogens zur Erfassung der Lebensqualität bei chronischen Rückenschmerzpatienten, der im Rahmen des Nationalen Forschungsprogramms Nr. 26 B „Chronifizierung von Rückenschmerzen“ [13] als Ergänzung zu den übrigen eingesetzten Fragebögen entwickelt wurde.

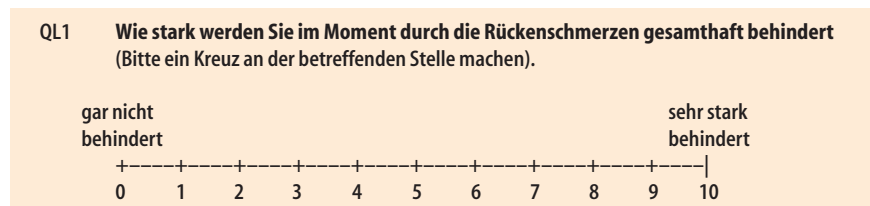
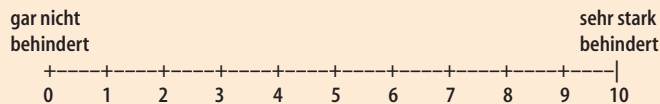


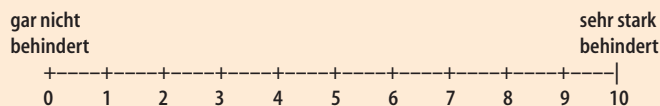
Abb. 2 ▲ Numerische Ratingskala QL1

QL3 Patienten haben uns häufig berichtet, dass sie bei verschiedenen Alltagsverrichtungen durch die Rückenschmerzen behindert oder beeinträchtigt sind. In welchem Ausmass fühlen Sie sich im Vergleich zu früher, als Sie noch gesund waren, behindert bei folgenden Tätigkeiten.
(Bitte ein Kreuz auf der Skala an der betreffenden Stelle machen).

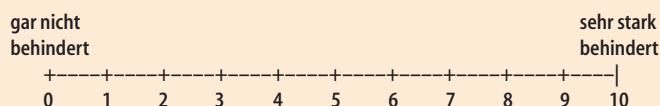
QL3.1 Beim Stehen:



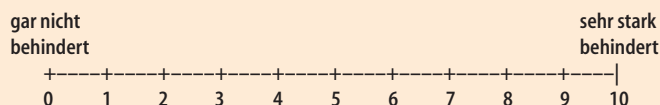
QL3.2 Beim Sitzen:



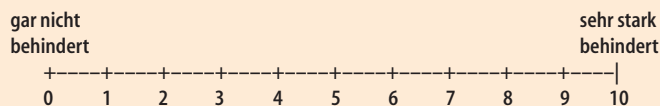
QL3.3 Beim Gehen auf ebener Erde:



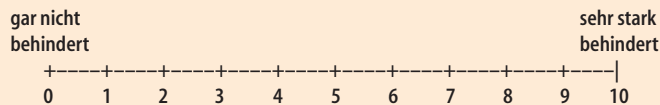
QL3.4 Beim Autofahren oder Benützen von Öffentlichen Verkehrsmitteln (Tram, Bus, Bahn, Flugzeug, etc.):



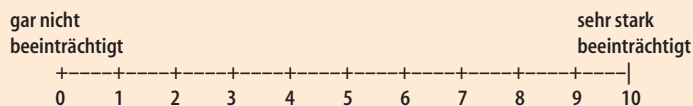
QL3.5 Beim Tragen leichter bis mittelschwerer Lasten (z.B. Einkaufstasche/Mappe):



QL3.6 Beim Tragen schwerer Lasten (mindestens 10kg, z.B. Eimer voll Wasser):

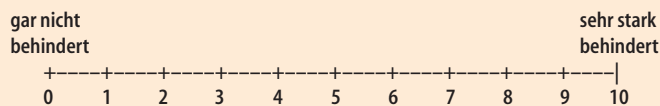


QL3.7 Beim Schlafen:



QL9 Die restlichen Fragen betreffen nur Personen mit einem festen Partner, mit welchem sie zusammenleben oder regelmässig zusammen sind. (Andernfalls leer lassen). Es sind Fragen zum Einfluss der Rückenschmerzen auf ihre Partnerschaft und die Sexualität.

QL10 **Vorerst ist wichtig, dass Sie uns angeben, wie stark Sie sich beim Geschlechtsverkehr durch ihre Schmerzen behindert fühlen.**
(Machen sie bitte ein Kreuz auf der Skala).



Gesamthafte Behinderung (QL1)

Die *gesamthafte Behinderung (QL1)* wird mittels einer 11-stufigen numerischen Ratingskala erfasst (0=gar nicht behindert bis 10=sehr stark behindert) (Abb. 2).

Die Vorteile numerischer Ratingskalen liegen in deren Einfachheit der Instruktion, in der hohen Plausibilität für den Laien sowie in der schnellen und wenig aufwendigen Durchführbarkeit [11]. Lediglich Patienten mit einer geringen Abstraktionsfähigkeit tun sich mit diesem Verfahren schwer [15]. Problematisch ist aber die Erfassung eines mehrdimensionalen Konstrukts wie Behinderung mittels eines einzigen Items (Single-Item-Verfahren): Es bleibt unklar, welche Komponenten in welchem Ausmass zu den verschiedenen Messzeitpunkten kovariieren. Eine Vergleichbarkeit der Messwerte ist dadurch erschwert.

Bereichsspezifische Behinderung (QL3)

Einen Kompromiss stellt das Instrument zur Erfassung der *bereichsspezifischen Behinderung (QL3)* dar (Abb. 3). Mittels 11-stufiger numerischer Ratingskalen (0=gar nicht behindert bis 10=sehr stark behindert) werden die Patienten nach dem Ausmass ihrer Behinderung bei 8 verschiedenen Alltagsverrichtungen befragt. Die bereichsspezifische Behinderung ergibt sich aus dem Mittelwert der Einzelwerte.

Psychometrische Eigenschaften

Statistik

Alle statistischen Berechnungen wurden mit dem Tabellenkalkulationsprogramm Excel (Version 5.0, Microsoft Corporation 1993–1994) sowie dem Statistikprogramm SPSS (Version 8, SPSS Inc., Chicago, IL) durchgeführt. Da die vorliegenden Variablen ordinalskaliert sind, kamen nichtparametrische Verfahren zur Anwendung [3, 4].

Die Qualität eines Messinstruments ist von der Art und der Zusammensetzung seiner Items abhängig. In einem 1. Schritt wurden deshalb die Instrumente

Abb. 3 ◀ Numerische Ratingskala QL3

Tabelle 1

Soziodemographische Daten der Stichprobe A

Stichprobe A	Alter ^a			Geschlecht ^b	
	n	MW	Std	Spanne	Männlich
345	44,63	11,62	17–84	204	121

^aBei 24 Patienten keine Angaben^bBei 20 Patienten keine Angaben

n Stichprobenumfang, MW Mittelwert, Std Standardabweichung

(soweit möglich) einer *Itemanalyse* (Rohwertverteilung, Itemschwierigkeit bzw. Zustimmungsrate, Trennschärfe, Homogenität sowie Überprüfung der Dimensionalität) unterzogen. In einem 2. Schritt erfolgte die Schätzung der *Reliabilität* nach der Test-Retest-Methode, ggf. auch nach der Split-half-Methode sowie mittels Konsistenzanalyse.

Zur Bestimmung der *Konstruktvalidität* wurden die zu evaluierenden Instrumente untereinander sowie mit verschiedenen anderen Instrumenten und Maßen korreliert (Rangkorrelationskoeffizient nach Spearman):

1. Deutsche Version des Oswestry disability questionnaire (ODQ) zur Bestimmung des funktionellen Niveaus von chronischen Rückenschmerzpatienten (Gerbershagen, pers. Mitteilung, Keel, pers. Mitteilung) [9]. Die Skala besteht aus 10 Subskalen:

- Einnahme von Schmerzmitteln,
- Körperpflege,
- Heben von Lasten,
- Gehen,
- Sitzen,
- Stehen,
- Schlafen,
- Sexual-,
- Gesellschaftsleben und
- Reisen.

Jede Subskala enthält 6 Aussagen zur Auswahl, die nach dem Grad der Funktionseinschränkung geordnet sind.

2. Körperliche Leistungsfähigkeit (Übung zur Kräftigung der Bauch- und Rückenmuskulatur: Anzahl der Wiederholungen, Rotationsbewegung an der Kraftmaschine: total kg rechts und links, Leistungstest am

Fahrradergometer: maximale Leistung in Watt) [13].

3. Fragebogen „Beschreibung der Schmerzen (PQ)“ [13] (Original: „short-form McGill pain questionnaire“ [14] mit Schmerzzeichnung von Capra et al. [6]).
4. Fragebogen „Allgemeine Verfassung (GW)“ [5, 13] (Original „psychological general well-being“ – PWGB – [8]) zur Erfassung von 6 Dimensionen der psychischen Befindlichkeit:

- (Nicht-)Ängstlichkeit,
- (Nicht-)Depressivität,
- Wohlbefinden,
- Selbstkontrolle,
- Gesundheit und
- Vitalität.

Ein hoher Gesamtsummenscore gibt ein hohes Ausmaß an subjektivem Wohlbefinden an.

5. Instrument zur Bestimmung der psychosozialen (Nicht-)Belastung aus dem Fragebogen „Beeinträchtigung der Lebensqualität (QL)“ (Deutschmann, pers. Mitteilung) [13]. Auch hier gibt ein hoher Wert ein hohes Ausmaß an psychosozialer Nichtbelastung an.

Die *Änderungssensitivität* der zu evaluierenden Instrumente wurde mittels Wil-

coxon-Test (Vergleich zwischen den Zeitpunkten t₁ und t₄ der multizentrischen Interventionsstudie bei chronischen Rückenschmerzen [13]) überprüft.

Stichproben

Die teststatistischen Kennwerte der zu prüfenden Messinstrumente basieren auf folgenden Stichproben:

1. Stichprobe A

Stichprobe A setzt sich aus

- 282 Patienten einer Interventionsstudie bei chronischen Rückenschmerzen [13] sowie
- 63 Rückenschmerzpatienten, die unabhängig von der Interventionsstudie ausschließlich zur Ermittlung der teststatistischen Kennwerte der zu prüfenden Messinstrumente, v. a. der Test-Retest-Reliabilität, rekrutiert wurden,

zusammen. Um die psychometrischen Eigenschaften der Instrumente möglichst präzise einschätzen zu können, wurden im Verlauf der Untersuchungen beide Stichproben zusammengefasst. In Tabelle 1 sind die Alters- und Geschlechtsverteilung der Gesamtstichprobe aufgeführt. Die 282 Patienten der ersten Teilstichprobe nahmen an einer multizentrischen Therapiestudie teil. Vergleichbar wurden ein traditionelles, mehr auf aktive und passive Einzelmaßnahmen basierendes Behandlungskonzept und ein neuartiges Gruppenbehandlungsprogramm, das aus einem Kraft- und Ausdauertraining besteht, kombiniert mit einem allgemeinen Fitnesstraining sowie unterstützenden pädagogischen und psychologischen Maßnahmen. Die Patienten wurden

- beim Eintritt in die Klinik (t₁),
- bei der Entlassung (t₂) sowie

Tabelle 2

Soziodemographische Daten der Stichprobe B

Stichprobe A	Alter ^a			Geschlecht ^b	
	n	MW	Std	Spanne	Männlich
41	45,9	15,5	18–78	21	20

n Stichprobenumfang, MW Mittelwert, Std Standardabweichung

Tabelle 3

Itemanalyse von RM zum Zeitpunkt t1

Behinderungsfragebogen (RM) Items und Skala	Rohwertverteilung			Schwierigkeitsindizes p	Trennschärfeindizes r_{it}	Homogenitäten r_{ii}
	n	MW	Std			
RM	281	13,73	4,70			0,17
1 (meistens zu Hause)				0,51	0,49	0,22
2 (bequeme Körperhaltung)				0,94	0,18	0,08
3 (gehen/langsam)				0,79	0,44	0,19
4 (Arbeiten/zu Hause)				0,64	0,44	0,20
5 (Treppe/mit Geländer)				0,67	0,49	0,21
6 (liegen, um auszuruhen)				0,82	0,40	0,18
7 (aufstehen/Sessel)				0,64	0,50	0,22
8 (Hilfe von anderen)				0,46	0,26	0,12
9 (anziehen/langsam)				0,65	0,50	0,22
10 (selten aufstehen)				0,29	0,22	0,10
11 (bücken oder niederknien)				0,73	0,37	0,17
12 (aufstehen/Stuhl)				0,60	0,55	0,24
13 (Schmerzen)				0,82	0,36	0,16
14 (drehen im Bett)				0,68	0,41	0,18
15 (Appetit)				0,16	0,26	0,12
16 (anziehen/Socken)				0,73	0,41	0,18
17 (gehen/kurze Strecke)				0,55	0,53	0,23
18 (schlafen)				0,71	0,41	0,18
19 (anziehen/Hilfe)				0,07	0,26	0,12
20 (meistens sitzen)				0,20	0,17	0,07
21 (schwere Arbeiten)				0,87	0,26	0,12
22 (Stimmung)				0,39	0,27	0,13
23 (Treppe/langsam)				0,75	0,49	0,21
24 (meistens im Bett)				0,14	0,32	0,14

n Stichprobenumfang, MW Mittelwert, Std Standardabweichung

- 3 (t₃) und
- 12 Monate nach Abschluss der Behandlung (t₄)

einer ausgedehnten Evaluation unterzogen. Dabei wurden auch die zu prüfenden Instrumente eingesetzt. Die 63 Patienten der 2. Teilstichprobe waren an der Orthopädischen und Rheumatologischen Universitätsklinik Basel sowie an der Rheumatologischen Abteilung des Bethesda-Spitals Basel zur stationären medizinischen Abklärung (mittels Myelographie oder Facetteninfiltration) oder Rehabilitation. Zur Schätzung der Test-Retest-Reliabilität wurden diesen Patienten die zu prüfenden Verfahren innerhalb von 24 h 2-mal zur Bearbeitung vorgelegt.

2. Stichprobe B

Für den Behinderungsfragebogen (RM) liegt zusätzlich ein wichtiges Ergebnis zur Konstruktvalidität aus

der Studie „Prädiktoren des Erfolgs bei stabilisierenden Wirbelsäuleneingriffen“ (Keel, pers. Mitteilung) vor. Ziel dieser Studie war es, den Einfluss psychosozialer Variablen auf den Operationserfolg bei stabilisierenden Eingriffen an der Lendenwirbelsäule zu untersuchen. Die Patienten wurden

- bei der Aufnahme (t₁),
- 6 (t₂) und
- 12 Monate nach Abschluss der Behandlung (t₃)

einer ausgedehnten Evaluation unterzogen. Dabei wurde auch der zu prüfende Fragebogen RM zusammen mit einer deutschen Version des Oswestry disability questionnaire (ODQ) (Gerbershagen, pers. Mitteilung) [9] eingesetzt. In Tabelle 2 sind die Alters- und die Geschlechtsverteilung dieser Stichprobe aufgeführt.

Ergebnisse

Behinderungsfragebogen (RM)

Bezüglich der *Testinstruktion* hatten manche Patienten Schwierigkeiten, zu entscheiden, ob sie ein Item bejahen oder verneinen sollten, wenn ihre Situation schlimmer war als darin beschrieben [Bemerkung eines Patienten zu Item 5 (Treppe/nur mit Geländer): „Ich kann überhaupt nicht mehr laufen, was soll ich jetzt ankreuzen?“]. Dem Konzept des Fragebogens entsprechend, dass die Anzahl der „trifft zu“-Antworten das Ausmaß der Behinderung anzeigt, sollten diese Items auch bei noch größeren Einschränkungen bejaht werden. Es scheint daher wichtig, die Patienten ausdrücklich darauf hinzuweisen, da sonst Patienten mit extremer Behinderung als wenig behindert imponieren würden. Für die Zukunft wäre daher zu überlegen, die Testinstruktion zu erweitern

Tabelle 4
Reliabilitätskoeffizienten von RM
zum Zeitpunkt t1

Behinderungsfragen (RM)	n	r
Split-half (Spearman-Brown)	281	0,83
Split-half (Guttman)	281	0,83
Konsistenzanalyse (Kuder-Richardson-Formel 20)	281	0,83
Test-Retest	62	0,85

oder bei bestimmten Items die Phrase „oder schlimmer“ zu ergänzen.

Die *Itemanalyse* erbrachte folgende Resultate (Tabelle 3): Mit durchschnittlich 14 von 24 möglichen Punkten bei einer Standardabweichung von 5 Punkten gaben die Patienten eine Behinderung an, die nur wenig über dem mittleren Wert lag. Hinsichtlich der Schwierigkeitsindizes zeigte sich, dass die Items 15 (kein Appetit), 19 (anziehen/Hilfe) und 24 (meistens im Bett) überdurchschnittlich selten und die Items 2 (bequeme Körperhaltung) und 21 (schwere Arbeiten) überdurchschnittlich häufig bejaht wurden. Die Berechnung der Trennschärfe ergab, dass knapp 1/3 der Items $<0,30$ liegt. Als wenig trennscharf erwiesen sich die Items 2 (bequeme Körperhaltung), 8 (Hilfe von anderen), 10 (selten aufstehen), 15 (Appetit), 19 (anziehen/Hilfe), 20 (meistens sitzen), 21 (schwere Arbeiten) und 22 (Stimmung). Die Gesamthomogenität des Fragebogens liegt mit 0,17 unterhalb der empfohlenen Norm von 0,20–0,40. Als ausgesprochen wenig homogen ($r < 0,20$) erwiesen sich 2/3 der Items. Eine Faktorenanalyse (PCA mit Varimax-Rotation, Eigenwerte >1) ergab eine 6-Faktoren-Lösung mit 2–6 Items/Faktor. Faktor 1 (von uns „Gehen/Treppensteigen“ genannt) fasst die Items 3, 5, 17 und 23 zusammen, Faktor 2 („Gelenkigkeit“, v. a. das Aufstehen/Anziehen betreffend) die Items 7, 9, 12 und 16, Faktor 3 („Sich Schonen“) die Items 2, 4, 6, 8, 11 und 21, Faktor 4 („psychische Belastung sowie Schlafprobleme“) die Items 13, 14, 18 und 22 sowie Faktor 5 („Mobilitätverlust“) die Items 1, 10, 20 und 24. Faktor 6 enthält die Items 15 und 19, die bereits durch niedrige Schwierigkeitsindizes und eine geringe Trennschärfe auffielen (Tabelle 3). Diese beiden Items trugen somit wenig zum Gesamtergebnis des Fragebogens bei.

Aufgrund der geringen Itemzahl der einzelnen Dimensionen (durchschnittlich 4 Items) schien es aber wenig sinnvoll, den Fragebogen nach Dimensionen auszuwerten. Der Anteil an der Gesamtvarianz lag bei den einzelnen Faktoren zwischen 5,7% und 10,4%.

Die Schätzungen für die *Reliabilität* lagen mit Werten von 0,83–0,85 recht hoch (Tabelle 4). Sie unterschieden sich somit nur geringfügig von denen des Originals [2, 17] sowie der deutsch-österreichischen Version des Roland & Morris disability questionnaire von Wiesinger et al. [19]. Heterogene Tests lassen in der Regel keinen hohen Konsistenzkoeffizienten erwarten. Die Schätzung der Reliabilität mittels Testwiederholung erbrachte aber keinen wesentlich höheren Wert.

Die Untersuchungen zur *Konstruktvalidität* ergaben mittelmäßige Korrelationen zu den numerischen Ratingskalen [mit Werten von $r=0,41$ ($p \leq 0,01$) für QL1 und $r=0,55$ ($p \leq 0,01$) für QL3]. Mit der deutschen Version des Oswestry disability questionnaire von Gerbershagen (Gerbershagen, pers. Mitteilung) korrelierte RM präoperativ mittelmäßig [$r=0,52$ ($p \leq 0,01$)]. Eine hohe Korrelation – wie sie für die Originalverfahren beschrieben wurde – zeigte sich postoperativ [$r=0,92$ ($p \leq 0,01$)] (Keel, pers. Mitteilung). Mittelmäßige Korrelationen ergaben sich zu den verschiedenen Maßen zur psychischen Befindlichkeit [z. B. Un-

terskala (Nicht-)Ängstlichkeit: $r=-0,41$ ($p \leq 0,01$) sowie Unterskala (Nicht-)Depressivität: $r=-0,49$ ($p \leq 0,01$) von „Allgemeine Verfassung (GW)“]. Mit den unterschiedlichen Messverfahren zur Erfassung der subjektiv erlebten Schmerzintensität sowie der klinisch beobachtbaren körperlichen Leistungsfähigkeit korrelierte der Fragebogen nur wenig [Werte von $r=0,12$ ($p > 0,05$)–0,36 ($p \leq 0,01$)]. Insgesamt kamen wir somit zu ähnlichen Ergebnissen, wie sie für die Originalversion [2, 17] beschrieben wurden. Dagegen nannten Wiesinger et al. für ihre Version wesentlich höhere Korrelationen zu den von ihnen ausgewählten Schmerz- und Leistungsmaßen [19].

Der Wilcoxon-Test zeigte für RM eine signifikante Veränderung zwischen den Zeitpunkten t1 und t4 [Z ($n=235$)= $-5,769$ ($p \leq 0,001$); Mittelwert/Standardabweichung: 0,58/0,20 (t1), 0,47/0,25 (t4)], was auf eine vorhandene *Änderungssensitivität* schließen lässt.

Numerische Ratingskalen

Die wichtigsten Ergebnisse für die *numerischen Ratingskalen* zur gesamthafter (QL1) und zur bereichsspezifischen Behinderung (QL3) lassen sich wie folgt zusammenfassen: Hinsichtlich der *Testinstruktion* fällt auf, dass einige Patienten zwischen 2 Zahlenwerten ankreuzten, was so nicht vorgesehen war. Hier

Tabelle 5
Itemanalyse von QL1 und QL3 zum Zeitpunkt t1

Fragebogen	Statistik			Schwierigkeitsindizes p	Trennschärfeindizes r_{it}	Homogenitäten r_{ii}
	n	MW	Std			
QL1	300	5,71	2,24	0,57		
QL3	312	5,47	1,80			0,44
1 (stehen)	310	5,83	2,59	0,58	0,59	0,43
2 (sitzen)	310	6,15	2,47	0,62	0,62	0,45
3 (ebenerdig gehen)	308	3,96	2,61	0,40	0,60	0,43
4 (Auto fahren)	310	5,26	2,67	0,53	0,69	0,49
5 (leicht tragen)	310	5,58	2,69	0,59	0,71	0,50
6 (schwer tragen)	311	7,51	2,34	0,75	0,64	0,46
7 (schlafen)	309	4,61	2,72	0,46	0,52	0,38
QL9 (Geschlechtsverkehr) ^a	243	4,63	2,58	0,46	0,52	0,38

^a20% der Patienten gaben an, nicht in einer festen Partnerschaft zu leben
n Stichprobenumfang, MW Mittelwert, Std Standardabweichung

Tabelle 6
**Reliabilitätskoeffizient von QL1
zum Zeitpunkt t1**

Gesamthafte Behinderung (QL1)	n	r
Test-Retest	63	0,77

wäre es denkbar, in einer zukünftigen Version anstelle eines Zahlenstrahls 11 von 0–10 durchnummerierte Kästchen in einer Reihe darzubieten, auch wenn sich das Problem dadurch nicht völlig eliminieren lässt. Jensen et al. schlugen daher in ihrer Publikation eine 21-Punkte-Box-Skala vor [12]. Des Weiteren gab es bei QL1 und QL3 immer wieder Patienten, die sichtlich Mühe hatten, der von ihnen erlebten Behinderung einen Zahlenwert zuzuordnen. Ob dieses Problem auf eine mangelnde Abstraktionsfähigkeit der Patienten zurückzuführen ist, wie es in der Literatur beschrieben wurde [15], konnte nicht eruiert werden.

Die *Itemanalyse* sowie die Schätzungen zur *Reliabilität* erbrachten folgende Ergebnisse: Bei QL1 gaben die Patienten im Durchschnitt eine mittlere Behinderung an (6 von maximal 10 Punkten mit einer Standardabweichung von 2 Punkten). Der Schwierigkeitsindex lag mit 0,57 im akzeptablen Bereich (Tabelle 5). Die Test-Retest-Reliabilität von 0,77 kann als zufrieden stellend bis gut bezeichnet werden (Tabelle 6). Auch bei QL3 nahm die durchschnittliche Behinderung (Mittelwert aus den 8 Subskalen) einen mittleren Wert ein (5 von 10 Punkten mit einer Standardabweichung von 2 Punkten). Mit Schwierigkeitsindizes von 0,40–0,75 lagen alle Subskalen im akzeptablen Bereich. Die Trennschärfeindizes von >0,50 sind als hoch zu bezeichnen. Das traf auch für die Gesamthomogenität der Skala zu: Der Wert von 0,44 liegt oberhalb der empfohlenen Norm von 0,20–0,40. Als extrem homogen erwiesen sich 6 der 8 Subskalen (Tabelle 5). Die Schätzungen für die Reliabilität bewegen sich zwischen 0,83 und 0,94 (Tabelle 7), was einer hohen bis sehr hohen Reliabilität entspricht.

Die Untersuchungen zur *Konstruktvalidität* ergaben mittelmäßige Korrelationen zu RM [QL1: $r=0,41$ ($p\leq 0,01$) und QL3: $r=0,55$ ($p\leq 0,01$)]. Untereinander

korrelierten beide Maße hoch [$r=0,65$ ($p\leq 0,01$)]. Für beide Skalen ließen sich außerdem geringe Korrelationen sowohl zu den von uns eingesetzten physikalischen Tests zur körperlichen Leistungsfähigkeit [Werte von $r=-0,09$ ($p>0,05$) bis $-0,28$ ($p\leq 0,01$)] als auch zu den verschiedenen Messmethoden zur Erfassung subjektiv erlebten Schmerzes und psychischer Befindlichkeit beschreiben [Werte von $r=0,13$ ($p\leq 0,05$) bis $r=0,37$ ($p\leq 0,01$)]. Ausnahmen bildeten die mittelmäßigen Korrelationen zur Erfassung der momentanen Schmerzintensität [$r=0,56$ ($p\leq 0,01$) für QL1 und $r=0,49$ ($p\leq 0,01$) für QL3] sowie zu der Subskala „Gesundheit“ im Fragebogen „Allgemeine Verfassung (GW)“ [$r=-0,49$ ($p\leq 0,01$) für QL1 und $r=-0,46$ ($p\leq 0,01$) für QL3].

Der Wilcoxon-Test zeigte für beide Maße eine hoch signifikante Veränderung zwischen den Zeitpunkten t1 und t4 [QL1: Z ($n=207$)= $-4,746$ ($p\leq 0,001$); Mittelwert/Standardabweichung: 5,71/2,24 (t1), 4,80/2,79 (t4) und QL3: Z ($n=229$)= $-4,253$ ($p\leq 0,001$); Mittelwert/Standardabweichung: 5,47/1,80 (t1), 4,73/2,32 (t4)], was auf eine vorhandene *Änderungssensitivität* schließen lässt.

Diskussion

Alle 3 Instrumente zur Erfassung rückenschmerzbezogener Behinderung erwiesen sich als reliabel und änderungssensitiv.

Untersuchungen zur Konstruktvalidität ergaben, dass sich die tatsächliche körperliche Leistungsfähigkeit in der subjektiv erlebten Behinderung, wie sie durch die vorliegenden Instrumente gemessen wird, nur wenig widerspiegelt. Dieses Ergebnis bestätigt die immer wieder zu beobachtende und in der Literatur häufig beschriebene Diskrepanz zwischen der objektiven funktionellen Einschränkung chronischer Schmerzpatienten und ihrer subjektiv erlebten Beeinträchtigung. Die mittelmäßigen Korrelationen zu psychosozialen Maßen, wie sie für RM, bezüglich der Subskala „Gesundheit“ von GW auch für QL1 und QL3 beschrieben wurden, deuten darauf hin, dass die subjektiv erlebte Behinderung v. a. vom momentanen psychischen Befinden abhängig ist. Ob sich in der auffällig hohen Korrelation zwischen der 11-stufigen numerischen Ratingska-

la zur Erfassung der Schmerzintensität und QL1 bzw. QL3 auch eine erhebliche Beeinflussung der subjektiv erlebten Beeinträchtigung durch die momentan erlebte Schmerzbelastung des Patienten widerspiegelt oder sogar etwas Ähnliches gemessen wird, lässt sich nicht definitiv beantworten. Die Tatsache aber, dass es sich bei allen 3 Instrumenten um die gleiche Skalenart handelt (11-stufige numerische Ratingskalen), die Korrelation zum Globalmaß QL1 höher ausfiel als zu der sehr viel differenzierteren bereichsspezifischen Befragung durch QL3 und die Korrelationen zu anderen Schmerzinstrumenten wie Schmerzzeichnung und -deskriptoren gering blieben, spricht eher dafür, dass viele Patienten bei der Einschätzung ihrer Behinderung die globale Schmerzeinschätzung zugrunde legten (Ankereffekt).

Eine wichtige Frage bei der Beurteilung neuer Instrumente ist, in welcher Beziehung diese zu bereits eingesetzten Verfahren stehen. Zur Rückenschmerzversion des Funktionsfragebogens Hannover (FFbH-R) [14], dem im deutschsprachigen Raum mittlerweile am weitesten verbreiteten Verfahren, liegen uns keine Daten vor, da der FFbH-R zur gleichen Zeit entwickelt wurde. In einer weiteren Validierungsstudie sollte auf jeden Fall der FFbH-R als Außenkriterium mit einbezogen werden. Dabei könnte auch die von Beurskens et al. [2] aufgeworfenen Diskussion über den Einfluss der Itemformulierung auf die Behinderungseinschätzung durch den Patienten wieder aufgegriffen werden: Während RM die Patienten fast ausschließlich nach dem aktuellen Tun befragt, sind die Items des FFbH-R und QL3 als Fähigkeitsfragen formuliert.

Tabelle 7
**Reliabilitätskoeffizienten QL3 zum
Zeitpunkt t1**

Bereichsspezifische Behinderung (QL3)	n	r
Split-half (Spearman-Brown)	237	0,83
Split-half (Guttman)	237	0,83
Konsistenzanalyse (Cronbach α)	237	0,86
Test-Retest	63	0,94

Resümee

Bezogen auf die einzelnen Instrumente lässt sich abschließend folgende Bewertung angeben:

Mit dem *Behinderungsfragebogen (RM)* liegt eine weitere deutschsprachige Version des im englischsprachigen Raum weit verbreiteten Originals vor, das die Erfassung der subjektiv erlebten Behinderung von Rückenpatienten ohne großen Zeitaufwand für Patient und Auswerter ermöglicht. Der Fragebogen spricht alltägliche Situationen an. Patienten, denen es schwer fiel, ihrer Behinderung einen bestimmten Zahlenwert oder eine bestimmte Abstufung zuzuordnen (wie dies bei QL1 und QL3 erforderlich ist), hatten mit diesem Fragebogen keine Schwierigkeiten. Der Fragebogen erwies sich als ein reliables, valides und änderungssensitives Instrument.

Numerische Ratingskalen wie QL1 und QL3 setzen beim Patienten eine gewisse Fähigkeit zu abstraktem Denken voraus. Auch wenn sich QL1 und QL3 hinsichtlich der Gütekriterien nur geringfügig unterscheiden, ist die Verwendung von QL3 der von QL1 vorzuziehen: Die 8 numerischen Ratingskalen zur Erfassung der bereichsspezifischen Behinderung (QL3) haben alle Vorteile einer einzelnen numerischen Ratingskala (Einfachheit der Instruktion, hohe Plausibilität für den Laien sowie schnelle und wenig aufwendige Durchführbarkeit), zeigen aber ein differenzierteres Bild von der Behinderung des Patienten, was einen Vergleich zu unterschiedlichen Messzeitpunkten möglich macht.

Ein Vergleich hinsichtlich der Testqualitäten von QL3 und RM ergibt für QL3 bessere Testqualitäten.

Literatur

1. Bergner M, Bobbitt RA, Carter WB, Gilson BS (1981) Sickness impact profile. Development and final revision of health status measure. *Med Care* 19:787
2. Beurskens AJ, Vet HC de, Köke AJ, Heijden GJ van der, Knipschild PG (1995) Measuring the functional status of patients with low back pain. Assessment of the quality of four diseasespecific questionnaires. *Spine* 20:1017
3. Bortz J, Döring N (1995) Forschungsmethoden und Evaluation. Springer, Berlin Heidelberg New York
4. Bühl A, Zöfel P (1999) SPSS für Windows Version 8: Einführung in die moderne Datenanalyse unter Windows. Addison-Wesley Publishing, Bonn
5. Bullinger M, Heinisch M, Ludwig M, Geier S (1990) Skalen zur Erfassung des Wohlbefindens. Psychometrische Analysen zum Profile of Mood States (POMS) und zum Psychological General Wellbeing Index (PGWI). *Z Diff Diagn Psychol* 11:53
6. Capra P, Mayer TG, Gatchel R (1985) Adding psychological scales to your back pain assessment. *J Musculoskelet Med* 2:41
7. Dillmann U, Nilges P, Saile H, Gerbershagen HU (1994) Behinderungseinschätzung bei chronischen Schmerzpatienten. *Schmerz* 8:100
8. Du Puy HJ (1984) The psychological general well-being (PGWB) index. In: Wenger NK, Mattson ME, Furberg CD, Ellinson J (eds) Assessment of quality of life in clinical trials of cardiovascular therapies. LeJacq Publishers, New York, p 170
9. Fairbank JCT, Couper J, Davies JB, O'Brien J (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66:271
10. Franz C, Pflingsten M (1992) Schmerz- und Verhaltensdiagnostik. In: Basler HD, Rehfisch HP, Zink A (Hrsg) Psychologie in der Rheumatologie. Springer, Berlin Heidelberg New York, S 113
11. Geissner E, Dalbert C, Schulte A (1992) Die Messung der Schmerzempfindung. In: Geissner E, Jungnitsch G (Hrsg) Psychologie des Schmerzes. Diagnose und Therapie. Psychologieverlagsunion, Weinheim, S 79
12. Jensen MP, Miller L, Fisher LD (1998) Assessment of pain during medical procedures: a comparison of three scales. *Clin J Pain* 14:343
13. Keel PJ, Wittig R, Deutschmann R, Diethelm U, Knüsel O, Löschmann C, Matathia R, Rudolf T, Spring H (1998) Effectiveness of in-patient rehabilitation for sub-chronic and chronic low back pain by an integrative group treatment program (swiss multicentre study). *Scand J Rehabil Med* 30:211
14. Kohlmann T, Raspe H (1996) Der Funktionsfragebogen Hannover zur alltagsnahen Diagnostik der Funktionsbeeinträchtigung durch Rückenschmerzen (FFbH-R). *Rehabilitation* 35:1
15. Melzack R (1987) The short-form McGill Pain questionnaire. *Pain* 30:191
16. Nilges P, Kröner-Herwig B, Denecke H, Glier B, Klinger R, Redegeld M, Weiß L (1995) Qualitätssicherung in der Therapie chronischen Schmerzes. Ergebnisse einer Arbeitsgruppe der DGSS zur psychologischen Diagnostik. VI. Verfahren zur Erfassung von Behinderung/Beeinträchtigung. *Schmerz* 9:242
17. 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
18. WHO (1980) International classification of impairments, disabilities, and handicaps. WHO, Genf
19. Wiesinger GF, Nuhr M, Qittan M, Ebenbichler G, Wölfl G, Fialka-Moser V (1999) Cross-cultural adaptation of the Roland-Morris questionnaire for German-speaking patients with low back pain. *Spine* 24:1099