Low Back Pain Rating scale: validation of a tool for assessment of low back pain

Claus Manniche a,*, Karsten Asmussen a, Birgitte Lauritsen a, Henrik Vinterberg a, Svend Kreiner b and Alan Jordan c

a Department of Rheumatology, Central Hospital, Hillergård, Hillerød (Denmark), b The Danish Institute for Educational Research, Copenhagen (Denmark), and c Department of Physiology and Nuclear Medicine, Frederiksberg Hospital, Frederiksberg (Denmark)

(Received 22 January 1993, revision received 22 November 1993, accepted 29 November 1993)

Summary
Low Back Pain Rating scale is an index scale which includes measurements of pain intensity, disability, and physical impairment. The scale was designed to monitor the outcome of clinical trials of low back pain treatment. It has been validated in 58 patients following first-time discectomy. The scale rating can be rapidly carried out and requires no special aids. With slight modification it can be used in office and telephone interviews, as well as postal questionnaires. These modifications only slightly reduce the quantity of information gathered. In the study, a high rater agreement (97.7%) was found without level difference between two observers using the scale. The validation process included: construct validity, criterion-related validity and item bias, relative to Global Assessments pronounced by the patient and an experienced clinician. Low Back Pain Rating scale has been shown to be valid and reliable in the assessment of low back pain.

Key words: Low back pain; Outcome assessment; Validation; Clinical trial

Introduction
Many treatment regimens exist for acute or chronic low back pain (Spitzer et al. 1987). In attempts to document treatment effectiveness the randomized clinical trial is regarded as the most important scientific instrument (Wulff 1981; Deyo 1983; Bloch 1987). Using randomized clinical trials, active and placebo treatments can be compared. A validated assessment instrument is clearly essential for monitoring the outcome of any controlled clinical investigation.

Low back pain patients often have no measurable physical abnormalities. Impaired spinal mobility, positive Straight Leg Raising Test, or sensory loss/motor paralysis in the lower limb, may appear with such insufficient regularity or relevance as to exclude them as outcome parameters in clinical trials (Välfors 1985).

"Inability to do normal work" or "work loss" have often been used as objective measurements, but they may be confounded by social factors (Waddell et al. 1984), and do not always significantly correlate with other effect parameters: e.g., pain scores, overall assessments, or disability scores (Waddell et al. 1988; Holmström and Moritz 1991).

A subjective assessment (the patient's overall assessment) is the classical way of measuring the efficacy of a treatment intervention (Howe and Frymoyer 1985). However, the success rate of surgery may vary as much as 30-40%, depending on questionnaire design.

Obviously, since an ideal and unequivocal parameter for registering low back pain does not exist, several authors (Fairbank et al. 1980; Million et al. 1982; Waddell et al. 1982; Roland and Morris 1983a, b; Lehman et al. 1983; Waddell and Main 1984a; Evans et Kagan 1986; Lawlis et al. 1989; Spratt et al. 1990; Greenough and Fraser 1992; Von Korff et al. 1992) have developed different point score systems in which some subjective and/or objective parameters are included. Most of these assessment systems have been
validated, but only the Oswestry Back Pain Questionnaire and the Waddell Score have been employed in numerous clinical trials (Fairbank et al. 1980; Waddell and Main 1984). In the authors’ opinion, the outcome assessment system chosen by a back researcher often reflects his nationality or previous experience with a particular effect parameter. Attempts at comparing assessment systems are rare (Greenough and Fraser 1992; Hsieh et al. 1992). Waddell (1988) states that when evaluating the degree of severity of low back pain, a differentiation must be made among 3 recordable clinical illness components: pain, disability, and physical impairment. These 3 illness components can be registered separately and often appear to be in relative agreement with one another. They may differ however (Waddell et al. 1988) due to underlying factors which influence the components in varying degrees. For example etiological/pathoanatomical conditions, psychological variations in the patient group, or different methods of measurement. Social and work-related factors also appear to be very important (Waddell et al. 1984b).

Taking these factors into account CM has designed a new effect parameter the Low Back Pain Rating scale (RS). RS includes the 3 clinical illness components suggested by Waddell (1988) and additionally has now been employed in several clinical trials (Manniche et al. 1988, 1991, 1993a,b,c,d; Christensen et al. 1993) since 1985. The RS was constructed as an index scale where the goal has been to create a compact, readily usable and simultaneously complete indirect measurement of low back pain. It was primarily devised for use in clinical trials, but may also be used as a status assessment in clinical practice. A primary goal of RS was to keep it user-friendly and time-efficient. With slight modification, it can be employed in the doctor/patient interview situation, as well as in postal questionnaires.

The aim of this study was to evaluate the relevance of the information given by the RS and to examine if summation of the individual illness components accurately reflects patient status.

**Principle of validation**

An index scale, such as the RS, is a compilation of several separate items into one summary score or classification of response patterns. A general definition of index scales imposes no limits on any individual item. Items may, for example, be responses to a graded ordinal scale, or measurements of quantitative responses. Criteria and methods of evaluating index scales require indirect measurements of unobservable latent variables (phenomena) and have been developed within the field of psychometry. These evaluation methods are equally applicable for the assessment of index scales, used as indirect measurements of low back pain. Bohrnstedt (1983) has provided an excellent introduction to the validity concept. The most important criteria of validity stated in terms of the RS are: (1) content validity — questions whether RS is based upon items that clearly and concisely cover all relevant aspects of low back pain; (2) construct validity — here we examine the theoretical foundation of the low back pain in question and if RS actually measures this. This is both a theoretical and an empirical question, and requires two steps during the validation process. Firstly the theoretical validation, which is a question of evaluating the relationship between the individual illness components on one hand, and the latent variable (low back pain) on the other hand (Fig. 1). Illness components must be marginally correlated, but conditionally independent of each other. Lastly, construct validity examines the possible correlation between different latent or manifest variables. In our study, factor analysis is used for the analysis of theoretical and criteria-related construct validity.

**Criterion-related validity**

This study considers criterion-related validity and the correlation between RS and certain important key variables related to back pain such as disability or pain intensity, from the point of view of construct validity.

The problem addressed by the validation procedures may be characterized as a question of the scale's theoretical foundation. In addition, the technical aspects of the RS as a measure of low back pain were examined by analysis of item bias and rater agreement.

**Item bias.** The analysis examines whether RS includes all available information regarding the patient's low back pain or whether there is more information to
be gained from studying the individual responses of the individual items in the scale. Item bias is examined by testing whether the individual components of the RS are indeed independent of a series of external criteria (Osterlind 1983).

Rater agreement. Rater agreement is closely related to the question of the reliability of the scale. How reliable is RS as a measuring instrument? To what extent will repeated measurements of the same phenomenon give the same results? We examined this question by parallel and independent measurements done by two observers. In the analysis, regard must be taken both to the level and correlation between the two measurements.

Low Back Pain Rating scale

In accordance with Waddell (1988) RS is constructed so as to include the three clinical illness components which constitute low back pain separately in point scales: pain, disability, and physical impairment (Fig. 1). Pain is divided into back pain and leg pain, in accordance with the Nordic Questionnaire (Kourinka et al. 1987). The combination of pain recording, functional outcome scales, measurement of endurance of the back muscles, flexibility of the spine, and use of analgesia guarantees a broad status of the patients low back illness. These are essentially the same elements that other back researchers have used fully or partially in their assessment systems (Fairbank et al. 1980; Million et al. 1982; Waddell et al. 1982; Roland and Morris 1983a,b; Lehman et al. 1983; Waddell and Main 1984a; Evans and Kagan 1986; Lawlis et al. 1989; Greenough and Fraser 1992), for which reason we therefore regard the RS index as content valid.

Pain. The most widely used scale for registering the patient’s perception of pain is the visual analogue scale (VAS) (Huskisson 1974), but several other scales have been found to be suitable (Jensen et al. 1986). In the RS, we use an 11-point box scale which has the advantage of being suitable for both visual and audio usage, and simplifies readings such that they can be done without the use of aids. This also makes data registration easier. In order to counteract an arbitrary registration of pain caused by the often erratic variation in time, the patient has been asked about the pain at the time of examination (0–10 points), the worst low back pain within the last 2 weeks (0–10 points), and the average level of the back pain during the same period (0–10 points). Low back pain and leg pain are recorded separately, each with a 0–30-point pain score. The pain dimension in total gives 0–60 points. Localization of the registered pain is defined on the basis of the illustration with text in the Nordic Questionnaire about low back pain (Kourinka et al. 1987).

Disability. Disability is measured by a questionnaire containing 15 daily tasks (Table I). For each question answered yes = 0 points, can be a problem = 1 point, no = 2 points, giving a total score of 0–30 points. The aim is to record both physical and psychological functional loss. The questions are posed in random succession. A design has been attempted which covers both slight disability as well as very severe disability.

Physical impairment. In several earlier studies (Hansen 1964; Biering-Sørensen 1984; Biering-Sørensen et al. 1993; Jørgensen and Nicolaisen 1986), registration of the endurance of the back muscles has been done using the following simple test. The patient is placed prone with legs strapped to a bench and the trunk is left unsupported from the level of iliac crest. The length of time that the patient can remain horizontal, and thus clear of the floor, is recorded. Zero seconds scores the maximum of 10 points. One point is removed from the maximum score for every 30-sec interval attained. Zero points are scored for a patient’s ability to maintain this position for 270 sec or more.

As a measure of back mobility, Macrae and Wright (1999) have proposed the modified Schober’s test. Later, others have used the same test (Reynolds 1975; Merrit et al. 1986; Lawlis et al. 1989). In RS, one scores 10 points for 0–19 mm, 8 points for 20–29 mm, 6 points for 30–39 mm, 4 points for 40–49 mm, 2 points for 50–59 mm and 0 points for scoring more than 59 mm.

The patient’s mobility is assessed in the following way. From a supine position on a flat couch 80 cm above the floor, the patient steps onto the floor next to

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DISABILITY INDEX</strong></td>
</tr>
<tr>
<td>1. Can you sleep at night without low back pain interfering?</td>
</tr>
<tr>
<td>2. Can you do your daily work without low back pain reducing your activities?</td>
</tr>
<tr>
<td>3. Can you do easy chores at home such as watering flowers or cleaning the table?</td>
</tr>
<tr>
<td>4. Can you put on shoes and stockings by yourself?</td>
</tr>
<tr>
<td>5. Can you carry two full shopping bags (10 kg in total).</td>
</tr>
<tr>
<td>6. Can you get up from a low armchair without difficulty?</td>
</tr>
<tr>
<td>7. Can you bend over the wash basin to brush your teeth?</td>
</tr>
<tr>
<td>8. Can you climb stairs from one floor to another without resting because of low back pain?</td>
</tr>
<tr>
<td>9. Can you walk 400 meters without resting because of low back pain?</td>
</tr>
<tr>
<td>10. Can you run 100 meters without resting because of low back pain?</td>
</tr>
<tr>
<td>11. Can you ride a bike or drive a car without feeling any low back pain?</td>
</tr>
<tr>
<td>12. Does low back pain influence your emotional relationship to your nearest family?</td>
</tr>
<tr>
<td>13. Did you have to give up contact with other people within the last 2 weeks because of low back pain?</td>
</tr>
<tr>
<td>14. If it was of present interest, do you think that are certain jobs which you would not be able to manage because of your back trouble?</td>
</tr>
<tr>
<td>15. Do you think that the low back pain will influence your future?</td>
</tr>
</tbody>
</table>
TABLE II

POINT SCORING OF LOW BACK PAIN RATING SCALE, DOCTOR'S GLOBAL ASSESSMENT AND PATIENT'S ASSESSMENT IN 58 PATIENTS

(in points) B = back pain; L = leg pain; D = disability; P = physical impairment; RS = Low Back Pain Rating scale.

| Patient (no.) | Sex (m/f) | Age (years) | C.M. a | K.A. b | |  |
|---------------|-----------|-------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-----------
the couch, in the shortest possible time, goes to the foot of the couch where he preforms a deep knee bend, and returns to the starting position. The score is 0 points for less than 10 sec, 2 points for 10–19 sec, 4 points for 20–29 sec, 6 points for 30–39 sec, 8 points for 40–49 sec and 10 points for more than 49 sec.

Use of analgesics/NSAID is recorded and scored from 0 = no use during a week, 2 = use of NSAID/non-narcotic analgesics up to 4 times a week, 4 = use of NSAID/non-narcotic analgesics more than 4 times a week, 8 = use of morphine/analogues up to 4 times a week and 10 = use of morphine/analogues more than 4 times a week. Combined physical impairment yields from 0 to 40 points.

All of these disease components are weighted so that in the total score they are distributed by 60 points for pain scoring, 30 points for the level of functioning, and 40 points for physical impairment. Combined, they form a rank scale where the person with a completely healthy back scores 0 points and the hopelessly disabled patient 130 points. Approximately 15 min are required to complete the RS. In postal questionaires a modified RS is used. Here we include ‘use of analgesics’ together with pain and disability dimensions. Back muscle endurance, spinal mobility and the patients mobility are obviously omitted. Total number of points is reduced to 100.

Method

All patients who underwent first-time lumbar surgery for disc prolapse, without re-operation, from 1983 to 1987 inclusive at Central Hospital, Hillerød (n = 305) were asked to complete and return a postal questionnaire regarding their current back status (response rate: n = 261/305, 86%) (Manniche et al. 1993d). The time after surgery ranged from 14 to 60 montha. Patients with persistent low back pain were invited to participate in a training study (Manniche et al. 1993b), and these patients underwent the following 1.5 h examination program in order to include them in the clinical trial. (1) Completion of Low Back Pain Rating scale 1 by C.M. (30 min); (2) completion of Low Back Pain Rating scale 2 by K.A. (30 min); (3) Doctor’s Global Assessment of operation outcome by a third physician (B.L.) (30 min); (4) Patient’s Global Assessment of operation outcome prior to medical examination.

Global assessment

Global Assessment is a graded evaluation scale based on the following classification: excellent (very satisfactory), good (satisfactory, little discomfort), fair (acceptable, some discomfort) or poor (unchanged or aggravated). Scores of RS range from 0 to 130 points, and the Doctor’s/Patient’s Global Assessments were registered. All examinations by doctors were blinded with regard to the other doctors’ or patients’ RS scoring or assessment scores. The RS was assessed with no special aids besides a bench, a fixation strap, a stop watch, and a measuring tape.

The aim was to make the doctor/patient situation at the time of the Doctor's Global Assessment identical with a standard clinical examination. Case files of earlier admissions including the operation note, X-ray report, the patient's oral statement regarding the case, and a current clinical examination formed the basis of an evaluation of the current back status. Thirty minutes was allotted for each doctor's examination, and the total examination session for each patient was completed in the course of 2 h on the same day.

Results

Fifty-eight patients (29 women; median age: 51 years, 10/90 percentil: 39–69) participated in the investigation. The complete demographic data are published elsewhere (Manniche et al. 1993b). Scoring in the RS sub-components and Doctor’s/Patient’s Global Assessment are shown in Table II. The correlation between RS score by C.M. (RS₁) and K.A. (RS₂) was R(S) = 0.97.

Fig. 2 shows the relationship between RS and Doctor’s Global Assessment. Compared with the Doctor’s Global Assessment, the specificity is 67%, sensitivity is 88%, predictive value for a negative result is 93% and the predictive value for a positive result is 50%, when RS scores are dichotomized by a cut-point total of 39
RS points (39 points = median RS scoring of the study population).

Results of the validation process

Construct validity. Theoretical validity of RS requires that the variation in the 4 illness components is determined by the latent low back pain variable. The 4 items have to be marginally correlated but conditionally independent (Fig. 1). The hypothesis may first be tested by an analysis of the so-called TETRAD equations, relating correlations of the 4 RS illness components to each other (Glymour et al. 1987). The second part of the analysis uses factor analysis to estimate the extent to which the variation of the separate illness components is accounted for by the latent low back pain variable. Finally, we examine whether the values of the illness components should simply be added to one another or whether single illness components should be weighted individually.

A unidimensional RS implies that correlations of the 4 components, low back pain (b), leg pain (l), disability (d), and physical impairment (p), will satisfy the so-called TETRAD equations:

\[
(r_{bl} \times r_{dp} - r_{bp} \times r_{ld})^2 = 0
\]

\[
(r_{bd} \times r_{lp} - r_{bp} \times r_{ld})^2 = 0
\]

\[
(r_{bd} \times r_{lp} - r_{bl} \times r_{dp})^2 = 0
\]

The results show that the unidimensional variable is within acceptable limits.

\[
(r_{bl} \times r_{dp} - r_{bp} \times r_{ld})^2 = 0.0210 \quad (p = 0.77)
\]

\[
(r_{bd} \times r_{lp} - r_{bp} \times r_{ld})^2 = 0.0302 \quad (P = 0.67)
\]

\[
(r_{bd} \times r_{lp} - r_{bl} \times r_{dp})^2 = 0.0092 \quad (P = 0.89)
\]

TETRAD assessments are the initial attempt to find evidence against theoretical validity. They must be followed by a more careful factor analysis. The results of the factor analysis are shown in Table III. There is only one eigenvalue greater than 1.0. This implies that the latent variable is unidimensional. The latent variable accounts for 65.9% of the total variation of the 4 components. We found that the Back Pain component was most expressive of the latent Low Back Pain variable.

Our data show that factor loading of the individual illness components results in essentially similar values. Therefore simple addition of the 4 illness components approximates the optimal. In other words, there is nothing to be gained from weighting individual illness components.

Doctor's and Patient's Global Assessment: criteria validity

Criteria validity requires a high degree of correlation between RS on the one hand, and the Doctor's Global Assessment and Patient's Global Assessment, on the other hand. The requirement of criteria validity is that RS is marginally as well as partially correlated to Doctor's and Patient's Global Assessments. Criteria validity may thus be understood as a question of which of the 3 assessments (RS, Doctor's Global Assessment or Patient's Global Assessment) best represents the independent latent low back pain variable. The preferable statistical model in this case is the so-called CG distribution (Lauritzen 1989). The problem regarding partial correlation can in this case be formulated as a question of whether 2 of the 3 variables are conditionally independent given the third variable – in other words, can the connection between the 2 analyzed

<table>
<thead>
<tr>
<th>Factor no.</th>
<th>Eigenvalue</th>
<th>Cummutat. %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.63</td>
<td>65.86</td>
</tr>
<tr>
<td>2</td>
<td>0.52</td>
<td>78.93</td>
</tr>
<tr>
<td>3</td>
<td>0.50</td>
<td>91.44</td>
</tr>
<tr>
<td>4</td>
<td>0.34</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Variable     | Factor loading | Communality |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Back pain</td>
<td>0.87</td>
<td>0.75</td>
</tr>
<tr>
<td>Leg pain</td>
<td>0.80</td>
<td>0.64</td>
</tr>
<tr>
<td>Disability</td>
<td>0.79</td>
<td>0.63</td>
</tr>
<tr>
<td>Phys. impairment</td>
<td>0.78</td>
<td>0.61</td>
</tr>
</tbody>
</table>

\(^a\) Eigenvalue. The latent dimension is usually taken to be equal to the number of eigenvalues that are \(> 1.0\).

\(^b\) Cumulative percentiles. The explained variation given inclusion of this factor.

\(^c\) Factor loading. The correlation between the manifest component and the unidimensional independent low back pain variable.

\(^d\) Communality. The explained variation of the independent low back pain variable.
TABLE IV
DISTRIBUTION OF PATIENTS WITH REGARD TO DOCTOR'S GLOBAL ASSESSMENT AND PATIENT'S GLOBAL ASSESSMENT
Observed counts * and estimated counts for every patient category.

<table>
<thead>
<tr>
<th>Doctor's Global Assessment</th>
<th>Patient's Assessment</th>
<th>Excellent/good</th>
<th>Fair/poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent/good</td>
<td>Observed</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Expected</td>
<td>20.994</td>
<td>18.006</td>
</tr>
<tr>
<td></td>
<td>RS mean</td>
<td>24.07</td>
<td>30.20</td>
</tr>
<tr>
<td>Fair/poor</td>
<td>Observed</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Expected</td>
<td>2.006</td>
<td>16.994</td>
</tr>
<tr>
<td></td>
<td>RS mean</td>
<td>44.30</td>
<td>59.44</td>
</tr>
</tbody>
</table>

* Expected counts are calculated under the condition that RS gives a complete explanation of the connection between Doctor's Global Assessment and Patient's Global Assessment, and under partial assumption of the variance homogeneity of RS across both Doctor's Global Assessment and Patient's Global Assessment. The common variance is equal to 133.78 (S.D. = 11.57).

variables be explained by referring to a third variable. This question can be tested by using &likelihood ratio tests, which result in a asymptotically chi-square distributed results; conditional independence of RS and Doctor's Assessment given Patient's Assessment, $P < 0.00005$; conditional independence of RS and Patient's Assessment given Doctor's Assessment, $P < 0.00005$; conditional independence of Doctor's Assessment and Patient's Assessment given RS, $P = 0.10$. This result shows that RS correlates strongly with both the Doctor's Global Assessment, and Patient's Global Assessment. These relationships to RS cannot be explained by a possible relation between the 2 Global Assessments. On the contrary, the results show that RS can explain the relationship between the 2 Global Assessments.

The analysis of the relationship between the 2 Global Assessments and RS has until now assumed that the variation of RS might depend on the Global Assessments. A statistical test of variance homogeneity reveals very acceptable results ($\chi = 9.49$, $df = 5$, $P = 0.091$). Apart from the different RS levels in the patient groups graded by the Global Assessments, the variation of RS scores shows a high level of regularity.

Table IV shows the degree to which RS is capable of predicting the relation between Doctor's and Patient’s Global Assessments. The table illustrates both the observed and expected number of patients in every patient category, as well as estimated mean scores in each group defined by the 2 Global Assessments.

**Item bias**

RS is defined as the direct sum of 4 different illness components. Factor analysis showed that this index appears to approximate the optimal index.

To verify that this approximation does not result in any obvious imbalances in relation to the 2 other criteria (Doctor's and Patient's Global Assessments), it is necessary to test the hypothesis that all of the illness components of RS are item unbiased in other words that the separate components are conditionally independent of the external criteria (Osterlind 1983).

These hypotheses, which include both qualitative and quantitative variables, can be tested by likelihood ratio tests for CG models, in the same way as was the case with criteria validity. The results disclose no basis of item bias (conditional independence of back pain and Global Assessment given RS, $P = 0.58$; conditional independence at leg pain and Global Assessment given RS, $P = 0.28$; conditional independence of disability and Global Assessment given RS, $P = 0.13$; conditional independence of physical impairment and Global Assessment given RS, $P = 0.63$).

**Rater agreement and reliability**

The analyses discussed in the previous paragraph refer only to the results from 1 of the 2 independent observers (C.M.). The results from the other observer (K.A.) are completely parallel in all respects, and analysis results in the same conclusions.

The remaining question regarding the scale's reliability can be evaluated by an analysis of rater agreement. The first step in this analysis is to calculate the correlation between the measurements, on the one hand, and an analysis of eventual scoring, on the other hand. Results of these calculations are shown in Table V for each of the singular components, and for the total RS score. A paired $t$ test was used to evaluate possible differences in the mean levels of the 2 raters' findings. The results shown in Table V indicate an extremely high degree of agreement in RS as measured by 2 different observers.

A reasonable measurement of agreement could therefore be: (expected variance - observed variance)/ (observed variance). In this case the calculation is 97.7%. Agreement between the 2 observers explains in other words almost 98% of the expected variation.

TABLE V
ANALYSIS OF CORRELATION AND IMBALANCES BETWEEN LOW BACK PAIN RATING SCALE 1 AND LOW BACK PAIN RATING SCALE 2 IN EVALUATION THE OUTCOME

<table>
<thead>
<tr>
<th></th>
<th>Difference</th>
<th>Paired $t$ test</th>
<th>Correlation</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low back pain</td>
<td>-0.26</td>
<td>-0.72</td>
<td>0.89</td>
<td>0.36</td>
<td>0.72</td>
</tr>
<tr>
<td>Leg pain</td>
<td>0.43</td>
<td>1.53</td>
<td>0.13</td>
<td>0.28</td>
<td>0.53</td>
</tr>
<tr>
<td>Disability</td>
<td>0.45</td>
<td>2.06</td>
<td>0.043</td>
<td>0.22</td>
<td>0.45</td>
</tr>
<tr>
<td>Phys. impairment</td>
<td>-0.10</td>
<td>-0.42</td>
<td>0.68</td>
<td>0.25</td>
<td>0.53</td>
</tr>
<tr>
<td>RS</td>
<td>0.52</td>
<td>1.01</td>
<td>0.32</td>
<td>0.51</td>
<td>0.89</td>
</tr>
</tbody>
</table>
Discussion

The requirement of an ideal effect parameter for use in controlled trials concerning low back pain must primarily be that the recording be as valid and reliable for patients with only slight back trouble, as for patients with severe back disease. The utilization of RS should be reasonably straight forward and not dependent upon a multitude of costly materials. This increases the possibility of including sufficient numbers of patients in controlled clinical trials.

Several clinical trials (Manniche et al. 1988, 1991, 1993a,b,c,d) have shown that the Low Back Pain Rating scale is rapidly performed (15 min/session) and requires no special aids. At the same time RS can, with slight modification, be used in office and telephone interviews, as well as in postal questionnaires. These modifications only slightly reduce the information gathered (Table III).

In the clinical setting RS can be used to determine the status of back patients. As an outcome assessment scoring system for patients having undergone first time disc surgery, Fig. 2 shows RS scoring values for sensitivity (88%), specificity (67%), predictive value negatives (93%) and predictive value positives (50%). These figures certainly approximate generally recognized and employed diagnostic methods of measurement within other areas of disease (Sax and Alto 1986; McCombe et al. 1989).

No 'gold standard' is available in the validation of assessment systems in low back pain. In spite of this comparisons have been made between existing assessment methods in only a few cases (Howe and Frymoyer 1985; Greenough and Fraser 1992). Such comparisons will most likely be exposed to either conscious or unconscious bias which result in inaccuracies. This is especially true if the developer of a new assessment method does the study himself. In the current study, the Global Assessment of 1 experienced clinician has been regarded as 'the gold standard'. This was done to avoid bias and also because we wanted to have the opportunity of relating RS to clinical reality. The circumstances regarding the Doctor's Global Assessment were as complete as possible. Previous medical records including X-ray reports, case history and clinical examination were all utilized in order to secure an outcome assessment which was as complete as possible.

A limiting factor in this study lies in the fact that relatively few patients were included. Scale evaluation will lose some credibility in this situation. To compensate for the limited patient numbers, RS was statistically tested by diverse methods. It has been the authors' wish to utilize a series of validating tests such as those used in psychometry for many years, despite the fact that many of these tests are not well known in physical medicine. Construct validity was tested by factor analysis. RS easily managed the demands that one usually makes of measurements of theoretical constructs. The criteria validity was tested according to partial correlation, which is a more demanding procedure than those normally used. The results of this test were also satisfactory. The value of RS is furthermore underscored by its ability to predict the relationship between Doctor's and Patient's Global Assessment.

RS has been intentionally designed as a user-friendly scale. It is the sum of 4 separate components. This simple indexing of the scale was found to approach the optimal weighted sum of the scales components in the factor analysis. To test if unsystematic imbalances were introduced in conjunction with the index construction, RS was also analyzed for item bias in relation to Doctor's Global Assessment and Patient's Global Assessment. There were no apparent problems here either. Finally RS was tested for reliability by analyzing rater agreement. A very high degree of rater agreement (97.7%) was found without level difference between the 2 observers. Only a small percentage of clinical daily examination procedures of low back patients reveal similar rater agreement (McCombe et al. 1989).

Even though the statistical analysis shows that RS reflects a satisfactory status of the patients low back pain problems, one can imagine situations in clinical research where summation of the separate illness dimensions is not desirable, e.g., a randomized study (Manniche et al. 1993a) during which behavioral manipulation is used as the only intervention to improve the disability index. We had no expectation of either pain scoring or physical impairment improvement and this was indeed the case. In this particular situation a cumulative index scale will present an incomplete non-specific outcome and may result in masking of clinically relevant improvements in the patient.

In spite of the limited patient numbers, it may be concluded that the Low Back Pain Rating scale is both valid and reliable in the assessment of low back pain.

At the present time a slightly modified form of RS is the only used as a quality control measurement in a large Fitness Chain in Denmark (Alan Jordan, Form og Figur, personal communication) and the Low Back Rating scale is included in a standardized, multi-center data-base recording, compiled by a group of low back pain researchers in Copenhagen, Copenhagen Back Research Association (COBRA). It is also being used as an outcome measurement instrument in several current clinical trials.

Acknowledgements

This study was supported by grants from Sygeplejerske Ethel Clausen Foundation. Our appreciation
Robert Cooper, M.D., and Jack Hoffmann, M.D., for their helpful comments during the preparation of this manuscript, and to secretary Ruth Reimann for her assistance during the data collection.

Appendix I

TETRAD EQUATIONS

A unidimensional factor analysis model binds the individual correlation coefficient values, which describe the relationship between the manifest responses.

Regard, for example, 4 different partial scales, A, B, C, and D, and note that a singular latent variable, S, can explain the partial scale’s variation and co-variation, such that the partial scales are conditionally independent, given the value of the latent variable. This conditional independence implies, that the partial correlation between the partial scales, controlled for the latent variable, must equal zero. For 2 manifest variables, A and B, the partial correlation controlled for S is expressed as a function of the ordinary correlation coefficients:

\[
r_{AB.s} = \sqrt{r_{AB} - r_{AS}r_{BS}} / (1 - r_{AS})(1 - r_{BS})
\]

The partial rank correlation is zero only if the numerator is zero.

\[(r_{AB} - r_{AS}r_{BS} = 0) \text{ or } (r_{AB} = r_{AS}r_{BS}) \]

Equivalent results will be obtained for all other combinations of 2 manifest variables.

(1) \(r_{AB} = r_{AS}r_{BS}\)
(2) \(r_{AC} = r_{AS}r_{CS}\)
(3) \(r_{AD} = r_{AS}r_{DS}\)
(4) \(r_{BC} = r_{BS}r_{CS}\)
(5) \(r_{BD} = r_{BS}r_{DS}\)
(6) \(r_{CD} = r_{CS}r_{DS}\)

Multiplying (1) by (6) gives:

\[r_{AB} \times r_{CD} = r_{AS}r_{BS} \times r_{CC}r_{DS}\]

In other words the same result as multiplying (2) by (5)

\[r_{AC} \times r_{BD} = r_{AS}r_{CS} \times r_{BS}r_{DS}\]

In other words

\[r_{AB} \times r_{CD} = r_{AC} \times r_{BD}\]

This is one of the Tetrads calculations utilized in the article. The other calculations can be interpreted the same way.

The likelihood ratio test is a general-purpose way of testing a null hypothesis, \(H_0\), against an alternative hypothesis \(H_1\). In this test one maximizes the likelihoods under both hypotheses and calculates \(-2L\) where \(L\) is the ratio of the maximized likelihoods under \(H_0\) and \(H_1\) respectively. It is known that this statistical test has optimal properties under very general conditions. It is asymptotically chi-square distributed, and it is equal to or asymptotically equivalent to several well-known statistical tests.

References


