Eur Spine J (2012) 21 (Suppl 2):S187–S195
DOI 10.1007/s00586-012-2255-6

ORIGINAL ARTICLE

How well do observed functional limitations explain the variance in Roland Morris scores in patients with chronic non-specific low back pain undergoing physiotherapy?

F. Caporaso · N. Pulkovski · H. Sprott · A. F. Mannion

Received: 1 February 2012 / Revised: 26 February 2012 / Accepted: 28 February 2012 / Published online: 20 March 2012

Abstract
Purpose Self-rated activity limitations in patients with non-specific chronic low back pain (cLBP) do not correlate well with performance in traditional tests of impairment (e.g. back strength, ROM, etc.). Tests using more "functional activities" have therefore been recommended as alternative "objective" outcome measures. We examined the relationship between a battery of such tests and self-reported activity limitations, before and in response to physiotherapy, and the influence of psychological factors on the relationship.
Methods 37 patients with cLBP took part (45 ± 12 years; 23 female, 14 male); 32 completed 9 weeks' physiotherapy. Before and after therapy, the patients completed the Roland Morris (RM) disability questionnaire and questionnaires to assess fear avoidance beliefs, catastrophising and psychological disturbance. They also performed eight simple functional tests (stair climb, prolonged flexion, stand to floor, lift test, sock test, roll-up test, pick-up test, fingertip-to-floor test).
Results Baseline RM scores were significantly ($p < 0.05$) correlated with all but one of the functional test scores (ranging from $r = -0.34$ (half-flexion) to 0.56 (pick-up test), and with a functional test index score for all tests together ($r = 0.60$, $p < 0.0001$). The correlation between the change-scores (after treatment) for RM and for the functional test index was 0.55 ($p = 0.001$). Psychological factors explained 7–23 % variance in RM scores (baseline, post-therapy, and change scores), beyond that which was explained by the functional tests. Effect sizes for patients with a self-rated "good global outcome" were 1.23 for RM and 0.75 for the functional test index; for those with a "poor outcome", they were −0.08 and 0.23, respectively.
Conclusion Moderately high correlations (for both absolute and change scores) were observed between the subjective and observed measures of activity limitation. This indicates that to some extent they are assessing the same underlying construct, but it also suggests that each is delivering a certain amount of unique information. Psychological factors explained some of the discrepancy between the two types of measure. Both were responsive to therapy, and their change scores reflected well the patients' global outcome ratings. The two methods of assessing activity limitations should serve to complement one another in the assessment of treatment outcome.

Keywords Activity limitations · Self-ratings · Observed function · Roland Morris · Chronic low back pain

Introduction

Low back pain (LBP) is a common problem in medical and physiotherapy practice [4, 26, 30]. Up to 84 % of the adult population in industrialised countries will suffer an episode of LBP at some point in their life [2, 3]. Most of these individuals do not seek medical care and are not disabled by their pain; instead, they recover spontaneously after a short period of time [2, 8, 30, 42, 44]. Best estimates place the prevalence of chronic non-specific LBP at approximately 23 %, with about 12 % of the population being

F. Caporaso · N. Pulkovski · H. Sprott
Division of Rheumatology, University Hospital Zürich, Zurich, Switzerland

A. F. Mannion (✉)
Spine Center Division, Department of Research and Development, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland
e-mail: anne.mannion@kws.ch

disabled by LBP [1, 2]. It has been suggested that perceived disability is more strongly associated with care-seeking than is pain intensity [8, 42].

Assessing the impact of back pain on physical performance, in terms of impairment (range of motion, strength, endurance), activity limitations (elementary and complex activities of daily living) and participation restrictions (work capacity, leisure activities and private life), can be a challenging task [10, 16]; however, it is a prerequisite for the targeted assessment and treatment of the individual patient. Many different kinds of assessment are available. Self-report questionnaires, typically including items concerned with activity limitations and participation restrictions [10, 33], are being used with increasing frequency in clinical practice to measure perceived LBP-related problems in daily functioning and their change after therapy. They are simple to administer and complete, making them especially useful for repeated assessments over a course of treatment and follow-up, and they are important for indicating how the patient perceives his/her own ability to function. However, self-rated disability scores may be influenced by perceptions, beliefs, and other psychological factors [11, 21, 43, 46], meaning that they may not always reflect the true capacity of the patient to perform the activities enquired about. Direct observation/measurement can be expected to yield more reliable estimates of performance capacity. When comparisons of the two are made, disability ratings derived from questionnaire assessments are often shown to be worse than those derived from observation [15, 31, 45]. However, the most relevant functional test(s) to employ when making such assessments remains the subject of discussion. Traditional physical tests that seek to measure impairment (e.g. tests of lumbar range of motion, back muscle strength, selective activation of certain abdominal muscles, etc.) have proven to be a poor indicator of the disability in everyday activities reported by patients with chronic non-specific LBP (cLBP) [5]. As such, they are of limited use as objective, clinically relevant outcome measures in assessing the effects of treatment. Moreover, they can be expensive and time-consuming. In recent years, more clinically relevant measures of physical performance have been developed, using batteries of tests of "functional activities" [12, 17, 18, 27, 34, 43]. The majority of these tests challenge trunk mobility, speed, and coordination of movement, which are often impaired in patients with cLBP. Such activities are typically also the focus of enquiry in self-rated disability questionnaires. According to clinimetric theory, if "objective" and "subjective" measures of activity limitations/disability are measuring the same underlying construct, they would be expected to correlate significantly, with coefficients ranging from 0.4 to 0.8 [38]. And indeed, studies carried out to date report correlations of 0.40–0.60 between self-rated activity limitations and performance in functional test batteries [16, 18, 27, 34]. However, all of these studies have only examined the relationship between subjective ratings and observed function on a cross-sectional basis. If these measures are to be used to monitor outcome, it is important to know whether the different methods are equally well able to detect change after therapy and whether they deliver concordant information (i.e. whether their change scores after therapy are also correlated).

The aim of this study was to examine the relationships between self-reported activity limitation and observed functional performance and their respective changes following a programme of physiotherapy. A secondary aim of the study was to quantify the variance in self-reported activity limitation that psychological variables (fear avoidance, psychological disturbance and catastrophising) explain over and above that explained by observed performance.

Methods

Patients

37 patients with cLBP (45 ± 12 years old; 23 women, 14 men; average duration of LBP, 8.6 ± 11.1 years) participated in the study, which was part of a larger investigation into various aspects of deep trunk muscle activation in cLBP [20, 24]. The patients were recruited from the outpatient departments of rheumatology, orthopaedics and neurology of local participating hospitals [one university hospital, two foundation (non-profit) hospitals] and a local practice of general practitioners. The inclusion criteria for the study were: non-specific LBP [1, 40] with or without referred pain (of a non-radicular nature) for at least 3 months and about to undergo physiotherapy; average pain intensity over the last 2 weeks ≥3 and ≤8 on a 0–10 visual analogue scale; good understanding (written and oral) of the German language; and willingness to comply with the study protocol. Exclusion criteria included factors reflecting the presence of serious spinal disorders, as described in LBP treatment guidelines [1, 40], as well as pregnancy within the last 2 years (potential for subsequent changes in abdominal muscle function) and prior participation in a programme of spine segmental stabilisation exercises. The study was approved by the local medical ethics committee. All participants gave their signed informed consent to participate after receiving verbal and written information about the study.

Therapy

Patients were referred for a 9-week programme of spine segmental stabilisation exercises, directed by physiotherapists who were specially trained in this therapy concept.

Attendance at physiotherapy was required once/week. The treatment approach was based on the methods described by Richardson et al. [28, 32], as detailed in Mannion et al. [20, 24]. Patients were also asked to perform home exercises comprising a sequence of 10 × 10 s repetitions, 10 times a day (total exercise time, approximately 20–25 min/day). In order to explain the rationale behind the treatment concept, and increase motivation for the programme, patients were given illustrated information brochures describing the exercises, their purpose, and how to perform them, and offering various tips and advice on how to integrate the exercises into their activities of daily living.

Questionnaires

Before and after therapy, the patients completed a questionnaire booklet containing questions on sociodemographic variables, pain history (duration of problem, length of time in treatment, work absence, etc.), the Roland and Morris Disability Questionnaire [RM; measures 24 activity limitations due to back pain (score 0–24: higher score, increased disability)] [7, 33]; the Fear Avoidance Beliefs Questionnaire [35, 41]; the Pain Catastrophising Questionnaire [25, 39]; and Psychological Disturbance [9, 22] given by the combined scores of the Modified Somatic Perception Questionnaire (heightened somatic awareness) [19] and Modified Zung Self-Rating Depression Questionnaire [47]. In each case, higher scores indicated a higher degree of the attribute being measured. All the questionnaires were available in German or had been adapted for the German language prior to the study.

After therapy a further question inquired about the global outcome of treatment (“overall, how much did the treatment you received in the last few months help?”) using a five-item Likert response scale, subsequently dichotomized for describing the success of the treatment into “good outcome” (“helped”, or “helped a lot”) or “poor outcome” (“helped only little”, “didn’t help”, “made things worse”) [23].

Functional performance tests

Performance was assessed using a battery of seven simple functional tests with previously documented reliability [17] in addition to the standard “fingertip-to-floor” trunk mobility test (Fig. 1). The battery included both quantitative and

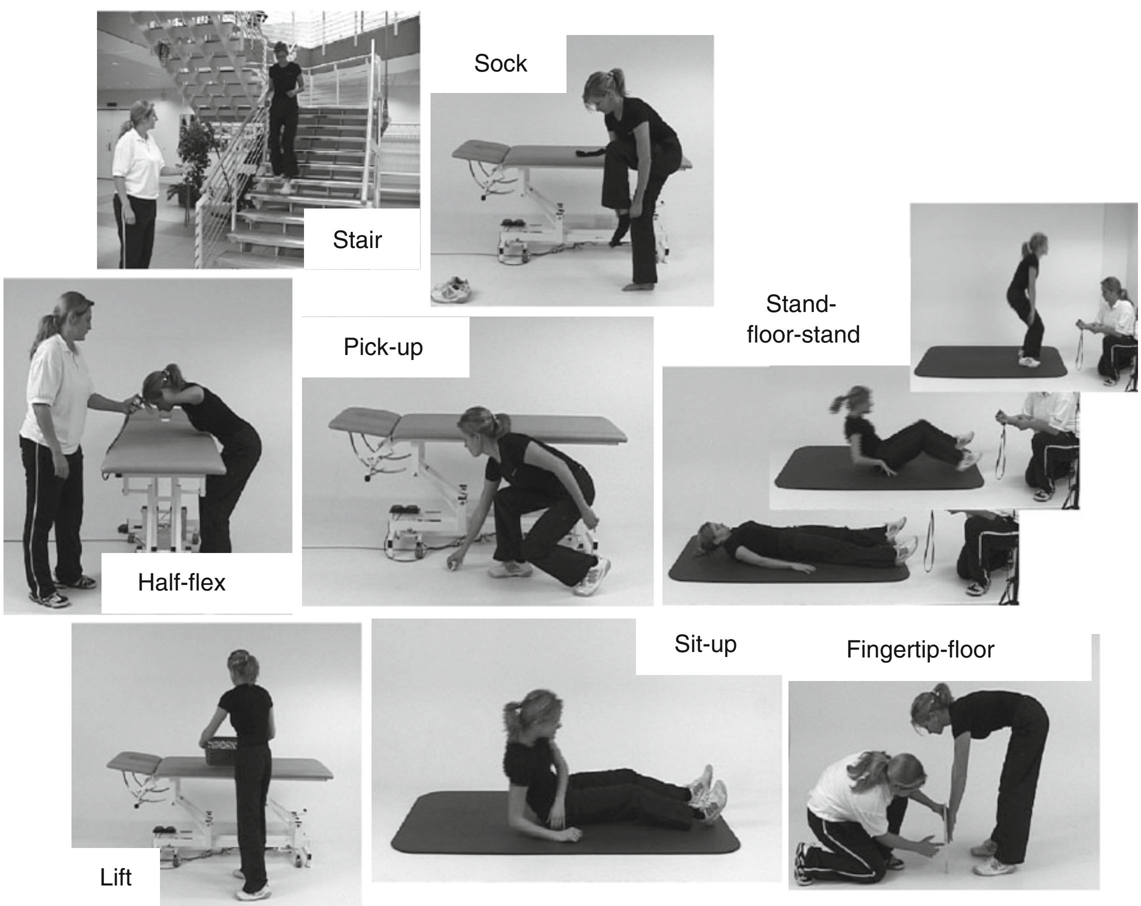

Fig. 1 Battery of functional tests

qualitative assessments of performance, with the latter being graded from 0 (patient can perform the activity without any difficulty) to 4 (cannot perform activity at all). The tests generally involved moving around and bending the trunk, some with lifting or support of the upper body weight, and comprised the following:

Quantitative assessments:

1. Stair climbing test: walking up and down 20 steps in a standard staircase as fast as safely possible (timed)
2. Half flexion test: bending forward as if washing one's hair in a sink. The position must be held for as long as possible (timed: max. 180 s)
3. Stand-to-floor test: lying down onto the floor from a standing position and getting back up to standing again (timed)
4. Finger-tip-to-floor test: bending forward in standing, knees straight, trying to reach the floor with the fingertips (movement quantified by measuring the distance of the fingertips to the floor)

Qualitative assessments:

1. Sock test: putting on a loose sock in a standing position; the leg with the worse performance is rated (movement quality: scored 0–4).
2. Pick-up-test: picking up a scrunched-up piece of paper from the floor in standing (movement quality: scored 0–4).
3. Lift test: five repetitions of lifting a box (size = $0.4 \times 0.4 \times 0.3$ m with evenly distributed content) from the floor to a table (height = 0.75 m) and back to the floor again. Total weight of box was 3 kg for women and 4 kg for men (movement quality: scored 0–4).
4. Sit-up test: sitting up from lying supine on a plinth into a long-sitting (legs outstretched) position (movement quality: scored 0–4).

The instructions and methods of evaluating test performance were standardized in a test manual, and the tests were administered by an independent investigator not involved in the treatment of the patients. The test duration was approximately 15 min.

To form a functional test index score for all eight tests, performance in each of the quantitative tests was firstly recoded as a 0–4 rating, based on the distribution of the raw scores for the group (detailed information available on request from the authors). The scores for all eight tests were then averaged to yield a functional test index score ranging from 0 to 4.

Statistical analysis

A sample size of approximately 35–40 patients allowed for the determination of a moderate correlation between the performance tests and RM disability scores (or their change scores) of $r = 0.5$ (i.e. 25 % shared variance) with a probability of 85 % (power) against the null hypothesis of $r = 0$, at a two sided significance level of 5 %, and allowing for $\sim$20 % drop-out.

Descriptive data are presented as means and standard deviations (SD), or medians and interquartile ranges (IQR), depending on the nature of the data (approximately normally distributed or not, when examined graphically). The strength of the relationships between variables was quantified using Spearman Rank correlation coefficients or Pearson Product–Moment correlation coefficients, as appropriate. Stepwise multiple regression was used to identify the psychological attributes (fear avoidance beliefs about physical activity and work, catastrophising, or psychological disturbance) that made a significant contribution to explaining the variance in RM scores, over and above that accounted for by the functional test index score. That is, the functional test index score was entered first as an independent variable, and the RM as dependent variable; in a second step the psychological variables were entered using forward conditional selection (with a probability-of-F-to-enter $\leq$0.05). The presence of collinearity was excluded by examining the tolerance values and variance inflation values for the independent variables in the final regression model; values <0.1 and >5, respectively, were considered to suggest problematic collinearity [13].

Differences between scores before and after treatment were examined using paired *t* tests. Responsiveness (for the RM and the performance test scores) was given by the standardized response mean [SRM = (post test mean−pre test mean)/SD changes].

Significance was accepted at the 5 % level. Following the reasoning of Perneger et al. [29], no corrections were made for multiple testing.

The analyses were carried out using Statview 5.0 (SAS Institute Inc, San Francisco, USA).

Results

All 37 patients completed the baseline assessments (though some questionnaire data was missing from one individual). 32/37 (86 %) patients (21 women, 11 men; 44 ± 12 years old; average duration of LBP, 7.7 ± 10.8 years; RM score 8.9 ± 4.7) completed the 9-week physiotherapy programme and the post-therapy questionnaire and performance assessments.

Baseline RM scores were significantly ($p < 0.05$) correlated with each of the functional test scores except for the sock test, with coefficients ranging from $r = -0.34$ (prolonged half-flexion test) to 0.56 (pick-up test) (first column Table 1, ranked by strength of correlation). The RM scores

Table 1 Correlations between performance tests and RM scores at baseline and at post-therapy, and for the change scores in each measure (from baseline to post-therapy)

Test	Baseline (pre-therapy) ($N = 36^{b}$)		Post-therapy ($N = 32$)		Change scores (baseline to post-therapy) ($N = 32$)	
	Correlation coefficient	*p* value	Correlation coefficient	*p* value	Correlation coefficient	*p* value
Sock test[a]	0.27	0.11	0.33	0.07	0.28	0.12
Prolonged half-flexion	−0.34	**0.04**	−0.65	**<0.0001**	−0.01	0.95
Sit-up test*[,a]	0.36	**0.03**	0.48	**0.008**	0.53	**0.003**
Stand to floor	0.40	**0.02**	0.51	**0.003**	0.26	0.16
Lift test[a]	0.49	**0.004**	0.45	**0.01**	0.27	0.15
Stair climb	0.49	**0.002**	0.52	**0.003**	0.23	0.21
Fingertip-to-floor test*	0.51	**0.001**	0.43	**0.01**	0.45	**0.01**
Pick-up test[a]	0.56	**0.001**	0.15	0.39	0.04	0.81
Functional test index score*	0.60	**<0.0001**	0.70	**<0.0001**	0.55	**0.001**

Bold highlighted *p* values indicate $p < 0.05$

* Significant correlation for each analysis (baseline, post-therapy and change scores)

[a] Spearman Rank correlation coefficients (Rho, corrected for ties) for the qualitative tests rated 0–4; Pearson's *r* for all others (quantitative tests)

[b] Missing questionnaire data for one patient at baseline

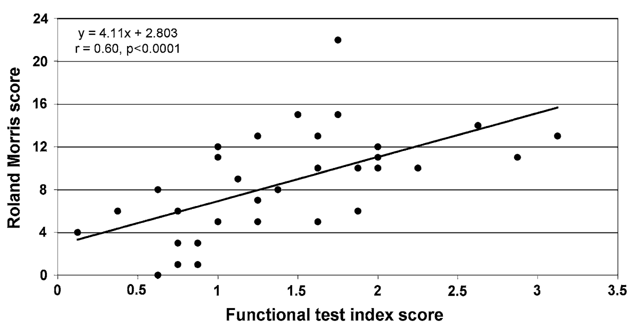

Fig. 2 Relationship between baseline RM scores and the functional test index score

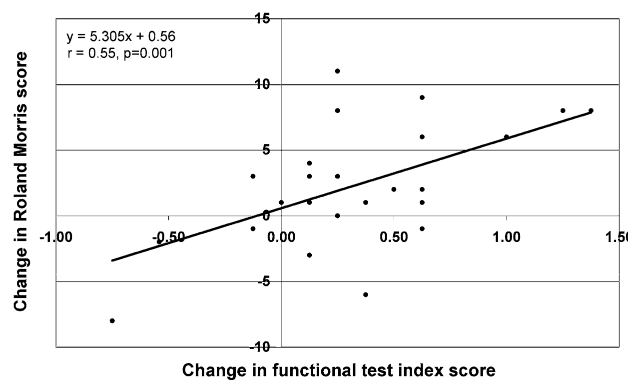

Fig. 3 Relationship between changes (pre- to post-therapy) in RM scores and functional test index scores

correlated most strongly with the functional test index score for all the tests ($r = 0.60$; $p < 0.0001$; Fig. 2).

Similar correlation coefficients ($r = 0.43–0.70$; $p < 0.05$; second column, Table 1) were recorded for these relationships post-therapy, with the exception of those involving the sock test and the pick-up test. Change scores (pre-therapy to post-therapy) for two of the individual tests (sit-up test and fingertip to floor test) and for the functional test index score (Fig. 3) showed a significant ($r = 0.45–0.55$; $p < 0.05$) correlation with the RM change scores (third column, Table 1).

The multivariate analysis of the baseline data, with RM scores as the dependent variable and the functional test index score as a force-entered independent variable, revealed that fear avoidance beliefs about work made an additional statistically significant contribution ($p < 0.001$), uniquely accounting for 19 % variance in baseline RM scores (Table 2). For the post-therapy multivariate model, psychological disturbance was the only additional significant variable, uniquely explaining 7 % ($p < 0.05$) variance in the post-therapy RM scores (Table 2). For the longitudinal (pre-therapy to post-therapy) multivariate model, the change in catastrophising was the only significant variable entered, explaining a further 23 % ($p < 0.001$) variance in the RM change-scores over above that explained by the change-scores for the functional test index (Table 2).

The standardised response mean (SRM) or "effect size" for the change in RM score, pre-treatment to post-treatment, was 0.54. The corresponding value for the functional test index was 0.73. In response to the global outcome question, 17/32 (53 %) patients were classified as having a "good outcome" and 15/32 (47 %) a "poor outcome". The SRM for the RM for the "good" group was 1.23 and for the "poor" group, −0.08. The corresponding values for the functional test index score were 0.75 and 0.23. In other words, the RM scores were better able than the functional

Table 2 Results of the final multiple regression models showing the ability of psychological variables to explain the variance in RM scores, over and above that explained by the functional test index scores

MODEL and variables selected for inclusion at each step	Standardised beta coefficient	R^2	Adjusted R^2	Step increase in adj. R^2 (%)	p value, R^2 step change
Baseline data					
Step 1—forced entry					
Functional test index score	0.472	0.36	0.34		0.0001
Step 2—stepwise selection					
FAB-work	0.457	0.55	0.53	19	0.001
Post-therapy data					
Step 1—forced entry					
Functional test index score	0.603	0.46	0.44		<0.0001
Step 2—stepwise selection					
Psychological disturbance (MSPQ and ZUNG score)	0.294	0.54	0.51	7	<0.05
Longitudinal (change-score) data					
Step 1—forced entry					
Change score, functional test index score	0.468	0.30	0.28		0.001
Step 2—stepwise selection					
Change score, catastrophising	0.500	0.54	0.51	23	0.001

test scores to differentiate between good and poor outcomes (wider separation between the SRMs in the good and poor groups).

Discussion

This study examined the relationship between self-reported activity limitation and observed performance in functional tests, and the changes in each following a programme of physiotherapy, in patients with cLBP. It also examined the role of various psychological factors in explaining the variance in self-report measures beyond that explained by observed performance.

Bivariate analyses revealed moderate correlations between self-reported activity limitation and observed performance, with a shared variance of approximately 30–50 %. This indicates that to some extent they are assessing the same underlying construct, but it also suggests that each is delivering a certain amount of unique information. The correlations were similar to those reported in previous studies on cLBP patients for other functional test batteries and disability instruments ($r = 0.40$–0.60; [16, 18, 27, 34]). The consistency of the findings across studies of only "moderate correlations" has prompted some authors to suggest that self-reported activity limitation is not just a measure of physical function but is likely influenced by other factors such as the patient's perceptions, beliefs, and fears regarding their abilities and limitations, and their lifestyle, emotional and social functions [10, 16]. The findings of a recent study on patients with acute LBP, in which stronger correlations were reported between psychological factors and subjective ratings than between psychological factors and observed function [43], suggested that this was indeed the case, and the present study confirmed the hypothesis in patients with chronic problems: when the psychological variables (fear-avoidance beliefs, psychological disturbance, and catastrophising) were entered into the multivariable models, one of them always explained a significant proportion of additional variance (7–23 %) in the self-reported limitations, beyond that explained by the measured performance. In clinical practice, the relative extent to which a patient's self-rated function is influenced by physical factors as opposed to psychosocial/beliefs factors is not always immediately apparent to the clinician; this highlights the benefit of using multiple/combined input sources when attempting to quantify function. Other factors that might explain the unique variance in each of the two measures (subjective and objective) include the fact that they do not have exactly the same content. Although they attempt to measure the same construct, some tests are not reflected in the RM items, and some items in the RM have no direct equivalent in the tests; depending on whether those respective items are of relevance to a given patient or not, the patient could end up with different disability ratings. Patients with LBP are a heterogeneous group, with differences in the manner and extent to which LBP compromises their function. Finally, there is a certain amount of measurement error in each method, which inevitably accounts for some of the unique variation in the scores of each. Though it might be useful to know which of the two is the more "accurate" or "recommended" measure, this is impossible to ascertain, since there is currently no "gold standard" measure of activity limitation.

According to Strand et al. [37], functional performance should be regarded as a global measure that is not sufficiently represented by a single test, but requires a combination of tests. Further, from the "clinimetrics" perspective, a greater number of items in an index leads to greater internal reliability of the construct being measured. In the present study, the index score combining all individual tests generally gave the best correlations with the RM scores. Individually, some of the tests showed a poor correlation with the RM scores, but analysis of the 8 items together revealed good internal consistency for the whole battery, with no improvement after removal of any given item (Cronbach's alpha 0.83–0.85; detailed results not shown). A scaled-down test battery would nonetheless be desirable, where time was of the essence in clinical practice or large numbers of patients were to be assessed in research studies. Based on the results of this small study, two of the most promising tests to include would appear to be the sit-up test and the fingertip-floor test, addressing a qualitative and quantitative dimension of function, respectively. Both of these items performed well in the Back Performance Scale of Strand et al. [36, 37] and previous studies have highlighted the superior clinical relevance of the fingertip-floor distance [6, 26, 37]. It may be advisable to include both qualitative and quantitative tests in any test battery, since the manner in which the movement is carried out, including potential deficits in motor control or balance, is not evidenced by timed-methods alone. The tests should also be acceptable and meaningful to patients and practitioners, as well as simple to administer and to interpret.

The RM and the functional test index showed comparable responsiveness for the whole group change-scores after therapy (RM scores, 0.54, and functional test index score, 0.73), with moderate effect sizes that were in keeping with those reported in the literature for the physiotherapeutic treatment of cLBP [14]. Although the SRM for the functional test index was slightly higher, at first sight suggesting it might be more responsive than the RM, it indicated a small positive effect (SRM, 0.23) even in the group that declared a poor global treatment outcome (i.e. those with a poor outcome still showed an improvement in function), and not such a high effect size in the good outcome group (SRM, 0.75). In contrast, the RM scores showed a more appropriate differentiation between good and poor global outcomes in terms of the corresponding SRMs (1.23 and −0.08, respectively). In other words, there was a notable improvement in RM scores for those with a good outcome and almost no change in score in the group with a poor outcome. This is a desirable characteristic of an outcome instrument, but it may possibly have resulted from the choice of external criterion used for assessing "global treatment outcome". Other types of (possibly more objective) external criteria for indicating success should be investigated in future studies.

Certain limitations of the present study are worth mentioning. We did not evaluate the reliability of the test battery although this was done during its initial development [17] and many of its component tests have proven reliable, valid and responsive in other studies [18, 36, 37]. The sample size was relatively small, but the participants were nonetheless representative of typical patients with non-specific cLBP as defined in the European guidelines for cLBP [1] and the main findings were comparable to those reported in other studies, as far as the correlations between objective and subjective measures was concerned [18, 27, 34, 37]. Further, many analyses were performed, which increases the likelihood of chance findings occurring. The results should be interpreted with caution until such times as they can be confirmed in larger groups of patients.

Conclusion

Moderately high correlations (for both absolute and change scores) were observed between the subjective and observed measures of activity limitation. This indicates that to some extent they are assessing the same underlying construct, but it also suggests that each is delivering a certain amount of unique information. Psychological factors explained some of the discrepancy between the two types of measure in both the cross-sectional analyses (before and after therapy) and also in relation to the change scores measured over time. Both were responsive to therapy, and their change scores reflected well the patients' global outcome ratings. The two methods of assessing activity limitations should serve to complement one another in the assessment of treatment outcome.

Acknowledgments Supported by a grant from the National Research Programme NRP 53 "Musculoskeletal Health—Chronic Pain" of the Swiss National Science Foundation (Project 405340-104787/2) and the Schulthess Klinik Research Funds. We would like to express our thanks to: Prof Beat A. Michel for providing the infrastructure to carry out this work within the Division of Rheumatology, University Hospital Zürich, Switzerland; Daniel Helbling, Deborah Gubler and Valeriu Toma for their assistance with the data collection; the physiotherapists and doctors of our departments at the University Hospitals of Zürich (USZ and Balgrist) and the Schulthess Klinik for completing the necessary documentation, referring patients into the study and carrying out the treatments; all the patients who took part in the study.

Conflict of interest None.

References

1. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G (2006) Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 15(Suppl 2):S192–S300

2. Balague F, Mannion AF, Pellise F, Cedraschi C (2011) Non-specific low-back pain. Lancet 2011 [Epub ahead of print]
3. Costa Lda C, Maher CG, McAuley JH, Hancock MJ, Herbert RD, Refshauge KM, Henschke N (2009) Prognosis for patients with chronic low back pain: inception cohort study. BMJ 339:b3829
4. Dagenais S, Tricco AC, Haldeman S (2010) Synthesis of recommendations for the assessment and management of low back pain from recent clinical practice guidelines. Spine J 10:514–529
5. Deyo RA (1988) Measuring the functional status of patients with low back pain. Arch Phys Med Rehabil 69:1044–1053
6. Enthoven P, Skargren E, Kjellman G, Oberg B (2003) Course of back pain in primary care: a prospective study of physical measures. J Rehabil Med 35:168–173
7. Exner V, Keel P (2000) Erfassung der Behinderung bei Patienten mit chronischen Rückenschmerzen. Schmerz 14:392–400
8. Ferreira ML, Machado G, Latimer J, Maher C, Ferreira PH, Smeets RJ (2010) Factors defining care-seeking in low back pain: a meta-analysis of population based surveys. Eur J Pain 14:e1–e7
9. Greenough CG, Fraser RD (1991) Comparison of eight psychometric instruments in unselected patients with back pain. Spine 16:1068–1074
10. Grotle M, Brox JI, Vollestad NK (2005) Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. Spine (Phila Pa 1976) 30:130–140
11. Grotle M, Vollestad NK, Veierod MB, Brox JI (2004) Fear-avoidance beliefs and distress in relation to disability in acute and chronic low back pain. Pain 112:343–352
12. Harding VR, Williams AC, Richardson PH, Nicholas MK, Jackson JL, Richardson IH, Pither CE (1994) The development of a battery of measures for assessing physical functioning of chronic pain patients. Pain 58:367–375
13. Heiberger RM, Holland B (2004) Statistical analysis and data display. Springer, New York
14. Keller A, Hayden J, Bombardier C, van Tulder M (2007) Effect sizes of non-surgical treatments of non-specific low-back pain. Eur Spine J 16:1776–1788
15. Kremer EF, Block A, Gaylor MS (1981) Behavioral approaches to treatment of chronic pain: the inaccuracy of patient self-report measures. Arch Phys Med Rehabil 62:188–191
16. Lee CE, Simmonds MJ, Novy DM, Jones S (2001) Self-reports and clinician-measured physical function among patients with low backpain: a comparison. Arch Phys Med Rehabil 82:227–231
17. Lüder S, Pfingsten M, Lüdtke K, Müller G, Strube J, Hildebrandt J (2006) Kann die Aktivitätskapazität von Patienten mit Rückenschmerzen objektiv und reliabel gemessen werden? Measuring the activity capacity of patients with low back pain. Physioscience 2:147–155
18. Magnussen L, Strand LI, Lygren H (2004) Reliability and validity of the back performance scale: observing activity limitation in patients with back pain. Spine (Phila Pa 1976) 29:903–907
19. Main CJ (1983) The modified somatic perception questionnaire (MSPQ). J Psychosom Res 27:503–514
20. Mannion AF, Caporaso F, Pulkovski N, Sprott H (2010) Goal attainment scaling as a measure of treatment success after physiotherapy for chronic low back pain. Rheumatology (Oxford) 49:1734–1738
21. Mannion AF, Junge A, Taimela S, Muntener M, Lorenzo K, Dvorak J (2001) Active therapy for chronic low back pain: part 3. Factors influencing self-rated disability and its change following therapy. Spine 26:920–929
22. Mannion AF, Müntener M, Taimela S, Dvorak J (1999) A randomised clinical trial of three active therapies for chronic low back pain. Spine 24:2435–2448
23. Mannion AF, Porchet F, Kleinstuck FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective: Part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 18:374–379
24. Mannion AF, Pulkovski N, Helbling D, Sprott H (2008) The influence of adherence with therapy on clinical outcome after a spine stabilisation exercise program in patients with chronic non-specific low back pain. International Society for the Study of the Lumbar Spine (Spine Week). Geneva, Switzerland
25. Meyer K, Sprott H, Mannion AF (2008) Cross-cultural adaptation, reliability, and validity of the German version of the Pain Catastrophizing Scale. J Psychosom Res 64:469–478
26. Michel A, Kohlmann T, Raspe H (1997) The association between clinical findings on physical examination and self-reported severity in back pain: results of a population-based study. Spine 22:296–303
27. Novy DM, Simmonds MJ, Lee CE (2002) Physical performance tasks: what are the underlying constructs? Arch Phys Med Rehabil 83:44–47
28. O'Sullivan PB (2000) Lumbar segmental 'instability' clinical presentation and specific stabilizing exercise management. Man Ther 5:2–12
29. Perneger TV (1998) What's wrong with Bonferroni adjustments. BMJ 316:1236–1238
30. Picavet HS, Struijs JN, Westert GP (2008) Utilization of health resources due to low back pain: survey and registered data compared. Spine (Phila Pa 1976) 33:436–444
31. Reneman MF, Jorritsma W, Schellekens JM, Goeken LN (2002) Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. J Occup Rehabil 12:119–129
32. Richardson CA, Jull GA (1995) Muscle control–pain control. What exercises would you prescribe? Manual Therapy 1:2–10
33. Roland M, Fairbank J (2000) The Roland–Morris disability questionnaire and the oswestry disability questionnaire. Spine 25:3115–3124
34. Simmonds MJ, Olson SL, Jones S, Hussein T, Lee CE, Novy D, Radwan H (1998) Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. Spine 23:2412–2421
35. Staerkle R, Mannion AF, Elfering A, Junge A, Semmer NK, Jacobshagen N, Grob D, Dvorak J, Boos N (2004) Longitudinal validation of the fear-avoidance beliefs questionnaire (FABQ) in a Swiss–German sample of low back pain patients. Eur Spine J 13:332–340
36. Strand LI, Anderson B, Lygren H, Skouen JS, Ostelo R, Magnussen LH (2011) Responsiveness to change of 10 physical tests used for patients with back pain. Phys Ther 91:404–415
37. Strand LI, Moe-Nilssen R, Ljunggren AE (2002) Back Performance Scale for the assessment of mobility-related activities in people with back pain. Phys Ther 82:1213–1223
38. Streiner DL, Norman GR (1995) Health Measurement Scales: a practical guide to their development and use. Oxford University Press Inc., Oxford
39. Sullivan M, Bishop S, Pivik J (1995) The pain catastrophising scale. Development and validation. Psychol Assess 7:524–532
40. Waddell G (1999) The back pain revolution. Churchill Livingstone, London
41. Waddell G, Newton M, Henderson I, Somerville D, Main CJ (1993) A fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 52:157–168
42. Walker BF, Muller R, Grant WD (2004) Low back pain in Australian adults. Health provider utilization and care seeking. J Manipulative Physiol Ther 27:327–335
43. Wand BM, Chiffelle LA, O'Connell NE, McAuley JH, Desouza LH (2010) Self-reported assessment of disability and

performance-based assessment of disability are influenced by different patient characteristics in acute low back pain. Eur Spine J 19:633–640

44. Wieser S, Horisberger B, Schmidhauser S, Eisenring C, Brugger U, Ruckstuhl A, Dietrich J, Mannion AF, Elfering A, Tamcan O, Muller U (2011) Cost of low back pain in Switzerland in 2005. Eur J Health Econ 12:455–467
45. Wijlhuizen GJ, Ooijendijk W (1999) Measuring disability, the agreement between self evaluation and observation of performance. Disabil Rehabil 21:61–67
46. Woby SR, Watson PJ, Roach NK, Urmston M (2004) Are changes in fear-avoidance beliefs, catastrophizing, and appraisals of control, predictive of changes in chronic low back pain and disability? Eur J Pain 8:201–210
47. Zung WW (1965) A Self-Rating Depression Scale. Arch Gen Psychiatry 12:63–70