

ORIGINAL REPORT

DIAGNOSTIC ACCURACY AND RELIABILITY OF MUSCLE STRENGTH AND ENDURANCE MEASUREMENTS IN PATIENTS WITH CHRONIC LOW BACK PAIN

Wolfgang Gruther, MD¹, Franziska Wick, MD¹, Birgit Paul, PhD^{1,2}, Christoph Leitner, MD¹, Martin Posch, PhD³, Michael Matzner, MD⁴, Richard Crevenna, MD¹ and Gerold Ebenbichler, MD¹

From the ¹Department of Physical Medicine and Rehabilitation, Medical University of Vienna, Vienna General Hospital, ²Faculty of Psychology, University of Vienna, ³Section of Medical Statistics and ⁴Department of Orthopedics, Medical University of Vienna, Vienna General Hospital, Vienna, Austria

Objective: Dynamometric trunk muscle strength and endurance tests are performed widely within the rehabilitation management of chronic low back pain. The aim of this study was to examine the accuracy and long-term reliability of these measurements in patients with chronic low back pain.

Design: Cross-sectional study.

Subjects: Thirty-two patients with chronic low back pain, 19 healthy controls and 15 patients with chronic headache matched for age, sex and body mass index.

Methods: Both patient groups and healthy controls performed isokinetic and isometric trunk extensor and flexor tests on a Biodex 2000 dynamometer. The Biering-Sørensen test served to examine back muscle endurance. Borg-Category-Ratio-Scales CR-10 rated participants' body experience immediately before and after the testing. Patients with chronic low back pain repeated measurements after 3 weeks.

Results: Among dynamometric tests, isokinetic measurements revealed the best area under the curve (AUC = 0.89) for the discrimination between patients with chronic low back pain and healthy controls. Reliability testing revealed highly significant learning effects for isometric trunk flexion and isokinetic measurements. The Biering-Sørensen test demonstrated excellent accuracy (AUC=0.93) and no learning effects. Borg-category-ratio-scale ratings were not associated with the observed changes.

Conclusion: In chronic low back pain dynamometric trunk muscle measures are limited to muscle function assessment purposes. Monitoring treatment outcome in these patients with these measures appears to be problematic because of learning effects. Based on our findings, we recommend the Biering-Sørensen test for management of chronic low back pain rehabilitation.

Key words: low back pain, muscle strength, rehabilitation, reliability of results, validity of results.

J Rehabil Med 2009; 41: 613–619

Correspondence address: Gerold Ebenbichler, Department of Physical Medicine and Rehabilitation, Medical University of Vienna, Währingergürtel 18-20, Vienna AT-1090, Austria. E-mail: gerold.ebenbichler@meduniwien.ac.at

Submitted May 8, 2008; accepted April 21, 2009

© 2009 The Authors. doi: 10.2340/16501977-0391

Journal Compilation © 2009 Foundation of Rehabilitation Information. ISSN 1650-1977

INTRODUCTION

In chronic low back pain (cLBP) rehabilitation practice, trunk muscle power production is usually measured with dynamometers, either in the isokinetic or isometric test mode (1–4), whereas trunk/back muscle endurance is measured best with the Biering-Sørensen test (5). The relevance of muscle functional measures to the rehabilitation management of these patients has been emphasized by findings from a recent international research project by the “International Classification of Functioning, Disability and Health” (ICF) – Research Branch of the World Health Organization (WHO), which has developed Core Sets for cLBP (6). These Core Sets represent a selection of ICF domains or categories from the whole classification and are intended to serve as minimal standards for the reporting of functioning and health in clinical studies and clinical encounters or as standards for multi-professional, comprehensive assessment (7). The Brief ICF Core Sets for cLBP identified from the whole ICF classification a total of 10 highly relevant body functional categories and ranked “muscle power functions”, i.e. muscle strength, endurance and muscle tone, as the second most important category (6).

To be clinically and scientifically useful as functional diagnostic and treatment monitoring tools, measurements must be both accurate and reliable (8). Measurements are accurate if they correctly classify cases with a certain condition and cases without the condition, expressed by the area under the receiver operating characteristic (ROC) curve, and reliable (reproducible), if they are stable over time and show an acceptable level of measurement variability (9).

Reliability of trunk muscle strength and endurance tests has been examined mainly in healthy individuals (2, 10–12), but there is reason to believe that reliability is lower in patients with cLBP. A few studies investigated reliability of trunk muscle strength measures in patients with cLBP and revealed conflicting results. One study demonstrated excellent reliability, when patients were re-tested after 2 days (13), whereas others observed significant improvements of muscle power, when measures were performed for a second time after either 2–3 days (10, 14, 15), or after 5–10 days (2), respectively. These “learning effects” may be reduced, when baseline measure-

ments were repeated after 1 or 2 days and the second value was used as the reference value (10, 13–15). In everyday clinical practice, however, collecting the baseline data twice on 2 different days seems unfeasible, and would be unlikely to qualify for reimbursement by healthcare providers. Similarly, evidence supporting the reliability of trunk muscle endurance measures, as examined by the Biering-Sørensen test seems conflicting. Whereas some studies demonstrated good reliability (16–19), others found low reliability for the Biering-Sørensen test in patients with cLBP (2, 3, 20). In contrast to the muscle strength measurements, none of the muscle endurance reliability studies observed any relevant learning effects.

Several studies tested the discriminative power of trunk muscle strength and endurance measurements between cLBP patients and healthy controls. Results, however, were inconsistent. Some studies identified significantly lowered trunk muscle power (10) and endurance (18) in patients with cLBP, whereas others did not (2).

Regardless of the sparse and conflicting evidence on their accuracy and reliability, both trunk muscle strength and endurance tests are widely performed in rehabilitation settings. The ICF Core Sets for cLBP would indeed strongly support the use of dynamometric trunk muscle measurements within the rehabilitative assessment of patients with cLBP. It is general knowledge that muscle strength and endurance measurements are semi-objective and therefore depend on the patient's adherence with the measurements. Both a lack of compliance and related feelings may significantly affect the outcome of these measurements. We are not aware of any study that has examined the accuracy of trunk muscle strength and endurance tests in patients with cLBP, or has ever examined the effect of patient-related feelings on the results of isometric or isokinetic muscle strength tests.

This study sought: (i) to examine for the first time the accuracy of trunk muscle strength and endurance measurements in patients with cLBP; (ii) to test their long-term reliability with test protocols that are clinically feasible in rehabilitation settings; and (iii) to examine the relationship between possible learning effects and test-related feelings, respectively.

METHODS

All the procedures described in this study were approved by the local ethics committee and informed consent was obtained from all participants. The study was conducted at the Department of Physical Medicine and Rehabilitation, Medical University of Vienna, Vienna General Hospital.

Study design

A case-control cross-sectional reliability study was performed comparing 3 groups (patients with cLBP, patients with chronic headache (cHA) and healthy controls) with matched pairs.

Study population

The study included 32 patients with non-specific cLBP, 19 controls and 15 patients with cHA. Patients and healthy controls were matched on a specific matching system (13 pairs: 2 cLBP, 1 control, 1 cHA; 6 further pairs: 1 cLBP, 1 control, in 2 of the pairs 1 cHA) for age (± 5 years), gender, athletic activity and body mass index (BMI) (± 4 kg/m²).

Formal sample size analysis for the main comparisons resulted in sample size estimates of 30 patients with cLBP and 15 controls for a case-control design with matched pairs (2cLBP: 1 control), providing a power of 0.87. Sample size was estimated using data from 2 previous studies that investigated differences between healthy controls and patients with LBP using isometric measurement techniques of the back extensor muscle torque (6, 7).

Patients were eligible for the study if they were between 18 and 60 years old and had a BMI below 30 kg/m².

The patients with cLBP had experienced LBP for at least 3 months, the pain intensity had to be equal to or exceed 3, measured on an 11-point visual analogue pain rating scale (VAS), and they had not sought healthcare advice for headaches within the last year, and had had no more than 5 mild headache episodes of a maximum duration of 2 days per year and no headache within the past 6 weeks.

The patients with cLBP were included according to the recommendations of the International Headache Society (21) (suffering from bi-lateral headache, rather more frontal than in the neck or shoulders, described as a continuing aching or dull pain, feeling of tightness, external pressure or cap-like pressure around the head and occurring on the average at least twice per week) and did not seek healthcare advice within the last year for back pain, had no more than 5 mild back pain episodes per year with a maximum duration of 2 days, and no back pain within the past 6 weeks.

All patients were referred to the Department of Physical Medicine and Rehabilitation, Medical University of Vienna, Vienna General Hospital for diagnostic evaluation and treatment. Patients with peripheral neurological symptoms, spinal fracture, infection or cancer as the aetiology of back pain and patients with organic diseases that may interfere with the physical fitness were excluded from the study. No participant in the study had undergone any surgery involving the lower back. None of the participants had any previous experience with muscle power testing. All patients were asked not to take any drugs, such as muscle relaxants, analgesics and psychochemicals, for at least 4 days before the day of testing.

The 19 controls were allowed to have up to 5 headaches/back pain per year (none of which can be judged more than mild in severity), did not seek healthcare advice for headaches or LBP within the last year, and should not perform sports on a regular basis more frequently than once a week.

Outcome measurements

Trunk muscle strength test. For static and dynamic trunk muscle measurements, subjects were seated on a Biodex 2000 dynamometer (Biodex Medical Systems, New York, USA). After a few trials for familiarization with the equipment and a rest interval of a few minutes, the study assistant encouraged all subjects and patients to perform maximum static and dynamic trunk flexions and extensions as quickly and as forcefully as possible in a standardized way. Participants performed a total of 3 repetitions during the isometric measurement at 20°, 60° and 100° of hip flexion and extension. Main outcome parameter was peak torque in Newton-metres. The isokinetic measurements were repeated 4 times at 90°/sec and with a range of motion from 20° to 100° during hip flexion and extension. Main outcome parameters were maximal power output (power) in Watts, work (work) in Joules, and peak torque (torque) in Newton-metres. If variability exceeded 15% tests were repeated after a rest interval of 5 min.

Back muscle endurance test. The Biering-Sørensen test measures how many seconds the participant is able to keep the unsupported upper part of the body in a horizontal position. In this test, the load is equal to the weight of the upper part of the body, with torque determined by the lever arm from the pubic symphysis to the upper body centre of gravity (5).

The participant was positioned prone over an examination table. The lower extremities were stabilized by 2 belts at the level of the hips and just below the knees. The iliac crests were positioned at the edge of the table with the trunk extended beyond the table and initially hanging flexed in 90°. The trunk then was raised to the horizontal position with hands crossed over the chest. The test was continued until the

participant could no longer control the horizontal posture, or until he or she reached limit of fatigue pain (5).

Assessment of Body/pain perception (Borg Category-Ratio-Scales). The Borg Category-Ratio (CR-10) scales with scale ranges from 0 (nothing at all) to 10 (extremely strong) were used for assessing patients' body experience, perceived exertion, tension, and fear of harm or injury or re-injury of the back before and after the trunk muscle strength measurements. The rating "extremely strong" was anchored as strong as a person has ever experienced a stimulus previously. These scales have been used extensively and have been shown to have a high intra-subject reliability (22).

Procedure. Each subject underwent a comprehensive clinical examination and assessment in a standardized way according to the ICF, including the examination of sociodemographic data and the assessment of pain activities of daily life and physical functioning by an experienced physiatrist. The Biering-Sørensen test was subsequently performed. After a short rest period, subjects were seated on the dynamometer for the muscle strength measurements.

Immediately before static and dynamic trunk muscle strength measurements the test subjects were given a written description of the following measurements by a psychologist. All patients and subjects were asked to rate their test-related feelings immediately before static and dynamic trunk muscle strength measurements. They were instructed to imagine the required tasks of movement and then to rate their feelings on a 0–10 Anticipation-scale. The 6 rating contents of the Borg CR-10 scales with the score 0 (nothing at all) to 10 (extremely strong) (22), were expressed as 4 values: fear of LBP, fatigue level, feeling of tension, fear of injury or re-injury.

All of the static and dynamic trunk muscle strength measurements, the Biering-Sørensen test and the ratings of test-related feelings were administered on 2 separate days, 2–3 weeks apart in order to test long-term reliability in patients with cLBP. This interval corresponds to the period of time an inpatient rehabilitation programme for cLBP would take in central Europe (23). Patients with cLBP did not receive any physical and rehabilitation medicine (PMR) treatment between the first and second day of examination. cHA and controls were not scheduled to repeat the testing on a second test day.

Statistics. The 3 repetitions during the isometric measurements and the 4 repetitions during the isokinetic measurements were summarized by their arithmetic mean. As an additional variable for each patient, the ratios between flexion and extension during the isokinetic measurements were calculated.

Borg CR-10 Scales about exertion, tension, fear of harm and injury or re-injury were correlated to the outcome measurements.

Testing accuracy of the tests. Discriminative power of the trunk muscle strength and endurance tests was investigated including the data from the first visit for all 3 groups (cLBP, cHA and controls).

First, an analysis of variance (ANOVA) with the independent factors "group", "age", "sex", "height" and the random factor "subject" were used to compare the mean differences between groups.

Secondly, a matched analysis (ANOVA with independent factors group and matching variable) was performed.

Third, logistic regression analysis that included the data for the cLBP and healthy controls served only for testing specificity and sensitivity. In this analysis the independent variables were "muscle power", "muscle endurance" and "sex" and the dependent factor was "group", respectively. The analysis was repeated with the additional dependent factor "sex". The respective ROC curves were plotted and the area under the curve (AUC) calculated (9, 24).

Reliability. Reliability was investigated based on data from cLBP whose dynamometric measurements were assessed on 2 different days. Intra-class correlation coefficients for repeatability were computed from separate mixed model analyses. Paired *t*-tests served to examine systematic differences of the outcome measures between the first and the second test day. 95% prediction intervals were computed to assess clinically relevant changes of the respective variables. Due to systematic time trend from day 1 to day 2, the intra-class correlation coefficients (ICC) (25) have limited interpretability. Repeatability was visualized with Bland-Altman plots (26).

RESULTS

Subject's characteristics and patient flow

All 47 patients and 19 healthy controls completed clinical and dynamometric measurements as described above. The socio-demographic variables of the patients and healthy controls are shown in Table I. A total of 21 out of the 32 patients with cLBP, who received no therapy repeated all the experiments on a second test day, 2.38 (± 0.9) weeks later. Those 11 patients who refused re-testing decided to start with their rehabilitation program ($n = 7$) or just refused to be tested a second time ($n = 4$). Epidemiological and clinical parameters including pain intensity ($p = 0.305$), pain duration ($p = 0.497$) and total score ($p = 0.477$) did not show any significant differences between drop-outs and those who completed all experiments.

Accuracy and discriminative power of muscle strength and endurance measures

All the isokinetic and isometric muscle strength tests as well as the Biering-Sørensen muscle endurance test differed

Table I. Subject's characteristics and level of physical impairment

	Patients with cLBP		Patients with cHA	Controls
	Mean (SD)			
	Day 1	Day 2	Mean (SD)	Mean (SD)
Age, years	43 (10)		43 (11)	42 (13)
Height, cm	168 (8)		169 (7)	171 (9)
Weight, kg	73.4 (14.8)		71.5 (14.5)	74.3 (13.6)
Body mass index, kg/m ²	25.7 (4.2)		24.9 (4.1)	25.3 (3.9)
Back pain duration, months	92.8 (94.1)			
Back pain, VAS, cm	5.3 (1.6)	4.5 (1.8)		
Headache duration, months			179 (143)	
Headache, VAS (11 pts)			6.0 (1.3)	
Biering-Sørensen test, sec	96 (75)	89 (55)	174 (79)	221 (54)

cLBP: chronic non-specific low back pain; cHA: chronic headache; Day 1: first test day; Day 2: second test day; SD: standard deviation; VAS: visual analogue scale.

Table II. Discriminative power of isometric and isokinetic strength tests and the Biering-Sørensen test between controls, patients with chronic low back pain (cLBP) and patients with chronic headache (cHA) (expressed as means with standard deviations within parentheses)

	Controls	cLBP	cHA	Controls vs cLBP <i>p</i> -value	cLBP vs cHA <i>p</i> -value	Controls vs cHA <i>p</i> -value
Isokinetic strength tests during back extension						
Torque, Nm	197.94 (73.48)	103.11 (89.43)	167.59 (103.73)	<0.001**	0.003**	0.271
Work, J	221.74 (85.88)	109.04 (112.62)	184.62 (114.6)	<0.001**	0.004**	0.297
Power, W	180.36 (73.87)	85.31 (91.38)	140.48 (88.92)	<0.001**	0.007**	0.155
Isokinetic strength tests during flexion						
Torque, Nm	100.96 (43.11)	58.75 (43.68)	73.06 (36.04)	<0.001**	0.046*	0.009**
Work, J	107.76 (47.84)	55.38 (48.28)	80.88 (42.8)	<0.001**	0.008**	0.057
Power, W	83.54 (40.48)	40.89 (37.53)	57.18 (31.86)	<0.001**	0.014*	0.008**
Isometric strength tests during back extension						
20°, Nm	192.40 (68.61)	163.38 (68.88)	196.82 (60.42)	0.045*	0.028*	0.757
60°, Nm	247.08 (99.14)	179.01 (86.88)	235.30 (85.65)	0.001**	0.007**	0.595
100°, Nm	260.92 (94.14)	178.92 (100.59)	223.77 (84.65)	<0.001**	0.037*	0.124
Isometric strength tests during flexion						
20°, Nm	44.91 (35.7)	32.85 (41.23)	30.21 (27.52)	0.098	0.833	0.116
60°, Nm	97.41 (49.38)	73.36 (51.05)	87.71 (41)	0.005**	0.093	0.358
100°, Nm	116.64 (46.97)	84.38 (46.71)	107.28 (29.43)	0.002**	0.019*	0.526
Endurance test						
Biering-Sørensen test, sec	220.68 (54.29)	96.19 (74.47)	174.00 (79.38)	<0.001**	<0.001**	0.120

* $p < 0.05$, ** $p < 0.01$, matched analysis.

highly significantly between healthy controls and patients with cLBP. Furthermore, significant group differences were observed when the results from the isometric back extension and flexion strength tests as well as the muscle endurance test were compared between patients with cHA and cLBP. The subsequent comparison between healthy controls and patients with cHA, however, did not reveal relevant significant differences between groups. The results of the ANOVA are shown in Table II.

Logistic regression analysis tested accuracy and revealed sex-adjusted areas under the ROC curve with AUC values between 0.6 and 0.89 for isometric and isokinetic trunk extension and flexion strength tests. Among these tests, the best AUC values were observed for the isokinetic back extensions and flexions, respectively. Accuracy testing of the back muscle endurance test revealed an AUC value of 0.93, indicating excellent accuracy. For this test this would mean that a cut-off point corresponding to a sensitivity of 0.9 would result in a specificity of 1. The results are presented in Figs 1–3.

Reliability of trunk muscle strength and endurance measures

Significant changes in the mean between the 2 study days were observed for all the isokinetic trunk flexion and extension tests and for all isometric trunk flexion strength measurements. The observed increases ranged between 45% and 160% of the initial value. No such significant changes in the mean were found for the isometric back extension tests and the Biering-Sørensen test, respectively. The ICC for the Biering-Sørensen test revealed a value of 0.59 and for the isometric back extension tests between 0.81 and 0.85, suggesting clinically acceptable-to-good reliability. The results are shown in Tables III and IV.

Relationship between pain intensity, test-related feelings and systematic changes of dynamometric measures

Per scale ratings of feelings and fear-associated beliefs were not associated with the observed changes in the mean between the 2 test days.

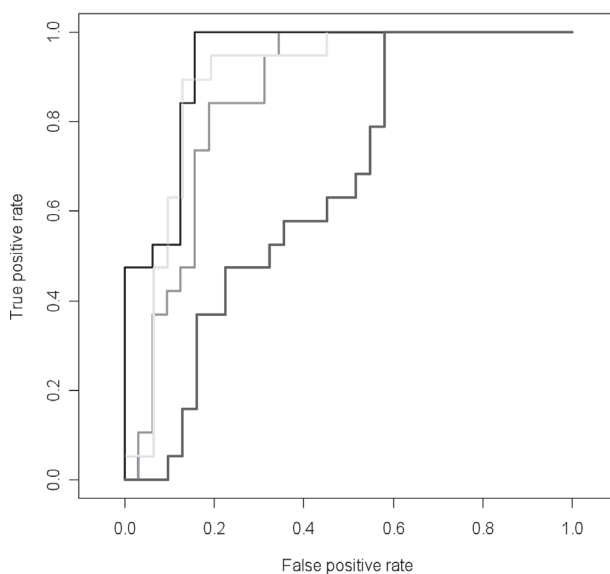


Fig. 1. Receiver operating characteristic plots of the sex-adjusted logistic regression for the discrimination between patients with chronic low back pain and healthy controls. ■ isokinetic peak torque during extension (area under the curve (AUC): 0.86); ■ isokinetic peak torque during flexion (AUC: 0.89); ■ Biering-Sørensen test (AUC: 0.93); ■ isokinetic ratio between flexion and extension peak torque (AUC: 0.66).

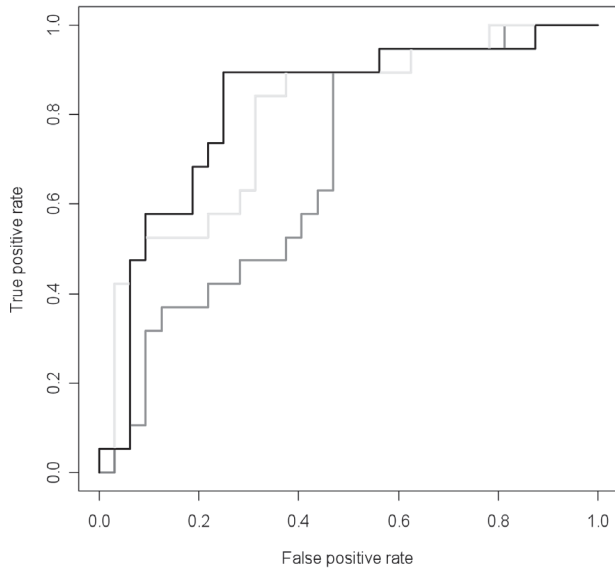


Fig. 2. Receiver operating characteristic plots of the sex-adjusted logistic regression for the discrimination between patients with chronic low back pain (cLBP) and healthy controls. ■ isometric peak torque during extension at 20° (area under the curve (AUC): 0.68); ▨ isometric peak torque during extension at 60° (AUC: 0.79); ■ isometric peak torque during extension at 100° (AUC: 0.82).

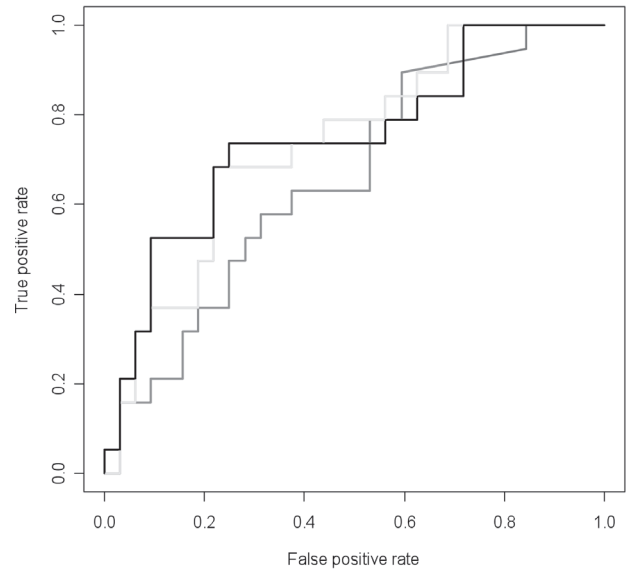


Fig. 3. Receiver operating characteristic plots of the sex-adjusted logistic regression for the discrimination between patients with chronic low back pain (cLBP) and healthy controls. ■ isometric peak torque during flexion at 20° (area under the curve (AUC): 0.66); ▨ isometric peak torque during flexion at 60° (AUC: 0.74); ■ isometric peak torque during flexion at 100° (AUC: 0.75).

DISCUSSION

To summarize: there are two major findings to report. Firstly, both isokinetic trunk muscle tests and the Biering-Sørensen back muscle endurance test revealed excellent diagnostic accuracy. This finding was corroborated by the results of the ANOVA, which tested differences between healthy controls and patients with cLBP or chronic headache patients, and the ROC. Secondly, reliability testing revealed major changes in

the mean for the isokinetic trunk muscle tests and the isometric trunk flexion tests. Such changes in the mean were not related to patients per scale ratings of feelings or fear. In contrast, the Biering-Sørensen back muscle endurance test revealed no significant changes in the mean and demonstrated clinically acceptable-to-good reliability.

It is widely accepted that a structured rehabilitation management, as presented in the Rehab-CYCLE® (27, 28), which

Table III. Indices of systematic changes in the mean of between-day analyses from patients with chronic low back pain (cLBP)

	Day 1 Mean (SD)	Day 2 Mean (SD)	Mean difference	p-value	95% prediction interval
Isokinetic strength tests during back extension					
Torque, Nm	87.74 (74.34)	117.96 (87.96)	30.23 (9.81)	0.006**	-65.71, 126.16
Work, J	89.83 (86.13)	118.14 (97.61)	28.32 (11.3)	0.021*	-82.26, 138.89
Power, W	69.98 (70)	93.59 (81.93)	23.61 (9.02)	0.016*	-64.63, 111.85
Isokinetic strength tests during flexion					
Torque, Nm	50.58 (35.57)	63.83 (38.13)	13.25 (4.53)	0.008**	-31.06, 57.55
Work, J	47.11 (39.37)	62.12 (43.44)	15.02 (5.06)	0.008**	-34.45, 64.48
Power, W	34.33 (30.64)	45.05 (31.76)	10.73 (3.84)	0.011*	-26.88, 48.34
Isometric strength tests during back extension					
20°, Nm	160.23 (58.08)	161.67 (68.05)	1.43 (7.85)	0.857	-75.40, 78.27
60°, Nm	171.74 (53.63)	182.28 (63.25)	10.54 (6.78)	0.136	-55.78, 76.86
100°, Nm	163.97 (62.85)	179.16 (73.78)	15.18 (8.75)	0.098	-70.40, 100.77
Isometric strength tests during flexion					
20°, Nm	28.15 (31.24)	35.23 (33.02)	7.08 (3.15)	0.036*	-23.74, 37.91
60°, Nm	69.67 (37.96)	81.87 (44.29)	12.21 (4.78)	0.019*	-34.59, 59.00
100°, Nm	77.54 (35.57)	95.22 (35.79)	17.68 (4.51)	0.001**	-26.41, 61.78
Endurance test					
Biering-Sørensen test, sec	81.57 (42.53)	88.81 (54.96)	7.24 (9.8)	0.469	-88.69, 103.17

p < 0.05*, p < 0.01**, paired t-tests.

SD: standard deviation; mean difference: between Day 1 and Day 2.

Table IV. Intraclass correlation coefficient values for chronic low back pain (cLBP) between days 1 and 2

Extension – torque	na	Flexion – torque	na
Extension – work	na	Flexion – work	na
Extension – power	na	Flexion – power	na
Extension – 20°	0.85	Flexion – 20°	na
Extension – 60°	0.85	Flexion – 60°	na
Extension – 100°	0.81	Flexion – 100°	na
Biering-Sørensen test, sec	0.59		

na: not applicable.

consists of 4 basic elements (assessment, assignment, realization and evaluation) will contribute to the optimization of a patient's rehabilitation outcome in a relevant way. In European countries the ICF Core Sets are regarded as a clinically feasible screening method for the assessment and outcome evaluation of patients with impaired functional health. The ICF Core Sets for cLBP suggest screening of muscle strength and endurance functions as being highly important.

To our knowledge, this is the first study to investigate the accuracy of dynamometric trunk muscle function measurements and the Biering-Sørensen test. Sensitivity analysis revealed excellent sensitivity and specificity for both the isokinetic trunk muscle tests and the Biering-Sørensen test. Our findings are supported by results from previous studies that observed a high discriminative power between healthy controls and patients with cLBP (5, 10, 18). As we also investigated for the first time the discrimination ability between patients with cLBP and patients with other pain conditions, such as chronic headache, our findings suggest that both the isokinetic trunk muscle tests and the Biering-Sørensen test qualify well as muscle function diagnostic tools within the rehabilitative assessment of patients with cLBP. Our findings seem to contrast with those of a recent study that did not find any group differences between healthy controls and patients with cLBP, when isokinetic measurements and the Biering-Sørensen test were performed (2). This is surprising as our study included relatively sedentary healthy controls, who were not athletic and did not performed any kind of sports on a regular basis. Such a discrepancy suggests that factors associated with dynamometric trunk muscle testing, such as subjects' and patients' anticipatory feelings and emotions, are likely to influence the discriminative power of these measurements.

So far only one study (2) seems to have investigated the reliability of both the isokinetic trunk muscle power measurements (at 60° and 150°/sec) and static back muscle endurance measurements using the Biering-Sørensen test in cLBP after an interval of 5–10 days and found significant learning effects. In the presented study we used a Biodex 2000 dynamometer. The test protocol followed the manufacturer's recommendations and tested isokinetic trunk muscle strength within a range of motion of 0–20–110° trunk/hip flexion/extension. We chose an angular velocity of 90°/sec, which was relatively slow, and refused to test at higher angular velocities of 150°/sec as, according to the personal experience of the researchers, testing at higher angular velocities was not feasible in all patients with LBP. Following the manufacturer's advice, tests were repeated if the coefficient of variation from a series of 4 trunk extensions and

flexions exceeded 15%. Furthermore, all patients and subjects performed training sessions with the test protocols in advance of the actual testing, and a break of sufficient duration was kept between the training and the test session, respectively. Despite all these quality assurances, we were unable to avoid significant changes in the mean when subjects were re-tested 2.5 weeks later. Both our findings and those of the Keller study (2), which used a different dynamometer, seem to support the notion that baseline assessment with dynamometric muscle strength tests have to be performed on 2 different tests days, with the second value being taken as the baseline reference value. This appears to be supported by findings that revealed no further changes in the mean in patients with cLBP, if they were re-tested on a third day a couple of days later (2). Thus, 2 baseline measurements recorded on 2 different days would be required if the isokinetic trunk muscle tests are to be used as an outcome measure in a rehabilitation programme. Unfortunately baseline recordings on 2 different test days may not be possible in daily practice in outpatient rehabilitation units, because these measures are time-consuming and it is unlikely that the second baseline measure would be paid for by healthcare providers. The isometric back tests, however, revealed either large 95% prediction intervals or significant changes in the mean. Therefore, none of these 2 muscle power measurement methods seem to be superior in rehabilitation outcome documentation of patients with cLBP.

It may be argued that the external validity of our findings was limited due to a relatively small and varying number of subjects in each of the 3 groups. The number of subjects was based on a power analysis in which there was an assumption that all the tests would be performed in a laboratory environment in advance of the study. A statistical power exceeding 0.8 deemed sufficient to provide external validity to our findings for the central European population. Nevertheless a future large multi-continental study will be needed to prospectively prove the validity of the muscle function assessment in different populations of patients with cLBP. For the main comparison between healthy controls and patients with cLBP, our study included a total of 19 matched pairs, with a total of 13 pairs agreeing in matching criteria according to a 1:2 ratio and further 6 pairs according to a 1:1 ratio, respectively. Such procedure is legitimate as the power of our statistical sensitivity analyses was not decreased, but rather increased. A total of 31.3% of the patients with cLBP refused to repeat the muscle strength and endurance tests on the second day. Such high drop-out rate may represent a limitation of the current investigation, even though the number of patients who repeated the muscle functional tests after 2–3 weeks was at least as high as in most of the reliability studies (29–32). A larger sample size may not change the estimate of reliability, but may serve to narrow the confidence intervals about reliability coefficients. Sufficient blinding of the different study groups was not achievable in this study. However, standardization of patients' information forms and the use of a standardized study protocol minimized the influence of any systematic bias. All dynamometric and muscle endurance examinations took place in the same room and the same examiners performed the tests and re-tests.

In conclusion, our findings suggest that dynamometric muscle strength measurements and isometric muscle strength measure-

ments are limited to muscle function diagnosis and treatment planning purposes in rehabilitative assessment. Monitoring the treatment outcome with these measures, as recommended in the Rehab-CYCLE[®] (27, 28), is problematic. The Biering-Sørensen test demonstrated both excellent diagnostic accuracy and acceptable reliability. Consequently, we recommend this test for the assessment of trunk muscle function in the rehabilitation management of patients with cLBP, and infer that the category "muscle function" in the Brief ICF Core Set for cLBP would be examined best by back muscle endurance tests.

ACKNOWLEDGEMENTS

We thank Monika Knötig, now at the Department Internal Medicine II at the Medical University of Vienna for her assistance with conducting this study. We are greatly indebted to Professor SH Roy, NeuroMuscular Research Center, Boston University (Boston, USA) for his valuable comments on the manuscript.

This study was supported by a grant from the Lorenz Böhler Forschungsfonds awarded to Gerold Ebenbichler.

This study was performed at the Joint and Bone Center (Center for Diagnosis, Research and Therapy of Musculoskeletal Disorders), Vienna Medical University.

REFERENCES

- Ebenbichler GR, Oddsson LI, Kollmitzer J, Erim Z. Sensory-motor control of the lower back: implications for rehabilitation. *Med Sci Sports Exerc* 2001; 33: 1889–1898.
- Keller A, Hellesnes J, Brox JI. Reliability of the isokinetic trunk extensor test, Biering-Sørensen test, and Astrand bicycle test: assessment of intraclass correlation coefficient and critical difference in patients with chronic low back pain and healthy individuals. *Spine* 2001; 26: 771–777.
- Mayer T, Gatchel R, Betancur J, Bovasso E. Trunk muscle endurance measurement. Isometric contrasted to isokinetic testing in normal subjects. *Spine* 1995; 20: 920–926; discussion 926–927.
- Nachemson A, Lindh M. Measurement of abdominal and back muscle strength with and without low back pain. *Scand J Rehabil Med* 1969; 1: 60–63.
- Biering-Sørensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. *Spine* 1984; 9: 106–119.
- Cieza A, Stucki G, Weigl M, Disler P, Jackel W, van der Linden S, et al. ICF Core Sets for low back pain. *J Rehabil Med* 2004; 36 Suppl 44: 69–74.
- Stucki G, Ewert T, Cieza A. Value and application of the ICF in rehabilitation medicine. *Disabil Rehabil* 2003; 25: 628–634.
- Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil* 2005; 84: 719–723.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29–36.
- Newton M, Thow M, Somerville D, Henderson I, Waddell G. Trunk strength testing with iso-machines. Part 2: Experimental evaluation of the Cybex II Back Testing System in normal subjects and patients with chronic low back pain. *Spine* 1993; 18: 812–824.
- Madsen OR. Trunk extensor and flexor strength measured by the Cybex 6000 dynamometer. Assessment of short-term and long-term reproducibility of several strength variables. *Spine* 1996; 21: 2770–2776.
- Delitto A, Rose SJ, Crandell CE, Strube MJ. Reliability of isokinetic measurements of trunk muscle performance. *Spine* 1991; 16: 800–803.
- Hutten MM, Hermens HJ. Reliability of lumbar dynamometry measurements in patients with chronic low back pain with test-retest measurements on different days. *Eur Spine J* 1997; 6: 54–62.
- Graves JE, Pollock ML, Carpenter DM, Leggett SH, Jones A, MacMillan M, et al. Quantitative assessment of full range-of-motion isometric lumbar extension strength. *Spine* 1990; 15: 289–294.
- Robinson ME, Greene AF, O'Connor P, Graves JE, MacMillan M. Reliability of lumbar isometric torque in patients with chronic low back pain. *Phys Ther* 1992; 72: 186–190.
- Dederling A, Roos af Hjelmsater M, Elfving B, Harms-Ringdahl K, Nemeth G. Between-days reliability of subjective and objective assessments of back extensor muscle fatigue in subjects without lower-back pain. *J Electromyogr Kinesiol* 2000; 10: 151–158.
- Evans K, Refshauge KM, Adams R. Trunk muscle endurance tests: reliability, and gender differences in athletes. *J Sci Med Sport* 2007; 10: 447–455.
- Latimer J, Maher CG, Refshauge K, Colaco I. The reliability and validity of the Biering-Sørensen test in asymptomatic subjects and subjects reporting current or previous nonspecific low back pain. *Spine* 1999; 24: 2085–2089.
- Ito T, Shirado O, Suzuki H, Takahashi M, Kaneda K, Strax TE. Lumbar trunk muscle endurance testing: an inexpensive alternative to a machine for evaluation. *Arch Phys Med Rehabil* 1996; 77: 75–79.
- Ljungquist T, Harms-Ringdahl K, Nygren A, Jensen I. Intra- and inter-rater reliability of an 11-test package for assessing dysfunction due to back or neck pain. *Physiother Res Int* 1999; 4: 214–232.
- Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. Headache Classification Committee of the International Headache Society. *Cephalalgia* 1988; 8 Suppl 7: 1–96.
- Borg G. Psychophysical scaling with applications in physical work and the perception of exertion. *Scand J Work Environ Health* 1990; 16 Suppl 1: 55–58.
- Kool JP, Oesch PR, Bachmann S, Knuesel O, Dierkes JG, Russo M, et al. Increasing days at work using function-centered rehabilitation in nonacute nonspecific low back pain: a randomized controlled trial. *Arch Phys Med Rehabil* 2005; 86: 857–864.
- Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993; 39: 561–577.
- Shrout P, Fleiss J. Intraclass correlation: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420–428.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
- Steiner WA, Ryser L, Huber E, Uebelhart D, Aeschlimann A, Stucki G. Use of the ICF model as a clinical problem-solving tool in physical therapy and rehabilitation medicine. *Phys Ther* 2002; 82: 1098–1107.
- Stucki G, Sangha O. Clinical quality management: putting the pieces together. *Arthritis Care Res* 1996; 9: 405–412.
- della Volpe R, Popa T, Ginanneschi F, Spidaleri R, Mazzocchio R, Rossi A. Changes in coordination of postural control during dynamic stance in chronic low back pain patients. *Gait Posture* 2006; 24: 349–355.
- Gabriel LS, Mu K. Computerized platform posturography for children: test-retest reliability of the sensory test of the VSR System. *Phys Occup Ther Pediatr* 2002; 22: 101–117.
- Hu MH, Hung YC, Huang YL, Peng CD, Shen SS. Validity of force platform measures for stance stability under varying sensory conditions. *Proc Natl Sci Counc Repub China B* 1996; 20: 78–86.
- Liston RA, Brouwer BJ. Reliability and validity of measures obtained from stroke patients using the Balance Master. *Arch Phys Med Rehabil* 1996; 77: 425–430.