PHYSICAL MEASUREMENTS AND QUESTIONNAIRES AS DIAGNOSTIC TOOLS IN CHRONIC LOW BACK PAIN

Pekka Rantanen

From The Orthopedic Hospital of The Invalid Foundation, Helsinki, Finland

The objective of this study was to assess the diagnostic value of common questionnaires and measures of physical performance in low back pain (LBP) syndrome. One hundred and fourteen patients with LBP classified according to the Quebec Task Force were compared with 50 patients with different pain syndromes but without apparent LBP. The discriminating value of each variable was estimated by calculating the area under the receiver operating characteristics (ROC) curve. The diagnostic value of the Million and Oswestry disability questionnaires was evident, with the area under the ROC curve varying between 0.73 and 0.88. The isometric trunk extension–flexion strength test with concomitant reaction-time test could not distinguish between patients (area under ROC curve 0.50–0.68). Sensitivity of pain drawing was excellent but specificity was low: 47% for men and 39% for women. In conclusion, disability questionnaires have discriminating power. The trunk muscle strength test does not perform well as a diagnostic tool. The area under the ROC curve and the use of other patients as controls make it easier to assess the diagnostic specificity of a particular method.

Key words: biomechanics, disability evaluation, low back pain, physical fitness.

J Rehab Med 2001; 33: 31–35

Correspondence address: Pekka Rantanen, Nuottalahdentie 32 F12, FIN-02230 Espoo, Finland. E-mail: loke@pp.inet.fi

(Accepted April 13, 2000)

INTRODUCTION

Low back pain (LBP) refers not to a diagnosis but to a clinical entity characterized by pain in the lumbar region which sometimes radiates to the lower extremities (1). Modern imaging techniques, such as computerized tomography and magnetic resonance imaging, provide much information concerning serious spinal pathology and nerve root compression. However, when these techniques are applied to common LBP they discriminate poorly between patients with back ache and normal subjects (2). Early blind laboratory screening is rarely useful. Medical history and clinical examination remain the main tools in the evaluation of common LBP (1).

Increasing interest has been focused on measurements of functional capacity (3). A battery of tests have been used for quantification of lumbar function. These include techniques for quantifying the mobility of lumbar spine, isolated trunk strength measurement systems, cardiovascular fitness measures, lifting capacity measurement and measurement of activities of daily living (4). Even slow psychomotor speed of reaction has been reported to be related to LBP (5). Countless different questionnaires and tests have been used to assess pain, depression and functional status (6, 7). The value of these different diagnostic methods has remained obscure.

A receiver operating characteristics (ROC) curve can be used to assess the diagnostic value of a particular method (8). If a diagnostic test produces a continuous measurement, then a convenient diagnostic cut-off must be selected to calculate the sensitivity and specificity of the test. The result of traditional sensitivity and specificity analysis depends on the cut-off criteria chosen. If both sensitivity and specificity are calculated for all possible cut-off values and these values are displayed graphically by plotting the sensitivity on the y-axis and the false positive rate (1 − specificity) on the x-axis, the result is the ROC curve (Fig. 1). The quantitative index most commonly used to describe the ROC curve is the area under the curve. If the area under the ROC curve is equal to one then a diagnostic test can be considered to be ideal, with no false-positive or false-negative results. In contrast, an area of 0.50 represents pure chance, with no accuracy in prediction or discrimination (9, 10).

The area under the ROC curve has not previously been used to measure the discriminating ability of different physical performance measurement methods employed in patients with common LBP. The purpose of the present study is to assess the diagnostic value of common questionnaires and measures of physical performance.

METHODS

The present study is part of a larger study focusing on the functional capacity of LBP patients (11). One hundred and fourteen consecutive patients (67 women, 47 men) with chronic LBP syndrome classified according to the Quebec Task Force (12) and 50 (31 women, 19 men) patients with different pain syndromes but without current LBP (controls) participated in this study (Table I). The patients were referred to the rehabilitation clinic with LBP or other pain by general practitioners. Final inclusion into either study group, as well as the diagnostic subgrouping, was confirmed by another clinician who was aware of the purpose of the study. There were no appreciable differences in age, height or weight between the different patient groups (Table II). Subjects completed the Oswestry questionnaire (13) and the Million Visual Analogue Scale (14) to assess subjective disability and Rimon's Brief Depression Scale (RBDS) (15) for evaluation of depression. Pain

 ISSN 1650–1977

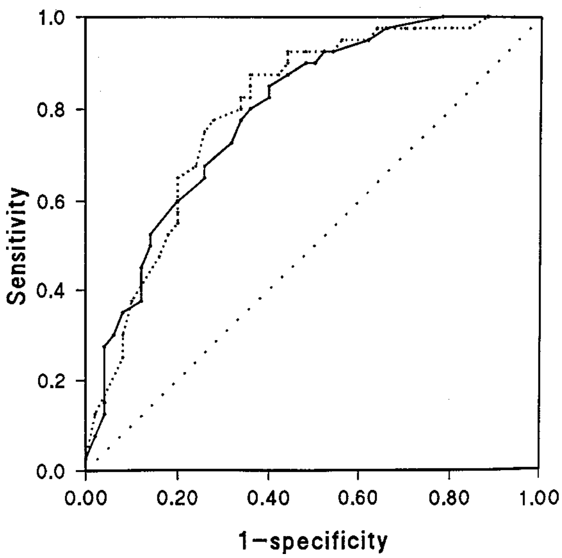

Fig. 1. Receiver operating characteristics curves for Oswestry (solid line; area under curve = 0.79) and Million (dotted line, area under curve = 0.80) disability indices. Genders have been combined.

drawing was used to assess the location of pain. The drawings were classified as either pain or no pain in the lumbar area.

The isometric trunk muscle strength measurement system was a modification of the method used by Biering-Sørensen (16). The sagittal rotation axis was set at the hip joint level, which was the most discriminating level for fulcrum between different patient groups in the previous study (11).

The reaction time was measured concomitantly with isometric muscle strength. The subject was instructed to exert the muscles immediately after an auditory signal. The reaction time was the interval between the auditory signal and the moment when the measured force exceeded 10 N. For analysing reaction time, the strain-gauge dynamometer was connected to a computer. Analogue/digital conversion was made by using a sampling rate frequency of 200 Hz. The measurements of physical performance and the analysis of pain drawing were blinded with regard to the classification.

The reliability of the tests of physical performance used was studied with 27 healthy volunteers by plotting the difference between consecutive test results against their mean (17). On the basis of these data, it was estimated that the test results would be stable in repeated measurement sessions with a 1-week interval.

The required sample size (17 subjects/group) was calculated by estimating the clinically significant difference between groups to be 150 N (30%) with a standard deviation (SD) of 150 in muscle strength measurements. Statistical significance was set at $p = 0.05$ and the risk of type II error at the 0.2 level. The BMDP software library for microcomputers (10) was used for the statistical analyses. The analysis of variance (ANOVA) was used in detecting differences between groups. The sensitivity of a test was the proportion of those with LBP and a positive test result (pain in the lumbar area in pain drawing or muscle strength below the cut-off point). The specificity of a test was the proportion of controls who gave a negative test result (no pain in the lumbar area in pain drawing or muscle strength above the cut-off point). The area under the ROC curve was calculated as the diagnostic value of a particular variable was assessed. Pearson's product moment correlation r was used as linear associations between variables were assessed. The 95% confidence interval for r was calculated (18). Otherwise, the results are expressed as means of groups with SDs.

The study design was accepted by the ethics committee of the University of Turku.

RESULTS

The difference between LBP patients and controls was evident when assessed with the Oswestry or Million disability indices (Table III). There were no between-group differences in terms of RBDS score, reaction time and trunk flexion strength. Trunk extension strength was lower in patients with LBP than in controls. According to the ROC curve analysis, the discriminatory power of physical measurements between groups remained low for both genders (Table III). ROC curves for Oswestry and Million scores for all the subjects are shown in Fig. 1. The

Table I. *Classification of patients with low back pain according to the Quebec Task Force (12) and diagnosis of controls*

	Men	Women
Classification of patients with low back pain		
1. Pain without radiation	17	20
2. Pain and proximal radiation	9	17
3. Pain and distal radiation	8	14
4. Pain and radiation with neurologic signs	6	4
5. Presumptive compression of a spinal nerve root on a simple roentgenogram	0	1
6. Confirmed spinal nerve root compression	5	1
7. Spinal stenosis	0	1
8. Post-surgical status, 1–6 months after intervention	0	3
9. Post-surgical status, >6 months after intervention		
9.1. Asymptomatic	0	0
9.2. Symptomatic	2	4
10. Chronic pain syndrome	0	0
11. Other diagnoses	0	2
Total	47	67
Diagnosis of controls		
Neck pain	5	16
Shoulder pain	1	1
Knee pain	0	1
Other musculoskeletal disorders	11	10
Neurological pain syndromes	2	3
Total	19	31

Table II. *Characteristics of patients with low back pain and controls. Mean values with SDs in parentheses*

	Men		Women	
	LBP patients ($n = 47$)	Controls ($n = 19$)	LBP patients ($n = 67$)	Controls ($n = 31$)
Height (cm)	178.0 (7.3)	179.2 (6.1)	164.4 (5.5)	163.3 (5.9)
Weight (kg)	82.8 (13.4)	81.9 (10.5)	70.0 (12.6)	74.1 (10.1)
BMI (kg/m^2)	26.1 (4.0)	25.6 (3.3)	25.9 (4.5)	27.8 (4.7)
Age (years)	43.8 (13.2)	43.4 (11.4)	47.7 (9.5)	47.3 (9.9)

Table III. *Average disability scores and measurements of physical performance in patients with low back pain and controls (C), the area under the receiver operating characteristics curve and the statistical significance of the difference between groups (*p*). SDs are given in parentheses*

	Men				Women			
	LBP	C	p	ROC	LBP	C	p	ROC
Oswestry	29.6 (15.3) $n = 46$	9.1 (9.0) $n = 18$	<0.001	0.88	30.0 (14.2) $n = 65$	17.6 (14.9) $n = 30$	<0.001	0.73
Million	47.7 (18.2) $n = 44$	21.4 (15.5) $n = 17$	<0.001	0.86	53.2 (15.6) $n = 62$	33.7 (22.2) $n = 22$	<0.001	0.76
RBDS	7.3 (4.8) $n = 43$	3.8 (3.4) $n = 18$	0.068	0.72	6.1 (4.3) $n = 60$	5.9 (4.1) $n = 31$	0.56	0.50
R_t	0.455 (0.072) $n = 47$	0.446 (0.054) $n = 19$	0.61	0.50	0.470 (0.082) $n = 67$	0.460 (0.067) $n = 31$	0.56	0.50
F_{ext}	441.6 (174.5) $n = 45$	568.1 (200.7) $n = 19$	0.014	0.68	283.2 (127.1) $n = 66$	354.7 (159.6) $n = 31$	0.019	0.63
T_{ext}	209.2 (87.5) $n = 45$	268.5 (93.0) $n = 19$	0.018	0.69	120.0 (55.4) $n = 66$	158.1 (71.0) $n = 31$	<0.005	0.66
F_{fl}	409.9 (165.6) $n = 47$	475.9 (133.2) $n = 19$	0.13	0.50	243.6 (97.2) $n = 67$	282.1 (117.7) $n = 31$	0.092	0.60
T_{fl}	194.0 (79.9) $n = 47$	225.3 (64.0) $n = 19$	0.13	0.50	103.4 (42.0) $n = 67$	125.3 (51.4) $n = 31$	0.028	0.64

RBDS = Rimon's Brief Depression Scale, Rt = reaction time, F = Force, T = Torque, fl = flexion, ext = extension.

association between depression and Million or Oswestry score remained low or moderate in both LBP patients and controls (Table IV).

The cut-off point of 23 in the Oswestry score gave a specificity of 90% and a sensitivity of 50%; the cut-off point of 15 gave a specificity and sensitivity of 78% in men. The cut-off point of 35 gave a specificity of 90% but a sensitivity of only 30%; the cut-off point of 22 gave a specificity and sensitivity of 67% in women. Therefore, the best physical measurement (strength in extension) with two different cut-off points gave a specificity of 90% but a sensitivity of only 26% and a specificity of 58% and a sensitivity of 51% in men only. In women, a specificity of 90% but a sensitivity of only 12% and a specificity of 58% and a sensitivity of 64% were achieved.

Pain drawings were available from 65 men and 96 women. The sensitivity of pain drawing to detect patients with LBP was 100%. However, the specificity was low: 47% in men and 39% in women (Table V).

Table IV. *Association between Oswestry or Million disability index and Rimon's Brief Depression Scale as determined by Pearson's correlation coefficient* r *(95% confidence interval in parentheses)*

Disability index	Patients with LBP	Controls
Oswestry		
Men	0.40 (0.12–0.63)	0.61 (0.18–0.84)
Women	0.38 (0.14–0.58)	0.68 (0.36–0.86)
Million		
Men	0.53 (0.27–0.72)	0.79 (0.49–0.92)
Women	0.12 (−0.14–0.36)	0.29 (−0.15–0.63)

LBP = low back pain.

DISCUSSION

The results indicate that only the subjective evaluation of disability using the Oswestry or Million indices achieved an acceptable discriminatory power between the two groups. High disability scores in LBP patients were not due to depression. Trunk muscle strength and reaction time tests had no diagnostic value, in spite of some statistically significant differences between groups.

Any analysis that involves sensitivity and specificity measurements requires an independent criterion, "the gold standard", by which the diagnosis is judged (8). If this gold standard is subject to error, then it will bias the estimates of sensitivity and specificity (9). The reason for using other pain patients as controls, "the gold standard", was the supposed similarity of pain, depression, malingering and litigation problems between these patients. Thus, the observed differences between groups were probably due to those features that really were specific to patients with LBP.

In the present study, LBP was used like a medical diagnosis although back pain is merely a perceived condition. The patients were originally referred for a specialist consultation due to a particular pain problem. The classification to either LBP or other pain was confirmed by a second clinician. Thus, the probability that the classification was seriously biased is low. The LBP group in this study represents a clinically significant LBP disorder and can be assumed to be a representative sample of

Table V. *Association between clinical classification of patients and pain drawing*

	Men			Women		
Pain drawing	LBP patients	Controls	Total	LBP patients	Controls	Total
Back pain	46	10	56	65	19	84
No back pain	0	9	9	0	12	12
Total	46	19	65	65	31	96

LBP = low back pain.

patients with moderate or severe chronic LBP. In addition, the controls were those patients who had not been considered to have LBP by physicians or even by themselves. However, the surprisingly low specificity of pain drawing leads to speculation that mild pain in the lower back is such a common symptom that most patients can overlook it if it does not cause any real disability. This is supported by a recent Canadian study: approximately half the population of Saskatchewan had had low intensity or low disability back pain during the preceding 6 months but only 11% had had serious, disabling back pain (19).

It is generally accepted that patients with LBP perform less well than normal subjects in trunk muscle strength tests (20). In the present study it was not possible to find any convenient cut-off point for diagnostic purposes. According to ROC curve analysis, the sensitivity and specificity of trunk-strength tests remained low and were independent of the cut-off point criteria chosen. The few examples of sensitivity and specificity calculations confirmed the results obtained by ROC curve analysis. An earlier attempt to find a useful diagnostic cut-off point for trunk strength measurements by using traditional sensitivity and specificity analysis produced similar results (21). One reason for the low diagnostic value of strength measurements is probably the considerable overlap of muscle strength data between patients with LBP and healthy controls (20).

It has been suggested that the psychomotor reaction time is increased in patients with chronic LBP, if it is compared to healthy controls (5). In the present study, a relationship between slow reaction time and LBP was not confirmed. The diagnostic value of the hand and choice reaction test for LBP was also previously shown to be low (22). It is probably true that factors such as fear responses, depression or anxiety, which may influence reaction time in patients with LBP (5), are similar in patients with different pain syndromes. Thus, it can be assumed that long psychomotor reaction time is not a characteristic feature for patients with LBP. This probably holds in spite of the fact that the reaction time measurement method used in the present study was not totally comparable to the method used in the previous study (5).

The presence of low back trouble was the main inclusion criterion in this study. Surprisingly, patients' subjective opinion about disability was a discriminating factor between groups and not perceived pain in the lower back region. This self-reported disability was not explained by depression, which was part of the original hypothesis. The physical performance of patients with LBP is obviously different from patients with other pain syndromes in clinical conditions. Unfortunately, there is a continuing lack of methods to document this difference. In the present study, simple physical performance tests were not discriminating.

In conclusion, during the diagnostic process of common LBP focus should be on disability instead of pain if certain serious disease processes have been excluded. It follows that the analysis and documentation of perceived disability due to LBP and more functional tests of physical performance should be developed. At the moment the best method for assessing the severity of common LBP is to ask the patient.

ACKNOWLEDGEMENT

This study was financially supported by The Academy of Finland.

REFERENCES

1. Roux EB, Vischer TL, Brisson PM. Medical approach to low back pain. Baillieres Clin Rheumatol 1992; 6: 607–627.
2. Waddell G. Biopsychosocial analysis of low back pain. Baillieres Clin Rheumatol 1992; 6: 523–557.
3. Hazard RG, Reeves V, Weisman G, Fleming BC, Pope MH. Dynamic lifting capacity: the relationship between peak force and weight as an indicator of effort. J Spinal Disord 1991; 4: 63–67.
4. Mayer TG. Assessment of lumbar function. Clin Orthop 1987; 221: 99–109.
5. Taimela S, Österman K, Alaranta H, Soukka A, Kujala UM. Long psychomotor reaction time in patients with chronic low-back pain: preliminary report. Arch Phys Med Rehabil 1993; 774: 1161–1164.
6. Gatchel RJ, Mayer TG, Capra P, Diamond P, Barnett J. Quantification of lumbar function. Part 6: The use of psychological measures in guiding physical function restoration. Spine 1986; 11: 36–42.
7. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993; 52: 157–168.
8. Beck JR, Shultz EK. The use of relative operating characteristics (ROC) curves in test performance evaluation. Arch Pathol Lab Med 1986; 110: 13–20.
9. Campbell MJ, Machin D. Medical statistics, a commonsense approach. Chapter 3. Probability and decision making. Chichester, UK: John Wiley & Sons, 1992: 32–43.
10. Dixon WJ, ed. BMDP Statistical Software Manual, Vols. 1 and 2. Berkeley, CA: University of California Press, 1992.
11. Rantanen P, Nykvist F. Optimal sagittal motion axis for trunk extension and flexion tests in chronic low back trouble. Clin Biomech, in press.
12. Quebec Task Force. Diagnosis of the problem (The problem of diagnosis). Chapter 3. Spine 1987; 12(7S): S16–S21.
13. Fairbank JC, Davies JD, Couper J, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980; 66: 271–273.
14. Million R, Haavik Nilsen K, Jayson MIV, Baker RD. Evaluation of

low back pain and assessment of lumbar corsets with and without back supports. Ann Rheum Dis 1981; 40: 449–454.
15. Keltikangas-Järvinen L, Rimon R. Rimon's brief depression scale, a rapid method for screening depression. Psychol Rep 1987; 60: 111–119.
16. Biering-Sørensen F. Physical measurements as risk indicators for low back trouble over a one year period. Spine 1984; 9: 106–119.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; i: 307–310.
18. Altman DG, Gardner MJ. Calculating confidence intervals for regression and correlation. In: Gardner MJ, Altman DG, eds. Statistics with confidence—Confidence intervals and statistical guidelines. Belfast, UK: BMJ, 1989.
19. Cassidy JD, Carroll LJ, Cote P. Saskatchewan health and back survey. The prevalence of low back pain and related disability in Saskatchewan adults. Spine 1998; 23: 1860–1867.
20. Newton M, Waddell G. Trunk strength testing with iso-machines. Part 1: Review of a decade of scientific evidence. Spine 1993; 18: 801–811.
21. 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.
22. Luoto S, Taimela S, Alaranta H, Hurri H. Psychomotor speed in chronic low-back pain patients and healthy controls: Construct validity and clinical significance of the measure. Percept Mot Skills 1998; 87: 1283–1296.