Scand J Rehab Med 26: 211–215, 1994

NON-DYNAMOMETRIC TRUNK PERFORMANCE TESTS: RELIABILITY AND NORMATIVE DATA

Hannu Alaranta, MD, PhD,[1] Heikki Hurri, MD, PhD,[1] Markku Heliövaara, MD, PhD,[2] Anni Soukka, MD[2] and Riitta Harju, PT[1]

From the [1] Rehabilitation Unit of The Invalid Foundation and [2] Social Insurance Institution, Helsinki, Finland

ABSTRACT. A sample of 508 male and female white-collar and blue-collar employees aged 35 to 54 years was evaluated clinically to determine the reliability of repetitive sit-ups, repetitive arch-ups, repetitive squatting, and static back endurance tests, to determine the normal values of these tests and to detect determinants for trunk muscle performance. All of the given tests had fairly good or even excellent test-retest reliability. Muscular performance capacity decreased with advancing age, particularly among blue-collar workers. Men showed greater muscle endurance in all the muscle tests, and blue-collar workers lower values in all tests. The repetitive tests, in particular, showed reduced values in those with previous low-back pain. Normative values of back endurance, repetitive squatting, sit-up and arch-up tests for different age, sex and occupational groups are presented.

Key words: trunk performance capacity, muscle tests, reliability, normative data, age, sex, occupation, low-back pain.

INTRODUCTION

Physical measures show a relatively weak correlation with subjective symptoms in low-back pain patients. Psychosocial and work environmental factors have proved to be better predictors of disability than physical factors (7). Nevertheless, a number of methods for measuring trunk muscle performance are in routine use today in the evaluation of disability or rehabilitation outcome. With new technology, it has become possible to measure dynamic trunk muscle strength more accurately than before (8, 12).

The high cost of the dynamometric methods, however, has raised questions about the real benefits of the various sophisticated dynamometric devices in clinical practice (19). In a study of 185 chronic low-back pain patients, Rissanen et al. (18) concluded that non-dynamometric tests are still useful in clinical practice. Unexpectedly, the arch-up and sit-up tests had even stronger associations with the perceived disability (Million index; 16) than the isokinetic tests. Among women, the isokinetic trunk flexion and extension tests ($r = 0.27$–0.46) and the non-dynamometric tests ($r = 0.29$–0.49) correlated equally with pain and disability, whereas in men the isokinetic flexion and extension tests ($r = 0.01$–0.32) had weaker correlations than the non-dynamometric tests ($r = 0.26$–0.40).

When comparing the clinical status with findings in time series and with observations by several persons, the lack of standards and normal values often becomes a problem. Reliable comparisons between different patient groups and within time series require normative data for age, sex and occupational groups. To establish this we undertook the present study. We also sought to find such strength measurements as would be applicable to both healthy people and chronic low-back pain patients. The present study was focused on trunk muscle and leg muscle endurance tests. Endurance is mechanically defined as the point where the contraction can no longer be maintained at a certain level (isometric fatigue, or static back endurance in the present study), or when repetitive work can no longer be sustained at a certain output (dynamic fatigue, measured via repetitive sit-up, arch-up and squatting tests in the present study) (15).

More specifically, the aims of the present study were: i) to determine the reliability of muscular trunk performance tests by non-dynamometric means, ii) to determine the normal values of non-dynamometric trunk performance tests for different age and sex groups, and iii) to assess the relation of non-dynamometric test results with previous low-back pain.

MATERIALS AND METHODS

Subjects

This study was part of a comprehensive clinical survey of 594 persons aged 35–54 years working in four defined occupational and gender groups for the Helsinki City Council. Of this sample, 508 persons (86%) attended a clinical evaluation at the Orthopaedic Hospital of the Invalid Foundation: 159 of 187 male truck or tractor drivers (blue collar workers) (85%), 95 of 112 male civil servants (white collar workers) (85%), 160 of 183 female office or school cleaners (blue collar workers) (87%), and 94 of 112 female clerks or civil servants (white collar workers) (84%). At the phase of the test all the subjects were capable of working.

Methods

History of low-back pain was obtained using a standard questionnaire (13). The subjects were asked to classify themselves into one of the following categories: 1. Never suffered from pain in the lower back; 2. Low-back pain more than 12 months ago; 3. Low-back pain during the previous 12 months but no disability due to the pain; 4. Low-back pain during the previous 12 months and limitation in daily tasks at work or in leisure time because of low-back pain.

Repetitive sit-up test: The subject was lying in supine position with the knees flexed at 90° and was held fixed by the tester in the ankle region. The subject moved the arms towards the knees. After having touched the knee caps with the thenar region, the upper trunk curled back down to the supine position.

Repetitive arch-up test: The subject was lying in a prone position with the arms along the sides, the inguinal region at the edge of the test-table, the upper trunk flexed downward at 45°, and was held fixed by the tester in the ankle region. The subject moved the upper trunk up to horizontal position and back down to 45° flexion.

Repetitive squatting test: In the initial position, the subject was standing with the feet 15 cm apart. The subject squatted until the thighs were horizontal and then returned to a standing position.

In all the above tests the repetition rate was one repetition per 2–3 seconds. The movement was repeated as many times as possible in peaceful but constant pace; the repetition maximum was 50 times.

Static back endurance test: The test was modified from the test by Biering-Sörensen (3). The subject was lying in prone position, fixed in the ankle region and with the upper limbs held along the sides. The upper trunk was unsupported horizontally, the inguinal region at the edge of the test-table. The number of seconds the subject was able to maintain the horizontal position was recorded; the duration maximum was 240 seconds.

Procedure for reliability and reproducibility: The intraobserver reproducibility was assessed with two measurements made at a one-year interval by the same physiotherapist. The 93 subjects were those without low-back pain either at baseline or during follow-up. The aim was to detect such measurements as would be sufficiently reproducible in healthy subjects over a period of several months at least.

The interobserver reliability was studied from measurements carried out by two trained physiotherapists at one-week intervals for 17 subjects in random order. None of the subjects suffered from low-back pain.

Table I. *Intraobserver reproducibility by the same tester one year apart among subjects without low-back pain (n = 93)*

	1. test		2. test			
	x̄	SD	x̄	SD	p^a	r^b
Repetitive sit-up	23	14	23	15	0.54	0.84
Repetitive arch-up	29	12	29	13	0.61	0.65
Repetitive squatting	30	14	30	15	0.47	0.87
Static back endurance (sec)	96	51	99	58	0.60	0.63

[a] Statistical significance of systematic shift (paired *t*-test).
[b] Reliability coefficient (22).

Statistical methods

The reliability and reproducibility of the measurements were assessed with the reliability coefficient (22). A coefficient greater than 0.75 is considered to represent excellent reliability, coefficients between 0.4 and 0.75 represent fair to good reliability, coefficients below 0.4 represent poor reliability (6, 14). The difference between the first and the second measurement was calculated for each individual and the presence of a systematic shift between the measurements was tested with the paired *t*-test.

The adjusted means of the measurements were estimated on the basis of a general linear model (4). Age (years), sex (male, female), occupation (white-collar, blue-collar) and the type/level of pain were entered in the multifactorial models as explanatory variables.

RESULTS

The one-year intraobserver reproducibility of the muscular measurements showed fairly high coefficient values ranging from 0.63 to 0.87 (Table I). The highest values were noted for repetitive sit-ups and squatting.

None of the tests showed statistically significant shifts between the two measurements.

The interobserver reliability coefficients were also fairly good or excellent, ranging from 0.66 to 0.95 (Table II). Repetitive squatting had the highest value. The back endurance test was the only one which showed a statistically significant shift between the two measurements.

Squatting and sit-up test measurements showed a significant decrease with aging even when adjusted for sex and history of low-back pain, unlike back endurance or arch-up measurements (Table III). Men had higher test results in all the tests, although back endurance just failed to reach statistical significance. In all the muscle tests, white-collar workers showed stronger muscle performance after adjustment for age and sex. All the measurements, particularly

Table II. *Interobserver reliability by two testers one week apart among subjects without low-back pain (n = 17 subjects; 34 measurements)*

	$\bar{x}$	r	$D\bar{x}$	DSD	t	p
Repetitive sit-up	27	0.91	0.35	6.53	0.32	0.75
Repetitive arch-up	33	0.83	2.06	7.16	1.68	0.10
Repetitive squatting	32	0.95	−0.33	4.99	0.38	0.71
Static back endurance (sec)	109	0.66	23.44	41.83	3.27	0.003

$\bar{x}$ = mean of all the measurements (2 × N).
r = reliability coefficient (22).
$D\bar{x}$ = mean difference (1.-2.).
DSD = deviation of Dx.
t = paired t-test (df = n − 1).
p = statistical significance of systematic shift.

sit-up- and arch-up-tests, were found to be lowest in those with a history of disabling low-back pain, independently of age, sex and occupation (Table III).

The normative values of trunk strength measurements are presented for different age, sex and occupational groups in Table IV.

DISCUSSION

A problem with trunk strength measurements in low-back pain patients is the wide individual variation in the relationship of trunk muscle performance capacity with low-back pain. This variation obviously decreases the value of strength measurement in the assessment of the disability (1, 5). Thus there is an urgent need for normative data, which was one goal the present study set out to reach.

The only comparable data for the reliability assessments presented here has been published for back endurance (3, 10, 11, 17). In the earlier studies as well as in the present study, the reliability of the back muscle endurance test was found to be good. The lack

Table III. *Adjusted mean values of back endurance (seconds) and repetitive squatting, sit-up and arch-up tests (number of repetitions) according to age, sex, occupation, and previous low-back pain*

Explanatory factor	n	Back endurance	Squatting	Sit-up	Arch-up
Age[1]					
35–39	123	90	33	28	27
40–44	136	93	30	24	26
45–49	107	103	29	23	28
50–54	109	88	24	19	25
p-value		0.20	<0.001	<0.001	0.30
Sex[2]					
Male	242	98	37	27	28
Female	233	89	21	20	25
p-value		0.06	<0.001	<0.001	0.008
Occupation[3]					
White-collar	181	109	34	29	30
Blue-collar	294	84	27	21	24
p-value		<0.001	<0.001	<0.001	<0.001
Previous low-back pain[4]					
Never any pain	116	100	31	23	29
Pain more than 12 months ago	46	85	29	23	27
Pain during previous 12 months, no disability	166	99	31	27	28
Disabling pain during previous 12 months	147	85	27	21	23
p-value		0.05	0.03	0.005	0.002

Class-specific mean values were adjusted for [1] sex and occupation; [2]age and occupation; [3]age and sex; [4]age, sex and occupation.

Table IV. *Mean values ($\bar{x}$) and standard deviations (SD) of repetitive sit-up, arch-up and squatting tests, and static back endurance test*

AGE	MALES						FEMALES					
	Blue collar		White-collar		All		Blue collar		White collar		All	
	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
Repetitive sit-up-test												
35–39	29	13	35	13	32	13	24	12	30	16	27	14
40–44	22	11	34	12	27	13	18	12	19	13	19	12
45–49	19	11	33	15	24	14	17	14	22	15	19	14
50–54	17	13	36	16	23	16	9	10	20	13	11	11
35–54	23	13	35	13	27	14	17	13	24	15	19	14
Repetitive arch-up-test												
35–39	26	11	34	14	29	13	28	13	27	11	27	12
40–44	23	12	36	14	28	14	25	14	20	11	23	13
45–49	24	13	34	16	28	15	25	15	31	16	27	15
50–54	21	11	35	17	26	15	18	14	26	14	19	14
35–54	24	12	35	15	28	14	24	14	26	13	24	14
Repetitive squatting test												
35–39	39	13	46	8	42	12	24	11	27	12	26	12
40–44	34	14	45	9	38	13	22	13	18	8	20	12
45–49	30	12	40	11	33	13	19	12	26	13	22	13
50–54	28	14	41	11	33	14	13	10	18	14	14	11
35–54	33	14	43	10	37	13	20	12	23	12	21	12
Static back endurance test (sec)												
35–39	87	38	113	47	97	43	91	61	95	48	93	55
40–44	83	51	129	57	101	57	89	57	67	51	80	55
45–49	81	45	131	64	99	58	90	55	122	73	102	64
50–54	73	47	121	56	89	55	62	55	99	78	69	60
35–54	82	45	123	55	97	53	82	58	94	62	87	59

of corresponding information concerning the other tests is astonishing because the tests presented here have been commonly applied in clinical practice. On the basis of this study, the short-term reliability and the long-term reproducibility of the non-dynamometric tests represent a sufficient standard.

The demand for reliability is obvious, but we argue that long-term reproducibility is also an important criterion when looking for suitable methods for follow-up studies or for methods for workplace health care. A test with wide variation within a year (even reliable over a period of two weeks) would not fulfill that demand. When a test is reliable over a long course of time among healthy people, it can as well apply to monitoring true changes due to disease process, if the test is sufficiently valid. We consider the tests presented here suitable in, for instance, workplace health care units, and, where ever it is desirable to avoid the expensive dynamometric methods.

Age, sex and occupation form the basic strata when presenting normative data. It is common knowledge that women are weaker than men, as also confirmed by the present results. However, when strength is adjusted for body weight, women may be as strong as men (20). Squatting and sit-up muscle tests especially showed that physical capacity was reduced with advancing age. It was, however, somewhat surprising that physical capacity deteriorated more rapidly among blue-collar than white-collar workers. The observation gives rise to the question whether we should more vigorously combat physical incapacity among blue-collar workers. Lately the role of psychosocial factors and work satisfaction have duly received more emphasis (2, 9), but we should certainly not ignore the basic physical capacity of the employees, for whom physical performance is still a crucial issue.

All the repetitive muscle tests presented here showed somewhat lowered mean values among the cases of disabling pain during the preceding year, but perhaps less than expected. Back endurance seemed least affected by previous low-back pain. That it does not have a very close association with the experience of low-back pain is of extra value for the test with respect to clinical applications.

Usually one presents the normative data for those subjects who are completely symptom-free. We considered it proper to pool all the subjects together because of the special interest in the trunk muscle strength of working people; most of them experience some symptoms in any case, because low-back pain is so common. It should also be noted that since the target population comprised working people, they cannot have any very severe spinal pathology or symptoms, and roughly one-third of the subjects were totally symptom-free.

Mechanical parameters of fatigue, such as failure to maintain a posture or pace of work, are highly subjective phenomena influenced by motivation. In particular integrated electromyography, as a non-invasive technique, may provide a more appropriate way to evaluate muscle endurance in the future. However, the present tests have proved to be quite convenient in a clinical practice contributing to diagnostic and therapeutic assessments. They are inexpensive, and aggravation of pain has occurred extremely seldom during the past four years that we have applied them on a regular basis among thousands of low-back pain patients as well as healthy subjects.

ACKNOWLEDGEMENT

The study was sponsored by The Finnish Work Environment Fund.

REFERENCES

1. Battie, M. C.: The reliability of physical factors as predictors of the occurrence of back pain reports. A prospective study within industry. Thesis, Gothenburg University, Sahlgren's Hospital, Göteborg, Sweden, 1989.
2. Bigos, S. J., Battie, M. C., Spengler, D. M., Fisher, L. J., Fordyce, W. E., Hansson, T. H., Nachemson, A. L. & Wortley, M. D.: A prospective study of work perceptions and psychosocial factors affecting the report of back injury. Spine *16*: 1–6, 1991.
3. Biering-Sörensen, F.: Physical measurement as risk indicators for low back trouble over a one-year period. Spine *9*: 143–148, 1984.
4. Cohen, J. & Cohen, K.: Applied multiple regression/correlation analysis for the behavioral sciences. Wiley, New York, 1975.
5. Deyo, R. A.: Measuring the functional status of patients with low-back pain. Arch Phys Med Rehabil *69*: 1044–1053. 1988.
6. Fleiss, J. L. & Cohen, J.: The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas *33*: 613–619, 1973.
7. Frymoyer, J. W.: Predicting disability from low-back pain. Clin Orthop *279*: 101–109, 1992.
8. Hasue, M., Masatoshi, F. & Kikuchi, S.: A new method of quantitative measurement of abdominal and back muscle strength. Spine *5*: 143–148, 1980.
9. Hurri, H.: The Swedish back school in chronic low-back pain. Part II. Factors predicting the outcome. Scand J Rehabil Med *21*: 41–44, 1989.
10. Hyytiäinen, K., Salminen, J. J., Suvitie, T., Wickström, G. & Pentti, J.: Reproducibility of nine tests to measure spinal mobility and trunk muscle strength. Scand J Rehabil Med *23*: 3–10, 1991.
11. Jorgensen, K. & Nicolaisen, T.: Trunk extensor endurance: determination and relation to low-back trouble. Ergonomics *30*: 259–267, 1987.
12. Kishino, N. D., Mayer, T. G., Gatchel, R. J., McCrate, M., Anderson, C., Gustin, L. & Mooney, V.: Quantification of lumbar function. Part 4: Isometric and isokinetic lifting simulation in normal subjects and low-back dysfunction patients. Spine *10*: 921–927, 1985.
13. Kuorinka, I., Jonsson, B., Kilbom, Å., Vinterberg, H., Biering-Sörensen, F., Andersson, G. & Jörgensen, K.: Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. Applied Ergonomics *18*: 233–237, 1987.
14. Landis, J. R. & Koch, G. G.: The measurement of observer agreement for categorial data. Biometrics *33*: 159–174, 1977.
15. Liang, M. H. & Katz, J. N.: Clinical evaluation of patients with a suspected spine problem, In: The Adult Spine: Principles and Practice (ed. J. W. Frymoyer), pp. 233–274, Raven Press, Ld., New York, 1991.
16. Million, R., Nilsen, K. H., Jayson, M. I. V. & Baker, R. D.: Evaluation of low-back pain and assessment of lumbar corsets with and without back supports. Ann Rheum Dis *40*: 449–454. 1981.
17. Nordin, M., Kahanovitz, N., Verderame, R., Parnianpour, M., Yabut, S., Viola, K., Greenidge, N. & Mulvihill, M.: Normal trunk muscle strength and endurance in women and the effect of exercises and electrical stimulation. Spine *12*: 105–111, 1987.
18. Rissanen, A., Alaranta, H., Sainio, P., Härkönen, H.: Isokinetic and non-dynamometric tests in low-back pain patients related to pain and disability index. Spine (in press) 1994.
19. Schoene, M. (ed): Back machines: a waste of money? The Back Letter *5*: 8, 1991.
20. Smidt, G. L., Blanpied, P. R. & White, R. W.: Exploration of mechanical and electromyographic response of trunk muscles to high intensity resistive exercise. Spine *14*: 815–830, 1989.
21. Triano, J. & Schultz, A.: Correlation of objective measure of trunk motion and muscle function with low back disability ratings. Spine *6*: 561–565, 1987.
22. Winer, B.J.: Statistical principles in experimental design. 2nd ed. Tokyo, McGraw-Hill, 1971.

Address for offprints:

Hannu Alaranta, MD, PhD
The Rehabilitation Unit of
The Invalid Foundation
Tenholantie 10
FIN-00280 Helsinki
Finland