

Reliability of Musculoskeletal Fitness Tests and Movement Control Impairment Test Battery in Female Health-Care Personnel with Re-Current Low Back Pain

Raija Pia Annika Taulaniemi^{1*}, Markku Juhani Kankaanpää², Kari Juhani Tokola¹, Hannu Antero Luomajoki³ and Jaana Helena Suni¹

¹UKK Institute for Health Promotion Research, Tampere, Finland

²Department of Physical and Rehabilitation Medicine, Tampere University Hospital, Finland

³Zurich University of Applied Sciences, Winterthur, Switzerland

Abstract

Background: Low back pain (LBP) is one of the most important causes of work absenteeism in health-care personnel. Low performance levels for fitness components, and movement control impairment (MCI) of the spine have been posited to be risk factors for persistent LBP. The aims of this study were to examine the long-term reliability (test I – test II) of selected motor and musculoskeletal fitness tests and the standard MCI test battery in female health-care personnel with nonspecific LBP, and to investigate associations between change in intensity of LBP and variation in test results in two measurement points.

Methods: The participants (n=47) were workers in geriatric wards. Long-term reliability of six field tests assessing motor abilities, flexibility and muscular strength, and the MCI test battery were studied. Mean test I – test II time-span was 18 days (SD 7.9). The estimates of reliability for interval-scale measurements (fitness) were typical error (s) of measurement and the coefficient of variation (CV). The reliability of nominal-scale measurements was analysed by the kappa coefficient.

Results: The lowest within-subject variation was found in running figure of eight test (s=0.22 s, CV=2.8%) and the highest for modified push-ups (s=1.04 repetitions, CV=12.2%). The kappa values for the dorsal pelvic tilt, sitting knee extension, the anterior pelvic tilt, and lying-prone knee flexion varied between 0.71 and 0.45 but were lower for rocking forwards and backwards (k=0.31) and the one-leg stance (k=0.16). Change in the pain intensity was associated only with variation in test I – test II results for modified push-ups (p=0.045).

Conclusions: All of the tests assessing musculoskeletal fitness and four of the MCI tests had an acceptable level of long-term reliability. The two tests for MCI that had poor reliability need to be better standardised in order to serve as reliable clinical measurement tools.

Keywords: Muscle strength; Motor ability; Movement control; Reproducibility; Repeatability; Patients with pain

List of abbreviations: ANOVA: Analysis of Variance; BMI: Body Mass Index; CI: Confidence Interval; CV: Coefficient of Variation; k: Cohen's Kappa Coefficient; LBP: Low Back Pain; MCI: Movement Control Impairment; NSLBP: Non-specific Low Back Pain; NRS: Numeric Rating Scale; r: Pearson's Correlation Coefficient; s: Typical (standard) Error of Measurement; VAS: Visual Analogue Scale

Introduction

The prevalence of LBP varies between occupational groups, with physically demanding jobs known to be a risk factor [1]. Among health-care workers, the one-year prevalence of LBP has been found to be as high as 45% to 77% [2]. Low back problems have often developed already during the nursing-school clinical training period [3]. Among nursing personnel, musculoskeletal disorders are the main cause of work absenteeism and early retirement, LBP being the leading cause. In Finland, only 48% of nursing assistants and 58% of nurses reach the old-age pension point [4].

In most low back patients (85–90%) the pain is classified as nonspecific low back pain (NSLBP). After an acute pain episode, the majority recover, but 50–70% find the pain recurring within the following year, and 10% will become chronic [5]. In the literature there is still some ambiguity in conceptions of the episodic or fluctuating nature of NSLBP [6]. LBP episodes are traditionally regarded as individual events with good prognosis. This opinion has been challenged in longer-term follow-up studies [7–9]. These studies

indicate, that majority of people with LBP do experience back pain off and on over long period of time [10]. Preventing new LBP episodes is considered to be of importance for prevention of persistent pain.

Whilst associations found between physical fitness and LBP are partly conflicting, there is increasing evidence that low performance levels for several elements of physical fitness are risk factors for LBP [11–15], and LBP patients tend to be more deconditioned and less physically active than healthy controls [16,17]. Accordingly, sufficient muscular strength and motor abilities might protect the back and prevent back problems whereas poor performance levels might indicate predisposition to persistent LBP especially in physically demanding work such as nursing [18,19]. Health-care workers are exposed to a high physical workload encompassing frequent bending

***Corresponding author:** Raija Pia Annika Taulaniemi, Urho Kaleva Kekkonen Institute for Health Promotion Research (UKK Institute), Kaupinpuistonkatu 1, 33500 Tampere, Finland, Tel: +358 3 2829 264; Fax: +3583 2829 200; E-mail: annika.taulaniemi@uta.fi

Received October 16, 2015; **Accepted** January 20, 2016; **Published** January 30, 2016

Citation: Taulaniemi RPA, Kankaanpää MJ, Tokola KJ, Luomajoki HA, Suni JH (2016) Reliability of Musculoskeletal Fitness Tests and Movement Control Impairment Test Battery in Female Health-Care Personnel with Re-Current Low Back Pain. J Nov Physiother 6: 282. doi:10.4172/2165-7025.1000282

Copyright: © 2016 Taulaniemi RPA, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and heavy lifting. Therefore, it is conceivable that movement control (i.e. motor skills), ability to stabilize the spine with trunk muscles and muscular endurance/strength are of importance in connection with the risk of developing LBP [5,13,20]. Consequently there is a clear need for reliable measurement methods for evaluation of musculoskeletal fitness and motor abilities in connection with physically demanding nursing work. Repeatability of the field tests for health-related fitness has been studied in healthy populations [21,22], but to our knowledge not in people with NSLBP.

Impairments in postural and movement control of the lumbar spine have been posited to be risk factors for prolonged LBP [23-25]. A significant difference in ability to actively control the movement of the low back has been found between patients with LBP and subjects without back pain, and between acute and chronic LBP patients [26,24], but to our knowledge the test-retest repeatability of the standard movement control impairment (MCI) test battery has not been studied.

Exercise is often recommended for LBP patients [27]. Reliable and valid measurement methods are essential for evaluating the effects of exercises. In addition, methods of measurement for screening the physical performance capacity of people who have a physically demanding job and who may be at risk for LBP are needed.

The standard measure of reliability is the test-retest repeatability within one week [28]. However, long-term reliability (with 2-3 weeks interval between measurements) of measurements for sensorimotor functions [29] and muscle strength and endurance measurements [30] has been studied in patients with chronic LBP, but not in people with minor LBP problems, who are still at work.

The aim of the study reported upon here was to examine the long-term reliability of selected motor and musculoskeletal fitness tests and the MCI test battery for female health-care personnel with recurrent NSLBP. Another aim was to ascertain whether change in pain intensity between two measurement points had any effect on the variation in the test results.

Materials and Methods

Participants

The participants (n=47) of the present reliability study were a subsample from a larger randomised controlled trial (the NURSE RCT, clinical trial registration NCT04165698). The both measurement sessions (test I and test II) in the present reliability study were conducted before the study interventions started, and no feedback, education or training was given to participants between measurements.

Written information about the NURSE RCT was disseminated by the head nurses to all health-care personnel working in the geriatric wards of two municipal hospitals and an old people's home in late 2011 in Tampere, Finland. Personnel with LBP who were willing to participate in the NURSE RCT filled in the screening questionnaire. To be eligible, personnel had to meet the following criteria:

- 1) Being a woman aged 30–55
- 2) Who had worked in her current job for at least 12 months and
- 3) Experienced LBP intensity of at least 2 on the numeric rating scale (NRS) employed (0–10) [31] within the preceding four weeks.

The exclusion criteria for the study were

- 1) Serious former back injury (fracture, surgery, or prolapsed disc),
- 2) Chronic LBP as defined by a physician or self-reporting of daily LBP

over the past seven months or longer, 3) A serious other disease or symptoms limiting participation in moderate-intensity neuromuscular exercise, 4) Engaging in neuromuscular-type exercise more than once a week, and 5) Pregnancy or recent delivery (< 12 months). Written informed consent was obtained from all participants.

More than 80% of the study subjects had irregular working hours, in two or three shifts, and 68% of them were assistant nurses. The majority of them perceived their health to be average or good, although 66% reported having at least three musculoskeletal pain sites. Thirty per cent rated their fitness level somewhat worse than that of persons of the same age and gender. Other background characteristics of the participants are reported in Table 1.

Study design

Long-term reliability was studied by means of test I – test II design. The average interval between pre-study screening and the first measurement point was approximately three weeks, and the mean number of days between the two measurements points was 18 (SD 7.9). The reliability of selected musculoskeletal fitness tests and the MCI test battery were studied in the analysis. Health screening was conducted prior to testing, in accordance with the safety model of the Health-related Fitness Test Battery for Middle-aged Adults [32].

At the first measurement point (test I), an experienced physiotherapist (with master's degree and 14 years' experience of clinical work) scored the performance of the subjects in six tests of MCI and six field tests of motor and musculoskeletal fitness in set order. At the second measurement point (test II), the same physiotherapist followed the test protocol as used before. The results from the first measurements occasion were not available for the physiotherapist at

	mean (SD)	n	%
Age (years)	47.9 (5.9)		
Profession			
assistant nurse		34	72
nurse		8	17
physiotherapist or PT assistant		5	11
Number of years of work	11.4 (9.4)		
Work characteristics			
three-shift work		27	57
two-shift work		13	28
regular day shifts		5	11
regular night shifts		2	4
Self-reported work ability (0–10 scale, 0=unable to work, 10 = best possible)	7.9 (1.4)		
Perceived health			
below average		-	
average		23	49
good		22	47
very good		2	4
Perceived fitness in comparison to persons with same age and gender			
somewhat worse		14	30
similar		24	51
somewhat better		8	17
much better		1	2
BMI	26.7 (4.5)		
Number of musculoskeletal pain sites			
≤ 2		16	34
3		13	28
≥ 4		18	38
Frequency of low back pain in test I			
daily		3	7
in some or most days of the week		29	71
recovered from the pre-study screening		9	22

Table 1: Baseline characteristics of the study sample (n=47).

the second measurement point. The measurements were conducted in the physiotherapy department of the geriatric hospital where most of the study subjects were working.

Test protocols and measurements

Motor and musculoskeletal fitness tests: The field tests of motor and musculoskeletal fitness assessed motor abilities, flexibility and muscular strength. The motor ability tests used included rhythm co-ordination and running a figure of eight [33]. Trunk side-bending [22] was used for assessing the range of motion of lateral flexion of the spine, and dynamic sit-ups [34], modified push-ups and one-leg squats [22] for determination of muscular strength.

The movement control impairment test battery: The MCI test battery [26], based on descriptions by Sahrman [35] and O’Sullivan [36], consists of six tests: 1) the waiters bow (flexion of the hips in upright standing without movement of the lower back), 2) dorsal tilting of the pelvis, 3) sitting knee extension, 4) rocking backwards and forwards in quadruped position, 5) prone-lying active knee flexion and 6) one-leg stance. The MCI test results were judged dichotomously by observation: the subject was noted either as not having motor control impairment (0) or as having impairments (1).

All individual measurements were conducted in standardised order for each study subject: MCI tests before fitness tests, and motor and flexibility tests before measurements of muscular strength. The test instructions, criteria for the MCI test battery, and motor and musculoskeletal fitness tests are described briefly in Supplementary information (1).

The measurements of LBP intensity: Intensity of LBP during previous four weeks was measured by NRS (1-10) [31] at pre-study screening, at the baseline (test I) and at the re-test session (test II) in order to get an overview of pain development. To clarify more precisely the intensity of LBP at both measurement points (test I and test II), pain intensity during past seven days was also measured via visual analogue scale VAS (where 0 = no pain and 100 = the worst possible pain) [37].

Statistical analysis

The mean, standard deviation (SD), and range of the test I–test II measurements of the fitness tests are presented as descriptive statistics in Table 2.

Estimates of repeatability for interval-scale measurements were calculated in a manner suggested by Hopkins [38]. 1) within-subject variation in terms of typical error (s) as the standard deviation of the test–retest difference divided by the square root of 2 ($s = sd_{diff} / \sqrt{2}$) and 2) the relative error measure coefficient of variation (CV) as typical error divided by the mean of two tests [$CV = s / \text{mean}(\text{test}_1 + \text{retest}_1) / 2$].

Systematic change in the mean is an important issue when subjects perform repeated series of test trials [38,28]. Therefore, the percentage changes in mean performance between the first and second test results were calculated too. The systematic changes in mean between the two measurement sessions, which were considered statistically significant if their 95% confidence interval (CI) did not include value zero, were also calculated.

The repeatability of the nominal-scale measurements in the MCI tests (0 or 1) was analysed by means of Cohen’s kappa coefficient (k).

For analysis of the possible effect of change in pain intensity on variation between the test I and test II results, new variables were calculated: 1) change in intensity of LBP, measured by VAS (0-100)

during last seven days 2) change between the first and second MCI tests (indicating that performance had deteriorated, remained unchanged, or improved), and 3) change in fitness-test results. Normality of distributions for numerical variables was confirmed. Associations between change in pain and nominal-class variables for change in MCI tests were analysed via the Kruskal-Wallis test. Kruskal-Wallis test was used instead of analysis of variance (ANOVA), because number of cases in some categorical classes was too small for ANOVA. Association between change in LBP and change in interval-scale fitness-test results were examined by means of Pearson’s correlation coefficient (r). All analyses were conducted with the SPSS statistical analysis package, version 22.

The criteria for sufficient reliability

At present, there are no standards delimiting acceptable measurement precision for monitoring physical fitness in healthy populations, nor for monitoring the motor or musculoskeletal fitness levels of people with recurrent or chronic LBP. Altman [39] has rated the kappa coefficient thus: ≤ 0.20 =poor, 0.21 – 0.40 =fair, 0.41 – 0.60 =moderate, 0.61 – 0.8 =good, and ≥ 0.80 =very good.

Results

The participants had sub-acute or periodic, but not chronic NSLBP. Mean of their pain intensity (NRS; 1-10 during last four weeks) decreased from 4.5 (SD 1.8) at the pre-screening time to 3.4 (2.4) at the first measurement point and further to 2.8 (2.7) at the second measurement point. At the first measurement point 71% of the study sample had LBP in some or most days of the week, but not daily, 7% had daily pain and 22% had recovered from the screening time and were pain-free (Table 1).

The mean number of days between the test I and test II measurements was 18, and the median LBP intensity, measured on VAS (0–100 mm) during last seven days, decreased from 26 ($Q_1 = 15$, $Q_3 = 51$) to 19 ($Q_1 = 4$, $Q_3 = 36$) in that time. The reduction in pain intensity was statistically significant ($Z = -2.77$, $p = 0.006$). The change in VAS between two measurement points was less than 15mm for 51% of the study subjects, for 38% the decrease in pain intensity was more than 15mm and for 11% the pain intensity increased more than 15 mm.

Descriptive results from test I and test II measurements of the motor and musculoskeletal fitness tests are presented in Table 2. In general, all subjects were able to perform the rhythm co-ordination,

Fitness test	n	mean	sd	min	max
Running figure of eight (seconds)					
1st measurement	46	8.1	0.65	6.8	9.9
2nd measurement	46	8.1	0.62	7.9	9.9
Rhythm coordination (points)					
1st measurement	47	13.7	2.9	6	16
2nd measurement	47	14.1	2.8	7	16
Trunk lateral flexion (cm)					
1st measurement	47	18.0	3.1	11.7	24.9
2nd measurement	47	18.1	3.3	10.6	25.7
Dynamic sit-up (repetitions)					
1st measurement	47	16.4	5.3	5	20
2nd measurement	47	17.5	4.7	5	20
One leg squat (repetitions)					
1st measurement	47	9.5	2.3	2	12
2nd measurement	47	9.6	2.7	2	12
Modified push-up (repetitions)					
1st measurement	46	7.9	2.7	0	15
2nd measurement	46	9.2	2.6	0	18

Table 2: Descriptive results of test I – test II measurements of the musculoskeletal fitness tests.

trunk lateral flexion, dynamic sit-up, and modified one-leg squat tests. On account of musculoskeletal problems other than LBP, one person was excluded from the modified push-up test and one from the running figure of eight.

Results of the test I – test II long-term reliability analysis for interval-scale fitness measurements are presented in Table 3. The least within-subject variation was found for the running figure of eight ($s = 0.22$ s, $CV = 2.8\%$) and the most for modified push-ups ($s = 1.04$ reps, $CV = 12.2\%$). Systematic changes in the mean between the two measurement sessions were detected for dynamic sit-ups and modified push-ups, which indicates a slight learning effect.

Results speaking to the level of test I – test II long-term reliability of the MCI tests are presented in Table 4. The best repeatability was found for the dorsal pelvic tilt and poorest for the one-leg-stance.

Associations between change in pain intensity and variation in test I – test II results are presented in Table 5. Only the modified push-up test seemed to be correlated with changes in pain. Lower pain levels implied a higher number of push-up repetitions ($r = -0.3$, $p = 0.045$, $n = 46$), for better test results when the pain had decreased, or on account of the learning effect. The change in pain had no effect on variation in any other test, including the MCI tests.

Discussion

The main goal for the study carried out was to evaluate the long-term reliability of selected motor and musculoskeletal fitness tests and the MCI test battery for female health-care workers with recurrent NSLBP. Another aim was to ascertain whether change in pain intensity between two measurement points had any effect on the variation in the test results.

Within-subject variation is the most important type of test-retest repeatability measurements: the less within-subject variation there is, the greater the precision of the individual measurements and the better the observation of changes [28,38]. A typical example of a systematic change in results of physical fitness testing is the learning effect (bias). Participants perform better in the second test session than in the first, because they have benefitted from the experience of the first session [28,38].

Long-term reliability of the motor and musculoskeletal fitness tests

The running figure of eight test requires both agility and power. This test showed the lowest intra-individual variation of all tests ($CV: 2.8\%$)

Fitness test	n	mean of two tests	Typical error (s)	Coefficient of variation (CV)	Change in the means:	(95% CI)	% †
Running figure of eight	46	8.09	0.22 s	2.8%	-0.43	(-0.55-0.14)	5
Trunk lateral flexion	47	18.06	1.36 cm	7.5%	0.06	(-0.54-0.65)	0.3
Rhythm coordination	47	13.33	1.08 points	7.7%	0.36	(-0.86-0.81)	2
Dynamic sit-up	47	16.93	1.9 rep.	11.2%	1.02	(0.24-1.80)*	6
One leg squat	47	9.52	1.13 rep.	11.9%	0.20	(-2.72-0.68)	2
Modified push-up	46	8.41	1.04 rep.	12.2%	1.5	(1.04-1.95)*	19

* learning effect, † percent change in the mean: change in the mean divided by the mean of the first test = mean (test₂-retest₁)/mean(test₁).

Table 3: Test I – test II repeatability of the motor and musculoskeletal fitness test.

Movement control impairment test	Test-retest (intra-rater) kappa, n = 47
Dorsal pelvic tilt	0.71 (0.50–0.92)
Sitting knee extension	0.56 (0.17–0.94)
Walters bow	0.53 (0.27–0.78)
Knee flexion in prone	0.45 (0.11–0.78)
Rocking forwards and backwards	0.31 (0.03–0.58)
One leg stance	0.16 (-0.11–0.44)

Table 4: The test I – test II repeatability (with 95% CI) of the movement control tests of the low back.

	Change in pain intensity			p-value
	n	r	Change in VAS; median (min and max)	
Change in motor and musculo-skeletal fitness tests				
Rhythm coordination	47	-0.26		0.15
Running figure of eight	46	0.17		0.27
Trunk lateral flexion	47	0.15		0.32
Dynamic sit ups	47	0.23		0.16
One leg squats	47	-0.06		0.7
Modified push-ups	46	-0.3		0.045
Change in MCI test results				
Walters bow				0.34
Deteriorated	4		3 (-11, 32)	
Remained unchanged	37		-7 (-76, 60)	
Improved	6		-11 (-38, 7)	
Pelvic tilt				0.41
Deteriorated	1		6	
Remained unchanged	41		-7 (-76, 60)	
Improved	5		-4 (-22, 22)	
One-leg stance				0.55
Deteriorated	7		-2 (-30, 36)	
Remained unchanged	28		-7 (-76, 60)	
Improved	12		-13 (-38, 56)	
Sitting knee extension				0.89
Deteriorated	3		1 (-37, 22)	
Remained unchanged	43		-7 (76, 60)	
Improved	1		-4	
Rocking forwards and backwards				0.83
Deteriorated	5		-4 (-47, 9)	
Remained unchanged	32		-7 (-76, 60)	
Improved	10		-7 (-66, 22)	
Knee flexion in prone				0.3
Deteriorated	2		-2 (-11, 7)	
Remained unchanged	40		-7 (-76, 36)	
Improved	5		1 (-19, 60)	

VAS = visual analogue scale 0-100 during past 7 days, MCI = movement control impairment, r = Pearson's correlation coefficient

Table 5: Associations between change in pain intensity (VAS in the second test – VAS in the first test) and change in motor and musculoskeletal fitness as well as MCI test results in test I – test II (statistically significant p-value in bold).

and a small change in the mean (5%). The results are in agreement with previous studies [21,40], and the test seems highly repeatable across populations. A high performance level in this test has been linked to high quality of life in elderly women [41], but its relevance in testing of people with LBP has not been studied. That says agility is a capacity that is needed in nursing duties such as transferring patients, who may not behave in the anticipated manner.

Trunk lateral bending assesses spinal mobility in the frontal plane.

The low intra-individual variation (CV: 7.5%), very small change in mean (0.3%), and narrow 95% CI (-0.54–0.65) indicate that the test is repeatable, and there is no systematic change in the mean. A reduced range of lateral bending has been shown to be a risk indicator for LBP among younger health-care workers [42] and adolescents [43].

Testing of rhythm co-ordination assesses ability to simultaneously co-ordinate the movements of the upper and lower limbs at a slower and faster rhythm [44]. The low intra-individual variation (CV: 7.7%) and change in mean (2%) indicate that the test is repeatable. Previous results in healthy volunteers, reported by Rinne et al. [33] seem only modest, and the intra-class correlation coefficient (ICC) was fairly low, 0.70. Changes in the central nervous system, responsible for interpreting sensory stimuli and motor responses, have been found in LBP patients [23,45,46], however, the relevance of rhythm co-ordination in motor ability testing of LBP patients is not known.

The dynamic sit-up test assesses the strength and endurance of the flexor muscles of the trunk. The intra-individual variation was barely acceptable (CV: 11.2%), and there was a small learning effect (change in mean 6%). In a cross-sectional sample [47], persons with LBP were shown to have less abdominal muscle endurance than asymptomatic controls did.

The one-leg squat test assesses the strength of the lower extremities. The intra-individual variation was in the inadequate-reliability margin (CV: 11.9%), but the change in mean was small (2%). No learning effect was detected. Strength of the lower extremities is an important capacity when one is lifting and transferring patients in nursing duties, along with keeping the back in neutral position, although scientific evidence of associations between leg strength and LBP is limited. Individuals with chronic LBP have been shown to have weaker gluteus medius muscles than control subjects without back pain [48].

The modified push-up test requires both upper-body muscular strength and trunk stabilisation. Of the musculoskeletal fitness tests, this one showed the least repeatability, with the highest intra-individual variation (CV: 12.2%). The large change in mean (19%) and 95% CI not including the value zero (1.04–1.95) indicate a learning effect. On the other hand, this was the only test for which we found statistically significant associations between reduced pain and better performance. The test is physically heavy to perform, and might create high compressive forces in the spine. Therefore, the variation in test I – test II results may be due to the learning effect, change in pain intensity, or both. Furthermore, the broad range of results (0–18 repetitions) indicates that performance capacity among nurses varies quite dramatically. A learning effect has been found also in physically active healthy adults [21] and in a less selected population [22]. Low fitness in the modified push-up test has been associated with poor perceived health, low back dysfunction, and pain among middle-aged subjects [49]. Also, poor endurance in the back musculature has been reported to be a risk factor for low back pain [14,50]. The modified push up test requires endurance of the back muscles for trunk stabilisation. Increased risk of low back pain has also been reported in young conscripts with a poor fitness level in trunk muscle endurance and aerobic performance [13].

All three musculoskeletal fitness tests assessing muscle strength and endurance had a CV > 11% and were borderline for having adequate long-term reliability. With the learning effect detected borne in mind, they can be used in intervention studies to evaluate changes in muscular strength.

Long-term reliability of the movement-control-impairment test battery

Repeatability was good for the dorsal pelvic tilt; moderate for sitting knee extension, waiter's bow, and knee flexion in prone position; fair for rocking forwards and backwards; and poor for the one-leg stance. The results indicate that subjects' performance may vary from day to day. There is a clear need for better standardisation in the two tests yielding the poorest repeatability results (rocking forwards and backwards and the one-leg stance).

Previously, good to excellent intra-observer reliability ($k=0.67-0.95$) for the same six MCI tests were reported by Luomajoki et al. [26]. The subjects in their study were either LBP patients ($n=27$) or persons without LBP ($n=13$). Measurements were videoed in a standardized manner, and for the analysis of intra-observer reliability, examiners rated the same videos two weeks apart. In our long-term reliability study, the participants performed the MCI tests in two different measurement points, and all subjects had had back pain within the preceding two months.

Monnier et al. [51] found poor to moderate intra-rater test-retest repeatability for six clinical movement-control tests employed with marines. Their test battery differed from that used in our study, but rating was still dichotomous by visual observation. The only similar test was the waiter's bow (or standing bow), with an intra-rater kappa coefficient of 0.48 for rater A and 0.39 for rater B (0.53 in our study).

Whether it is more appropriate in clinical MCI tests to use quantitative outcome variables or dichotomous ones may be a subject worthy of discussion. Quantifying test results might enable the rater to obtain more information, which may be more useful for diagnosis and evaluation in a clinical setting [52].

Variation in MCI test I – test II repeatability results can be caused by day-to-day variation in test performance or difficulties in rating of the performance. According to Enoch et al. [53], it is difficult to estimate visually how much the lumbar region is moving during MCI tests without using any technical equipment. If the test is rated dichotomously 'yes/no' or 'can/cannot', much information is hidden between the two endpoints. Also, there is no clear consensus on when the test is passed / not passed or on where the dichotomous cut-off points should be [53].

In clear 'yes or no' cases, it is quite easy to place a person in the correct category, but there are many more complicated cases. If, for example, the MCI test performance starts correctly but at the very end of the range of motion the person loses control of the lumbar spine (maybe because of muscle tightness or restricted hip mobility), it is more difficult for the rater to decide whether the performance is correct or not. Two thirds of the study subjects had three or more musculoskeletal pain sites, which may have had an effect on the variation of the results. For instance, pain in the wrists or knees can affect weight transfer when one is rocking forwards and backwards in quadruped position.

For bringing higher reliability to the MCI test battery, we suggest either better standardisation of those MCI tests with poor to fair reliability or adding of a third band between the dichotomous ratings currently used (movement-control impairment and no impairment). This new 'in-between' class could cover all of the less clear cases, which are more difficult to classify without hesitation.

Ability to control the position and movement of the lumbar spine is important for back health. Patients with chronic LBP have more movement control impairments [24], lower tactile acuity in

the back region [54], and significantly poorer ability, on average, to sense a change in lumbar position [55] relative to healthy subjects. Measurement of the neuro-motor control of the spine in patients with NSLBP is considered to be important, but there seems to be lack of reliable and feasible measurement methods [56]. Clinical screening tests for assessing MCI in people with NSLBP include a high risk of bias [52].

Associations between change in pain and variation in test results

Ostelo [37] has evaluated the minimal important change in VAS as 15mm. The change in pain was less than 15mm for 51% and more than 15mm for 49% among our study sample. Pain is often presupposed to influence performance negatively [57]. Our hypothesis was that variation in pain causes variation in test results – i.e., less pain means better performance, and vice versa. The results of our study, however, do not support this hypothesis. Change in intensity of pain in persons with NSLBP had almost no effects in terms of variation of the test I – test II results for motor and musculoskeletal fitness tests, and for MCI tests. Our results are in agreement with results presented by Leitner et al. [29], who found in their long-term reliability study on patients with chronic LBP, that changes in pain intensity are not associated with changes in postural stability measurements.

Long-term reliability is not often analysed in repeatability studies. It has been recommended that for basic repeatability assessment the retest time should be within one week from the first test [28]. Longer periods between measurements are considered to admit more intra-individual changes in fitness, tiredness, and other such factors affecting physical performance. So far, to our knowledge, only two studies [29,30] with interval of 2-3 weeks between tests have investigated long-term reliability of measurements of physical parameters, such as muscular strength and endurance, and sensorimotor functions like postural control in people with LBP. That time interval corresponds the period of time in inpatient rehabilitation program for patients with chronic LBP in central Europe. Fitness or neuro-motor capacity of people with LBP seldom changes without any intervention in a few weeks, but LBP and pain intensity often seems to fluctuate from day to day. The periodic nature of pain in many people with low back trouble might influence patients' adherence or motivation with the measurements [30,57]. This topic needs further investigation with larger study samples.

Limitations of the study

The main limitation of the study was, that we did not standardise the measurement time in relation to work shifts. Tiredness after work could have caused variation in test results in comparison with measurements, which were conducted before work-shift or on the day off. Given the irregular working hours encountered with shift work, the arrangement of test sessions suitable for the same tester was very challenging.

Strengths of the study

All measurements took place twice in the same locations, with same examiner and same equipment. Prior to the first testing session, the written test instructions and criteria were specified, reviewed, and made the subject of drills by two members of the research group and the rater. The statistical methods and outcome measures agree with expert recommendations [28,38]. We showed that selected motor and musculoskeletal fitness tests are reliable, safe and feasible to use in evaluation of motor and musculoskeletal fitness components in female health-care personnel with LBP and having a physically demanding

job. We also showed that some of the widely used tests in the MCI test battery need better standardisation.

Implications

For fitness or MCI tests to be applied in clinical practice, the tests must be reliable and also safe, economical, simple, and easy to administer. All of the musculoskeletal and motor fitness tests studied in the present study fulfil the above-mentioned demands; however, the MCI test battery seems to have some deficits with respect to long-term reliability (for the one-leg stance test and rocking forwards and backwards).

The results provide useful information in selection of measurement methods for intervention studies aimed at reducing LBP and for clinical work in patients with LBP. The results indicate a need for further development of the MCI test battery, which is widely used by physiotherapists in evaluation of patients with LBP.

Acknowledgements

The authors wish to thank the Social Insurance Institution of Finland and the Tampere University Hospital for financial support for the NURSE RCT.

References

1. Sterud T, Tynes T (2013) Work-related psychosocial and mechanical risk factors for low back pain: a 3-year follow-up study of the general working population in Norway. *Occup Environ Med* 70: 296-302.
2. Karahan A, Kav S, Abbasoglu A, Dogan N (2009) Low back pain: prevalence and associated risk factors among hospital staff. *J Adv Nurs* 65: 516-524.
3. Cheung K (2010) The incidence of low back problems among nursing students in Hong Kong. *J Clin Nurs* 19: 2355-2362.
4. http://www.kela.fi/documents/10180/1630875/Statistical_Yearbook_of_the_Social_Insurance_Institution_2011.pdf/0578574a-2b93-4ca5-ac0f-650f12d15141
5. Choi BK, Verbeek JH, Tam WW, Jiang JY (2010) Exercises for prevention of recurrences of low-back pain. *Cochrane Database Syst Rev*.
6. Macedo LG, Maher CG, Latimer J, McAuley JH, Hodges PW, et al. (2014) Nature and determinants of the course of chronic low back pain over a 12-month period: a cluster analysis. *Phys Ther* 94: 210-221.
7. Kolb E, Canjuga M, Bauer GF, Laubli T (2011) Course of back pain across 5 years: a retrospective cohort study in the general population of Switzerland. *Spine (Phila Pa 1976)* 36: 268-273.
8. van Oostrom SH, Monique Verschuren WM, de Vet HC, Picavet HS (2011) Ten year course of low back pain in an adult population-based cohort—the Doetinchem cohort study. *Eur J Pain* 15: 993-998.
9. Hestbaek L, Leboeuf-Yde C, Engberg M, Lauritzen T, Bruun NH, et al. (2003) The course of low back pain in a general population. Results from a 5-year prospective study. *J Manipulative Physiol Ther* 26: 213-219.
10. Dunn KM, Hestbaek L, Cassidy JD (2013) Low back pain across the life course. *Best practice & research Clinical rheumatology* 27: 591-600.
11. Hsu P, Suni J (2012) Predictive validity of health-related fitness tests on back pain and related disability: a 6-year follow-up study among high-functioning older adults. *Journal of physical activity & health* 9: 249-258.
12. Heneweer H, Picavet HS, Staes F, Kiers H, Vanhees L (2012) Physical fitness, rather than self-reported physical activities, is more strongly associated with low back pain: evidence from a working population. *Eur Spine J* 21: 1265-1272.
13. Taanila HP, Suni JH, Pihlajamäki HK, Mattila VM, Ohrankammen O, et al. (2012) Predictors of low back pain in physically active conscripts with special emphasis on muscular fitness. *The spine journal* 12: 737-748.
14. Strøyer J, Jensen LD (2012) The role of physical fitness as risk indicator of increased low back pain intensity among people working with physically and mentally disabled persons: a 30-month prospective study. *Spine (Phila Pa 1976)* 33: 546-554.
15. del Pozo-Cruz B, Gusi N, Adsuar JC, del Pozo-Cruz J, Parraca JA, et al. (2013) Musculoskeletal fitness and health-related quality of life characteristics among

- sedentary office workers affected by sub-acute, non-specific low back pain: a cross-sectional study. *Physiotherapy* 99: 194-200.
16. Hodselmans AP, Dijkstra PU, Geertzen JH, van der Schans CP (2010) Nonspecific chronic low back pain patients are deconditioned and have an increased body fat percentage. *Int J Rehabil Res* 33: 268-270.
 17. Ryan CG, Grant PM, Dall PM, Gray H, Newton M, et al. (2013) Individuals with chronic low back pain have a lower level, and an altered pattern, of physical activity compared with matched controls: an observational study. *The Australian journal of physiotherapy* 255: 53-58.
 18. Carragee E, Alamin T, Cheng I, Franklin T, Hurwitz E (2009) Does minor trauma cause serious low back illness?. *Spine (Phila Pa 1976)* 31: 2942-2949.
 19. Hodder JN, Holmes MW, Keir PJ (2010) Continuous assessment of work activities and posture in long-term care nurses. *Ergonomics* 53: 1097-107.
 20. Rasmussen CD, Jorgensen MB, Clausen T, Andersen LL, Stroyer J, et al. Does self-assessed physical capacity predict development of low back pain among health care workers? A 2-year follow-up study. *Spine (Phila Pa 1976)* 38: 272-276.
 21. Suni JH, Rinne MB, Ruiz JR (2014) Retest Repeatability of Motor and Musculoskeletal Fitness Tests for Public Health Monitoring of Adult Populations. *J Nov Physiother* 4:194.
 22. Suni JH, Oja P, Laukkanen RT, Miilunpalo SI, Pasanen ME, et al. (1996) Health-related fitness test battery for adults: aspects of reliability. *Arch Phys Med Rehabil* 77: 399-405.
 23. Pijnenburg M, Caeyenberghs K, Janssens L, Goossens N, Swinnen SP, et al. (2014) Microstructural integrity of the superior cerebellar peduncle is associated with an impaired proprioceptive weighting capacity in individuals with non-specific low back pain *PLoS one* 9: e100666.
 24. Luomajoki H, Kool J, de Bruin ED, Airaksinen O (2008) Movement control tests of the low back; evaluation of the difference between patients with low back pain and healthy controls *BMC Musculoskelet Disord* 9: 170.
 25. Hadizadeh M, Mousavi SJ, Sedaghatnejad E, Talebian S, Parnianpour M (2014) The effect of chronic low back pain on trunk accuracy in a multidirectional isometric tracking task. *Spine (Phila Pa 1976)* 39: E1608-15.
 26. Luomajoki H, Kool J, de Bruin ED, Airaksinen O (2007) Reliability of movement control tests in the lumbar spine. *BMC Musculoskelet Disord* 8:90.
 27. Balague F, Mannion AF, Pellise F, Cedraschi C (2012) Non-specific low back pain. *Lancet* 379: 482-491.
 28. Hopkins WG, Schabert EJ, Hawley JA (2001) Reliability of power in physical performance tests. *Sports Med* 31: 211-234.
 29. Leitner C, Mair P, Paul B, Wick F, Mittermaier C, et al. (2009) Reliability of posturographic measurements in the assessment of impaired sensorimotor function in chronic low back pain. *J Electromyogr Kinesiol* 19: 380-390.
 30. Gruther W, Wick F, Paul B, Leitner C, Posch M, et al. (2009) Diagnostic accuracy and reliability of muscle strength and endurance measurements in patients with chronic low back pain. *Journal of rehabilitation medicine: official journal of the UEMS European Board of Physical and Rehabilitation Medicine* 41: 613-619.
 31. Dionne CE, Dunn KM, Croft PR, Nachemson AL, Buchbinder R, et al. (2008) A consensus approach toward the standardization of back pain definitions for use in prevalence studies. *Spine (Phila Pa 1976)* 33: 95-103.
 32. Suni JH, Miilunpalo SI, Asikainen TM, Laukkanen RT, Oja P, et al. (1998) Safety and feasibility of a health-related fitness test battery for adults. *Phys Ther* 78: 134-148.
 33. Rinne MB, Pasanen ME, Miilunpalo SI, Oja P (2001) Test-retest reproducibility and inter-rater reliability of a motor skill test battery for adults. *Int J Sports Med* 22: 192-200.
 34. <http://gih.diva-portal.org/smash/get/diva2:399894/FULLTEXT01>
 35. Sahrman SA (2001) *Diagnosis and treatment of movement control impairment syndromens.* (1stedn), St Louis Mosby, USA.
 36. O'Sullivan P (2005) *Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism.* *Manual therapy* 10: 242-55.
 37. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, et al. (2008) Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 33: 90-94.
 38. Hopkins WG (2000) Measures of reliability in sports medicine and science. *Sports Med* 30: 1-15.
 39. Altman DG (1991) *Practical Statistics for Medical Research.* CRC Press, USA.
 40. Vartiainen MV, Rinne MB, Lehto TM, Pasanen ME, Sarajuuri JM, et al. (2006) The test-retest reliability of motor performance measures after traumatic brain injuries. *Adv Physiother* 8: 148-152.
 41. Karinkanta S, Heinonen A, Sievanen H, Uusi-Rasi K, Kannus P (2005) Factors predicting dynamic balance and quality of life in home-dwelling elderly women. *Gerontology* 51:116-121.
 42. Adams MA, Mannion AF, Dolan P (1999) Personal risk factors for first-time low back pain. *Spine (Phila Pa 1976)* 24: 2497-2505.
 43. Jones MA, Stratton G, Reilly T, Unnithan VB (2005) Biological risk indicators for recurrent non-specific low back pain in adolescents. *Br J Sports Med* 39: 137-140.
 44. Rinne MB, Pasanen ME, Vartiainen MV, Lehto TM, Sarajuuri JM, et al. (2006) Motor performance in physically well-recovered men with traumatic brain injury. *Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine* 38: 224-229.
 45. Henry DE, Chiodo AE, Yang W (2011) Central nervous system reorganization in a variety of chronic pain states: a review. *PM & R : the journal of injury, function, and rehabilitation.* 2 3: 1116-1125.
 46. Tsao H, Danneels LA, Hodges PW. ISSLS prize winner: Smudging the motor brain in young adults with recurrent low back pain. *Spine (Phila Pa 1976)* 36: 1721-1727.
 47. del Pozo-Cruz B, Mocholi MH, del Pozo-Cruz J, Parraca JA, Adsuar JC, et al. (2014) Reliability and validity of lumbar and abdominal trunk muscle endurance tests in office workers with nonspecific subacute low back pain. *Journal of back and musculoskeletal rehabilitation* 27: 399-408.
 48. Penney T, Ploughman M, Austin MW, Behm DG, Byrne JM (2014) Determining the activation of gluteus medius and the validity of the single leg stance test in chronic, nonspecific low back pain. *Arch Phys Med Rehabil* 95: 1969-1976.
 49. Suni JH, Oja P, Miilunpalo SI, Pasanen ME, Vuori IM, et al. (1998) Health-related fitness test battery for adults: associations with perceived health, mobility, and back function and symptoms. *Arch Phys Med Rehabil* 79: 559-569.
 50. Cho KH, Beom JW, Lee TS, Lim JH, Lee TH, et al. (2014) Trunk muscles strength as a risk factor for nonspecific low back pain: a pilot study. *Annals of rehabilitation medicine.* 38: 234-240.
 51. Monnier A, Heuer J, Norman K, Ang BO (2012) Inter- and intra-observer reliability of clinical movement-control tests for marines. *BMC Musculoskelet Disord* 13: 263.
 52. Carlsson H, Rasmussen-Barr E (2013) Clinical screening tests for assessing movement control in non-specific low-back pain. A systematic review of intra- and inter-observer reliability studies. *Manual therapy* 18: 103-110.
 53. Enoch F, Kjaer P, Elkjaer A, Remvig L, Juul-Kristensen B (2011) Inter-examiner reproducibility of tests for lumbar motor control. *BMC Musculoskelet Disord* 12:114.
 54. Luomajoki H, Moseley GL (2011) Tactile acuity and lumbopelvic motor control in patients with back pain and healthy controls. *Br J Sports Med* 45: 437-440.
 55. Taimela S, Kankaanpää M, Luoto S (1999) The effect of lumbar fatigue on the ability to sense a change in lumbar position. A controlled study. *Spine (Phila Pa 1976)* 24: 1322-1327.
 56. Villafane JH, Zanetti L, Isgro M, Cleland JA, Bertozzi L, et al. (2015) Methods for the assessment of neuromotor capacity in non-specific low back pain: Validity and applicability in everyday clinical practice. *Journal of back and musculoskeletal rehabilitation* 28: 201-214.
 57. Huijnen I, Verbunt J, Wittink H, Smeets R (2013) Physical performance measurement in chronic low back pain: measuring physical capacity or pain-related behaviour? *European Journal of Physiotherapy* 15:103-110.