J Musculoskelet Neuronal Interact 2017; 17(3):197-208

Journal of Musculoskeletal and Neuronal Interactions JMNI

Original Article

Intertester and intratester reliability of a movement control test battery for patients with knee osteoarthritis and controls

P.T. Kaukinen[1,2], J.P. Arokoski[3], E.O. Huber[4], H.A. Luomajoki[4]

[1]Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland; [2]Department of Physical and Rehabilitation Medicine, Kuopio University Hospital, Kuopio, Finland; [3]Department of Physical and Rehabilitation Medicine, Helsinki University Hospital, Helsinki, Finland and University of Helsinki, Helsinki, Finland; [4]Institute of Physiotherapy, School of Health Professions, Zurich University of Applied Sciences, Winterthur, Switzerland

Abstract

Objectives: To develop a test battery of movement control (MC) tests and assess its intertester and intratester reliability. **Methods:** 29 subjects with knee OA with mean age of 64.7 (SD 8.7) years and 12 controls without either knee pain or previous diagnosis of OA (mean age 36.6 (SD 16.2) years) were included. Two experienced physiotherapists rated the filmed test performance of six MC tests blinded to the patients and to each other on 3-point scale as correct, incorrect or failed. Weighted kappa coefficient (wK) with 95% confidence interval (95%CI) and the percentage of agreement were calculated for each test. **Results:** One-leg stance, one-leg squat 30 degrees and step down tests showed moderate to excellent inter- and intratester reliability with wK ranging between 0.43-0.85 for intertester and 0.51-0.80 for intratester reliability. The reliability of the 90 degrees squat test, small squat and step up tests was poor (wK ranging between 0.09-0.50). **Conclusions:** One-leg stance test, one-leg squat 30 degrees and step down test are reliable in the subjects with knee OA and controls. Further studies are needed to evaluate the discriminative validity of the reliable tests.

Keywords: Movement Control, Reliability, Knee, Osteoarthritis

Introduction

Knee osteoarthritis (OA) is a disease of multifactorial origin with pain and disability[1-3]. It is characterized by unbalanced equilibrium between dynamic degenerative and regenerative structural changes seen in all joint tissues around the knee, including cartilage, bone, synovium, and periarticular soft tissues[4-6]. However, the structure-symptom discordance have been widely noted with approximately 50% of persons with structural changes in plain radiographs consistent with OA being asymptomatic[3,5,7,8] suggesting that other processes including peripheral and central neurophysiological mechanisms contribute to pain and disability[9].

Besides muscles strength, other aspects of muscle function are also affected by the OA disease process, including activation patterns and proprioceptive acuity[10]. A recent study showed consistent differences in knee proprioception between groups with and without knee OA across all knee movement directions (varus, valgus, flexion and extension)[11]. In dynamic knee stability during activities movement control seems to be a critical factor[12].

Movement control (MC) is defined as an ability to perform active movements while remaining balanced alignment of the body and lower extremities with an appropriate muscle response[13]. MC impairment means a reduction of active control of movement[14,15]. The underlying hypothesis is that impaired MC and lack of awareness of maladaptive movement patterns sustains pain[16,17]. In acute pain the central nervous system can change the motor behavior of the body to remove further threat of tissue damage and to support healing which has short-term benefits[18]. However, in the chronic phase the motor response may be less meaningful with potential consequences to the quality of movement and load on tissues, which, for its part, may lead to tissue irritation and continuous pain[18,19]. Furthermore, concerning knee OA,

The authors have no conflict of interest.

Corresponding author: Päivi Kaukinen, Kuopio University Hospital, PL 100, FIN-70029 KYS, Finland
E-mail: paivi.kaukinen@kuh.fi

Edited by: F. Rauch
Accepted 26 June 2017

Table 1. Inclusion criteria for knee OA subjects and exclusion criteria for all subjects.

Inclusion criteria for knee OA subjects:	
1	Pain within and/or around the knee (WOMAC pain subscale for 24 h prior to study entry >0)
2	Kellgren-Lawrence 1-4 radiographic knee OA[1]
3	Pain within the last year in and/or around the knee occurring on most days for at least a month
4	Age > 18 years
Exclusion criteria for knee OA subjects and controls:	
1	Other present pain problem or a central or peripheral nervous system condition causing sensory dysfunction (e.g., stroke, multiple sclerosis, spinal cord disorders or peripheral nerve lesion)
2	History of radicular pain from lower back (L3-S1) during the previous year
3	History of total arthroplasty of knee or hip joint
4	Severe depression or other severe psychiatric disorder
5	Indemnity problems
6	Cognitive impairment
7	Fibromyalgia or other widespread pain disorder
8	Any other unstable disease (e.g., cancer, acute trauma or infection)
9	Inflammatory arthritis or other systemic connective tissue disease
10	Inability to come to the hospital for evaluations
11	Inadequate knowledge of the Finnish language (inability to understand the standardized instructions)
12	Current use of analgesics or drugs that modify central pain modulation (e.g. tricyclic antidepressants, serotonin-norepinephrine reuptake inhibitors and gabapentinoids)
13	Inability to be without any pain medication for 24 h before the investigation

[1]Please, see Materials and methods. The radiographs were evaluated according to Kellgren-Lawrence (K-L) grading in which 0 means no OA and 4 refers severe OA[33].

observational data support the importance of loading in the progression of the disease[19].

Several studies have assessed the tests of MC of the lumbar spine[22,23], hip[24], neck[15] or head and eyes coordination[25] showing good reliability and validity. It has been suggested that movement control impairment (MCI) represents a subgroup of low back pain and that group-specific treatment interventions may be more efficient than non-specific interventions[26]. To our knowledge MC tests have not yet been investigated in subjects with painful knee OA.

There is no gold standard for MCI assessment in lower extremities. It has been suggested that the diagnosis of MCI should be based on the visual observation of active movements and functional activities in different starting positions[17,27-30]. Therefore, we developed a test battery of six motor tasks. Because clinical tests are vulnerable for mistakes, bias and different perceptions, it is important to evaluate their reliability. Thus, the aim of this study was to assess intertester and intratester reliability of MC tests in patients with knee OA and subjects without either knee pain or previous diagnosis of OA.

Materials and methods

Study design

Six active movement control tests were developed based descriptions by Sahrmann, Cook and Comerford and Mottram[16,17,30], and in accordance with earlier studies in low back pain[23], and intertester and intratester reliability study was conducted. The test performance by each participant was filmed in a standardized manner. Two experienced physiotherapists rated the test performance blinded to the patients and to each other on 3-point scale as correct, incorrect or failed. The study was conducted according to Helsinki Declaration and approved by The Ethics Committee of Kuopio University Hospital. All participants gave written informed consent prior enrollment.

Study sample

The sample size requirement for comparing coefficients of intertester agreement was calculated by selecting the level of significance as alpha=0.01 and power [beta=0.80] for testing hypothesis 0 (HO): kappa (k)1 ≥0.4 versus H1: k1 ≤0.4, the required sample size for group testing would be 36 cases for good [k index 0.40] strength of agreement[31] [31]. The sample size was set as n=40 to cater for a potential drop- out rate of 10%.

It was considered important to include in to the study sample subjects who would perform the tests well in order to increase variability and thus avoid a possible bias of the results of too many incorrect or failed test performance[31].

Twenty-nine (29) subjects with knee OA and 12 volunteers without either knee pain or previous diagnosis of OA (later in the text: controls) participated. The inclusion and exclusion

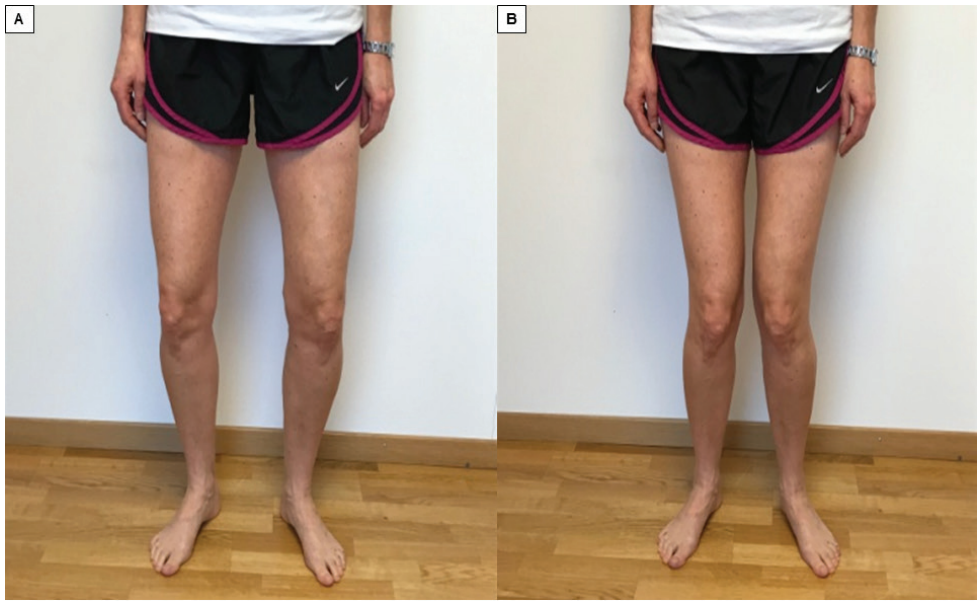

Figure 1. A-B. Small Squat to 30 Degrees Knee Flexion (frontal visual evaluation, right and left sides separately). Standardized instructions: "First, take some steps staying on the place you are standing, and then remain standing on both feet forward, feet aligned about hip-width apart. Then from this position, perform a small squat with flexing your knees approximately 30 degrees (so that you cannot see your foot tips). Repeat the movement four times, and on the fourth squat, please stay in the squat position for ten seconds." **A)** Correct performance (0 points): Pelvis stays horizontal, hips are aligned, knee stays aligned, knee gap is aligned to foot and foot stays straight. **B)** Incorrect performance (1 point): Pelvis is not horizontal, hips do not stay symmetrical, knee is giving away and knee gap is not pointing to the foot but e.g. medially) and foot is giving in (eg. there is a strong pronation).The performance is unsecure. Failed performance (2 points): the subject cannot perform the movement.

criteria are presented in Table 1. Knee OA diagnosis was based on clinical symptoms and knee radiographs[32]. The knee OA subjects were recruited via local primary health care providers and from the outpatient clinics of orthopaedics and physical and rehabilitation medicine of Kuopio University Hospital, and the controls from colleagues, friends and family members of the research team.

Evaluation of knee radiographs

The standard plain radiograhps from the symptomatic knee joint(s) were obtained from the knee OA subjects on clinical basis as a part of usual clinical practise. The radiographs were evaluated by two readers experienced in clinical radiographic evaluation (PTK and JPA) according to Kellgren-Lawrence (K-L) grading in which 0 means no OA and 4 refers severe OA[33].

Test protocol

At the beginning of the test visit, the knee OA subjects completed a validated Finnish version of the Western Ontario and McMaster Universities Arthritis Index (WOMAC) questionnaire[34,35]. Knee pain severity of both knees was recorded by using 100 mm visual analogue scale (VAS).

Two trained physicians (PTK and JPA), who were not involved in test performance rating, performed the MC testing protocol using standardized instructions including a demonstration of each test. The standardized instructions, and criteria for correct, incorrect and failed performance for the tests are presented in Figures 1-6. Each test performance was explained and demonstrated by the examiner (either PTK and JPA) to the subject and after that the subject was allowed to try the test once before the actual test performance. The tests were as follows: small squat to 30 degrees knee flexion (Figure 1), squat to 90 degrees knee flexion (Figure 2), one-leg stance (Figure 3), small squat on one-leg stance (Figure 4), step up (Figure 5) and step down (Figure 6). The test performance using both index knee (i.e. more symptomatic knee in the subjects with knee OA and randomly selected knee in the control subjects) and contralateral knee (i.e. less symptomatic or asymptomatic knee) were filmed except for squat to 90 degrees knee flexion -test in which the performance was videoed and evaluated from the side of the index knee and thus the performance of the contralateral knee could not be assessed. The order of the tests was randomized for each participant. Half of the participants were tested by one examiner and the other half by the other. The same examiner (either PTK or JPA) videotaped the test performances. Videos were taken

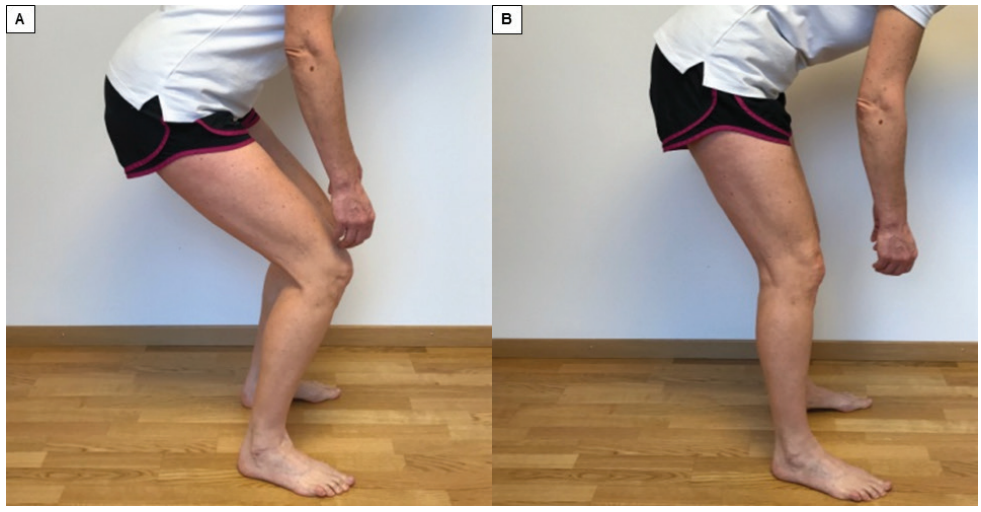

Figure 2. A-B. Squat to 90 Degrees Knee Flexion (visual evaluation from the side; only the site of the index knee was evaluated). Standardized instructions: "First, take some steps stationary staying broadside between the lines marked on the floor so that your more symptomatic (or randomly selected) knee is closest to the videocamera, and then remain standing on both feet forward, feet aligned about hip-width apart. Then, from this position, perform a squat with shifting your pelvis and hips backwards and downward to approximately 90 degrees of flexion of the hips and knees so that your fingertips touch the knee caps. The position of the spine should not change during the movement. Repeat the movement four times, and on the fourth squat, please stay in the squat position for ten seconds". **A)** Correct performance (0 points): Hip is flexing without flexion of the lumbar spine. There is 90 degrees flexion in both hip and knee joints. Movement is easy. **B)** Incorrect performance (1 point): Lumbar spine is flexing strongly. There is not enough flexion in the hip or knee joint but both joints flex at least 30 degrees. Failed performance (2 points): There is not at least 30 degrees flexion in the hip or knee joint, or the subject cannot perform the movement.

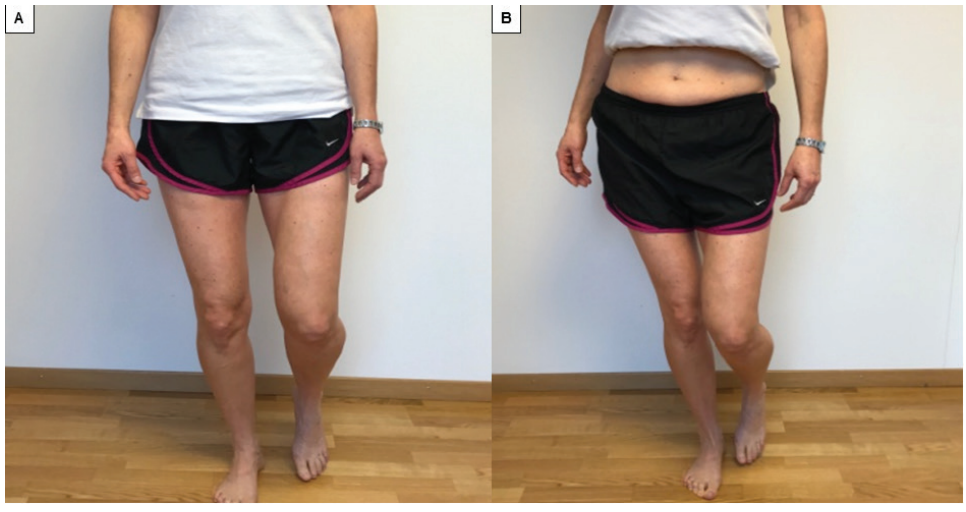

Figure 3. A-B. One-Leg Stance (frontal visual evaluation, right and left sides separately). Standardized instructions:"First, stand between the lines marked on the floor facing forward (towards the videocamera), feet aligned about hip-width apart. Then perform a one-leg stance with your right leg (the left foot has to be entirely lifted in the air) for ten seconds. The pelvis and the upper part of the body should not move and stay straight. Afterwards the test is performed with your left leg". **A)** Correct performance (0 points): Stance is stable and easy. Alignment of the hip, knee and foot are good (no varus or valgus). No compensating movements. **B)** Incorrect performance (1 point): Insecurity or helping with hands in the air. Pelvis does not stay horinzontal. Alignment of the hip, knee or ankle/ foot is not good (eg. there is varus or valgus in the knee, or the hip is tilting in adduction or the foor is giving in). The subject is tipping slightly to the ground with other foot. Failed performance (2 points): The subject cannot perform the movement or cannot stay on one leg for at least 5 seconds.

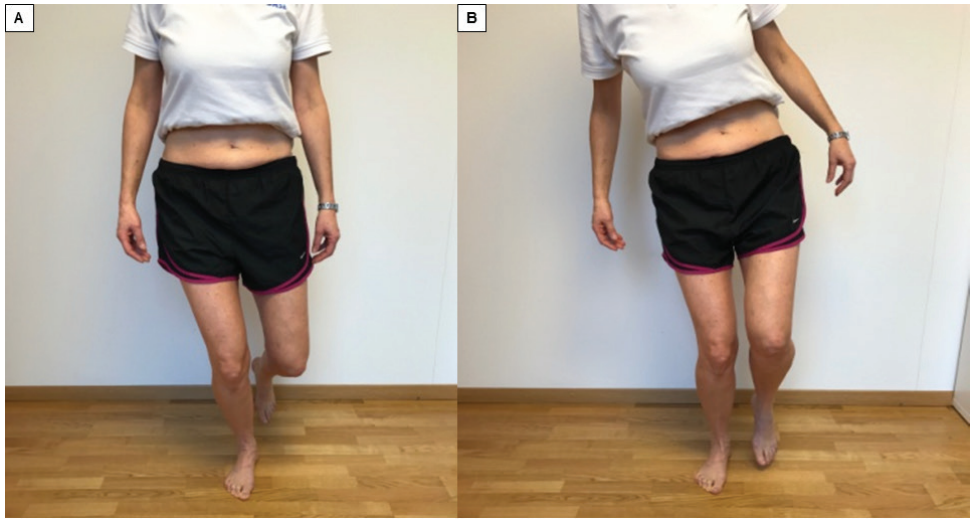

Figure 4. A-B. Small Squat on One-Leg Stance (frontal visual evaluation, right and left sides separately). Standardized instructions:"First, stand between the lines marked on the floor facing forward (towards the videocamera), feet aligned about hip-width apart. Then, from this position, take a one-leg stance position with your right leg (lifting your left foot entirely in the air) and perform a small squat with flexing your knee approximately 30 degrees (so that you cannot see your foot tip). The pelvis and the upper part of the body should not move and stay straight. Repeat the movement four times, and on the fourth squat, please stay in the squat position for ten seconds. Afterwards the test is performed with your left leg". **A)** Correct performance (0 points): Movement is stable and easy. Alignment of the hip, knee and foot are good (no varus or valgus). There are not compensating movements. **B)** Incorrect performance (1 point): Insecurity or helping with hands in the air. Pelvis does not stay horinzontal. Alignment of the hip or knee or ankle / foot is not good (eg. there is varus or valgus in the knee or the hip is tilting in adduction or the foot is giving in). The subject is tipping slightly to the ground with other foot. Failed performance (2 points): The subject cannot perform the movement or cannot stay on one leg stance.

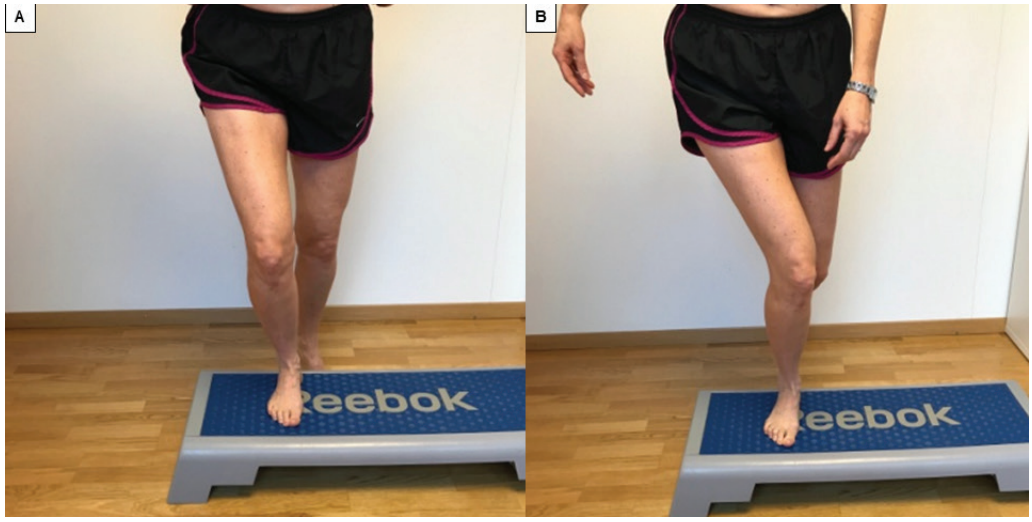

Figure 5. A-B. Step Up (frontal visual evaluation, right and left sides separately). Standardized instructions: "First, stand behind the aerobic step facing forward (towards the videocamera), feet aligned about hip-width apart. Then, step up and down (backwards) on the step starting with your right leg. Repeat the movement four times. Then, perform the movement starting with your left leg". **A)** Correct performance (0 points): Movement is stable and easy. Alignment of the hip, knee and foot are good (no varus or valgus). There are not compensating movements. **B)** Incorrect performance (1 point): Insecurity or helping with hands in the air. Pelvis does not stay horinzontal. Alignment of the hip or knee or ankle / foot is not good (eg. there is varus or valgus in the knee or hip is tilting in adduction or the foot is giving in). Failed performance (2 points): The subject cannot perform the movement.

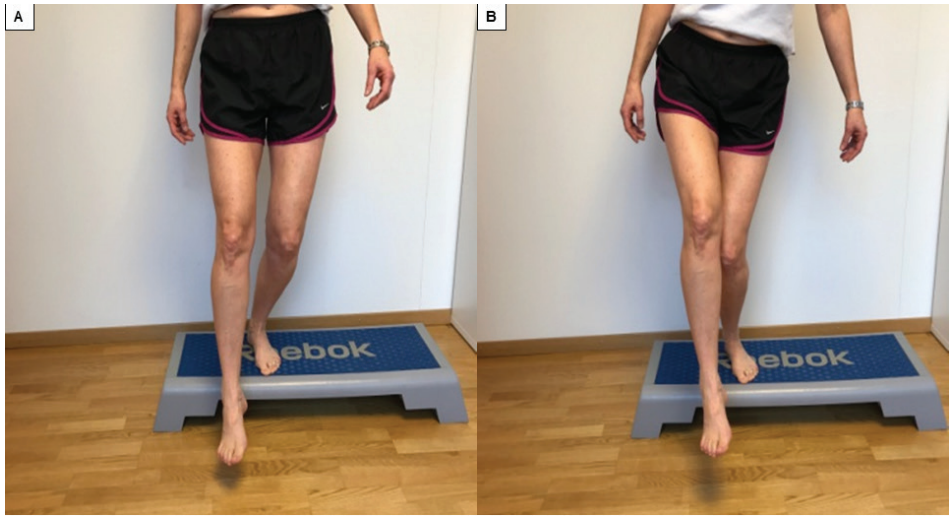

Figure 6. A-B. Step Down (frontal visual evaluation, right and left sides separately). Standardized instructions: "First, stand on the aerobic step facing forward (towards the videocamera), feet aligned about hip-width apart. Then, step down with your right leg so that the heel touches the floor first. Then, take a couple of steps forward, turn around and walk behind the aerobic step, step up and repeat the movement starting with your right leg four times. Afterwards, perform the movement starting with your left leg". **A)** Correct performance (0 points): Movement is stable and easy. Alingnment of the hip, knee and foot are good (no varus or valgus). There are not compensating movements. **B)** Incorrect performance (1 point): Insecurity or helping with hands in the air. Pelvis does not stay horinzontal. Alignment of the hip or knee or ankle / foot is not good (eg. there is varus or valgus in the knee or the hip is tilting in adduction or the foot is giving in). Failed performance (2 points): The subject cannot perform the movement.

anonymously, without showing the face, so that the person could not be identified. Each test was videoed separately and it took approximately 10 minutes per participant to perform the test set. Participants wore underwear so that posture and movements of lumbar spine, pelvis, hips and lower extremities could be observed.

All participants were without any pain medication for at least 24 h before the testing visits, and any unfamiliar sporting activity was prohibited for one week prior the measurements to avoid delayed onset muscle soreness. Bedside examination including motor, sensory and deep tendon reflex testing to exclude (subclinical) polyneuropathy was performed[36] at the beginning of the first visit and the participants with abnormal findings were excluded.

Rating of the test performance

Two experienced physiotherapists, EOH and HAL, blinded to the medical history and to each other rated the test performance in random order. Both raters are experienced physiotherapist with 30 years working experience. They reviewed the protocol thoroughly with each other and practised the ratings with videos. One physiotherapist (HAL) re-rated the test performances after 2 months period blinded to his earlier rates. The test performance was rated on 3-point scale as correct, incorrect or failed (i.e. subject could not perform the test). The criteria for correct and incorrect performance are presented in Figures 1-6.

Statistical analysis

Mean values and standard deviations (SDs) were calculated for all continuous variables. The number of incorrect or failed test performances was calculated for each test for index knee and contralateral knee with the exception of the squat to 90 degrees knee flexion in which only the performance of the index knee could be evaluated. For the controls the index knee was randomly selected. The sum (0-6) of incorrect test performances, named as MC score, was calculated for the index knee and the contralateral knee in both groups and an average of incorrect tests was calculated separately for knee OA subjects and controls. Paired samples t test was used to compare MC score between the index knee and contralateral knee in both groups. Because these groups were neither age- or sex -matched nor equal in size, no comparisons between the groups were performed.

In the reliability analyses the data from both groups (subjects with knee OA and controls) were treated as a whole cohort. For intertester and intratester reliability of each test, weighted kappa (wK) coefficient with 95% confidence interval (95%CI) and the percentage of agreement were calculated for each test. WK coefficients were interpreted according to Landis and Koch[37] (0.81-1 excellent agreement, 0.61-0.80 good, 0.41-0.60 moderate, 0.21-0.40 fair and <0.20 poor agreement). The definition of acceptable reliability was set on wK≥0.4 with lower bound of 95% CI≥0.2. Intraclass correlation coefficients (ICCs) with 95% confidence interval were

Table 2. Characteristics of study subjects.

	Knee OA subjects (n=29)	Controls (n=12)
Age (years), mean (SD)	65 (9)	37 (16)
Sex		
Female, n (%)	22 (76)	4 (33) (33.3)
Male, n (%)	7 (24)	8 (67) (66.7)
Body mass index (kg/m²), mean (SD)	29 (5)	27 (2) 27.2 (1.5)
Duration of knee pain (years), mean (SD)	7 (7)	
Knee pain, VAS (mm) , mean (SD)		
Index knee	34 (26)	
Contralateral knee	8 (13)	
WOMAC		
Pain subscale (mm), mean (SD)	35 (17)	
Function subscale (mm), mean (SD)	27 (16)	
Radiographic grades (Kellgren-Lawrence) of the index knee, n (%)		
1	4 (14)	
2	14 (48)	
3	10 (34)	
4	1 (3)	

Table 3. Number of incorrect or failed test performances for each test and MC score[1].

	Index[2] Knee		Contralateral[3] Knee	
Test	**Knee OA subjects n=29**	**Controls n=12**	**Knee OA subjects n=29**	**Controls n=12**
Small Squat to 30 Degrees Knee Flexion				
Incorrect, n (%)	7 (24)	3 (25 25.0)	8 (28 27.6)	3 (25)
Failed, n (%)	0	0	0	0
Squat to 90 Degrees Knee Flexion				
Incorrect, n (%)	9 (31)	1 (8)	na	na
Failed, n (%)	0	0	na	na
One-Leg Stance				
Incorrect, n (%)	10 (35)	3 (25)	11 (38)	3 (25)
Failed, n (%)	4 (14)	1 (8)	3 (10)	1 (8)
Small Squat on One-Leg Stance				
Incorrect, n (%)	11 (38)	3 (25)	13 (45)	5 (42)
Failed, n (%)	14 (48)	1 (8)	8 (28)	1 (8)
Step Up				
Incorrect, n (%)	11 (38)	1 (8)	8 (28)	1 (8)
Failed, n (%)	0	0	0	0
Step Down				
Incorrect, n (%)	18 (62)	0	16 (55)	3 (25)
Failed, n (%)	0	0	0	0
MC Score, mean (SD)	2.9 (2)	1.1 (2)	2.6 (2)	1.5 (2)

[1] MC Score = the sum (0-6) of incorrect test performances.
[2] Index Knee = symptomatic (or more symptomatic) knee in the subjects with knee OA and randomly selected knee in the controls.
[3] Contralateral Knee = asymptomatic (or less symptomatic) knee in the subjects with knee OA and contralateral knee in the controls
na = not assessable: in the squat to 90 degrees knee flexion test only the performance of the index knee could be evaluated.

Table 4. Intertester and intratester reliability of the movement control tests and the percentage (%) of agreement.

	Intertester Reliability		Intratester Reliability	
Test	Weighted Kappa (95% CI[1])	% Agreement	Weighted Kappa (95% CI[1])	% Agreement
Small Squat to 30 Degrees Knee Flexion				
Index Knee	0.38 (0.06-0.71)	78	0.28 (-0.05-0.62)	78
Contralateral Knee	0.30 (-0.02-0.62)	78	0.35 (0.05-0.66)	81
Squat to 90 Degrees Knee Flexion				
Index Knee	0.09 (-0.09-0.28)	76	0.50 (0.17-0.82)	85
Contralateral Knee	na[2]		na[2]	
One-Leg Stance				
Index Knee	0.60 (0.42-0.79)	85	0.65 (0.43-0.86)	78
Contralateral Knee	0.63 (0.44-0.82)	85	0.71 (0.50-0.91)	81
Small Squat on One-Leg Stance				
Index Knee	0.52 (0.34-0.69)	76	0.73 (0.56-0.90)	88
Contralateral Knee	0.43 (0.23-0.62)	73	0.66 (0.46-0.85)	88
Step Up				
Index Knee	0.41 (0.12-0.71)	81	0.36 (0.05-0.67)	78
Contralateral Knee	0.38 (0.03-0.73)	83	0.38 (0.03-0.73)	81
Step Down				
Index Knee	0.85 (0.69-1.00)	93	0.80 (0.62-0.99)	90
Contralateral Knee	0.55 (0.30-0.80)	78	0.51 (0.24-0.77)	76

[1] *CI= confidence interval*
[2] *na= not assessable*

used to assess inter- and intratester reliability of MC score using two-way mixed effects analysis of variance, type consistency with single measures. ICC values were interpreted according to guidelines established by Shrout and Fleiss[38] where values >0.75 indicate excellent reliability, 0.6-0.75 is good reliability, 0.4-0.59 is fair reliability, and <0.4 is poor reliability.

Statistical tests were performed two-tailed and at the 5% significance level. All statistical calculations were performed using the Statistical Package for the Social Sciences (version 22, SPSS Inc., Chicago, IL, USA).

Results

29 subjects with knee OA with mean age of 65 (SD 9) years (22 [76%] female) and 12 controls (mean age 37 (16) years, 33% female) were included. The demographic data of study subjects are shown in Table 2. The mean duration of symptoms in knee OA subjects was 7 (SD 7) years and 4 (14%) had Kellgren-Lawrence (K-L) 1, 14 (48%) K-L 2, 10 (34%) K-L 3 and 1 (3%) K-L 4 radiographic OA. Mean pain in the more symptomatic knee was 34 (26) mm and in the contralateral knee 8 (13) mm. 16 (55%) of the knee OA subjects had bilateral symptoms.

The number of incorrect or failed test performances for each test and MC score for both index knee and contralateral knee are presented in Table 3. On average, knee OA subjects performed 2.9 out of 6 tests (SD1.5) incorrectly with their more symptomatic (index) knee compared to 2.6 (1.6) of their less symptomatic/ asymptomatic (contralateral) knee *(p=0.103)*. In the control group, on average, 1.1 (SD 1.6) out of 6 tests in the index knee and 1.5 (1.7) in the contralateral knee were performed incorrectly *(p=0.1875)*.

Intertester reliability

The weighted Kappa (wK) values with 95% CI for intertester reliability and the percentage of agreement of each movement control test are shown in Table 4. One-leg standing test (wK for the index knee 0.60, 95% CI 0.42-0.79 and for the contralateral knee 0.63, 0.44-0.82) showed good and one-leg squatting test moderate (wK for the index knee 0.52, 95% CI 0.34-0.69 and for the contralateral knee 0.43, 0.23-0.62) reliability in both sides and step down test showed excellent reliability in the index i.e. more sym ptomatic knee (wK 0.85, 95% CI 0.69-1.00) but only moderate reliability in the contralateral knee (wK 0.55, 95% CI 0.30-0.80). Other tests showed fair or poor intertester reliability (wK range 0.09-0.41).

Table 5. Quality appraisal of diagnostic reliability (QAREL) checklist[48].

Question	Answer	Explanation
Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied?	Yes	The subjects with knee OA represented the subjects to whom the results are intended to be applied. Both primary and secondary care patients were included with different stages of the disease. The control group, however, was not age- and gender-matched.
Was the test performed by the raters who were representative of those to whom the authors intended the results to be applied?	Yes	
Were raters blinded to the findings of other raters during the study?	Yes	
Were raters blinded to their own prior findings of the test under evaluation?	Yes	
Were raters blinded to the results of the accepted reference standard or disease status for the target disorder (or variable) being evaluated?	Yes	Raters were blinded to the disease status of the subjects.
Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design?	Yes	
Were raters blinded to additional cues (e.g. history, imaging, blood tests) that were not part of the test?	Unclear	It might be possible to conclude from the videotapes whether the subject is having knee OA for example due to deformity of the knee.
Was the order of examination varied	Yes	Randomized
Was the stability (or theoretical stability of variable being measured taken into account when determining the suitability of the time-interval between repeated measures?	Yes	In this study the test-retest (intratester reliabilty) was evaluated from one videoed test performance in two separate sessions. Thus, potential variablity over time (for example due to present pain or as a consequence of learning were eliminated.
Was the test applied correctly and interpreted appropriately?	Yes	The tests are described in detail including rationale, procedure and interpretation. Interpretation of the tests were discussed and agreed by the raters.
Was appropriate statistical measures of agreement used?	Yes	

Intraclass correlation coefficient (ICC) for intertester reliability of the sum score of the tests was 0.71 (95% CI 0.51-0.83) for the index knee and 0.64 (0.42-0.79) for the contralateral knee showing good reliability.

Intratester reliability

The weighted Kappa (wK) values with 95% CI for intratester reliability and the percentage of agreement of each movement control test are shown in Table 4. One-leg standing test (wK for the index knee 0.65, 95% CI 0.43-0.86 and for the contralateral knee 0.71, 0.50-0.91) showed moderate and one-leg squatting test good (wK for the index knee 0.73, 95% CI 0.56-0.90 and for the contralateral knee 0.66, 0.46-0.85) reliability in both sides and step down-test showed excellent reliability in the index i.e. more symptomatic knee (wK 0.80, 95% CI 0.62-0.99) but only moderate reliability in the contralateral knee (wK 0.51, 95% CI 0.24-0.77). Other tests showed fair or poor intratester reliability (wK range 0.28-0.50).

ICC for intratester reliability of the sum score of the tests was 0.68 (0.47-0.82) for the index knee and 0.71 (0.51-0.83) for the contralateral knee showing good reliability.

Discussion

Our study evaluated the intertester and intratester reliability of movement control impairment (MCI) tests of lower extremities in subjects with knee OA and controls without either knee pain or previous diagnosis of OA. Three out of six tests (one-leg stance, one-leg squat 30 degrees, step down test) showed moderate to excellent inter- and intratester reliability, and both the intertester and intratester reliability for sum score of the six tests (MC score) was good.

To our knowledge this is the first study to assess the reliability of MC tests in knee OA. However, similarities can be found in earlier studies evaluating the reliability of movement control tests in lower extremities, hip or back in healthy subjects[39-43], in hip OA[24] and in low back pain[22,44].

Asprion et al.[24] assessed the reliability of movement control tests in the subjects with hip OA in almost identical setting than ours (movement control test battery consisting of five tests: small squat 30 degrees, squat to 90 degrees, one-leg stance, small single leg squat and step up tests) and found acceptable intertester (weighted kappa, wK 0.52-0.71) and intratester (wK for experienced rater 0.56-0.87 and for less experienced rater 0.35-0.61, respectively)

reliability. They suggested the use of squat to 90 degrees, one-leg stance, small single leg squat and step up tests in evaluating subjects with hip OA[24]. Better reliability in the 90 degrees squat test and step up test in their study may reflect the differences in our study populations (hip OA versus knee OA) and better sensibility of these tests to reveal movement control impairment in the subjects with hip OA. It can be hypothesised that by hip OA patients the control and differentiation between hip and low back is much more critical than by patients with knee problems.

In low back pain, intratester reliability of movement control tests has been varied from moderate to very good and intertester reliability from poor to very good[23]. One-leg stance test has shown constantly very good agreement[22-44]. Despite the similarity of the test performance in the one-leg stance test, the differences between the use of test are obvious: in our study the alignment and movement of lower extremity, i.e. the presence of excessive medial rotation of the involved femur, knee hyperextension or excessive ankle pronation were assessed, in addition to pelvic tilt[44]. The rotation[44] or lateral shift of the pelvis[22] was not assessed in our study like it has been done by patients with low back pain.

In knee OA the one-leg stance test has been used as a measure of postural balance with high repeatability and validity[45,46] with findings of postural deficits in the knee OA subjects compared to healthy age-matched controls[47,48]. To our knowledge, the control of movement in one-leg standing position has not been investigated in this population. Even if a minority of our study subjects failed one-leg stance test (i.e. was not able to stand in one-leg for 10 seconds), it is, important to notice that also poor balance may have affected the subjects' performance and might be an independent contributor to the subjects' pain problem[17].

The reliability of one-leg squat test has been evaluated in healthy subjects with good-to-excellent intertester reliability[39-41,43] and fair-to-excellent intratester reliability[40,41,43]. Good intertester and intratester reliability of one-leg squat test was also reported by Asprion et al.[24] in the hip OA subjects. The evaluation criteria for test performance have been variable with the criteria of Asprion et al.[24], Poulsen et al.[41] and Crossley et al.[40] being closest to ours.

It is worth to note that a substantial part (48% compared to 38% with incorrect performance, respectively) of knee OA subjects failed (i.e. were unable to perform) the one-leg squat test in our study. This may be due to aggravating pain during the activity forcing the subjects to discontinue their performance. In the future, assessment of the impact of movement impairment modifications on the pain and subject's performance would show whether the activity might be influenced with such a manner or if the pain severity is too high for this kind of activity for at least a subgroup of knee OA patients[17]. Movement control impairment is not characterised with intensive pain and extensive limitations[17]. Based on our results, in evaluating MC in subjects with knee OA, the high-level activities such as one-leg squat might be most useful in subjects with earlier stage of the disease with less disabling pain.

With the remaining three tests (small squat to 30 degrees, squat to 90 degrees knee flexion and step up tests) with poor reliability, the result just shows the complexity of the clinical evaluation of the MC[23]. Majority of the subjects with knee OA performed these tests correctly and one explanation of poor reliability of the remaining tests might be that they were too easy to perform and thus incapable to reveal movement control impaiment[17].

Our study has both strengths and limitations. When observed through the QAREL tool[49] (Table 5), the study protocol was well designed and thus, decreases the risk of bias. The subjects with knee OA represented the subjects to whom the results are intended to be applied. Both primary and secondary care patients were included with different stages of the disease. The control group was not age- and sex -matched, but we considered important to include in to the study sample subjects who would perform the tests well in order to increase variability of the results and thus avoid a possible bias of the results of too many incorrect or failed test performance[31]. However, as a consequence, even if there seemed to be a difference between the groups in the average number of incorrect tests (the subjects with knee OA performed on average 2.9 tests incorrectly with their index knee compared to 1.1 incorrect tests in healthy volunteers), we were not able to assess the discriminative validity of MC tests reliably because the groups were neither age- and gender-matched nor equal in size.

Also, it is worth to mention as a limitation that intratester reliability could be affected by a systematic error as the analysis was based on re-ratings by only one, not both, raters.

In our study the intratester reliability was evaluated from one videoed test performance in two separate sessions and thus the effect of pain variability and learning on variability of the results was eliminated. However, in the future studies it is important to investigate the test-retest reliability of the tests through separate testing sessions because in clinical practice such tests are commonly used for follow-up evaluations[42].

Conclusions

One-leg stance test, one-leg squat 30 degrees and step down -test are reliable in the subjects with knee OA and healthy volunteers. The three further tests showed only poor reliability. Use of the three reliable tests in the subjects with knee OA may increase our clinical knowledge of the mechanisms of pain and help us to create more efficient, patient-specific treatments. Further research is needed to evaluate the discriminative validity of these tests.

Acknowledgements

We thank Tuomas Selander, M.Sc, for his advice for the statistical analyses. Furthermore, we would like to thank all participants of the study and the person in the photographs. She gave her written consent allowing their use for publication.

References

1. Arden N, Nevitt MC. Osteoarthritis: epidemiology. Best practice & research. Clin Rheumatol 2006;20:3-25.
2. Hawker GA, Stewart L, French MR, Cibere J, Jordan JM, March L, et al. Understanding the pain experience in hip and knee osteoarthritis-an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16:415-422.
3. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013;21:1145-1153.
4. Loeser RF, Goldring SR, Scanzello CR, Goldring MB. Osteoarthritis: a disease of the joint as an organ. Arthritis Rheum 2012;64:1697-1707.
5. Hunter DJ, Guermazi A, Roemer F, Zhang Y, Neogi T. Structural correlates of pain in joints with osteoarthritis. Osteoarthritis Cartilage 2013;21:1170-1178.
6. Glyn-Jones S, Palmer AJ, Agricola R, Price AJ, Vincent TL, Weinans H, et al. Osteoarthritis. Lancet 2015; 386:376-387.
7. Kim IJ, Kim HA, Seo YI, Jung YO, Song YW, Jeong JY, et al. Prevalence of knee pain and its influence on quality of life and physical function in the Korean elderly population: a community based cross-sectional study. J Korean Med Sci 2011;26:1140-1146.
8. Davison MJ, Ioannidis G, Maly MR, Adachi JD, Beattie KA. Intermittent andO constant pain and physical function or performance in men and women with knee osteoarthritis: data from the osteoarthritis initiative. Clin Rheumatol 2016;35(2):371-9.
9. Fingleton C, Smart K, Moloney N, Fullen BM, Doody C. Pain sensitization in people with knee osteoarthritis: a systematic review and meta-analysis. Osteoarthritis Cartilage 2015;23(7):1043-56.
10. Bennell KL, Wrigley TV, Hunt MA, Lim BW, Hinman RS. Update on the role of muscle in the genesis and management of knee osteoarthritis. Rheum Dis Clin North Am 2013;39:145-176.
11. Cammarata ML, Schnitzer TJ, Dhaher YY. Does knee osteoarthritis differentially modulate proprioceptive acuity in the frontal and sagittal planes of the knee? Arthritis Rheum 2011;63:2681-2689.
12. Williams GN, Chmielewski T, Rudolph K, Buchanan TS, Snyder-Mackler L. Dynamic knee stability: current theory and implications for clinicians and scientists. J Orthop Sports Phys Ther 2001;31:546-566.
13. Elsig S, Luomajoki H, Sattelmayer M, Taeymans J, Tal-Akabi A, Hilfiker R. Sensorimotor tests, such as movement control and laterality judgment accuracy, in persons with recurrent neck pain and controls. A case-control study. Man Ther 2014;19:555-561.
14. Luomajoki H. Movement Control Impairment as a Subgroup of Non-specific Low Back Pain. Evaluation of Movement Control Test Battery as a Practical Tool in the Diagnosis of Movement Control Impairment and Treatment of this Dysfunction. Publications of the University of Eastern Finland. Dissertations in Health Sciences 24; 2010.
15. Patroncini M, Hannig S, Meichtry A, Luomajoki H. Reliability of movement control tests on the cervical spine. BMC Musculoskelet Disord 2014;15:402.
16. Comerford MJ, Mottram SL. Movement and stability dysfunction-contemporary developments. Man Ther 2001;6:15-26.
17. Sahrmann S editor. Movement System Impairment Syndromes of The Extremities, Cervical and Thoracic Spine. Considerations for Acute and Long-Term Management: Mosby Inc., an affiliate of Elsevier Incl.; 2011.
18. Hodges PW. Pain and motor control: From the laboratory to rehabilitation. J Electromyogr 2011;21:220-228.
19. Block JA, Shakoor N. Lower limb osteoarthritis: biomechanical alterations and implications for therapy. Curr Opin Rheumatol 2010;22:544-550.
20. Sterling M, Jull G, Wright A. The effect of musculoskeletal pain on motor activity and control. J Pain 2001;2:135-145.
21. Jull G, Trott P, Potter H, Zito G, Niere K, Shirley D, et al. A randomized controlled trial of exercise and manipulative therapy for cervicogenic headache. Spine 2002; 27:1835-43.
22. Luomajoki H, Kool J, de Bruin ED, Airaksinen O. Reliability of movement control tests in the lumbar spine. BMC Musculoskelet Disord 2007;8:90.
23. Carlsson H, Rasmussen-Barr E. Clinical screening tests for assessing movement control in non-specific low-back pain. A systematic review of intra- and inter-observer reliability studies. Man Ther 2013;18:103-110.
24. Lenzlinger-Asprion R, Keller N, Meichtry A, Luomajoki H. Intertester and intratester reliability of movement control tests on the hip for patients with hip osteoarthritis. BMC Musculoskelet Disord 2017;18:55-017-1388-5.
25. Della Casa E, Affolter Helbling J, Meichtry A, Luomajoki H, Kool J. Head-eye movement control tests in patients with chronic neck pain; inter-observer reliability and discriminative validity. BMC Musculoskelet Disord BMC 2014;15:16.
26. Vibe Fersum K, O'Sullivan P, Skouen JS, Smith A, Kvale A. Efficacy of classification-based cognitive functional therapy in patients with non-specific chronic low back pain: a randomized controlled trial. Eur J Pain 2013;17:916-928.
27. Loudon JK, Wiesner D, Goist-Foley HL, Asjes C, Loudon KL. Intrarater Reliability of Functional Performance Tests for Subjects With Patellofemoral Pain Syndrome. J Athl Train 2002;37:256-261.
28. Hegedus EJ, Cook C, Hasselblad V, Goode A, McCrory DC. Physical examination tests for assessing a torn meniscus in the knee: a systematic review with meta-analysis. J Orthop Sports Phys Ther 2007;37:541-550.
29. Gaasbeek RD, Groen BE, Hampsink B, van Heerwaarden RJ, Duysens J. Valgus bracing in patients with medial compartment osteoarthritis of the knee. A gait analysis study of a new brace. Gait Posture 2007;26:3-10.

30. Cook G. Movement: Functional Movement Systems: Screening, Assessment, and Corrective Strategies: On Target Publications; 2010.
31. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257-268.
32. Altman R, Asch E, Bloch D Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986;29:1039-49.
33. Kellgren JH, Jeffrey MR, Ball J. The epidemiology of chronic rheumatism. Atlas of Standard Radiographs of Arthritis. F. A. Davis; 1963.
34. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-1840.
35. Koli J, Multanen J, Häkkinen A, Kiviranta I, Kujala U, Heinonen A. et al. Reliability of the Finnish versions of WOMAC and KOOS forms for knee osteoarthritis. Physiotherapy 97: Supp S1, D203-204, 2011.
36. England JD, Gronseth GS, Franklin G, Miller RG, Asbury AK, Carter GT, et al. Distal symmetrical polyneuropathy: definition for clinical research. Muscle Nerve 2005; 31:113-123.
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-174.
38. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-428.
39. Ageberg E, Bennell KL, Hunt MA, Simic M, Roos EM, Creaby MW. Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. BMC Musculoskelet Disord 2010;11:265.
40. Crossley KM, Zhang WJ, Schache AG, Bryant A, Cowan SM. Performance on the single-leg squat task indicates hip abductor muscle function. Am J Sports Med 2011; 39:866-873.
41. Poulsen DR, James CR. Concurrent validity and reliability of clinical evaluation of the single leg squat. Physiother Theory Pract 2011;27:586-594.
42. Monnier A, Heuer J, Norman K, Ang BO. Inter- and intra-observer reliability of clinical movement-control tests for marines. BMC Musculoskelet Disord 2012;13:263.
43. Harris-Hayes M, Steger-May K, Koh C, Royer NK, Graci V, Salsich GB. Classification of lower extremity movement patterns based on visual assessment: reliability and correlation with 2-dimensional video analysis. J Athl Train 2014;49:304-310.
44. Roussel NA, Nijs J, Truijen S, Smeuninx L, Stassijns G. Low back pain: clinimetric properties of the Trendelenburg test, active straight leg raise test, and breathing pattern during active straight leg raising. J Manipulative Physiol Ther 2007;30:270-278.
45. Curb JD, Ceria-Ulep CD, Rodriguez BL, Grove J, Guralnik J, Willcox BJ, et al. Performance-based measures of physical function for high-function populations. J Am Geriatr Soc 2006;54:737-742.
46. Yelnik A, Bonan I. Clinical tools for assessing balance disorders. Clinical neurophysiology 2008;38:439-445.
47. Sanchez-Ramirez DC, van der Leeden M, Knol DL, van der Esch M, Roorda LD, Verschueren S, et al. Association of postural control with muscle strength, proprioception, self-reported knee instability and activity limitations in patients with knee osteoarthritis. J Rehabil Med 2013; 45:192-197.
48. Hatfield GL, Morrison A, Wenman M, Hammond CA, Hunt MA. Clinical Tests of Standing Balance in the Knee Osteoarthritis Population: Systematic Review and Meta-analysis. Phys Ther 2016;96:324-337.
49. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol 2010;63:854-861.