RELIABILITY AND VALIDITY OF HIP STABILITY ISOMETRIC TEST (HipSIT): A NEW METHOD TO ASSESS HIP POSTEROLATERAL MUSCLE STRENGTH

Gabriel Peixoto Leão Almeida, PT, PhD Student$^{1,2}$
Helena Larissa das Neves Rodrigues, PT Student$^2$
Bruno Wesley de Freitas, PT$^2$
Pedro Olavo de Paula Lima, PT, PhD$^{1,2}$.

$^1$Physical Therapy Department, School of Medicine, Federal University of Ceará, Fortaleza, CE/ Brazil.
$^2$Knee Research Group, University of Ceará, Fortaleza, CE/ Brazil.

This study was approved by the Ethics Committee at the Federal University of Ceará with protocol number 1.000.404

This research was supported by PIBIC/CNPq. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Corresponding author:
Gabriel Peixoto Leão Almeida.
Alexandre Baráuña Street, 949 – 1º andar – Rodolfo Teófilo;
Postal Code: 60430-160; Fortaleza, CE, Brazil;
Telephone: 55 85 33668632
E-mail: gabriel_alm@hotmail.com

Manuscript word count: 3230
Abstract word count: 248
ABSTRACT
STUDY DESIGN: Cross-sectional study.

BACKGROUND: The HipSIT evaluates the strength of the hip posterolateral stabilizers in a position that favors greater activation of the gluteus maximus and gluteus medius and lower activation of the tensor fascia lata.

OBJECTIVES: To check the validity and reliability of the Hip Stability Isometric Test (HipSIT) and to evaluate the HipSIT in women with patellofemoral pain (PFP).

METHODS: The HipSIT was evaluated with a hand-held dynamometer. During testing the participants were side-lying, with the legs positioned at 45° of hip flexion and 90° of knee flexion. Participants were instructed to raise the knee of the upper leg while keeping the heels in contact. To establish reliability and validity, 49 women were tested, with the HipSIT, by 2 different evaluators on day one, and then again 7 days later. The strength of the hip extensors, abductors, and external rotators were also evaluated. Twenty women with unilateral PFP were also evaluated.

RESULTS: The HipSIT has excellent intra- and inter-rater reliability. The standard error of measurement was .01kgf/kg, and the minimal detectable change was .027 kgf/kg. The HipSIT showed good validity in the isolated hip abduction, external rotation and extension (P < .01). Women with PFP showed a 10% deficit in the HipSIT results for the symptomatic limb (P = .01).

CONCLUSION: The HipSIT showed excellent inter- and intra-rater reliability, moderate-to-good validity in women, and was able to identify strength deficits in women with patellofemoral pain.

Keywords: hip, muscle strength dynamometer, reproducibility of results, validation studies.
INTRODUCTION

Studies\textsuperscript{3, 8, 9, 17, 29} have reported that excessive hip internal rotation and adduction range of motion (ROM) can be prevented by proper alignment of the lower limbs during daily activities and sports. The ability to maintain proper alignment depends on the strength and proper activation of the hip abductor, external rotator, and extensor muscles\textsuperscript{3, 8, 9}. Weakness of these muscles in the lower limbs are found in different conditions such as osteoarthritis,\textsuperscript{5, 12} knee arthroplasty,\textsuperscript{23} patellofemoral pain (PFP),\textsuperscript{25, 31} anterior cruciate ligament injury,\textsuperscript{13} iliotibial (IT) band syndrome,\textsuperscript{21} femoroacetabular impingement (FAI),\textsuperscript{20} ankle sprain,\textsuperscript{11} and low back pain,\textsuperscript{7} among others.

Hip abductor, external rotator, and extensor muscle strengths are evaluated separately.\textsuperscript{25, 31} However, the movements of the human body are three-dimensional (3D), and this method of evaluation does not reproduce the functional demands of the hip stabilizer muscles. Furthermore, proper implementation of isolated hip muscle strength testing requires repetitive training, testing, rest periods, and positioning adjustments for the tests.\textsuperscript{12, 23, 25} Clinically, these tests provide valuable information, but they are tiring for the patient, and their execution is time-consuming.

The Hip Stability Isometric Test (HipSIT) was developed to allow a three-dimensional evaluation of gluteal muscle strength, which makes it more functional when compared to the uniplanar assessment of the hip muscles.

Selkowicz et al.\textsuperscript{27} evaluated the electromyographic activity of the gluteus maximus (Gmax), gluteus medius (Gmed), and tensor fascia lata (TFL) in 11 exercises. The authors found that the clam exercise produced the greatest electromyographic activation of the gluteus complex in relation to the TFL. This exercise is performed with the subject in the lateral decubitus position, lower limbs in
45° of hip flexion and 90° of knee flexion. The individuals are instructed to separate their knees while maintaining contact between their heels, moving the superior hip into abduction, external rotation, and extension.\textsuperscript{27, 33} The HipSIT was developed based on the body’s position in the clam exercise.\textsuperscript{27, 33}

A functional, reliable, valid, and quick test to assess the strength of the hip abductors, external rotators, and extensors could help clinicians and researchers in clinical decision making. The HipSIT offers a unique assessment of the strength of the entire posterolateral hip musculature without the need to evaluate each muscle alone. Thus, the objectives of this study were: 1) to assess the intra- and inter-rater reliability of the HipSIT and to evaluate its validity by comparing with isolated strength tests of the hip abductor, extensor, and external rotators, and 2) to determine whether the HipSIT can detect hip strength asymmetries in young female athletes with PFP.

METHODS

Participants

This study was a cross-sectional analysis of 49 physically active women (Table 1). Recreational physical activity was defined as any physical activity in which a participant was engaged for at least 30 minutes per day or for at least 150 minutes per week.\textsuperscript{22} This group was selected to evaluate intra- and inter-rater reliability and the validity of the HipSIT. Data were reported according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).\textsuperscript{14} In addition, the HipSIT was applied to 20 young female athletes with unilateral PFP. The inclusion criteria for the group with PFP were the following: pain specifically located around the patellofemoral joint; reproduced or reported pain with at least two of the following...
going up or down stairs, squatting, kneeling, sitting for prolonged periods, isometric
contraction of the quadriceps, jumping, running, and pain on palpation of the lateral
and/or medial facet of the patella; report of insidious onset of pain lasting at least
three months; pain level of at least three on the visual analog scale (VAS) of pain
during the last week \(^2\); and reporting a maximum of 86 points on the Anterior Knee
Pain Scale (AKPS). The AKPS value was used to exclude patients with PFP without
clinically relevant reduction of functional capacity and the minimal detectable change
for the scale is 13 (maximum = 100 points) \(^3\). In both study groups, participants with
the presence of trauma or surgery of the lower limbs and lower back, pain during
testing, and neurological disorders that compromised the tests were excluded.

All participants were from 18 to 30 years of age and were available for testing
at the Human Movement Analysis Laboratory, Federal University of Ceará. Prior to
participation, the objectives, procedures, and risks of the study were explained to
each participant. This study was approved by the Ethics Committee at the Federal
University of Ceará with protocol number 1.000.404. All of the participants provided
written informed consent before participating in the study.

**Instruments**

An evaluation form containing questions related to anthropometric and clinical
features, sporting activities, and injury history was administered. The Lower
Extremity Functional Scale (LEFS) \(^1\) was applied to the population without PFP, and
the AKPS \(^1\) was applied to the population with PFP to assess the functional status
of the lower limbs of the participants. The strength of the hip muscles was evaluated
using a hand-held dynamometer (Nicholas Manual Muscle Tester; Lafayette
Instrument Company, Lafayette, IN, USA).
Lower Extremity Functional Scale was developed based on the model suggested by the World Health Organization (WHO). There are five possible responses (0–4) for each of the 20 questions, with possible scores from 0 (worst) to 80 (best). This scale was culturally adapted to Brazilian Portuguese with excellent reliability and validity for knee injury patients.\textsuperscript{19}

Anterior Knee Pain Scale assesses pain and symptoms in patients with PFP. This scale was translated and culturally adapted to the Brazilian Portuguese language,\textsuperscript{10} and the score ranges from 0 (worst) to 100 (best).

Muscle strength was evaluated using the hand-held dynamometer. This instrument has been widely used to measure muscle strength because of its ease of use. Studies have shown that this equipment has excellent intra- and inter-rater reliability and validity compared with the gold standard isokinetic dynamometer.\textsuperscript{18, 28} This method has good-to-excellent intra-rater and inter-rater reliability for measurement of hip strength.\textsuperscript{1, 26} The positioning of the hand-held dynamometer in this study was based on parameters established in the literature.\textsuperscript{1} Muscle strength (kgf) data were normalized by the body mass (kg) of each participant (strength [kgf]/body mass [kg]).

**Procedures**

The women participated in two sessions of data collection. In the first session, they completed the evaluation form and the specific questionnaires. In addition, HipSIT and isolated hip abductor, external rotator, and extensor muscle strengths were evaluated. In the first session, HipSIT was evaluated by two blinded evaluators (1A and 2). In the first session, the sequence of evaluation of the hip muscles was
randomized using Random Allocation Software (v. 1.0.0). Participants were again evaluated with the HipSIT by only one evaluator (1B) after a week.

To mitigate the influence of the rater, a strap was used for all tests. Participants were instructed to push the dynamometer as hard as they could for 5 sec. They performed one practice trial, rested for 30 seconds, and then performed the measured trials. Two tests were performed with a 30 seconds rest between each trial. Mean values were calculated for each participant. Participants rested for one min before changing the muscle group. When compensation was identified, values were discarded, and a new evaluation was done after 20 sec.¹

The two researchers have three years of experience in measuring muscle strength of the lower limbs with the hand-held dynamometer. Before starting the data collections, the researchers held meetings to standardize verbal stimuli and positioning. The evaluators were blinded to the results of the HipSIT between themselves (inter-rater reliability) and between a first and second evaluation (intra-rater reliability). For HipSIT evaluation of patients with (PFP), the evaluator was blinded in relation to the symptomatic and asymptomatic limb.

The HipSIT was performed with the participant in side-lying, with both legs positioned at 45° of hip flexion and 90° of knee flexion, with the limb to be tested superior to the opposing limb (Figure 1). The participant was instructed to lift the knee of the superior leg while keeping the heels in contact, such that the hip was in 20° of abduction. The center of the dynamometer was laterally positioned 5 cm above the knee joint interline. After positioning, the proper performance of the test was demonstrated and the participant was asked to perform the movement with the greatest possible force by separating the knees without the feet losing contact.
Isometric hip abductor force was measured with participants in side-lying, with the lower hip and knee positioned at 45° of flexion. The tested hip was abducted (20°) and extended (10°), and with neutral rotation (absence of internal or external rotation), the knee was extended. The dynamometer was positioned 5 cm proximal to the lateral malleolus midpoint.

Isometric hip extensor force was measured with subjects in a prone position. The not to be assessed limb was fully extended, while the assessed limb was in 10° of hip extension, 10° of hip external rotation, and 90° of knee flexion. The dynamometer was positioned over the posterior thigh, 5 cm proximal to the popliteal crease.

Isometric hip external rotator force was tested with participants in the sitting position, with the hip and knees flexed to 90°. The dynamometer was positioned over the distal-medial tibia, 5 cm proximal to the medial malleolus midpoint. Straps were used to prevent subjects from adducting the hip.

The isometric hip posterolateral complex force consisted of the sum of the force values of the three hip stabilizer muscles, divided by 3 ([abductors + extensors + external rotators] ÷ 3).\(^1\).

Statistical analysis

The normality of distribution of the data was determined by using the Shapiro-Wilk test. Descriptive statistics (mean and standard deviation) were used to describe the anthropometric and clinical characteristics and the outcome variables. The first step of the analysis was to calculate intra- and inter-rater reliability, which was done...
through the degree of consistency for intraclass correlation coefficients (ICC$_{2,1}$).

Analysis of variance (ANOVA) was used to compare the means found in the three evaluations of the HipSIT that each subject was submitted (1A, 1B and 2).

Reliability coefficients were interpreted as follows: < .69 indicated poor inter-rater reliability; .70–.79 indicated fair inter-rater reliability; .80–.89 indicated good inter-rater reliability; and .90 to 1.0 indicated excellent inter-rater reliability.$^6$ We used three measures of agreement: the Bland and Altman plots, the standard error of measurement (SEM), and the minimal detectable change (MDC) with 95% confidence intervals (CI). The SEM was calculated by dividing the standard deviation (SD) of the mean differences between the two measurements by the square root of 2 (SD differences ÷ $\sqrt{2}$), and the MDC was calculated using the formula of MDC = 1.96 x $\sqrt{2}$ x SEM. The SEM reflects the absolute error of the instrument, and the MDC reflects the smallest within-person change in a score that can be interpreted as a “real” change, above the measurement error of an individual.$^{30}$ Limits of agreement (LOA) were calculated as the SD of the individual differences between raters multiplied by 1.96. Both the SEM and MDC were also presented as percentages by dividing SEM and LOA by the average score of HipSIT 1A and 1B (subsequent week).

Validity was analyzed using Pearson’s correlation coefficient to check the strength of the relationship between the HipSIT results for the force of the abductors, external rotators, extensors, and posterolateral complex of the hip, with coefficient values set as follows: < .5 indicated poor validity, .5–.75 indicated moderate-to-good validity, and > .75 indicated excellent validity. The Bland-Altman plots were designed in order to assess the agreement between HipSIT and other measures of hip strength.
The PFP group was used to verify differences in HipSIT between the symptomatic and asymptomatic limbs. It was not used for comparisons between groups. The same statistical procedures were performed in patients with PFP by considering two assessments in sequence of the HipSIT. Finally, the paired \( t \)-test was used to compare HipSIT results between limbs with PFP and limbs without PFP in young female athletes. Significance was established at the 5% level for all statistics. Calculations were performed using SPSS version 17.0 for Windows (SPSS, Inc., Chicago, IL, USA).

RESULTS

Reliability and validity of the HipSIT

The anthropometric and clinical features of the 49 women evaluated to analyze the reliability and validity of the HipSIT are presented in Table 1.

| Insert Table 1 |
| Insert Figure 2 |

The values from the HipSIT of each subject are shown in Figure 2. HipSIT 1A represents the values obtained for first evaluator, HipSIT 2 shows the values obtained for the second evaluator and HipSIT 1B represents the third evaluation (retest) performed after one week by the first evaluator. Forty-nine subjects performed the HipSIT 1A and 2, and 46 performed HipSIT 1A and 1B. Analysis of variance (ANOVA) did not find significant differences between the three evaluations \( (P = .58) \).

The intra- and inter-rater reliability of the HipSIT showed excellent reliability indices. The ICC\(_{2,1}\) intra-rater was .981 (CI 95%, .966–.99) and the inter-rater was
The LOA ranged from -.047 to .025 (kgf/kg) for the intra-rater and inter-rater, as shown by the Bland and Altman plots (Figure 3). The SEM was .013 kgf/kg (4.7%), and the MDC was .036 kgf/kg (13.2%).

The validity of the HipSIT with the isolated tests for abduction, external rotation, and extension of the hip showed good validity, with Pearson’s correlation coefficients equal to .535 \( (P < .01) \), .536 \( (P < .01) \), and .514 \( (P < .01) \), respectively.

The validity of the HipSIT for the average of the three muscle groups assessed was .65 \( (P < .01) \) (Figure 4). The Bland-Altman plot between the HipSIT and the other variables of hip strength presented a mean difference from .04 to .12 (kgf/kg) (Figure 5).

HipSIT in women with patellofemoral pain

The anthropometric and clinical features of young female athletes with PFP are shown in Table 2. The HipSIT showed excellent reliability indices between the first and second evaluation (ICC = .991 [CI 95%, .978–.997]) in women with PFP. The LOA ranged from -.030 to .039 (kgf/kg). The SEM was .012 kgf/kg (4%), and the MDC was .034 kgf/kg (11.4%). Compared to the limb with no PFP, the limb with PFP showed a 10% deficit in the results obtained in the HipSIT \( (P = .01) \) (Table 2).
DISCUSSION

The results from this study supports that HipSIT is a valid test for evaluating the muscle strength of the posterolateral hip stabilizers, presented excellent intra- and inter-rater reliability, moderate-to-good validity in measurements of isolated muscle strength. The differences found between the HipSIT and the isolated hip strength tests are within clinically acceptable values as demonstrated by the Bland-Altman plot.

The measurement of strength with the hand-held dynamometer is widely used in scientific research and clinical practice because it is easy to handle, is portable, and has good reliability and good validity compared with the isokinetic dynamometer, which is considered the gold standard for evaluations of strength. Studies\textsuperscript{15,18} have shown that the hand-held dynamometer has good-to-excellent reliability for assessing the strength of the lower limb muscles, especially the hip muscles. The HipSIT also provides excellent levels of intra- and inter-rater reliability to assess the strength of the hip muscles. The SEM and MDC presented similar values for the populations with and without PFP. These parameters help in the evaluation, interpretation, and monitoring of the evolution when using the HipSIT as a measure of outcome.

The physical position adopted in the HipSIT was based on the clam exercise and is suitable to evaluate the function of the Gmax and Gmed as hip stabilizers, decreasing the influence of the TFL,\textsuperscript{4,27,33} which is an important internal rotator muscle of the hip.\textsuperscript{27} In this position, there is high electromyographic activity of the Gmax and Gmed muscles, with low electromyographic activity of the TFL.\textsuperscript{27} Willcox and Burden\textsuperscript{33} evaluated the electromyographic activity of the Gmax, Gmed, and TFL
using the clam exercise at three different degrees of hip flexion: 0°, 30°, and 60°. It was found that, in all three positions, the electromyographic activity of the Gmax and Gmed was higher than of the TFL. Moreover, Berry et al. compared the Gmax, Gmed, and TFL in subjects performing side stepping with a resistive band around the ankle while maintaining a standing position and in a squat. In the squatting position, the activation of the Gmax and Gmed was superior to the TFL. Therefore, the literature supports the use of the position adopted in our study.

HipSIT is more practical for assessing the strength of the stabilizing muscles because it considers muscular action as 3D. Besides being a reliable test, it proved to be a valid test when compared with isolated measures of the extension, abduction, and external rotation of the hip and with the average of these three evaluations. However, HipSIT evaluates three hip muscle groups in concomitance; therefore, a deficit in only one muscle group may not be detectable because of the overlap of strength of the other muscle groups. For example, if the patient has weakness of the hip abductors and has adequate or superior strength of the external rotators and extensors, the HipSIT presents symmetrical data between the limbs. The differences found between the HipSIT and the isolated hip tests may be due to the changes in the patient’s position during the isolated tests, because this influences the activation and the lever arm of the muscles.

Weakness of the hip muscles has been associated with PFP. We found a strength deficit of 10% in the HipSIT results for the limb with PFP compared to the limb without PFP. This is similar to findings in previous studies, where deficits ranged from 12%–27% in abductors, 7%–52% in extensors, and 5%–36% in external rotator muscles.
The limitations of this study serve as guidance to determine future studies. Although evaluators were blinded to the symptomatic and asymptomatic limbs in patients with PFP, just knowing who has PFP can be considered a bias. Another limitation is that it was only evaluated in women, so it is unknown if the same results would be observed in men. Prospective cohort studies need to elucidate the HipSIT relationship with the development of lower limb injuries, especially PFP and injuries to the anterior cruciate ligament. The relationship of the HipSIT with biomechanical changes of the lower limb has not yet been evaluated. Such information can guide preventive and therapeutic approaches to lower limb injuries.

**CONCLUSION**

The HipSIT was found to have excellent intra- and inter-rater reliability in assessments of the hip muscles strength in women; good validity for measures of strength in the extensors, external rotators, and abductors of the hip; and the ability to identify strength deficits in women with PFP. In conclusion, the HipSIT can be widely used in clinical practice and scientific research because it is reproducible, valid, more functional and faster than other tests for analyzing each isolated plane of hip movement.
KEY POINTS

Findings: The Hip Stability Isometric Test was found to be a method with excellent reproducibility and good validity for measuring the strength of the extensor, external rotator, abductors, and posterolateral hip complex.

Implications: The HipSIT is a more functional way of evaluating the strength of the hip-stabilizing muscles and is faster than assessing these muscles in three isolated planes of motion.

Caution: We cannot determine a cause-and-effect relation of the HipSIT and injuries or infer its relationship with biomechanical abnormalities of the lower limbs.
REFERENCES


Figure 1. Hip Stability Isometric Test.
Figure 2. Crude data from the HipSIT 1A (first evaluator), HipSIT 2 (second evaluator) and HipSIT 1B (subsequent week for first evaluator). The horizontal axis shows each individual and the vertical axis show the recorded strength (kgf/kg) for each session.

Figure 3. Bland and Altman plots for intra-examiner (A) and inter-examiner (B). HipSIT 1A (first evaluator), HipSIT 2 (second evaluator) and HipSIT 1B (subsequent week for first evaluator).
Figure 4. Correlation between hip stability isometric test with A) hip abductor isometric test with ($r = .535$, $P < .01$); B) hip extensor isometric test ($r = .514$, $P < .01$); C) hip external rotation isometric test ($r = .536$, $P < .01$); and D) hip posterolateral isometric test ($r = .65$, $P < .01$).
Figure 5. Bland-Altman plots representing comparisons between strength (kgf/kg) of the HipSIT and hip (A) abduction, (B) extension, (C) external rotation, and (D) posterolateral complex.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>21.38 ± 3.4</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>58.19 ± 9.18</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.63 ± 0.05</td>
</tr>
<tr>
<td>BMI (Kg/m²)</td>
<td>21.74 ± 2.75</td>
</tr>
<tr>
<td>LEFS (0 – 80)</td>
<td>74.39 ± 3.8</td>
</tr>
<tr>
<td>Dominance (%)</td>
<td>91.8% Right</td>
</tr>
<tr>
<td>Abduction (kgf/kg)</td>
<td>0.31 ± 0.07</td>
</tr>
<tr>
<td>Extension (kgf/kg)</td>
<td>0.19 ± 0.06</td>
</tr>
<tr>
<td>Lateral rotation (kgf/kg)</td>
<td>0.14 ± 0.03</td>
</tr>
<tr>
<td>Posterolateral (kgf/kg)</td>
<td>0.21 ± 0.04</td>
</tr>
<tr>
<td>HipSIT (kgf/kg)</td>
<td>0.27 ± 0.07</td>
</tr>
</tbody>
</table>

SD, standard deviation; BMI, body mass index; HipSIT, Hip Stability Isometric Test; LEFS, Lower Extremity Functional Scale.
Table 2. Characteristics of the Women with Patellofemoral Pain (n = 20).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>22.09 ± 3.02</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>70.9 ± 11.7</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.72 ± 0.13</td>
</tr>
<tr>
<td>BMI (Kg/m²)</td>
<td>23.5 ± 3.3</td>
</tr>
<tr>
<td>VAS (0 - 10)</td>
<td>4.63 ± 1.5</td>
</tr>
<tr>
<td>AKPS (0 – 100)</td>
<td>74.36 ± 8.4</td>
</tr>
<tr>
<td>HipSIT Limb with PFP (kgf/kg)</td>
<td>0.27 ± 0.08</td>
</tr>
<tr>
<td>HipSIT Limb without PFP (kgf/kg)</td>
<td>0.30 ± 0.09</td>
</tr>
</tbody>
</table>

SD, standard deviation; BMI, body mass index; VAS, visual analogue scale; AKPS, Anterior Knee Pain Scale; PFP, patellofemoral pain; HipSIT, Hip Stability Isometric Test.