Testing the Hip Abductor Muscle Strength of Older Persons Using a Handheld Dynamometer

Daniel H. Awwad, MBBS[1], Jonathan D. Buckley, PhD[2],
Rebecca L. Thomson, PhD[2], Matthew O'Connor, MBBS[1],
Tania A. Carbone, BSc[3], and Mellick J. Chehade, MBBS, PhD[1,3]

Abstract

Objective: To investigate the reliability of a clinically applicable method of dynamometry to assess and monitor hip abductor muscle strength in older persons. **Design:** Bilateral isometric hip abductor muscle strength measured with a handheld dynamometer, patients supine with the contralateral hip positioned directly against a wall for stabilization. Reliability determined by comparing intra-assessor and inter-assessor results and comparison to a criterion standard (stabilized dynamometer with patients in the standing position). **Setting:** UniSA Nutritional Physiology Research Centre. **Participants:** Twenty-one patients older than 65 years were recruited from the Royal Adelaide Hospital. **Main Outcome Measures:** Intraclass correlation coefficients (ICCs), bias, and limits of agreement calculated to determine reliability. **Results:** Intra-assessor and inter-assessor ICCs were high (0.94 and 0.92-0.94, respectively). There was no intra-assessor bias and narrow limits of agreement ($\pm 2.4\%$). There was a small inter-assessor bias but narrow limits of agreement (0.6%-0.9% and $\pm 2.3\%$, respectively). There was a wide variation comparing results to the criterion standard ($\pm 5.0\%$-5.2% limits of agreement), highlighting problems attributed to difficulties that the test population had with the standing position used in the criterion standard test. **Conclusions:** Testing older persons' hip abductor muscle strength while in the supine position with optimal pelvic stabilization using a handheld dynamometer is highly reliable. While further studies must be done to assess patients with specific pathologies, this test has potential application to monitor and evaluate the effects of surgical interventions and/or rehabilitation protocols for a variety of conditions affecting hip abductor function such as hip fractures and arthritis.

The hip abductor muscles including gluteus medius, minimus, and tensor fascia lata are vital to ambulation. Their well-established function is to contract during single-leg stance to maintain the level of the contralateral pelvis.[1,2] Abductor dysfunction carries significant morbidity for the affected patient causing limp, pain, and instability.

Abductor dysfunction can occur after hip fracture and subsequent fixation. Internal fixation using intramedullary nails may cause direct damage to hip abductors, while the extramedullary dynamic hip screw device for the treatment of hip fracture may cause shortening of the femoral neck and thus a reduced femoral neck offset.[3] In total hip arthroplasty, femoral offset and alteration of the abductor lever moment arm have been implemented in abductor dysfunction.[4] Assessing hip abductor function quantitatively can assist us in studying the impact of surgical interventions,[5] determining patients' functional abilities and postoperative rehabilitation.[6-9]

[1] Discipline of Orthopaedics and Trauma, University of Adelaide, Adelaide, South Australia, Australia
[2] Nutritional Physiology Research Centre, University of South Australia, Adelaide, South Australia, Australia
[3] Department of Orthopaedics and Trauma, Royal Adelaide Hospital, Adelaide, South Australia, Australia

Corresponding Author:
Daniel H. Awwad, Discipline of Orthopaedics and Trauma, University of Adelaide, Adelaide, South Australia 5000, Australia.
Email: danny.awwad@gmail.com

However, there is no simple, clinically applicable, reliable, quantitative test of hip abductor function available. Classically, hip abductors have been tested clinically by manual muscle testing, that is, the patient pushing against the assessors hand and the assessor rating the muscle power from 0 to 5 (Medical Research Council classification).[10] This method has many flaws, including a subjective rating system that leads to variable results between examiners[8,11] and poor validity.[11] It is insensitive and hence unable to distinguish small improvements in strength, especially in those rated 4 or 5.[12] Quantitative muscle testing using dynamometers is an alternative. These devices can either be handheld or stabilized. Using the more clinically practical handheld dynamometers has not provided reliable results.[13,14] This has predominantly been attributed to examiners being unable to achieve a "mechanical advantage" when testing the relatively strong lower limbs of young, physically active people.[13-16] However, with a stabilized device, so that the examiner is not required to oppose the force of the patient, reliable results have been obtained.[13] Therefore, if the examiner was able to achieve a "mechanical advantage" when using a handheld device, then reliable results may be achieved.[13,14]

Dynamometric testing of the hip abductors can be done in the side-lying, supine, or standing positions. The supine position is the simplest clinically, most practical for testing patients post hip fracture and also, importantly, eliminates gravity. In this position, the patient lies in the center of the bed with a band around the hips for stabilization.[2] However, studies thus far have shown it to be the most unreliable position, primarily attributed to failure to adequately stabilize the contralateral side.[2] The current criterion standard test for hip abductor strength in young healthy patients is the side-lying position with the use of a stabilized dynamometer.[2] In this position, the surface of the bed stabilizes the contralateral side. However, in an older population post hip fracture fixation or arthroplasty, this position may not be suitable due to pain caused by lying directly on the affected hip and the need to abduct against gravity. In this population, the currently existing recommendation is to test patients in the standing position with a stabilized dynamometer.[2]

The overall goal of the study was to develop and validate, in a "normal" population, a simple clinical method to assess and monitor hip abductor muscle strength in older populations. We hypothesized that the supine position with the handheld dynamometer using a wall to stabilize the contralateral hip would improve test reproducibility. The specific aims of this study were to determine the intra-assessor and inter-assessor reliability of measuring hip abductor muscle strength in a new testing position using a handheld dynamometer in a "normal" elderly population representative of those who may have hip fractures or arthritis. The study also aimed to compare the accuracy of this technique with the current recommended criterion standard test for this population (standing stabilized dynamometer).
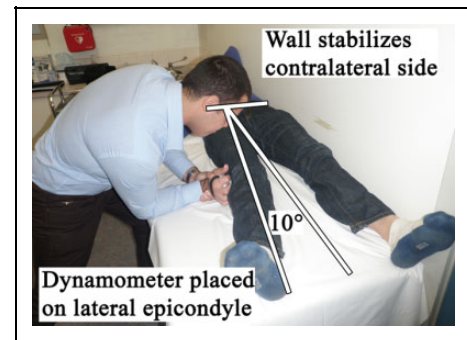

Figure 1. Setup for testing hip abductor muscle strength with a handheld device in the new position.

Methods

A prospective comparison to a criterion standard was undertaken. Testing was conducted at the Nutritional Physiology Research Centre at the University of South Australia. The instruments used were a MicroFET 2 handheld dynamometer (Hoggan Health Industries, West Jordan, Utah) and a Biodex System 4 Quick-Set isokinetic dynamometer (Biodex Medical Systems, Inc, Shirley, New York). Both assessors were male medical students aged 23 years. Ethics approval for this study was granted by the University of Adelaide Human Research Ethics Committee (H-135-2011) and the University of South Australia Human Research Ethics Committee. All participants provided written informed consent.

Patients were recruited from outpatient clinics at the Royal Adelaide Hospital and through posters at various locations around the hospital. Inclusion criteria were patients aged 65 years or over with a clinical diagnosis of osteoporosis (T score <2.5 and/or a proven previous fragility fracture) and able to provide independent consent. Patients were excluded if they had a preexisting lower limb functional limitation, including previous lower limb fragility fractures.

The method was devised to be simple to use clinically as well as comfortable for patients. Testing with the handheld device was done with the patient in the supine position (Figure 1). The opposite hip and leg to that being tested was positioned directly against a wall for stabilization of that side. With the patients leg at 10° abduction, the dynamometer was placed on the lateral epicondyle. The examiner was braced, so that they were able to oppose the force of abduction from the patient while keeping the dynamometer stationary. The patients were allowed 2 to 3 practice contractions at a submaximal level. The patients were then instructed to abduct with maximal force against the dynamometer for approximately 5 seconds before relaxing. The patient was then rotated, so that the contralateral hip abductors could be tested, and the procedure was repeated. The secondary assessor then tested each leg, followed by the primary assessor testing each leg for a second time. Patients were then tested with the Biodex isokinetic dynamometer according to the standard protocol (in the standing position with a board, supported by the examiner, to stabilize the contralateral side; Figure 2). Each patient performed
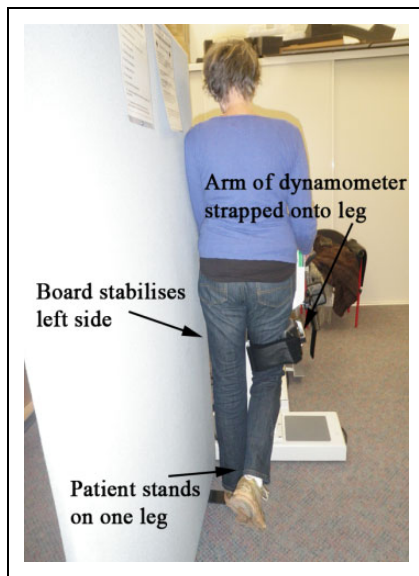
Figure 2. Setup for testing hip abductor muscle strength with the criterion standard test.

3 isometric muscle contractions at an angle of 10° abduction. Each contraction was held for 5 seconds and was followed by a 30-second rest.

The best of 3 maximal isometric contractions were recorded for each test. Raw data were collected in newton using the handheld dynamometer and the distance from the greater trochanter to the lateral epicondyle was measured to calculate the torque (Torque [N·m] = force [N] × moment arm [m]). The criterion standard machine measured torque directly. These were both standardized to normalized torque by the equation: Normalized torque (%) = torque (N·m)/weight (N) × 100.

Sample size was calculated using Lehr's formula. For a power of 80%, to detect a difference of 0.6% normalized torque (10 N with an average lever length of 0.4 m and average body weight of 65 kg) at 5% significance, with the estimate of 1.5% normalized torque for standard deviation (24 N, 0.4 m lever, and 65-kg person), a minimum of 23 independent measurements were needed.

Reliability was assessed by calculating the intraclass correlation coefficient (ICC). Various methods of calculating ICC exist and have differing uses in assessing correlation. Two of these are the ICC of agreement and ICC of consistency. The ICC of agreement was used in the analysis as it has been shown to better measure change in health status, whereas ICC of consistency measures distinction between assessors (however, both have been reported in Table 1).[17] The Bland-Altman test was also used to assess reliability through examining the bias and 95% limits of agreement between tests.

Results

Twenty-one patients were recruited and each leg was tested independently, thus providing 42 independent measurements. All patients recruited underwent all stages of testing and have been included in the final results. All were female, mean age 74 $\pm$ 6.5 years, height 158 $\pm$ 9 cm, weight 65.1 $\pm$ 12 kg, and lever arm 40 $\pm$ 3 cm.

Table 1. Table of Results.

| Comparison	Test	n	Torque Mean (SD), N·m	Normalized Torque Mean (SD), %	Bias (SD), %; P Value	95% Limits of Agreement, %	Intraclass Correlation Coefficient
Intra-assessor comparison	Primary assessor test 1	42	59.8 (20.5)	9.6 (3.5)	−0.338 (1.2); P = .08	±2.4	0.94
	Primary assessor test 2	42	61.8 (21.4)	9.9 (3.6)			
Inter-assessor comparison	Primary assessor test 1	42	59.8 (20.5)	9.6 (3.5)	−0.906 (1.16); P < .001	±2.3	0.92
	Secondary assessor test 1	42	65.5 (21.4)	10.5 (3.6)			
	Primary assessor test 2	42	61.8 (21.4)	9.9 (3.6)	−0.568 (1.14); P = .002	±2.3	0.94
	Secondary assessor test 1	42	65.5 (21.4)	10.5 (3.6)			
Comparison to gold standard	Primary assessor test 1	42	59.8 (20.5)	9.6 (3.5)	−0.911 (2.61); P = .03	±5.2	0.79
	Gold standard	42	65.1 (25.7)	10.5 (4.5)			
	Primary assessor test 2	42	61.8 (21.4)	9.9 (3.6)	−0.573 (2.5); P = .15	±5.0	0.84
	Gold standard	42	65.1 (25.7)	10.5 (4.5)			
	Secondary assessor test 1	42	65.5 (21.4)	10.5 (3.6)	−0.005 (2.62); P = .99	±5.2	0.83
	Gold standard	42	65.1 (25.7)	10.5 (4.5)			
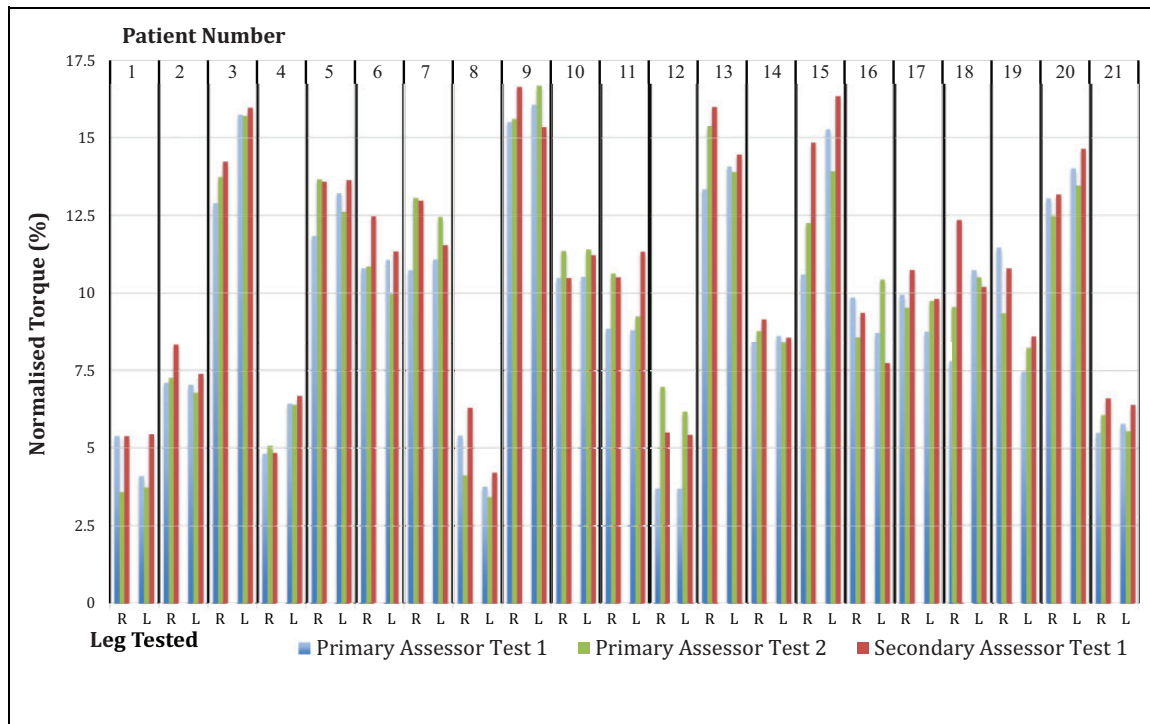
Figure 3. Results.
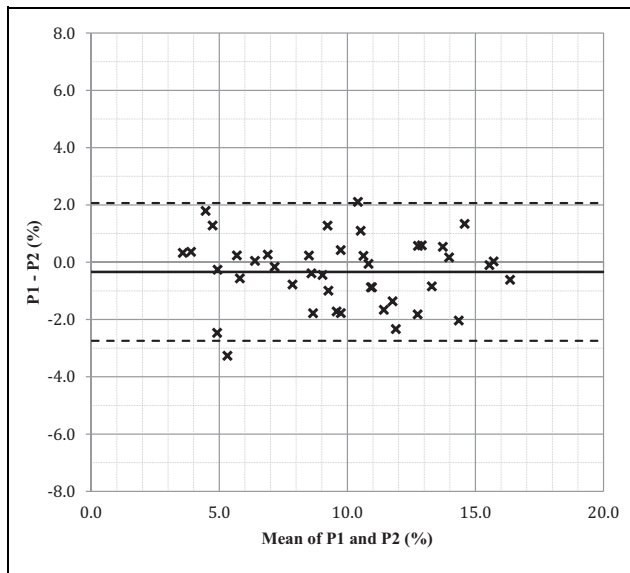
Figure 4. Intra-assessor comparison—primary assessor test 1 (P1) versus primary assessor test 2 (P2). Bias −0.3%, limits of agreement ±2.4%.

Figure 5. Inter-assessor comparison 1—primary assessor test 1 (P1) versus secondary assessor test 1 (S1). Bias −0.9%, limits of agreement ±2.3%.

Results are shown in Table 1 and Figure 3. There was no systematic bias between the 2 tests undertaken by the primary assessor (P1 vs P2: bias −0.3%, $P > .05$) and narrow limits of agreement at ±2% (Figure 4). The intra-assessor ICC was strong at 0.94. Comparing both the first and second test of the primary assessor (P1 vs S1 and P2 vs S1) to that of the secondary assessor, significantly lower values were shown on both occasions (bias −0.9% and −0.6%, respectively, $P < .05$).
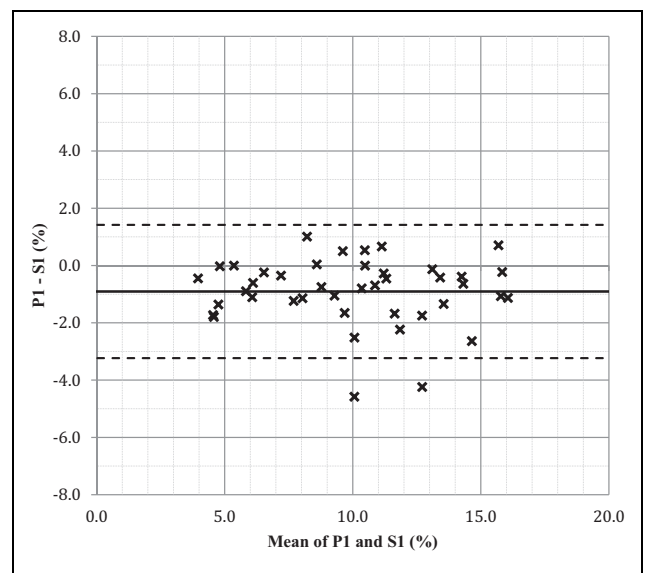
There were narrow limits of agreement of ±2.3% on both occasions (Figures 5 and 6). The inter-assessor ICC of agreement remained high at 0.92 and 0.94.

When assessing the accuracy of the handheld technique by comparing to the criterion standard test, we found on average lower mean values compared with the primary assessor (P1 vs CS: −0.9%, $P < .05$ and P2 vs CS: −0.6%, $P > .05$). However, there were much larger limits of agreement at 5.2% and 5%
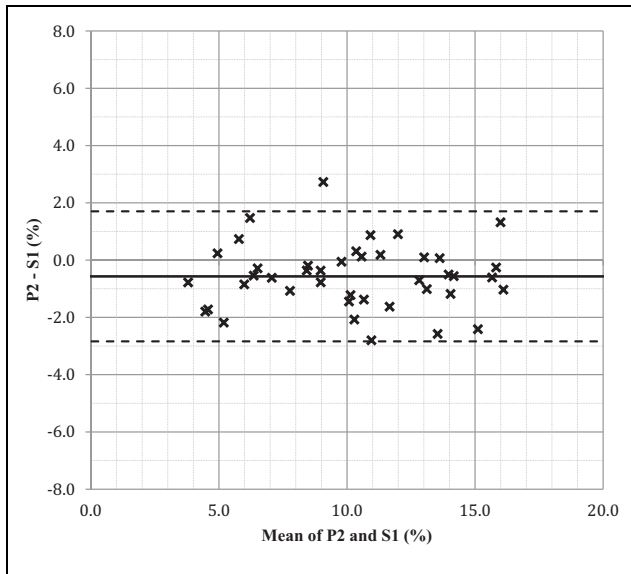
Figure 6. Inter-assessor comparison 2—primary assessor test 2 (P2) versus secondary assessor test 1 (S1). Bias −0.6%, limits of agreement ±2.3%.
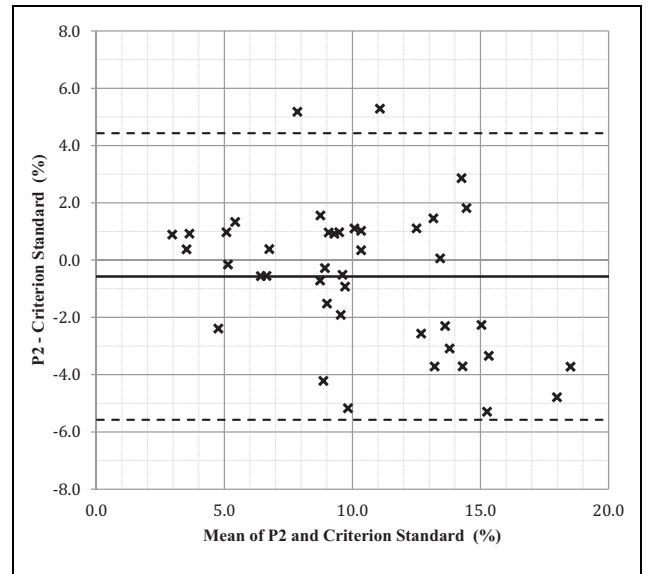

Figure 8. Comparison to criterion standard 2—primary assessor test 2 (P2) versus criterion standard. Bias −0.6%, limits of agreement ±5.0%.
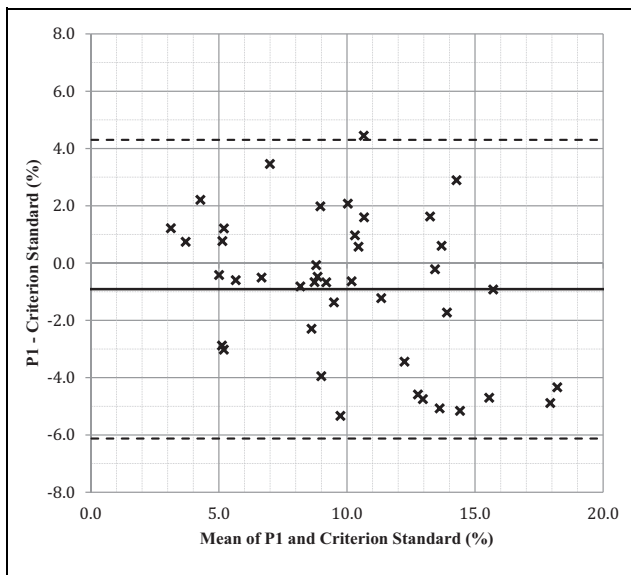

Figure 7. Comparison to criterion standard 1—primary assessor test 1 (P1) versus criterion standard. Bias −0.9%, limits of agreement ±5.2%.
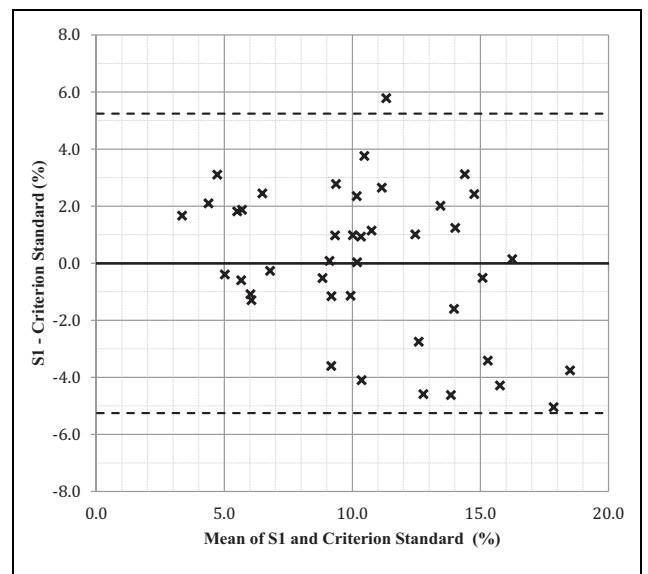

Figure 9. Comparison to criterion standard 3—secondary assessor test 1 (S1) versus criterion standard. Bias −0.0%, limits of agreement ±5.2%.

(Figures 7 and 8). The ICC was 0.79 and 0.84. There was no systematic bias between the secondary assessor and the criterion standard (S1 vs CS: −0.005%, $P > .05$), however, again the limits of agreement were wide at 5.2% (Figure 9). The ICC of agreement was 0.77.

Discussion

Our goal was to develop and validate a clinical test of hip abductor function that is objective, reproducible, and simple to use in a "normal" elderly population having abductor

dysfunction related to specific conditions such as hip osteoarthritis or post hip fracture.

The intra-assessor and inter-assessor ICCs were both high at 0.92 to 0.94 (interpreted according to the following criteria: >0.90 high, 0.89-0.80 good, 0.79-0.70 fair, and <0.69 poor reliability).[18] This result showed greater reliability compared with previous studies. A study by Widler et al published in 2009 found testing patients in the supine position with a stabilized device (as opposed to the more clinically practical handheld method used in this study) resulted in intra-assessor ICC

of 0.83.[2] While Widler et al agreed that the supine position had the advantage of avoiding the influence of gravity on strength assessment, they did not support the use of the position due to the poor reliability, lowest maximal contraction strength, and highest electromyographic ratio. These poor outcomes in the supine position were attributed to the poor body stabilization achieved using an abdominal belt. The side-lying and standing positions were favored due to the stabilization of the contralateral side by the examination table and wall, respectively, with ICCs of 0.90 and 0.88. The higher reliability in our study can be attributed to the optimized method of stabilization of the contralateral side in the supine position. In another study where a handheld device was used in the side-lying position, the intra-assessor and inter-assessor ICCs were 0.91 and 0.68, respectively.[13] The study involved healthy young participants (aged 22-31 years). As the patients examined in our study were both older and weaker than those included in the study by Krause et al,[13] the examiners were better able to oppose the force of abduction resulting in higher reliability.

Although there were statistically significant biases between the 2 assessors, the degree of systematic bias was quite small at 0.6% to 0.9%. The 95% limits of agreement were narrow for the inter-assessor comparisons ($\pm 2.4\%$ on both occasions).

There was only fair to good agreement (ICC range: 0.79-0.84) between the results of the handheld device and the criterion standard test using the isokinetic dynamometer. However, more importantly, a greater variability in results was seen when testing with the isokinetic dynamometer with 95% limits of agreement $\pm 5\%$ to $\pm 5.2\%$ (or approximately double that of the intra-assessor and inter-assessor measurements). The higher variability in results may have been due to the problems of positioning and balance encountered when testing this population in a standing position. Most of the older persons found the positioning awkward and had difficulty balancing on one leg while abducting with the other. Some patients who were able to stand on one leg found that the opposite hip to that being tested would fatigue first as the standing leg was necessary to stabilize the pelvis in that position. Some patients were unable to remain standing on one leg, so they placed the leg being tested on the ground, using it as a lever to push against the dynamometer rather than truly raising the leg off the ground and abducting. Due to this high variability between patients with respect to these issues of positioning and balance, there is limited validity of using the isokinetic dynamometer in the standing position to test the accuracy of the handheld device.

Although isokinetic dynamometry is recognized as the criterion standard for muscle strength assessment, we believe that it is suboptimal for testing hip abduction in the standing position in older persons with compromised power and balance. From our study, we have seen highly reliable results using a handheld dynamometer in the supine position using this new method of stabilization of the contralateral side. Future studies would be warranted to investigate further modifications of and comparisons with our supine "patient-stabilized" method by adding a mechanical positioner for the dynamometer (as in the study by Widler et al). Although less clinically applicable, it may prove to be a better research "criterion standard" for this patient population.

Study Limitations

A limitation of the study was that all patients were female. While females were not specifically selected for, the inclusion criteria of osteoporosis led to a higher likelihood of female selection. In any test of strength, results are effort-based and variables such as pain, cognition, and fatigue may impact overall results. Although not tested for, the greater mean strength recorded by the secondary assessor may have been due to a learning or practice effect with the task becoming more effective with repeated efforts. The order of assessors could be randomized in future studies to negate this potential learning effect. Fatigue could have had an impact as patients completed all of the tests on the same day within 1 hour.

Study patients had no real-time feedback during the test with the handheld device and were blinded to the results. As a consequence, it took longer to learn what was required and they were not able to challenge themselves to improve. Kim and Kramer studied knee extension in young, healthy participants and showed that higher torques were achieved when patients had visual feedback.[19] It is possible to attach the MicroFET 2 handheld dynamometer to a computer and view a real-time graph of the force, which may improve results.

Both assessors in the study had similar physical characteristics. In the unlikely situation where there may be a weaker assessor unable to match an older patient's hip abductor force, the test may not be as reliable. Conversely, to maximize reliability, it would be important to instruct an examiner to match rather than overcome a patient's abduction force.

Conclusions

We have found that testing hip abductor muscle strength using a handheld dynamometer in this novel testing position is highly reliable in this older population. It is a relatively simple and inexpensive portable test with immediate potential clinical and research applications. Further studies are required to validate the test in patients with hip pathology including those post hip fracture fixation or arthroplasty. Potential future applications include monitoring patient rehabilitation and comparing muscle dysfunction from various conditions and surgical interventions.

Authors' Note

Declaration of Conflicting Interests

References

1. Flack NA, Nicholson HD, Woodley SJ. A review of the anatomy of the hip abductor muscles, gluteus medius, gluteus minimus, and tensor fascia lata. *Clin Anat*. 2012;25(6):697-708.
2. Widler KS, Glatthorn JF, Bizzini M, et al. Assessment of hip abductor muscle strength. A validity and reliability study. *J Bone Joint Surg Am*. 2009;91(11):2666-2672.
3. Paul O, Barker JU, Lane JM, Helfet DL, Lorich DG. Functional and radiographic outcomes of intertrochanteric hip fractures treated with calcar reduction, compression, and trochanteric entry nailing. *J Orthop Trauma*. 2012;26(3):148-154.
4. McGrory BJ, Morrey BF, Cahalan TD, An KN, Cabanela ME. Effect of femoral offset on range of motion and abductor muscle strength after total hip arthroplasty. *J Bone Joint Surg Br*. 1995; 77(6):865-869.
5. Downing ND, Clark DI, Hutchinson JW, Colclough K, Howard PW. Hip abductor strength following total hip arthroplasty: a prospective comparison of the posterior and lateral approach in 100 patients. *Acta Orthop Scand*. 2001;72(3): 215-220.
6. Perez MM, Llusa M, Ortiz JC, et al. Superior gluteal nerve: safe area in hip surgery. *Surg Radiol Anat*. 2004;26(3):225-229.
7. Hardy AE, Synek V. Hip abductor function after the Hardinge approach: brief report. *J Bone Joint Surg Br*. 1988;70(4):673.
8. Frese E, Brown M, Norton BJ. Clinical reliability of manual muscle testing. Middle trapezius and gluteus medius muscles. *Phys Ther*. 1987;67(7):1072-1076.
9. Cahalan TD, Johnson ME, Liu S, Chao EY. Quantitative measurements of hip strength in different age groups. *Clin Orthop Relat Res*. 1989;(246):136-145.
10. Medical Research Council. *Aids to Examination of the Peripheral Nervous System: Memorandum No 45*. London, UK; Pendragon House; 1978.
11. Sapega AA. Muscle performance evaluation in orthopaedic practice. *J Bone Joint Surg Am*. 1990;72(10):1562-1574.
12. Wadsworth CT, Krishnan R, Sear M, Harrold J, Nielsen DH. Intrarater reliability of manual muscle testing and hand-held dynametric muscle testing. *Phys Ther*. 1987;67(9):1342-1347.
13. Krause DA, Schlagel SJ, Stember BM, Zoetewey JE, Hollman JH. Influence of lever arm and stabilization on measures of hip abduction and adduction torque obtained by hand-held dynamometry. *Arch Phys Med Rehabil*. 2007;88(1):37-42.
14. Agre JC, Magness JL, Hull SZ, et al. Strength testing with a portable dynamometer: reliability for upper and lower extremities. *Arch Phys Med Rehabil*. 1987;68(7):454-458.
15. Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session of strength assessment. *Phys Ther*. 1986; 66(2):206-209.
16. Click Fenter P, Bellew JW, Pitts TA, Kay RE. Reliability of stabilised commercial dynamometers for measuring hip abduction strength: a pilot study. *Br J Sports Med*. 2003;37(4):331-334.
17. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006; 59(10):1033-1039.
18. Currier DP. *Elements of Research in Physical Therapy*. 3rd ed. Lippincott Baltimore, MD: Williams and Wilkins; 1990.
19. Kim HJ, Kramer JF. Effectiveness of visual feedback during isokinetic exercise. *J Orthop Sports Phys Ther*. 1997;26(6): 318-323.