ORIGINAL ARTICLE

Reliability of stabilised commercial dynamometers for measuring hip abduction strength: a pilot study

P Click Fenter, J W Bellew, T A Pitts, R E Kay

See end of article for authors' affiliations

Correspondence to:
Assistant Professor Click
Fenter, School of Allied
Health, Department of
Physical Therapy, 1501
Kings Hwy, Shreveport, LA
71130-3932, USA;
pclick@lsuhsc.edu

Accepted 5 September
2002

Br J Sports Med 2003;37:331-334

Background: Reliable quantification of hip abductor strength in a clinical setting is challenging. **Objectives:** To examine the intrarater and interrater reliability of three commonly used commercial dynamometers in the measurement of hip abduction. **Methods:** Supine gravity minimised measures of unilateral hip abduction strength were recorded in 10 women (mean (SD) age 23.5 (1.9) years) using three different commercially available dynamometers. Measurements were repeated over a three day period with a different device used on each day. **Results:** Intrarater reliability ranged from 0.880 to 0.958 across the three devices, and measures of interrater reliability ranged from 0.899 to 0.948. **Conclusion:** Commercially available dynamometers can be used to quantify hip abduction strength with good to excellent reliability. A previously undescribed method of quantifying hip abduction strength in a clinical setting using readily available instrumentation is presented.

The hip abductors are key stabilisers of the pelvis during ambulation and are of primary concern when dealing with lower extremity pathology.¹⁻³ They function as lateral hip stabilisers during loading response and single limb support during gait.⁴ Periodic evaluation of muscle strength is an essential part of a progressive treatment programme during rehabilitation.^{2,5} With the increased emphasis on the quantification of muscular capability, it is important that a consistent, reliable method be developed that will allow objective measurement of hip abduction in the clinical setting.

Although manual muscle testing has classically been the most widely used method to assess muscle strength, it is fraught with error and poorly distinguishes differences in strength.^{1,3,5} Traditional manual muscle testing uses a five point grading scale and may vary from one evaluator to another⁶ depending on the testing styles and judgment of the examiner. It also does not allow small improvements to be detected, leading to questionable reliability between testers.³ Mulroy *et al*⁶ tested the ability of male and female clinicians to effectively test knee extension strength using manual muscle testing. Sex specific differences in the tester's ability to match quadriceps strength and detect weakness were found because the push force of female examiners was only 40-60% of normal knee extension force of the normals tested.⁶

Dynamometers commonly used for the assessment of grip strength have been studied and found to render reliable results.^{7,8} These devices are usually hand held and record the force produced by loading through tension or compression⁵; they are easily adjusted, portable, and will read up to 890 N of force.⁹ Although the use of hand held dynamometers (HHDs) is well established for testing grip strength, their use for the lower extremities is not well accepted as they have produced variable results. Agre *et al*¹⁰ reported hip abduction reliability coefficients (*r*) ranging from 0.49 to 0.92. The hip abduction tests of Andrews *et al*⁷ showed *r* = 0.71 for the non-dominant side and 0.72 for the dominant side. Bohannon and Andrews⁸ used HHDs to test the hip flexors, knee extensors, and ankle dorsiflexors of various diagnostic groups. They suggested that their use on patients with pathology may be easier than on healthy subjects. Agre *et al*¹⁰ reported that their prototype portable muscle dynamometer was more reliable for testing upper extremities than lower extremities. Interrater and intrarater reliability ranged from *r* = 0.85 to 0.99 for the upper

extremities and *r* = 0.49 to 0.81 for the lower extremities. Lower extremity scores showed a 30-40% variation in day to day measurement, with off centre loading identified as the problem. Kramer *et al*¹¹ found a belt resisted method to be superior to an examiner resisted method of HHD in young and old female subjects.

Stabilising and meeting the muscular force exerted by healthy, strong subjects with a hand held device when testing lower extremity muscle groups can be difficult for clinicians who are not physically strong.^{12,13} Nadler *et al*¹⁴ developed a device to stabilise a commercial dynamometer. However, subjects must be placed in an anti-gravity position on their side, requiring them to lift the weight of the lower extremity and overcome gravity before any pressure is placed on the dynamometer. Although they reported high reliability, this could result in inaccurate measurements particularly in the manual grade of "2" because this grade by definition requires the weight of the leg be eliminated.

The use of an HHD mounted for use in the supine, gravity minimised position has not been previously reported. This potential application may permit measurement of muscle forces at levels undetected by gravity dependent positions as may be seen in special populations such as seniors and subjects with muscle pathologies and neurogenic weakness. The intent of this study was to determine the interrater and intrarater reliability of three different commercially available HHDs used to measure hip abduction in a gravity minimised position.

METHODS

Subjects

Participants were a convenience sample of 10 healthy women aged 20-29 years old. Before participation, written informed consent was obtained from each. The institutional review board for the protection of human subjects at Louisiana State University Health Sciences Center, Shreveport approved the study. Excluded from the study were women who perform

Abbreviations: HHD, hand held dynamometer; ICC, intraclass correlation coefficient

Figure 1 Hip abduction measurement with a Dial Push-Pull Dynamometer.

Figure 2 Hip abduction measurement with a Jamar Hydraulic Hand dynamometer.

heavy manual labour or strenuous resistance training and those with acute illness, residual pain, disability, cardiovascular symptoms, or any neuromuscular disease or musculoskeletal pathology that would affect their ability to generate maximal strength or in which a strength test is contraindicated.

Measurements

The support used to stabilise the dynamometer was made of steel and mounted on a folding, portable treatment table for ease of transport (figs 1 and 2). Industrial grade Velcro was used to secure each of the HHDs to the stabilised upright. Muscle strength, defined as the maximal voluntary force that subjects were able to exert under specified testing conditions, was measured using three different dynamometers: (a) the Microfet 2 Load Cell Dynamometer (Hoggan Health Industries, Draper, Utah, USA), which records peak force in digital format from 3.6 to 89 N in 0.89 N increments, and 90 to 445 N in 4.45 N increments¹⁵; (b) the Jamar Hydraulic Hand Dynamometer (JHHD) (Sammons Preston, Bolingbrook, Illinois, USA), which measures maximum force up to 890 N in analogue format⁹; (c) the Dial Push-Pull Gauge (DPPG) (Chattillon, Greensboro, North Carolina, USA) dynamometer which measures up to 445 N in analogue format.¹⁶ Measurements were recorded to the nearest 22 N for the JHHD and DPPG. Before testing, each dynamometer was calibrated using known weight values up to 445 N and each showed an accuracy of greater than 99%. The Microfet 2 and the DPPG both

Table 1 Measurements of hip abductor strength (N) used to calculate interrater means and standard deviations for each dynamometer

	M		DP		HH	
	A	B	A	B	A	B
Mean	155.9	152.8	198.3	200.4	219.6	215.7
SD	36.0	37.1	32.6	32.4	40.5	34.3

A, Tester A; B, tester B; M, Microfet; DP, Dial Push-Pull Gauge; HH, Jamar Hydraulic Hand Dynamometer.

Table 2 Interrater correlation coefficients (ICCs)

Device	ICC	95% CI
MicroFet	0.948	0.876 to 0.979
Dial Push Pull	0.899	0.764 to 0.959
Hand Held	0.914	0.800 to 0.965

CI, Confidence interval.

had a 0.33 cm dense foam cover over the component that received force from the subject. Dense 0.33 cm foam was placed on the handle of the HHD to provide comfort to the subject during participation.

Subjects were tested in a gravity minimised, supine position with an HHD attached to a stationary device stabilised at the edge of a portable examination table. Subjects were positioned so that the dynamometer was 5 cm proximal to the lateral femoral condyle of the right limb. The same placement was used for each subject during all three tests. To stabilise the pelvis, a belt was placed across the participant's anterior superior iliac spines and around the table during the testing procedure. Care was taken not to allow subjects to rotate the pelvis or perform internal rotation, external rotation, or flexion at the hip. Use of the upper extremities to stabilise the trunk was permitted. Maximum effort was used to perform a "make" test,¹⁷ in which a subject exerted a maximal isometric force against the dynamometer for two to four seconds on each of the three stabilised dynamometers. The tester who explained the procedure initiated the process with verbal instructions. Then each subject received two practice sessions on each dynamometer before performing the actual test. Testing consisted of subjects performing two trials, tester A and tester B, for a total of four trials with a two minute rest between each trial. Measurements were taken over three days, with a single device being tested on each day. All subjects were tested using their right leg. The dynamometers were tested in random order.

Analysis

Statistical Package for Social Sciences 10.1 for Windows was used for data analysis. Mean (SD) age, weight, and height of the subjects were calculated. To determine intrarater and interrater reliability, intraclass correlation coefficients (ICCs) were calculated. In addition, 95% confidence intervals were calculated. ICCs (2,1) were calculated and interpreted according to established criteria for judgment;¹⁸ high reliability, 0.90–0.99; good reliability, 0.80–0.89; fair reliability, 0.70–0.79; poor reliability, less than 0.69. For intrarater reliability, each tester recorded two scores from each of 10 subjects permitting reliability calculations across 10 pairs of scores. Reliability comparisons were made between the 20 scores of tester A and the 20 scores of tester B.

RESULTS

The 10 subjects had a mean (SD) age of 23.5 (1.91) years, weight of 62.2 (11.5) kg, and height of 25.6 (1.2) m.

Table 3 Measurements of hip abductor strength (N) used to calculate intrarater means and standard deviations for each dynamometer

	M				DP				HH			
	A	A	B	B	A	A	B	B	A	A	B	B
Trial	1	2	1	2	1	2	1	2	1	2	1	2
Mean	151.2	160.6	151.2	154.4	198.4	198.2	199.7	201.1	220.2	219.1	210.2	221.3
SD	40.7	32.2	36.3	39.8	30.8	36.1	32.5	34.1	41.2	42.0	33.3	36.1

A, Tester A; B, tester B; M, Microfet; DP, Dial Push-Pull Gauge; HH, Jamar Hydraulic Hand Dynamometer.

Table 4 Intrarater correlation coefficients (ICCs)

Device/tester	ICC	95% CI
MicroFet		
A	0.880	0.577 to 0.969
B	0.942	0.796 to 0.985
Dial push pull gauge		
A	0.898	0.638 to 0.974
B	0.958	0.843 to 0.989
Hand held		
A	0.895	0.631 to 0.973
B	0.932	0.149 to 0.998

CI, Confidence interval; A, tester A; B, tester B.

Table 1 gives the mean and SD values of hip abduction strength used to calculate interrater reliability recorded by each tester with each device. The mean values for each device differed by no more than 3.92 N for comparisons between testers. Table 2 gives the ICCs for interrater reliability across the three devices. All showed high reliability ranging from 0.899 to 0.948. Table 3 gives the mean and SD values of hip abduction strength recorded by each tester with each device used to calculate intrarater reliability. The mean values ranged from 151.2 N to 221.3 N. The average difference of the paired means for each tester across all three devices was 42.7 N. Table 4 gives the ICCs for intrarater reliability of each tester across the three devices; they ranged from good reliability (0.880) to high reliability (0.958).

DISCUSSION

Both intrarater and interrater comparisons in this study showed high reliability for hip abduction strength measurement. These findings are in agreement with other studies measuring lower extremity strength with HHDs.^{10–12} Bohannon¹⁷ tested lower extremity muscle groups (including hip abduction) with an HHD for reliability and found high test-retest correlations (0.84–0.99); however, analysis of variance showed significant differences in repeated dynamometer test scores for hip abduction ($p = 0.006$). Nadler *et al*¹⁴ assessed the reliability of a commercial dynamometer attached to an anchoring station and reported correlation coefficients ranging from 0.94 to 0.98 for hip abduction and extension strength, establishing high reliability. However, the anchoring station described by Nadler *et al* appears to be cumbersome, and it would be difficult to position weaker patients, as the ability to lift the weight of the lower extremity would be required to test hip abduction.¹⁵ The weaker values of hip abduction strength would be eliminated from testing because of the anti-gravity position. This lack of sensitivity in quantifying lesser grades of muscle strength limits the clinical applicability of such a device.

Other studies have reported poor reliability when measuring lower extremity muscles because of the inability of examiners to stabilise the dynamometer against maximal volitional force.^{7–8} Lower extremity test-retest reliability has been noted to be present when stabilisation is required by the examiner.¹¹

It has further been suggested that subjects may be hesitant when applying maximum force to an examiner and thus not perform as well in this situation.¹¹ Bohannon,¹² in 1997, submitted reference values for extremity muscle strength obtained by HHD for adults (20–79 years); however, the single tester was a relatively young man (36 years) thus limiting those evaluators who are older, weaker, or less fit. Stabilisation of the dynamometer allows measurement of hip abductor strength without examiner intervention, which may allow those who are physically weak and/or inexperienced to obtain reliable results. Replicating the exact position for a test is difficult and may affect results because of alteration in the muscle length, strength, and biomechanical relation of the muscles being tested.¹⁰ With proper stabilisation of the dynamometer and accurate positioning of the subject, this limitation has been minimised in our study, as evidenced by the data. These results indicate that, when properly stabilised, this system may have greater clinical application than that of Nadler *et al*¹⁴ because of the ability to record hip abductor strength in the gravity minimised position.

The data suggest that these three commercially available dynamometers can be used with good to excellent reliability for the measurement of hip abduction. This finding is in general agreement with previously reported reliability data for hip abduction. This study differs from those of Andrews *et al*,⁷ Agre *et al*,¹⁰ and Bohannon¹⁷ in that reliability was calculated using ICCs whereas previous investigations reported Pearson r coefficients. Although this difference may initially seem trivial, the use of ICCs is favoured.¹⁹ Portney and Watkins¹⁹ assert that the Pearson r coefficient is not a true measure of reliability because it cannot differentiate the variance components that are the result of true error. Although it reflects the covariance in scores, it does not account for agreement in correlation. For example, suppose two subjects were measured for an imaginary variable on each of two days. On day 1, subject 1 scores 5, and subject 2 scores 7. On day 2, subject 1 scores 7 but subject 2 scores 9. The use of the Pearson r coefficient to assess reliability would render a value of 1.0, a perfect correlation, although it is clear from this example that the scores varied within subjects from day 1 to day 2. In contrast, an assessment of reliability using an ICC yields a correlation value of only 0.5 (possible range 0.0–1.0), illustrating the vast discrepancy between coefficients used to reflect reliability.

Reporting reliability using a Pearson r coefficient instead of ICCs often leads to erroneous interpretations of data, as evidenced by the example above. The Pearson r coefficient reflects the rank order of the subjects, and as subject 2 was two “units” greater than subject 1, a perfect correlation was rendered. As asserted by Portney and Watkins,¹⁹ Pearson r does not account for the agreement in correlation between subject scores over the two measurement days. Using ICCs would account for variability between subject scores and is thus a more sensitive marker of true reliability. In this manner, the data reported in this investigation provide new and more accurate information that may assist the clinician in quantifying hip abduction strength.

Take home message

Stabilisation of commercially available dynamometers offers a novel approach to objective measurement of hip abduction strength. It is a reasonable alternative to both manual muscle testing and manual dynamometry.

Conclusion

Dynamometric muscle testing that is mechanically stabilised is a reasonable alternative to both manual muscle testing and manual dynamometry. The HHD stabilisation device was specially designed to allow highly reliable measures of hip abduction strength in a clinical setting. The table and device allow accurate assessment of strength independent of the examiner, thus reducing measurement error resulting from examiner strength or skill. The stabilisation procedures presented here offer a new and easily replicated approach which may help to decrease error in the measurement of hip abduction strength. Accurate assessment of hip abductor strength is of clinical importance because of the role of these muscles in stabilising the trunk during ambulation. Further research is required to determine the optimum stabilisation and which instruments are the most accurate for measuring hip strength.

Authors' affiliations

P Click Fenter, J W Bellew, T A Pitts, R E Kay, Louisiana State University Health Science Center, Shreveport, LA, USA

REFERENCES

- 1 Beasley WC. Influence of method on estimates of normal knee extensor force among normal and postpolio children. *Physical Therapy Review* 1956;**36**:21–41.
- 2 Frese E, Brown M, Norton BJ. Clinical reliability of manual muscle testing: middle trapezius and gluteus medius muscles. *Phys Ther* 1987;**67**:1072–6.
- 3 Marino M, Nicholas J, Gleim G, et al. The efficacy of manual assessment of muscle strength using a new device. *Am J Sports Med* 1982;**10**:360–4.
- 4 Lyons K, Perry J, Gronley JK, et al. Timing and relative intensity of hip extensors and abductor muscle action during level and stair ambulation. *Phys Ther* 1983;**63**:1598–605.
- 5 Wadsworth CT, Krishnan R, Sear M, et al. Intratester reliability of manual muscle testing and hand-held dynamometric muscle testing. *Phys Ther* 1987;**67**:1342–7.
- 6 Mulroy SJ, Lassen KD, Chambers SH, et al. The ability of male and female clinicians to effectively test knee extension strength using manual muscle testing. *Journal of Sports Physical Therapy* 1997;**26**:192–9.
- 7 Andrews AW, Thomas MW, Bohannon RW. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. *Phys Ther* 1996;**76**:248–59.
- 8 Bohannon RW, Andrews AW. Interrater reliability of hand-held dynamometry. *Phys Ther* 1987;**67**:931–2.
- 9 Sammons Preston. *Jamar Hydraulic Hand Dynamometer*. Sammons Preston, 2001.
- 10 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**:454–8.
- 11 Kramer JF, Vaz, MD, Vandervoort, AA. Reliability of isometric hip abductor torques during examiner- and belt-resisted tests. *J Gerontol* 1991;**46**:M47–51.
- 12 Bohannon R. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. *Arch Phys Med Rehabil* 1997;**78**:26–32.
- 13 Bohannon RW, Corrigan, D. Reliability and validity of measurements of knee extension force obtained by hand-held dynamometry from community dwelling elderly women. *Issues on Aging* 2001;**24**:3–11.
- 14 Nadler SF, DePrince ML, Hauesien N, et al. Portable dynamometer anchoring station for measuring strength of the hip extensors and abductors. *Arch Phys Med Rehabil* 2000;**81**:1072–6.
- 15 Hogan Health Industries Inc. *Microfet 2*. Draper, UT: Hogan Health Industries Inc.
- 16 John Chatillon & Sons Inc. *Dial Push-Pull Gauge*. Greensboro, NC: John Chatillon & Sons Inc.
- 17 Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session of strength assessment. *Phys Ther* 1986;**66**:206–9.
- 18 Blesh TE, Fess EE, Moran CA. Cited by Currier DP in *Elements of research in physical therapy*. 3rd ed. Baltimore, MD: Williams & Wilkins, 1990.
- 19 Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. 2nd ed. Upper Saddle River, NJ: Prentiss-Hall Inc, 2000.