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

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 dynameters. 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.

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
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 increments; (b) the Jamar Hydraulic Hand Dynamometer (JHHD) (Sammons Preston, Bolingbrook, Illinois, USA), which measures maximum force up to 890 N in analogue format; (c) the Dial Push-Pull Gauge (DPPG) (Chatillon, Greensboro, North Carolina, USA) dynamometer which measures up to 445 N in analogue format. 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 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, 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: 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

<table>
<thead>
<tr>
<th>Device/tester</th>
<th>M</th>
<th>A</th>
<th>B</th>
<th>B</th>
<th>DP</th>
<th>A</th>
<th>A</th>
<th>B</th>
<th>B</th>
<th>HH</th>
<th>A</th>
<th>A</th>
<th>B</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Mean</td>
<td>151.2</td>
<td>160.6</td>
<td>151.2</td>
<td>154.4</td>
<td>198.4</td>
<td>198.2</td>
<td>199.7</td>
<td>201.1</td>
<td>220.2</td>
<td>219.1</td>
<td>210.2</td>
<td>221.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>40.7</td>
<td>32.2</td>
<td>36.3</td>
<td>39.8</td>
<td>30.8</td>
<td>36.1</td>
<td>32.5</td>
<td>34.1</td>
<td>41.2</td>
<td>42.0</td>
<td>33.3</td>
<td>36.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4 Intrarater correlation coefficients (ICCs)

<table>
<thead>
<tr>
<th>Device/tester</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microfet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.880</td>
<td>0.577 to 0.969</td>
</tr>
<tr>
<td>B</td>
<td>0.942</td>
<td>0.796 to 0.985</td>
</tr>
<tr>
<td>Dial push pull gauge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.898</td>
<td>0.638 to 0.974</td>
</tr>
<tr>
<td>B</td>
<td>0.958</td>
<td>0.843 to 0.989</td>
</tr>
<tr>
<td>Hand held</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.895</td>
<td>0.631 to 0.973</td>
</tr>
<tr>
<td>B</td>
<td>0.932</td>
<td>0.149 to 0.998</td>
</tr>
</tbody>
</table>

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

Table 3 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. Bohannon 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 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.

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. Lower extremity test-retest reliability has been noted to be present when stabilisation is required by the examiner.

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. Bohannon, 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. 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 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, Agre et al, and Bohannon 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. Portney and Watkins 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 9. The use of the Pearson r coefficient to assess reliability would render a value of 1.0, a perfect correlation, 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 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, 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 a highly reliable measurement 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.

References