Effects of Examiner Strength on Reliability of Hip-Strength Testing Using a Handheld Dynamometer

David A. Krause, Mathew D. Neuger, Kimberly A. Lambert, Amanda E. Johnson, Heather A. DeVinny, and John H. Hollman

Context: Hip-muscle impairments are associated with a variety of lower-extremity dysfunctions. Accurate assessment in the clinical setting can be challenging due to the strength of hip muscles relative to examiner strength. Objective: To examine the influence of examiner strength and technique on manual hip-strength testing using a handheld dynamometer. Design: Repeated measures. Setting: Research laboratory. Participants: 30 active adults (age 24 ± 1.4 y). Interventions: Three examiners of different strength performed manual muscle tests (MMT) in 2 different positions for hip extension, abstraction, and external rotation using a MicroFet handheld dynamometer. Examiner strength was quantified via a 1-repetition-maximum leg press and chest press with a Keiser A420 pneumatic resistance machine. Main Outcome Measures: Intrarater reliability (ICC3,1), interrater reliability (ICC2,1), and measured torque values. Results: Intrarater reliability for all measurements ranged from .82 to .97. Interrater reliability ranged from .81 to .98. Main effects for hip extension revealed a significant difference in torque values between examiners and between techniques. For the short-lever hip-abduction and seated hip-external-rotation tests, there was a significant difference between examiners. There was no significant difference in measured torque values between examiners with the long-lever hip-abduction or the prone hip-external-rotation tests. Conclusions: MMT of the hip may be performed with high reliability by examiners of different strength. To obtain valid MMT measurements of hip muscles, examiners must consider their own strength and testing techniques employed. The authors recommend a long-lever technique for hip abduction and a prone position for testing hip external rotation to minimize the influence of examiner strength. Both positions appear to provide mechanical advantages to the examiner compared with the alternative techniques. The authors are unable to recommend a preferred hip-extension-testing technique to minimize the influence of examiner strength.

Keywords: hip, muscle strength, physical examination

Accurate assessment of hip-muscle strength is indicated for a variety of lower-extremity presentations. Hip-muscle weakness is proposed to contribute to many-lower extremity conditions including patellofemoral pain, iliotibial band pain, and anterior cruciate ligament injuries. Ireland et al1 reported that women with patellofemoral pain were more likely to demonstrate hip-abduction and -external-rotation weakness than age-matched women without patellofemoral pain. Tyler et al6 tested the hip abductors and adductors in professional ice hockey players to determine if preseason performance correlated to the incidence of adductor strains. They found that adduction strength was 95% of abduction strength in uninjured players but only 78% of abduction strength in injured players. Nadler et al10 found, in general, that Division I athletes with a history of lower-extremity or low-back pain had a difference in hip strength compared with athletes without previous injury. In a prospective study of 98 high school running athletes, Finnoff et al1 found that athletes who developed patellofemoral pain had a lower baseline hip external-to-internal strength ratio than noninjured athletes, suggesting that a higher external-to-internal-rotation strength ratio may protect athletes from the development of patellofemoral pain.

The assessment of hip-muscle performance in athletes and active individuals may be accomplished through a variety of ways ranging from sophisticated instrumented testing equipment to simple manual techniques performed in a clinical or training-room setting. Handheld dynamometry is a quick and simple clinical means to quantify strength, with application to a variety of patient and client populations. An advantage of handheld dynamometry over traditional manual muscle testing using a grading system is the ability to detect and quantify strength deficits that may otherwise go undetected. A limitation of manual muscle testing with or without a handheld dynamometer (HHD) is the examiner’s strength relative to the strength of a tested muscle group. To perform a valid manual muscle test, the examiner’s force must be sufficient to match the force produced by the tested
Effects of Examiner Strength on Hip-Muscle Testing

As handheld dynamometry is used as a screening tool for hip-muscle performance in active populations, the influence of examiner strength when performing muscle tests requires further investigation.

The purpose of this study was to examine the effects of examiner strength on hip-extension, -abduction, and -external-rotation-strength testing techniques using an HHD in active male and female adults. We hypothesized that torque values would be influenced by both the examiner and the testing technique used. Specifically, we hypothesized that weaker examiner strength would result in less recorded force compared with stronger examiners with short-lever hip-extension and hip-abduction tests and seated hip external rotation and that torque values with long-lever hip-extension and -abduction tests and the prone hip-external-rotation test would not be influenced by examiner strength. Second, we hypothesized that all testing techniques would be reliable both between and within examiners. The goal of this study was to recommend testing procedures that can be effectively conducted in the clinical setting by examiners of different strength.

Methods

Design

We used a repeated-measures reliability study design. Independent variables were examiner and testing technique. The dependent variable was torque.

Subjects

Participants were recruited over a 3-month period via flyers and word of mouth in an academic setting. Thirty subjects (11 men and 19 women) ranging from 18 to 30 years old participated (mean age 24 ± 1.4 y, mean height 173.1 ± 9.1 cm, mean weight 70.9 ± 13.9 kg). Subjects with a history of hip or knee injury requiring medical treatment within the past 2 years were excluded from the study. All subjects reported they were free of any hip or knee condition that would affect the ability to perform the hip-muscle tests. An a priori power analysis concluded that testing 25 participants would provide over 80% power to detect a 20-point difference in test–retest reliability between examiners. Before participation, all subjects provided written informed consent. This study was approved by the institutional review board of Mayo Clinic.

Examiners

Three examiners performed the strength tests. We specifically selected examiners of varied physical characteristics and strength (Table 1). Examiners were second-year doctor-of-physical-therapy students who had completed a course in manual muscle testing. Before formal testing, a physical therapist with over 25 years clinical experience and board certified as an orthopedic specialist reviewed manual muscle tests and had examiners practice and demonstrate proper technique to ensure that all examiners were consistent and comfortable with test procedures and positioning of the HHD and were aware of and monitored for potential compensations by subjects during test performance. The supervising physical therapist was not present during subject testing. Maximal voluntary strength of the examiners’ lower extremities and upper extremities was quantified by measuring a 1-repetition-maximal effort (1-RM) for a leg press and chest press with Keiser A420 pneumatic resistance machines and integrated software (Keiser Sport, Fresno, CA). LeBrasseur et al19 investigated the reliability of muscle-strength testing using the Keiser A420 and found that a 1-RM chest press and leg press demonstrated excellent test–retest reliability (ICC .95–.99) in younger and older men. For the recumbent leg press, examiners started in a position of approximately 90° of knee flexion and 90° of hip flexion, and then extended to 0° of knee flexion. A standardized protocol consisted of warm-up repetitions, followed by a sequence of progressively increased resistances approaching the examiner’s 1-RM. The resistance was adjusted, and 1-RM attempts were performed, separated by 30-second rest periods, until the 1-RM was determined. For the seated chest press, examiners were positioned so the machine’s handles were aligned with the midline of the sternum in the horizontal plane and the anterior aspect of the chest in the frontal plane. The examiners’ knees were positioned at 90° of flexion with feet on the floor. The same process as with the leg press was used to determine the examiners’ chest-press 1-RM.

Procedures

Before testing, leg length was measured by the examiner who served the role of recorder. These measurements were used to torque values. Measurements were taken from the anterosuperior iliac spine (ASIS) to the apex of the medial malleolus and from the ASIS to the medial joint line of the knee.1 The intertester reliability for this measurement is reported as high as r = .98.20 Marks were then placed on the skin 5 cm to the medial malleolus for consistent placement of the HHD. Estimated torque was calculated with the equation Torque = force × moment arm. Moment-arm length for torque calculation was based on placement of the HHD. For long-lever tests, the moment arm was the ASIS to the medial malleolus minus 5 cm.1 For short-lever tests, the distance from the ASIS to the medial joint line was used.1 For

Table 1  Examiner Demographics

<table>
<thead>
<tr>
<th></th>
<th>Examiner</th>
<th>Examiner</th>
<th>Examiner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>female</td>
<td>female</td>
<td>male</td>
</tr>
<tr>
<td>Height (ft, in)</td>
<td>5, 3</td>
<td>5, 8</td>
<td>6, 0</td>
</tr>
<tr>
<td>Weight (lb)</td>
<td>123</td>
<td>155</td>
<td>170</td>
</tr>
<tr>
<td>Chest press (N)</td>
<td>470</td>
<td>675</td>
<td>1062</td>
</tr>
<tr>
<td>Leg press (N)</td>
<td>1828</td>
<td>2274</td>
<td>2787</td>
</tr>
</tbody>
</table>
Hip external rotation, the value used was the distance from the medial joint line to the medial malleolus minus 5 cm.\textsuperscript{1}

Hip-strength values were obtained with use of a MicroFet 2 HHD (Hoggan Health Industries, Inc, West Jordan, UT) with a custom-curved, foam transducer pad for patient comfort. Hip muscles assessed included the hip extensors, abductors, and external rotators. All tests with the exception of prone hip external rotation were consistent with those described by Hislop and Montgomery.\textsuperscript{21} A prone hip external rotation was a modification of the seated test. Two techniques of each muscle group were performed. Test sequence and examiner order were randomized. The leg tested was determined by asking subjects which leg they would use to kick a ball. Make tests were performed for all manual tests.\textsuperscript{9,22} A make test is one in which the examiner maintains a static resistance while the subject exerts maximal effort into the resistance. The examiner who performed the first sequence of tests provided testing instructions. After the examiner explained each test procedure, subjects were allowed a practice trial for familiarization before each individual test. Subjects were instructed to increase their force over a 5-second time interval. A standard cue of “ready, set, go” was used by all examiners to start the testing process. Verbal encouragement via the examiner saying the word \textit{push} repeatedly in a loud voice was provided during the muscle contraction. Subjects performed 2 maximal contractions for each tested motion.\textsuperscript{9,23} The greatest value on the HHD of the 2 tests was recorded. Examiners were blinded to results. A fourth examiner who served the role of recorder viewed the display on the HHD, recorded the result, and then reset the display to zero. Tests for each individual position were performed by 3 examiners (a total of 6 tests per test position). Each examiner completed all tests before testing by the next examiner. Once finished with testing, the examiner left the room and the next examiner per randomization completed all tests in the same order. Testing was performed in a single session. Between tests, a period of 30 seconds for rest was allowed. Between examiners, subjects rested for 3 to 5 minutes. To minimize effects of examiner fatigue, no more than 3 subjects were tested per day. Tests were repeated 1 week later to establish reliability.

The specific tests are illustrated in Figure 1. Short-lever hip extension was performed with the subject prone.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{(A) Hip extension—short lever, (B) hip extension—long lever, (C) hip abduction—short lever, (D) hip abduction—long lever, (E) hip external rotation—seated, and (F) hip external rotation—prone.}
\end{figure}
with test knee flexed to 90°. For additional stability, the subject grasped the examination table. The hip was extended enough to raise the knee off the table. The HHD was positioned on the distal posterior thigh. Static resistance was applied after attainment of the test position. For long-lever hip extension the subject was prone with knees extended. The subject maintained an extended knee and extended the extremity to clear the knee off the table. Once the test position was established, resistance was applied on the posterior leg 5 cm proximal to the level of the medial malleolus of the test leg. All subjects in our pool were able to extend their hip to the test position without the need for modification to accommodate for hip-flexor tightness.

Short-lever hip abduction was performed with the subject side-lying. The non-test limb was positioned in 30° to 45° of hip flexion and 90° of knee flexion to provide stability. The test limb was abducted so the leg was approximately parallel to the table. Resistance with the dynamometer was provided just proximal to the lateral joint line of the tested knee. For long-lever hip abduction the subject was in the same side-lying position. The examiner provided resistance proximal to the lateral malleolus.

Seated external rotation was performed with the subject positioned on the edge of the treatment table with a rolled towel placed under the distal thigh. Resistance with the HHD was provided 5 cm proximal to the medial malleolus. For prone external rotation, the subject was positioned on the treatment table with the test knee flexed to 90°. Resistance was provided 5 cm proximal to the medial malleolus.

Statistical Analysis

Intrarater reliability was examined using model 3 form k intraclass correlation coefficient (ICC3,1) as described by Shrout and Fleiss.24 Interrater reliability was examined using ICC2,1. We used t tests to assess differences between tests (α = .05).

A 2-way mixed-model analysis of variance (ANOVA) was conducted to test the null hypotheses that there would be no difference between hip-strength-testing techniques and no difference in torque recorded by examiners of different strengths. An α level of .05 was used for all statistical comparisons to minimize probability of type I error. Simple-effects tests were used to analyze techniques by examiner interactions, and all post hoc tests were conducted with Bonferroni corrections for multiple comparisons.

Results

Thirty subjects completed testing. Examiner gender, height, weight, and strength values established on the Keiser A420 bench- and leg-press strength tests are presented in Table 1. All muscle tests were reliable. Intrarater reliability (ICC3,1) ranged from .82 to .97 (Table 2). Interrater reliability (ICC2,1) ranged from .81 to .98 (Table 3).
Figure 2 — Hip-extension torque values (main effects). Torque values for short-lever hip extension were significantly less than long-lever torque values ($P < .001$). Torque values for examiner 1 were significantly less than those measured by examiner 2 ($P < .001$) and examiner 3 ($P < .001$). Torque values for examiners 2 and 3 were not significantly different.

Figure 3 — Hip-abduction torque values. Short-lever torque values were significantly greater than long-lever torque values ($P < .001$). For the short-lever test, torque values for examiner 1 were significantly less than those measured by examiner 3 ($P = .001$). Values for examiner 1 were not significantly different than examiner 2 ($P = .052$). Torque values for examiners 2 and 3 were not significantly different. Torque values for the long-lever test were not significantly different between examiners.

abduction, there was no significant difference in torque values measured by the 3 examiners (Figure 3).

For hip external rotation, torque values with the seated test were significantly greater than prone values ($P < .001$). With the seated tests, values measured by examiner 1 were significantly less than values measured by examiner 2 ($P < .001$). Examiner 1’s measurements were not different from those recorded by examiner 3 ($P = 1.00$). Values measured by examiner 2 were significantly greater than those of examiner 3 ($P = .035$). There was no significant difference in torque values among the 3 examiners for prone hip external rotation (Figure 4).

**Discussion**

Our results support our initial hypothesis that torque values would be influenced by examiner strength and testing technique and that all testing techniques would be reliable both between and within examiners. We analyzed our reliability based on a reliability scale that defined excellent reliability as .75 and greater, fair to good reliability as .40 to .75, and poor reliability as less than .40. Based on these reference values, our calculated values for both intrarater reliability (range.82–.97) and interrater reliability (range.81–.98) for all techniques were excellent.
Effects of Examiner Strength on Hip-Muscle Testing

Based on examiner-by-technique interactions identified, we recommend preferred positions for performing hip-abduction and hip-external-rotation testing. We are unable to suggest a preferred position for hip extension as there was no examiner-by-technique interaction. We recommend using a long lever for hip abduction and a prone position for testing hip external rotation to minimize the potential influence of examiner strength. Both positions appear to provide mechanical advantages to the examiner compared with the alternative techniques. For hip abduction, resistance at the distal leg provides a long-lever advantage for the examiner, possibly making the test less demanding to perform. For hip external rotation, the prone position may provide an opportunity for the examiner to assume a better position to perform the test.

The leg and chest press provided a basis to compare the relative strength of the 3 examiners and its influence on performing hip-muscle tests. Based on the 1-RM chest press and leg press, examiner 1 had the lowest strength of the 3 examiners, with examiner 2 having less strength than examiner 3. Strength differences appear to be a factor contributing to differences in torque results of both hip-extension tests examined, short-lever hip abduction, and seated hip external rotation. Examiner 1 obtained the lowest torque values of the 3 examiners for tests of hip extension, short-lever hip abduction, and seated hip external rotation. This suggests that these specific tests are physically challenging to perform. Other possible influencing factors include the examiner’s height and weight. In addition to lesser strength, examiner 1 was the shortest and lightest of the 3 examiners. The influence of height was likely minimal, as the examination table was adjustable, allowing the examiner to be positioned for mechanical advantage. Increased weight could be used to the advantage of the examiner for better stabilization as compared with an examiner of lesser weight. Examiner strength differences did not influence the long-lever hip-abduction test or the prone hip-external-rotation test, as evidenced by no significant differences in torque values recorded between examiners with these tests.

Our finding of excellent ICCs while also noting significant differences in torque values between examiners with some tests can be explained by the nature of the mathematical analysis. The general equation for calculating an ICC is a ratio of between-subjects variance minus within-subject variance divided by between-subjects variance. If between-subjects variance is large and within-subject variance is relatively small, then the magnitude of the ICC will approach 1.00. In our data, within-subject variance was represented by examiner-to-examiner differences in the hip-torque values measured on individual subjects. In some of the tests that were conducted, the examiner-to-examiner differences were statistically significant ($P < .05$). Nevertheless, the variance between examiners on individual subjects was considerably smaller than the variance between subjects, so the ICC values approached 1.00. Our data illustrate how measurements may be relatively reliable but not necessarily valid. Measurements taken by stronger and weaker testers can have a relatively high magnitude of reliability as assessed with the ICC—as our data show—but that does not necessarily mean that strength measurements taken by stronger and weaker testers will be equivalent.

The torque results were consistent, but given that there was a significant difference with some tests, the results did not always accurately reflect the measurement of interest, which was the force-producing capacity of the tested muscle group. Instead, some values may represent the maximum amount of force the examiner could resist for a specific muscle test. Findings of examiner strength’s influence on manual muscle testing are reported by others.

Figure 4 — Hip-abduction torque values. Short-lever torque values were significantly greater than long-lever torque values ($P < .001$). For the short-lever test, torque values for examiner 1 were significantly less than those measured by examiner 3 ($P = .001$). Values for examiner 1 were not significantly different than examiner 2 ($P = .052$). Torque values for examiners 2 and 3 were not significantly different. Torque values for the long-lever test were not significantly different between examiners.
Mulroy et al\textsuperscript{17} reported that a female examiner with a lower maximal vertical push force than a male counterpart could not detect quad weakness as accurately as the male examiner. Similarly, Thorborg et al\textsuperscript{18} investigated the influence of gender and upper extremity strength on manual-muscle-test values. They found that a female examiner obtained lower strength-test values for movements in all planes of the hip than those obtained by a male examiner. Our findings do not consistently support these findings. With the long-lever hip-abduction test and the supine hip-external-rotation tests, we did not find a significant difference between examiners. While other hip tests in this study found a significant difference between examiners, the torque measured by 1 of the female examiners was not significantly less than the torque measured by the male examiner, and thus we did not find a consistent gender effect among our examiners. An examiner's strength and the specific testing technique appear to have more influence on measured torque than gender does.

We chose to use the make test for performing muscle tests. A break test, in contrast, is one in which the examiner provides a level of force sufficient to overcome the force produced by the subject. Stratford and Balsor\textsuperscript{25} found higher reliability for the make test (.95) than for the break test (.87) when examining elbow flexors. They theorized that the lower reliability with the break test was due to additional skill required on the part of the examiner to administer the test. The examiner must overcome the force produced by the subject and record the force at the specific point when movement is detected. Bohannon\textsuperscript{22} found similar reliability between the 2 tests, again testing elbow flexors, and concluded that one test is not superior to the other. Both studies report that the break test results in greater force than the make test.\textsuperscript{22,25} This may be due to the eccentric activation of the muscle with the break test. This suggests that an examiner may need greater strength to perform a valid break test than for the make test.

Kramer et al\textsuperscript{26} tested the reliability of hip manual muscle testing for abduction in young and elderly women using handheld dynamometry and belt-resisted manual muscle testing, which would essentially be a make test. Although both methods were found to be reliable, belt-resisted testing demonstrated higher reliability. In addition, higher torque values were obtained using belt resistance. Therefore, examiners of lesser strength might consider using belt fixation to ensure maximal torque production in manual muscle testing and thus a more valid test.

The difference in the degree of hip flexion between the 2 techniques of hip external-rotation testing likely influences the demands on specific hip muscles. Calculating moment arms using anatomical specimens in various hip-flexion angles, Delp et al\textsuperscript{27} reported that hip position changes the moment arm of the various components of the gluteus maximus, piriformis, quadratus femoris, obturator internus, and obturator externus. While they reported a greater moment arm for the gluteus maximus in the prone position, we recorded greater force in the seated position. Other factors may have contributed to differences we found. While we monitored for substitution patterns, the seated position may allow more opportunity for subjects to position their trunk or pelvis to generate more force. Other potential influencing factors include differences in patient stabilization and the ability of the subject to visually observe the seated test.

Torque values for long-lever hip abduction were different than we expected. We expected calculated torque values to be as great or greater than those with the short lever. We attempted to keep the position of the hip consistent and thus do not think that a difference in the amount of hip abduction explains this difference. Possible explanations include the long-lever test's being an inherently more challenging test for the subject, thus requiring greater demands to the hip musculature for stabilization. Further investigation with electromyography is needed to confirm this hypothesis.

Clinicians should consider their own strength and the strength of the patient when performing manual muscle tests of the hip. This is particularly relevant when therapists of different strengths are co-creating a patient and manual muscle testing is performed throughout the patient's episode of care. Differing dynamometry values may be obtained from different therapists that do not correlate to actual differences in patient strength. This compromises the ability to show the strength gains or strength deficits of a patient. While we found all manual-muscle-testing techniques in this study to be reliable both within and between examiners, significant differences in torque values were found between techniques. As torque values depend on the specific test used, therapists should remain consistent with tests used throughout an episode of care.

A limitation of this study could be the influence of the patients' perceptions of the examiners' strength on their efforts in the manual muscle test. Because examiners were visibly different in height, weight, and strength, subjects may have given different efforts for different examiners. Another limitation of this study was the homogeneity of the subject pool. All subjects were active and overall healthy young adults. This would represent an athletic, active population but may not represent results in other patient populations. Finally, our torque calculations are estimates of actual torque, as we used the ASIS as a reference for our leg-length measurements. While the ASIS does not represent the actual axis of rotation at the hip, as the limb measurement was consistent for a specific individual across the 3 examiners, we do not believe that our methods change our suggestions for preferred testing positions and the relative influence of examiner strength.

This study examined the intrarater and interrater reliability of torque production of the hip muscles among examiners of different strengths under different testing techniques. Future research should examine the validity of different techniques between multiple examiners of different strengths on a single subject. In addition, comparing manual-muscle-testing techniques against a gold-standard strength test, such as a Biodex, could further investigate technique validity.
Conclusion

Hip manual muscle testing may be performed with excellent reliability by examiners of different strength. The validity of some tests is influenced by examiner strength. To obtain valid manual-muscle-testing measurements of hip-muscle groups using an HHD, examiners must consider their own strength and the testing techniques employed. We recommend using tests that provide a mechanical advantage to the examiner. We recommend using a long lever for hip abduction and a prone position for testing hip external rotation to minimize the potential influence of examiner strength.

References

