Interrater Reliability of Knee Muscle Forces Obtained by Hand-held Dynamometer from Elderly Subjects with Degenerative Back Pain

Jessica C. Dunn, PT, MS,[1,3] Maura D. Iversen, PT, MPH, SD[2,3,4]

[1]Centennial Healthcare, Charlwell House
[2]Department of Physical Therapy, School for Health Sciences, Simmons College
[3]RBB Multipurpose Arthritis & Musculoskeletal Diseases Research Center, Department of Medicine, Division of Rheumatology, Immunology & Allergy, Section on Clinical Research, Brigham & Women's Hospital
[4]Harvard Medical School, Boston, MA

This study was supported by grants from the Foundation for Physical Therapy (MDI) and the National Arthritis Foundation (MDI).

ABSTRACT

Background and Purpose: We determined the inter-rater reliability of measurements of knee extensor and flexor muscle strength obtained with a Nicholas Manual Muscle Tester (NMMT) from older adults with chronic low back pain. **Methods***:* Subjects were 25 (17 female and 8 male) patients (68.2±7.7 years) with degenerative back pain for > 6 months. Two female physical therapists measured subjects' isometric knee extensor and flexor muscle force. Knee position was standardized between measures and raters. The order of testers was performed at random. Intraclass correlation coefficients (ICCs) were used to describe the inter-rater reliability for the total sample and for subjects classified into high and low symptom groups according to the Lumbar Spinal Stenosis-symptom severity scale. We compared the inter-rater reliability of measurements obtained from patients with high and low symptoms. **Results***:* Inter-rater reliability was high for the total sample (0.87-0.93) and both symptoms groups (high symptom [ICC=0.89-0.94]; low symptom [ICC=0.80-0.92]. There was no significant difference in the inter-rater reliability of measures obtained from the high and low symptom groups. **Conclusions:** Inter-rater reliability is good to high for measurements of knee extensor and flexor muscle force obtained by hand held dynamometry. The reliability is not influenced by symptom severity.

Key Words: hand-held dynamometry, muscle force, knee

INTRODUCTION

Muscle strength is associated with functional performance, work productivity, and efficiency of movement.[1] The measurement of muscle strength, therefore, is an important component of patient evaluation and diagnosis.[1,2] Manual muscle testing (MMT) is the most frequently used technique for quantifying strength as it is cost effective, quick, and useful for comparing strength between and within individuals.[3]

Although MMT is widely used to quantify strength, its reliability and sensitivity have been questioned.[3] For example, Iddings et al,[4] examined the inter-rater reliability of MMT scores obtained by 10 examiners using a 13-integer scale (an expansion of the traditional 6-point MMT scale which incorporated plus/minus values to provide more descriptive strength measurements). They found a 45% agreement among the 10 examiners on all grades and a 96% agreement when researchers included plus or minus one grade (a 3 integer difference). Both Lilienfeld et al and Iddings et al concluded that MMT demonstrated good reproducibility in the zero to fair strength ranges.[4,5] However, the subjectivity of MMT resulted in poor reproducibility within the good to normal grades.[4]

Beasley[6] compared knee extensor force measurements obtained by testers who used MMT and a manual force gauge. The testers obtained bilateral measurements of knee extensor muscle force from healthy children and children with poliomyelitis. Subjects who they gave a 'normal' knee extensor strength grade bilaterally using MMT demonstrated nearly a 20% difference in extensor muscle force via manual force gauges (mean stronger side = 48.1 lb. versus mean weaker side = 40.9 lb.). Beasley also noted that patients who received a MMT rating of less than normal knee extensor strength unilaterally demonstrated nearly a 50% strength deficit compared to the opposing limb [mean (normal) strong side = 49.0 lb; mean (good) weak side = 31.6 lb]. He, therefore concluded that therapists using MMT were only able to detect weakness in the knee extensor muscle group when the loss in strength was ≥ 50%.[6]

Researchers and clinicians have attempted to standardize testing procedures to enhance the reproducibility of MMT.[7] However, procedural variations still exist, which in turn, decrease the reliability of muscle force measures by affecting the lever arm and the amount of the resistance applied. Factors which influence the quantification of force include: point of force application, magnitude and duration of resistance offered, direction of force provided, test position, type of test used (make vs. break), prestretching, number of practice sessions, and encouragement.[8,9] Patients' age, pain tolerance, type of disease, voluntary control, occupation, and level of fitness are other attributes that may influence physical therapists' grading of a patient's strength.[9] The testers', as well as the patients' physical characteristics affect strength scores.[10]

Concern about the subjectivity of muscle force measurement has stimulated the development of strength measuring

Address correspondence to: Dr Maura D Iversen, Department of Physical Therapy, School for Health Sciences, Simmons College, 300 The Fenway, Boston, Massachusetts 02115. Phone: 617-521-2633, Fax: 617-521-3032 (iversen@simmons.edu).

devices, such as the hand-held dynamometer (HHD).[3] Hand-held dynamometers can be used to precisely document muscle force in actual units (ie, Newtons, kg, or lbs). This precision allows clinicians to quantitatively monitor disease progression, and to potentially verify the effectiveness of physical therapy strengthening regimens.[10] Hand-held dynamometers are relatively inexpensive (approximate cost $800-1500). Also, HHDs are convenient, weigh less than 2 pounds, require no set-up, and can be used in many environments.[1,10]

Researchers have analyzed the reliability of measurements obtained by HHDs from upper and lower extremity muscles of healthy individuals[11-13] and patients with neurological and orthopedic disorders.[14,15] Studies analyzing knee extensor and flexor muscle strength have produced inconsistent results.[12-15] Researchers speculate that patient/examiner strength, the anthropometric characteristics of the examiner, standardization procedures, and pain symptoms may influence reliability results.[10,13,15-17]

Bohannon and Andrews evaluated the inter-rater reliability of lower extremity muscle force measurements obtained by HHD from 30 patients with neurological disorders.[14] They described high inter-rater reliability using Pearson product-moment correlations (r) and t-tests. Inter-rater reliability for measurements taken with the Chatillon force gauge was 0.90 when testing knee extensor force in a weak population.[14] In contrast, Kimura et al found poor inter-rater reliability using a HHD when testing the knee extensors in 12 healthy subjects.[13] They quantified knee extensor strength using the Chatillon HHD and MicroFet HHD. The mean inter-rater reliability using intraclass correlation coefficients (ICCs) was 0.32 for the measurements obtained with the Chatillon HHD and 0.41 for the measurements obtained with the MicroFet. Agre et al conducted one of the few studies that examined the inter-rater reliability of knee flexor strength measures using a HHD.[12] Three examiners measured the knee flexor strength of 4 healthy subjects who sat with their limb placed in 90˚ of flexion. The reliability of the measures ranged from moderate to high (r= 0.49 to 0.95).[12] However, Pearson correlations estimate association rather than agreement and therefore, the results should be interpreted with caution. Bohannon[18] found high test-retest reliability of measures of knee flexor strength using a HHD from Spark Instruments. Knee flexion strength was measured in 30 individuals with neurological issues using *make tests* and a standardized testing procedure. Test-retest reliability coefficients for knee flexion measures was 0.98.[18] In sum, while hand-held dynamometry is a practical alternative to MMT, questions remain about the reliability of hand held dynamometry, particularly among specific populations. Therefore, further studies are warranted.

The purpose of this study was to determine the inter-rater reliability of knee extensor and flexor strength measurements obtained with a HHD from patients with symptomatic low back pain of degenerative origin. To the best of our knowledge, researchers have not analyzed the inter-rater reliability of lower extremity strength measurements using HHD in this population. We anticipated large variability in muscle force production in this sample. The second purpose was to determine if the reliability of measures differed for patients with more versus less severe symptoms. We hypothesized that reliability would be moderate to high (0.55-0.90) for the whole sample, as we believed that low back pain symptoms may decrease muscle force production through inhibition or psychological factors impacting the subjects' willingness to exert muscle force.[17,19,20] We further hypothesized that inter-rater reliability would be higher for patients with more severe symptoms as these subjects may exert less muscle force than subjects with fewer symptoms.

METHODS

Subjects

This cross-sectional study involved the secondary analysis of data from a project entitled, "Enhancing Function In Community-Dwelling Persons with Chronic Low Back Pain: A Randomized Controlled Trial of An Endurance Training Program,"(M Iversen, PT, MPH, SD, unpublished data, 2000-2002) which was approved by Brigham & Women's Hospital Institutional Review Board (2000-P-0022921/1; BWH). In this analysis we used baseline questionnaire data and measures of lower extremity strength obtained using a HHD.

Twenty-five patients with symptomatic chronic low back pain of degenerative origin were recruited from a large tertiary institution and participated after providing an informed consent. Patients were recruited if they were 55 years or older, had a history of back, buttock, and/or leg pain exacerbated by lumbar extension, were able to speak and understand English, were able to exercise safely, were not currently enrolled in physical therapy or participating in an exercise training program, had no cognitive impairments, had no medical problems other than back pain that limited their function, and did not have low back surgery in the past year or epidural steroid injections over the last 6 months. Subjects completed questionnaires on demographics and general health status prior to muscle testing. Patients were inexperienced with muscle strength testing using a HHD.

Most (68%) of the patients were female. Twenty-three (92%) were Caucasian, one subject was Hispanic, and one subject was African-American. The subjects' age was 68.2$\pm$7.5 years (range 57-83). Only 8 patients (32%) worked full time at the time of measurement. More than half (52%) reported that fatigue limited their ability to work or perform activities of daily living at least sometimes. Fourteen patients (56%) had chronic low back pain (CLBP) without radiation, 4 patients (16%) had CLBP with either radiating pain or neural deficits, 6 patients (24%) had neurogenic claudication due to lumbar spinal stenosis (LSS), and 1 patient (4%) had sciatica due to a herniated disk.

Raters

Two female physical therapists with different anthropometric characteristics (rater 1: height: 157.5 cm; rater 2: height: 172.9cm) participated in this study. The raters had different levels of clinical experience (rater 1: 16 yrs; rater 2: 1 yr), as well as differing experience using the HHD. One had 3 years of experience using the HHD; the second had 1 year of experience using the HHD. Both were trained to use the HHD following the study protocol. The rater with less experience trained 2 months prior to the initiation of data collection using the Nicholas Manual Muscle tester (NMMT). She practiced until her force measurements were shown by ICCs to have high intrarater reliability.

Instrumentation

Lower extremity muscle force was assessed with a HHD, the NMMT (model 01160, Lafayette Instrument Company, Lafayette, Indiana). This HHD registers 0.0 to 199.9 kg with a precision of 0.1 kg.

Procedures

Subject stratification

Subjects were stratified into high and low symptom groups based on their LSS symptom severity score derived from the LSS-questionnaire outlined in Table 1.[21] The developers of the LSS symptom severity scale assessed its construct validity by comparing the pain domain of the symptom severity scale to pain measured on a visual analogue scale[22] and the Sickness Impact Profile (SIP).[23] They found high internal consistency for the LSS-symptom severity and functional capacity scales (Cronbach's alphas from 0.64 to 0.92) and high test-retest reliability (r = 0.82 to 0.96). The scales were also highly responsive (correlations between change scores ranged from 0.96 to 1.07).[21] Likert scales were used to assess subjects' symptom severity with 6 questions. Subjects' responses included one of the following choices scored from a 1- 5 scale: none, mild, moderate, severe, very severe. The question regarding balance disturbance only had 3 choices (never, sometimes, often) and was converted into a 1-3-5 scale. The unweighted mean of these 7 items was taken to obtain a patient score between 1 and 5. The high symptom group was operationally defined as any individual who scored $\geq$ 2.5 on the LSS-short form. The low symptom group was operationally defined as any individual who scored < 2.5 on the LSS-short form.[21] Fourteen patients (56%) were stratified into the high symptom group.

Standardization procedures

The *make test*, operationally defined as a maximal effort exerted against a stationary HHD, was used to quantify subjects' knee extensor and flexor isometric muscle force production.[14] Subjects were seated with a small towel roll placed distally to support the femur in 90º of hip flexion. The subject's knee was placed in 70º knee flexion using a goniometer. The distance from the tibial tuberosity to the superior aspect of the medial malleolus was measured and recorded in cm. We then multiplied this distance by 0.6 and a mark placed at that distance on the anterior tibia to identify the placement of the force plate during each force measure by each rater. A small towel was folded in quarters (24.5 x 24.5 cm) and placed over the mark to provide cushioning from the dynamometer's force plate.

We used the following standard verbal cues while measuring knee extensor muscle force: "Please kick as hard as you can. Kick slowly and smoothly. Do not lean backwards when testing. One, two, three, kick!" The nontesting rater monitored the patient's position. When the forces generated were minimal, the rater was able to stabilize the subject during testing. If the rater was unable to stabilize the patient, the nontesting rater stabilized the testing limb to decrease the chances of muscle substitution (Figure 1 and Figure 2). We used the following cues during quantification of knee flexor muscle strength: "Pull your heel underneath the bed as hard as you can. Do not lift your thigh off the bed during testing. Pull slowly and smoothly. One, two, three, pull!" (Figure 3 and Figure 4).

Testing order

The NMMT was reset to 0 prior to each muscle force measure and the joint angle was remeasured between trials. A

Table 1. Components of the Lumbar Spinal Stenosis Symptom Severity Scale[21]

In the past month, how would you describe the pain you have had on average? Include the pain in the back, buttocks, and pain that goes down the legs.
1 Very severe
2 Severe
3 Moderate
4 Mild
5 Very Mild
6 None

During the past month, how often have you had back, buttock, or leg pain?
1 Every minute of the day
2 Everyday for most of the day
3 Everyday, for at least a few minutes
4 At least three times a week
5 At least once a week
6 None

In the last month, on a typical day, how would you describe:
The pain in your back or buttocks?
1 I have no back or buttock pain
2 Mild
3 Moderate
4 Severe
5 Very Severe

The pain in your legs or feet?
1 I have no pain in my legs or feet
2 Mild
3 Moderate
4 Severe
5 Very Severe

Numbness or tingling in your legs or feet?
1 I have no numbness in my legs or feet
2 Mild
3 Moderate
4 Severe
5 Very Severe

Weakness in your legs or feet?
1 I have no weakness in my legs or feet
2 Mild
3 Moderate
4 Severe
5 Very Severe

In the past month, have you had any problems with your balance?
1 No, I have had no problems with balance.
2 Yes, sometimes I feel my balance is off, or that I am not sure-footed.
3 Yes, often I feel my balance is off, or that I am not sure-footed

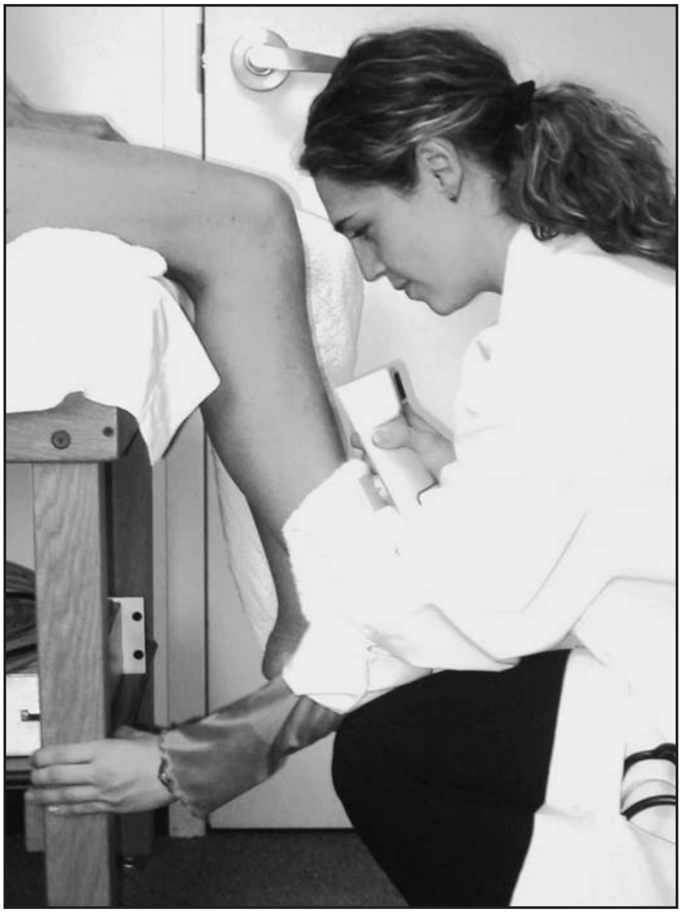

Figure 1. Rater 1 measuring isometric knee extensor muscle force. Rater 1 kneels while she measures right knee extensor strength. She stabilizes her right elbow against her right anterior superior iliac spine to assist her in stabilizing the dynamometer. With her other hand, she grips the leg of the plinth to stabilize her body during testing.

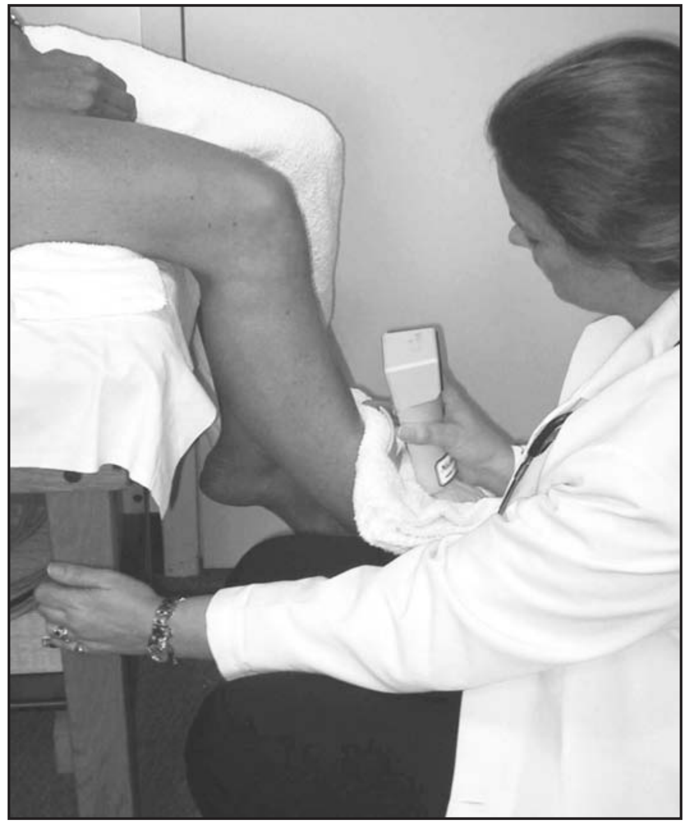

Figure 2. Rater 2 measuring isometric knee extensor muscle force. Rater 2 sits on a stool while measuring right knee extensor muscle force. She grips the leg of the plinth to stabilize her body during testing.

single practice trial was performed before testing to ensure each patient understood the commands. Each subject's muscle groups were tested in the following order: right knee extensors (RKE), right knee flexors (RKF), left knee extensors (LKE), left knee flexors (LKF). Each contraction was held for 4 seconds. Three measurements were taken per rater for each muscle group with a 30 second rest between each measure. The knee joint angle was reassessed between each measure using a goniometer. After a 3-minute rest period, the other rater repeated the procedure. The subject and the opposing rater were blinded from the muscle force results. The order of testing by rater was performed at random to reduce fatigue and testing effects.

Analysis

Data analysis was performed using the SAS Statistical Package.[24] The subjects' muscle force measurements from each rater were averaged separately to obtain the average force (kg) for each muscle group for the entire sample and for the LSS severity groups (Table 2). Standard deviations of the mean muscle force measures and standard errors of measurement (SEM) were reported for the entire sample and for the LSS severity groups (Table 2). The standard error of measurement, calculated as specified in Portney and Watkins,[25] is a measure of error expressed in the actual units of measurement that occurs in any one measure.

We used a mixed model analysis of variance (ANOVA) to account for nesting within subjects. This model was run for each muscle group. The random effects were the subject and rater; the order of testing was a fixed effect. We repeated the analysis with subjects classified into high and low symptom groups (Table 2). The variance components from the ANOVA were used to calculate the ICCs and 95% confidence intervals for the ICCs using the model 2 formula specified in Portney and Watkins[25] for each muscle group (Table 2).

Next, we examined whether the inter-rater reliability of our force measures differed between the high and low symptom groups. We tested whether the ICCs differed significantly from one other using the results of the F-tests from the ANOVA. We calculated the power based on our sample size of 25 and the differences in correlations of force scores between raters for the high and low symptom groups. With an alpha set at 0.05, the power to detect differences in force measures between the raters in the two groups was 0.95.

RESULTS

Our sample demonstrated diverse strength characteristics (Table 2). The inter-rater reliability of the NMMT across the sample ranged from 0.87 to 0.93 (Table 2). The inter-rater reliability (ICC) for the high symptom group ranged from 0.89

Figure 3. Rater 1 measuring isometric knee flexor muscle force. Rater 1 measures right knee flexor muscle force while squatting. She uses both hands to stabilize the dynamometer while rater 2 stabilizes the subject's thigh against the plinth. This is to prevent muscle substitution during testing.

Figure 4. Rater 2 measuring isometric knee flexor muscle force. Rater 2 sits on a stool while measuring the subject's right knee flexor muscle force. She uses one hand to stabilize the dynamometer and the other to stabilize the subject's thigh to prevent muscle substitution during testing.

to 0.94. In the low symptom group, the ICCs ranged from 0.80 to 0.92 (Table 2). In the low symptoms subgroup the ICCs for the knee extensor muscle force measures were slightly lower (0.80 and 0.81) in the compared to the high symptom subgroup (0.89-0.94) as were the ICCs for the knee flexor force measures (0.88-0.92 versus 0.93, respectively). However, examination of the 95%confidence intervals of the ICCs for each muscle group revealed significant overlap, indicating a lack of difference in the reliability coefficients between the subgroups. The F-tests confirmed the lack of difference in raters' reliability of measures between the subgroups (F-tests range 0.00-2.7; p-values range 0.12-0.99).

DISCUSSION

Researchers have hypothesized that patient strength and symptoms influence the reliability of measures of muscle force production using a HHD.[12,15,18] To test this hypothesis, we grouped patients into high and low symptom groups based on the assumption that pain symptoms may decrease muscle force production through inhibition,[26] or psychological aspects of pain such as fear or anxiety.[17,19,20] We hypothesized that our reliability would be higher in the high symptom group as these subjects may exert less muscle force than subjects with fewer symptoms. However, our inter-rater reliability was similar in the two groups which appeared to have similar strength characteristics. Therefore, it is unclear how patient/examiner strength influences the reliability of handheld dynamometry.

Although the mean muscle force measures for the LSS severity groups were similar, the standard deviations of the knee extensor force measures in the high symptom group were nearly twice those of the low symptom groups, indicating a wide range of muscle force abilities. One explanation for this variability in knee extensor force measures may be that these muscles are able to generate more muscle force than the hamstrings.[27] Kilmer and colleagues found the inter-rater reliability of force measures to be significantly higher when testing weaker subjects with hereditary motor and sensory neuropathies than healthy subjects.[28] No description of the raters' experience or strength abilities was provided. The researchers reported that the inter-rater reliability of measures of knee extensor muscle force in the weaker group was 0.89. The ICCs for knee extensor muscle force dropped to 0.53 when testing healthy subjects. These researchers concluded that the low reliability coefficients were due to the examiners' inability to stabilize the dynamometer during testing and the inexperience of the second measuring rater.

Limitations of this study should be noted. First, averaging muscle force measures may have biased the results toward the null, resulting in small differences between the groups. Second, subjects were stratified into severity groups based upon self-report. Self-reported questionnaires are a measure of patients' perceived functional performance. Therefore, potential exists for either an over-estimate or under-estimate of physical function, leading to random misclassification of

Table 2. Statistics Relevant to the Interrater Reliability of Knee Extension and Flexion Measurements Obtained by Hand-held Dynamometry from 25 Patients with Chronic Degenerative Low Back Conditions

Muscle Groups	Rater 1 Mean kg ± SD (SEM[e])	Rater 2 Mean kg ± SD (SEM[e])	ICC (95% CI)
All Patients (n=25)			
RKE[a]	18.4 ± 6.5 (1.30)	17.8 ± 6.8 (1.35)	0.92 (0.83-0.96)
RKF[b]	13.3 ± 4.7 (0.95)	12.9 ± 4.6 (0.92)	0.91 (0.81-0.96)
LKE[c]	15.8 ± 6.1 (1.22)	15.7 ± 5.2 (1.04)	0.87 (0.72-0.94)
LKF[d]	13.1 ± 5.3 (1.03)	12.6 ± 5.1 (1.02)	0.93 (0.84-0.97)
High Symptom Group (n=14)			
RKE[a]	18.0 ± 8.1 (2.15)	17.8 ± 8.6 (2.29)	0.94 (0.84-0.98)
RKF[b]	13.4 ± 5.4 (1.43)	12.7 ± 5.4 (1.46)	0.93 (0.79-0.98)
LKE[c]	15.9 ± 7.4 (1.97)	15.9 ± 6.1 (1.62)	0.89 (0.68-0.96)
LKF[d]	13.2 ± 5.9 (1.58)	13.0 ± 5.9 (1.60)	0.93 (0.79-0.98)
Low Symptom Group (n=11)			
RKE[a]	18.9 ± 4.1 (1.24)	17.9± 3.7 (1.11)	0.81 (0.44-0.94)
RKF[b]	13.3 ± 4.0 (1.21)	13.0 ± 3.5 (1.06)	0.88 (0.60-0.96)
LKE[c]	15.6 ± 4.4 (1.31)	15.6 ± 4.2 (1.26)	0.80 (0.42-0.94)
LKF[d]	12.8 ± 4.2 (1.26)	12.0 ± 3.9 (1.17)	0.92 (0.74-0.98)

[a]RKE, Right knee extension
[b]RKF, Right knee flexion
[c]LKE, Left knee extension
[d]LKF, Left knee flexion
[e]SEM, Standard errors of measurement (kg)

subjects into symptom groups. For example, subjects may have self-reported increased pain symptoms and weakness as compared to their premorbid state, but still produced high mean muscle force measures. Finally, since we analyzed the inter-rater reliability of measurements obtained with the NMMT (model 01160), we are unable to generalize our results to other brands and/or types of HHDs or to the assessment of different muscle groups.

In spite of our limitations, we found the procedures used in this study to be a reliable method of quantifying isometric knee extensor and flexor muscle force in elderly subjects with degenerative low back pain. Our protocol aimed to standardize subject biomechanical factors that influenced muscle force production such as knee joint angle and lever. In addition, tester order was randomly selected to reduce the effects of subject learning and fatigue. Consistent verbal cues were used to decrease motivational influences.[29] Although our protocol standardized subject biomechanical factors, the raters' testing positions varied. Each rater used testing positions that were suited to her anthropometric characteristics. Rater 1 felt she had a better mechanical advantage when she kneeled testing the knee extensors (Figure 1). Rater 2 felt more comfortable sitting on a stool (Figure 2). When testing the knee flexors, rater 1 squatted while holding the dynamometer with both hands as rater 2 stabilized the subject's thigh against the plinth to prevent muscle substitution (Figure 3). When rater 2 tested the same muscle group, she sat on the stool and held the dynamometer with one hand while the other hand stabilized the patient's thigh (Figure 4). Despite these differing positions, the inter-rater reliability of knee extensor and flexor force measures were high. Despite the examiners' different levels of clinical expertise and experience using the NMMT, our results indicate the NMMT can be used to reliably measure muscle force in a diverse sample when a standardized protocol is used.

SIGNIFICANCE

To the best of our knowledge, this is the first study to analyze the inter-rater reliability of measurements obtained with the NMMT from older adults with symptomatic chronic low back pain of degenerative origin. This instrument reliably measured muscle force when a standardized protocol and proper examiner training were implemented. We recommend therapists record patients' tibial length measurements to ensure consistency between measures and raters. The testing procedures used in this study are easily reproduced in a timely manner within a clinical setting. Dynamometers can provide clinicians with useful information and enable clinicians to rate the effectiveness of strengthening programs based on pre- and post-strengthening scores. Examiners using a standardized protocol and the NMMT can reliably quantify lower extremity muscle force.

REFERENCES

1. Wessel J, Kaup C, Fan J, et al. Isometric strength measurements in children with arthritis: Reliability and relation to function. *Arthritis Care Res.* 1999;12:238-246.
2. Guide to Physical Therapist Practice. 2nd ed. *Phys Ther.* 2001;81:9-744.
3. Bohannon RW. Quantitative testing of muscle strength: issues and practical options for the geriatric population. *Topics in Geriatric Rehabilitation.* 2002;18:17-35.
4. Iddings DM, Smith LK, Spencer WA. Muscle testing: part 2. Reliability in clinical use. *Phys Ther Rev.* 1960;41:249-256.
5. Lilienfeld AM, Jacobs M, Willis M. A study of the reproducibility of muscle testing and certain other aspects of muscle scoring. *Phys Ther Rev.* 1954;34:279-292.
6. Beasley WC. Influence of method on estimates of normal knee extensor force among normal and postpolio children. *Phys Ther Rev.* 1956;36:21-41.

7. Daniels L, Worthingham C. *Muscle Testing: Techniques of Manual Examination*. Philadelphia, Pa: WB Saunders; 1986.
8. Nicholas JA, Sapega A, Kraus H, Webb JN. Factors influencing manual muscle tests in physical therapy. *J Bone Joint Surg*. 1978;60-A:186-190.
9. Wakim KG, Gersten JW, Martin GM. Objective recording of muscle strength. *Arch Phys Med Rehabil*. 1950;23:90-100.
10. Wadsworth CT, Nielsen DH, Corcoran DS, et al. Interrater reliability of hand-held dynamometry: effects of rater gender, body weight, and grip strength. *J Orthop Sports Phys Ther*. 1992;16:74-81.
11. Byl NN, Richars S, Asturias J. Intrarater and interrater reliability of strength measurements of the biceps and deltoid using a hand-held dynamometer. *J Orthop Sports Phys Ther*. 1988;9:399-405.
12. 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;68:454-457.
13. Kimura IF, Jefferson LM, Gulick DT, Coll RD. Intra- and intertester reliability of Chatillon and MicroFet hand-held dynamometers in measuring force production. *J Sport Rehab*. 1996;5:197-204.
14. Bohannon RW, Andrews AW. Interrater reliability of hand- held dynamometry. *Phys Ther*. 1987;67:931-933.
15. Hayes KW, Falconer J. Reliability of hand-held dynamometry and its relationship with manual muscle testing in patients with osteoarthritis in the knee. *J Orthop Sports Phys Ther*. 1992;16:145-149.
16. Mulroy SJ, Lassen KD, Chambers SH, Perry J. The ability of male and female clinicians to effectively test knee extension strength using manual muscle testing. *J Orthop Sports Phys Ther*. 1997;26:192-199.
17. Linton SJ. The relationship between activity and chronic back pain. *Pain*. 1985;21:289-294.
18. Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session of strength assessment. *Phys Ther*. 1986;68:206-208.
19. Wolfe F. Determinants of WOMAC function, pain and stiffness scores: Evidence for the role of low back pain, symptom counts, fatigue and depression in osteoarthritis, rheumatoid arthritis and fibromyalgia. *Rheumatology* (Oxford). 1999;38:355-361.
20. Sokka T, Kankainen A, Hannonen P. Scores for functional disability in patients with rheumatoid arthritis are correlated at higher levels with pain scores than with radiographic scores. *Arthritis Rheum*. 2000;43:386-389.
21. Stucki G, Daltroy L, Liang M, et al. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine* 1996;21:796-803.
22. Price D. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain*. 1983;17:45-56.
23. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The sickness impact profile: Development and final revision of a health status measure. *Med Care*. 1981;19:787-805.
24. SAS Institute: *SAS/STAT User's Guide*, Version 6. 4th ed. Cary, NC: SAS Institute Inc.; 1990.
25. Portney LG, Watkins MP. *Foundations of Clinical Research Applications to Practice*. Upper Saddle River, NJ: Prentice Hall; 1993.
26. Fredericks CM, Saladin LK. *Pathophysiology of the Motor Systems: Principles and Clinical Presentations*. Philadelphia, Pa: F.A. Davis Company; 1996.
27. Murray MP, Baldwin JM, Sepic SB, Downs WJ. Maximum isometric knee flexor and extensor muscle contractions normal patterns of torque versus time. *Phys Ther*. 1977;57:637-643.
28. Kilmer DD, McCrory MA, Wright NC, et al. Hand-held dynamometry reliability in persons with neuorpathic weakness. *Arch Phys Med Rehabil*. 1997;78:1364-1368.
29. Jones RE. Reliability of muscle strength testing under varying motivational conditions. *JAPTA*. 1963;42:240-243.