

Hand-Held Dynamometry: Reliability of Lower Extremity Muscle Testing in Healthy, Physically Active, Young Adults

Brent M. Kelln, Patrick O. McKeon, Lauren M. Gontkof,
and Jay Hertel

Context: Hand-held dynamometry (HHD) has been shown to be a reliable, objective way to obtain strength measurements in elderly and physically impaired subjects. **Objective:** To estimate the intratester, intertester, and intersession reliability of HHD testing of lower extremity movements in young, healthy subjects. **Design:** Repeated measures. **Setting:** Sports medicine laboratory. **Participants:** Nine males and eleven females (Mean age = 26 years). **Measurements:** Strength measures of 11 right lower extremity movements were taken by 3 different testers on 2 separate days using a HHD. **Results:** Intratester ICC range was .77 to .97 with SEM range of .01 to .44 kg. Mean intertester ICC range was .65 to .87 with SEM range of .11 to 1.05 kg. Mean intersession ICC range was .62 to .92 with SEM range of .01 to .83 kg. **Conclusions:** HHD has the potential to be a reliable tool for strength measurements in healthy, strong subjects; however, there are noteworthy limitations with movements where subjects can overpower the testers.

A hand-held dynamometer (HHD) is a portable device that can be used to obtain more discrete, objective measures of strength during manual muscle testing (MMT) than can be achieved via traditional MMT.¹⁻⁸ Traditional MMT offers a subjective measure of strength based on a 0 to 5 scale^{3,9-11} that is largely dependent on the providers' educational and clinical experience.^{12,13}

There is a great deal of literature supporting the validity and reliability of HHD use to measure strength in cases where subjects are elderly, weak, or physically impaired in some way,^{4,7,14-17} and the testers in the given studies are physically strong, highly skilled, and experienced in HHD application.^{3,18-21} Lacking is work to support whether the former studies can be generalized to a young healthy population and possibly to practical use in a sports medicine setting with testers of wide ranging physical abilities, clinical backgrounds, and MMT or HHD application experience.

The purpose of this study is to determine the extent of intratester, intertester, and intersession reliability of HHD testing of eleven lower extremity movements

At the time of this study, the authors were with the Department of Kinesiology at the University of Virginia in Charlottesville. Brent Kelln is now with the Naval Health Clinic in Hawaii, and Patrick McKeon is now with the Department of Rehab Sciences at the University of Kentucky. E-mail: bkelln@hawaii.rr.com.

in young, healthy subjects using testers with varied levels of clinical experience and physical strength.

Methods

Subjects

Twenty subjects (Gender: 9 males and 11 females; Age: ranged from 19 to 45 with a mean of 28.1 years and a standard deviation of 7.4 years; Height: ranged from 160 to 183 cm with a mean of 170.7 cm and a standard deviation of 6.1 cm; Weight: ranged from 56.7 to 120.7 kg with a mean of 73.7 kg and a standard deviation of 15.4 kg) with healthy right lower extremities participated. None of the subjects experienced any significant complications from the study itself or from outside influences during the testing period. All twenty were able to complete both days of testing.

Instrumentation

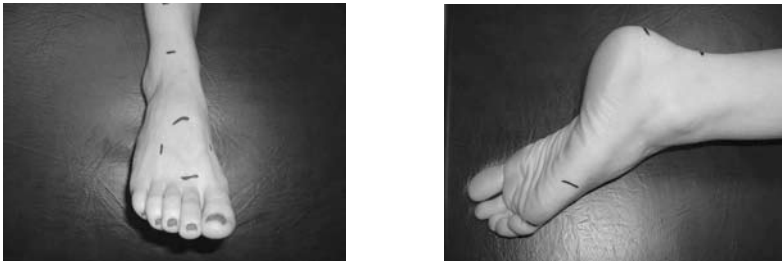
Two MicroFET 2 (Hoggan Health Industries, Inc., Draper, UT) HHDs were used in the study. Subjects were matched with a given dynamometer on day 1 of the study and followed with the same dynamometer on day 2. This was done as an extra measure of consistency, although there was no reason to suspect detectable differences from one unit to the other given that both offered equal and accurate measures during prestudy calibration. The use of two dynamometers was done simply to speed the process of data collection.

Testers

There were three testers involved with the study. Their respective backgrounds were as follows: Tester 1: a 39-year-old male physical therapist (height: 185 cm, weight: 88 kg) with 13 years of clinical experience. Tester 2: a 23-year-old female athletic trainer (height: 168 cm, weight: 59 kg) with one year of clinical experience. Tester 3: a 30-year-old male athletic trainer (height: 188 cm, weight: 93 kg) with 8 years of clinical experience. Unlike most of the previous HHD studies, a conscious effort was made to include testers with varied backgrounds and physical abilities to more thoroughly challenge the intended measures of reliability.

Procedure

Approval to conduct the research project was granted by the University of Virginia IRB. After completing the required informed consent agreement and screening form questionnaire without evidence of contraindication for participation, subjects were randomly assigned to one of two plinths along with a corresponding HHD. Subjects were then marked with a permanent felt-tip marker to ensure consistent dynamometer placement between testers and between testing days: (refer to Figures 1 and 2). The strength of eleven right lower extremity movements was then challenged isometrically by each of the three testers in turn. We chose to use the “make” rather than the “break” technique as described in previous studies.^{2,4,6,7,14,15,17–19,21–26} No significant rest periods took place between challenges provided by a given



Figures 1 and 2 — Subjects marked with a permanent marker for consistency in dynamometer placement between testers and between testing days.

tester; however, a 5 minute rest period was given during the transition from one tester to the next. Subjects were required to ramp into a 3 to 5 second maximal effort 3 times for each of the 11 movements with each of the 3 testers on both day 1 and day 2. Testers were matched to subjects in a manner that ensured application of a different tester order from day 1 to day 2. Movements tested included hip flexion, extension, abduction, adduction, internal rotation, and external rotation; knee flexion and extension; and ankle dorsiflexion, inversion, and eversion: (refer to Figures 3 through 13). As demonstrated in the figures, testing positions were chosen with the intent to offer testers the greatest possible mechanical advantage over the subjects.

Statistical Analysis

Data collected from the 11 movements tested were analyzed via intraclass correlation coefficient (ICC). Both ICC (2,1) intratester, intertester, and intersession reliability measures and ICC (2,K) intertester and intersession reliability average measures were obtained. Corresponding standard error of measurements (SEM) were also employed. SPSS version 10.1 (SPSS Inc., Chicago, IL) was used to process the data.

Results

Intratester reliability measures as indicated in Table 1 show high ICC values and low SEM values. The intratester ICC values ranged from .77 to .97 with averages from day 1 and day 2 for testers 1, 2, and 3, respectively, being .93, .91, .86 and .92, .96, .92. Meanwhile, intratester SEM values ranged from .01 kg to .44 kg. Given these results, it would appear that for a given tester, regardless of experience or physical ability, HHD is a very reliable measurement tool for repeated measures performed on the same day with the same subject.

Intersession reliability measures presented in Table 2 as one might expect, demonstrate lower, but still very good ICC values and higher SEM values than we found with the intratester evaluation. The mean intersession ICC values ranged from .62 to .92 with average mean intersession ICC values for testers 1, 2, and 3 being



Figures 3–13 — Movements tested included hip flexion, extension, abduction, adduction, internal rotation, and external rotation; knee flexion and extension; and ankle dorsiflexion, inversion, and eversion. As demonstrated in the figures, testing positions were chosen with the intent to offer testers the greatest possible mechanical advantage over the subjects.

.81, .79, and .81, respectively. SEM changes with regard to movements in which subjects had a greater mechanical and/or strength advantage over the testers (ie, ankle movements and knee extension) began to emerge in intersession reliability data when compared with the intratester data. Analysis of the intertester mean data combined with intersession mean values in Table 3 provide further illustration of differing SEM values for the more challenging movements along with declining ICC values.

Comments and Discussion

A prerequisite for quality MMT measures, and likewise HHD measures, is adequate force generating capacity by the testers performing the measurements.^{2,6,8,11,18,19,21,24,27–30} The results of this study lend clear support for this notion. As long as testers have a mechanical advantage over a subject along with the necessary strength to isometrically resist a given movement's maximal contraction, reliable measures can be expected.^{2,6,21} These reliable results with HHD appear to be achievable for a given tester regardless of experience in manual muscle testing and/or use of a HHD.^{6,8,17}

Intersession reliability results reveal that from day 1 to day 2, HHD in the hands of the same tester is still a reliable measure in a young healthy population; however, we note the size of the differences that begin to show up between testers

Table 1 Intratester Reliability ICC and SEM Values Obtained for Each Tester With Each Movement on Testing Days 1 and 2

Motion	Intratester Reliability – Day 1						Intratester Reliability Day - 2					
	Tester 1		Tester 2		Tester 3		Tester 1		Tester 2		Tester 3	
	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM
Ankle inversion	.92	.25	.90	.13	.82	.25	.94	.10	.91	.15	.90	.09
Ankle rversion	.94	.04	.93	.11	.93	.19	.94	.15	.94	.08	.91	.15
Ankle dorsiflexion	.94	.10	.94	.10	.87	.37	.92	.10	.90	.23	.89	.26
Hip adduction	.94	.07	.92	.10	.77	.15	.87	.06	.86	.05	.93	.05
Hip abduction	.96	.04	.96	.06	.92	.10	.97	.04	.97	.05	.94	.03
Hip flexion	.92	.06	.91	.09	.91	.10	.96	.05	.92	.09	.94	.11
Hip extension	.89	.06	.86	.03	.77	.06	.79	.13	.92	.09	.93	.10
Hip external rotation	.96	.04	.93	.01	.90	.13	.94	.04	.92	.05	.95	.07
Hip internal rotation	.92	.07	.91	.10	.84	.11	.96	.08	.94	.05	.95	.05
Knee flexion	.93	.08	.91	.13	.93	.05	.94	.06	.96	.03	.95	.06
Knee extension	.93	.44	.82	.32	.82	.22	.92	.21	.82	.09	.88	.12

Note. SEM values are represented in kilograms.

Table 2 Intersession Reliability and Intersession Reliability Mean ICC and SEM Values for Each Tester With each Movement

Motion	Intersession Reliability						Intersession Reliability Mean					
	Tester 1		Tester 2		Tester 3		Tester 1		Tester 2		Tester 3	
	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM
Ankle inversion	.85	.67	.84	.28	.77	.35	.85	.83	.85	.28	.78	.37
Ankle eversion	.87	.47	.88	2.23	.89	.15	.86	.62	.88	.24	.92	.11
Ankle dorsiflexion	.90	.26	.83	.29	.81	.63	.92	.27	.82	.25	.83	.63
Hip adduction	.74	.11	.70	.15	.78	.15	.67	.02	.62	.12	.80	.13
Hip abduction	.86	.07	.91	.10	.86	.10	.82	.01	.90	.10	.85	.06
Hip flexion	.75	.20	.74	.19	.79	.19	.65	.26	.66	.18	.74	.16
Hip extension	.76	.14	.77	.11	.72	.15	.79	.10	.73	.08	.69	.11
Hip external rotation	.86	.05	.85	.10	.88	.12	.82	.03	.84	.12	.89	.05
Hip internal rotation	.82	.20	.82	.24	.82	.23	.77	.24	.78	.30	.81	.29
Knee flexion	.86	.10	.91	.11	.83	.12	.84	.05	.92	.08	.78	.11
Knee extension	.89	.49	.70	.40	.77	.68	.92	.36	.70	.39	.79	.81

Note. SEM values are represented in kilograms.

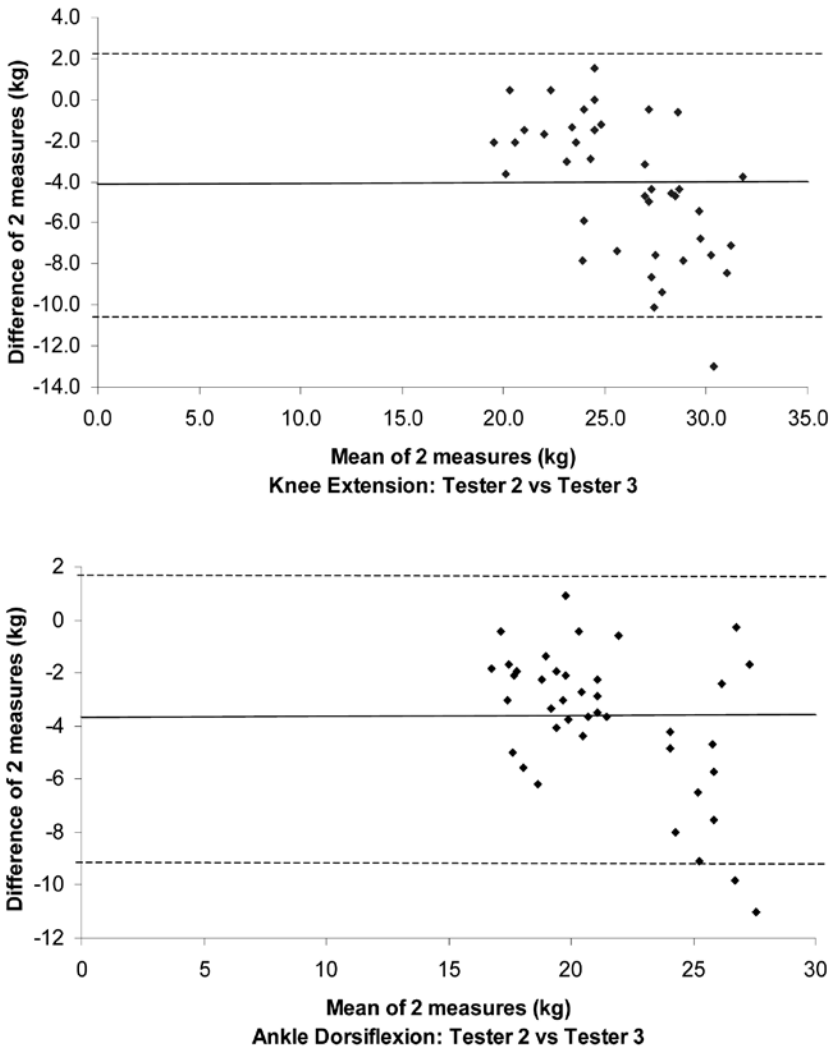
Table 3 Intertester Mean ICC and SEM Values Along With Intersession Mean Force and Standard Deviation Values Obtained for Each Tester With Each Movement

Motion	Intertester Mean		Intersession Mean Force Values		
	ICC	SEM	Tester 1	Tester 2	Tester 3
Ankle inversion	0.74	0.62	21.9 ± 5.4	19.5 ± 4.3	22.0 ± 4.4
Ankle eversion	0.84	0.52	21.8 ± 5.9	19.5 ± 4.9	22.4 ± 5.7
Ankle dorsiflexion	0.86	0.45	21.4 ± 6.7	19.5 ± 5.4	23.3 ± 6.6
Hip adduction	0.76	0.24	10.1 ± 2.3	10.6 ± 3.0	10.6 ± 3.2
Hip abduction	0.87	0.22	11.9 ± 3.3	13.0 ± 3.7	12.9 ± 3.5
Hip flexion	0.84	0.11	13.3 ± 3.8	13.1 ± 3.4	13.1 ± 3.8
Hip extension	0.65	0.64	12.2 ± 2.6	14.3 ± 2.6	12.3 ± 2.9
Hip external rotation	0.85	0.16	10.0 ± 2.7	11.0 ± 2.5	10.5 ± 3.9
Hip internal rotation	0.80	0.49	9.1 ± 3.9	9.1 ± 3.5	11.2 ± 4.4
Knee flexion	0.85	0.25	12.8 ± 2.9	13.1 ± 3.4	14.4 ± 2.8
Knee extension	0.77	1.05	28.3 ± 7.3	23.9 ± 5.0	28.1 ± 7.1

Note. SEM and mean force values are represented in kilograms. Tester 1: 39-year-old male (height: 185 cm, weight: 88 kg). Tester 2: 23-year-old female (height: 168 cm, weight: 59 kg). Tester 3: 30-year-old male (height: 188 cm, weight: 93 kg).

on the same subjects. Although intertester mean ICC values had an average of .80, all of the ankle movement and knee extension SEM values are noticeably larger than those from movements that were easy for the testers to control. Comparing the values for those same challenging movements in the intersession mean values section, we found a striking difference between tester 2 and the other 2 testers. On average, tester 2's measures for the 4 movements in question were 6 to 7.3 kg less than the other 2 testers with a range from 4 to 9.7 kg. No significant differences were found among the other, less challenging movements. Tester 1 and 3 were of similar physical ability and were likely able to consistently offer more resistance than was tester 2. Data from Table 3 allow us to identify differences between the three testers that occurred both between subjects and between testing days 1 and 2. To further illustrate this point, Bland-Altman plots (Figures 14 and 15) were generated comparing tester 2 and 3's knee extension and dorsiflexion measurement results. As force increases, a trend for greater difference between testers can be seen. With increasing level of movement difficulty, it would appear that HHD became less and less reliable.

Intratester reliability, regardless of how challenging any of the movements were, was very high in this study for all three testers. Even though tester 2 got significantly different results with the ankle and knee extension measures than did testers 1 and 3, she was consistent when compared with herself. This is an important point to consider when interpreting HHD results and may be a clinical reliability issue if difficult measures are being taken and reported by more than one clinician on the same patient.^{3,6}



Figures 14 and 15 — Bland-Altman plots were generated comparing tester 2 and 3's knee extension and dorsiflexion measurement results.

When subject strength is clearly beyond a tester's capability to control, HHD does not appear to be indicated.^{11,21,24,29,31} An example of this that we encountered during pilot work for this project involved attempts to measure plantar flexion. Despite numerous alterations in subject and tester position, we could not adequately resist plantar flexion strength to create an isometric contraction. Consistent with previous work attempting ankle movement HHD,^{3,32} our efforts to control the short

lever arm combined with the overpowering strength of the plantar flexor muscles were in vain. The same would be expected if one were to attempt using HHD to gain baseline strength measures on healthy, intercollegiate athletes. Some of the measures, where the tester had the mechanical advantage over the athlete, would work out; however, many would be questionable at best. If the person doing the testing did not give out first, the other limitation may be the dynamometer itself.^{19,21,32} In our case, the MicroFET 2 had a maximum capacity of 68 kg. None of the subjects in our study came close to reaching this maximum and if they had, there is no doubt that the testers would have failed to control such a movement.

The design of our study methods attempted to minimize the ability of subjects to overpower the testers. We were more concerned with developing positioning to allow for reliable repeated measures of muscle strength, rather than with positioning to allow the recording of absolute maximum force output. We were not interested in obtaining measures of discrete muscle contribution. By having all testers and subjects perform each measure in the same manner, we were able to adequately control for extraneous factors such as gravity. Consequently, as shown in Figures 3 through 13, we maximized lever arms for the movements being tested and positioned subjects in a manner that would both isolate the movement and offer greater mechanical advantage to the testers. In spite of these modifications, we still encountered difficulty with knee extension and the ankle movements. Worth noting at this point is that both hip internal rotation and extension intertester mean SEM values also were noticeably larger than the easy to control movements. The difference is likely due to the variability used by the subjects when performing maximal efforts on these relatively awkward movements. Hip extension made more so by the choice to maximize the lever arm with dynamometer placement over the Achilles insertion rather than at the distal posterior thigh. Many of the subjects required repeated verbal cueing to refrain from using knee flexion strategies while performing maximal hip extension. Even so, differences in hip internal rotation and extension are not evident in the intersession mean values.

Aside from limitations regarding mechanical advantage and strength, it may be worth mentioning the issue of patient comfort as a potential limitation. Even though the dynamometer is padded, it does not and cannot conform to a given body part like a tester's hand. A common subject complaint was tenderness over the dynamometer placement sites (especially after repeated trials). Consistent with the Brinkman study,²⁹ our testers experienced stabilization difficulty when attempting to control the dynamometer while resisting challenging movements in the stronger subjects. Some of the previous studies employed the use of straps, alternate positions, and/or a research assistant to help with difficult movement stabilization.^{19,25,27} Such measures seem to defeat the purpose of HHD use in favor of stationary isokinetic testing units. Another question that arises is whether the stronger subjects "let up" a bit if they felt themselves overpowering the tester.^{4,6} This may well have added to the differences noted in our intersession mean values for the more challenging movements when subjects felt themselves breaking tester 2's level of resistance. Regardless of these potential limitations, randomization of subject assignment provided adequate means of control as evidenced by the reliability data that supports their influence on our sample population as statistically inconsequential.

Conclusion

By no means do the findings of this study suggest that HHD has no place in sports medicine. One potential application could be for objective documentation of strength gains for athletes in early stages of recovery following an injury or surgical intervention. Beyond a certain point of recovery, however, such populations would likely be better served by other conventional, stationary isokinetic strength testing systems.

HHD is easy to use, portable, inexpensive^{2,13,20,24,32} and requires little training for proficient application.⁸ Evidence from this study, given the population used and modified testing procedures that were employed, suggests that the level of tester experience has little or no bearing on intratester reliability measures; however, movements with increased subject strength or mechanical advantage over testers produced weaker intertester reliability measures. Consequently, use of hand-held dynamometry as a primary strength measurement tool for healthy, strong individuals is not recommended.

References

1. Bohannon RW. Manual muscle test scores and dynamometer test scores of knee extension strength. *Arch Phys Med Rehabil.* 1986;67(6):390–392.
2. Stratford PW, Balsor BE. A comparison of make and break tests using a hand-held dynamometer and the Kin-Com. *J Orthop Sports Phys Ther.* 1994;19(1):28–32.
3. Merlini L, Mazzone ES, Solari A, Morandi L. Reliability of hand-held dynamometry in spinal muscular atrophy. *Muscle Nerve.* 2002;26(1):64–70.
4. Roebroec ME, Harlaar J, Lankhorst GJ. Reliability assessment of isometric knee extension measurements with a computer-assisted hand-held dynamometer. *Arch Phys Med Rehabil.* 1998;79(4):442–448.
5. Edwards RH, McDonnell M. Hand-held dynamometer for evaluating voluntary-muscle function. *Lancet.* 1974;2(7883):757–758.
6. Wikholm JB, Bohannon RW. Hand-held dynamometer measurements: tester strength makes a difference. *J Orthop Sports Phys Ther.* 1991;13(4):191–198.
7. Kwok CK, Petrick MA, Munin MC. Inter-rater reliability for function and strength measurements in the acute care hospital after elective hip and knee arthroplasty. *Arthritis Care Res.* 1997;10(2):128–134.
8. Bohannon RW, Wikholm JB. Measurements of knee extension force obtained by two examiners of substantially different experience with a hand-held dynamometer. *Isokinet Exerc Sci.* 1992;2(1):5–8.
9. Bohannon RW. Alternatives for measuring knee extension strength of the elderly at home. *Clin Rehabil.* 1998;12(5):434–440.
10. Bohannon RW. Measuring knee extensor muscle strength. *Am J Phys Med Rehabil.* 2001;80(1):13–18.
11. Scott DA, Bond EQ, Sisto SA, Nadler SF. The intra- and interrater reliability of hip muscle strength assessments using a handheld versus a portable dynamometer anchoring station. *Arch Phys Med Rehabil.* 2004;85(4):598–603.
12. Click Fenter P, Bellew JW, Pitts T, Kay R. A comparison of 3 hand-held dynamometers used to measure hip abduction strength. *J Strength Cond Res.* 2003;17(3):531–535.
13. Trudelle-Jackson E, Jackson AW, Frankowski CM, Long KM, Meske NB. Interdevice reliability and validity assessment of the Nicholas Hand-Held Dynamometer. *J Orthop Sports Phys Ther.* 1994;20(6):302–306.

14. Dunn JC, Iversen MD. Inter-rater reliability of knee muscle forces obtained by hand-held dynamometer from elderly subjects with degenerative back pain. *JGPT*. 2003;26(3):23–29.
15. Roy MA, Doherty TJ. Reliability of hand-held dynamometry in assessment of knee extensor strength after hip fracture. *Am J Phys Med Rehabil*. 2004;83(11):813–818.
16. Taylor NF, Dodd KJ, Graham HK. Test-retest reliability of hand-held dynamometric strength testing in young people with cerebral palsy. *Arch Phys Med Rehabil*. 2004;85(1):77–80.
17. Wang CY, Olson SL, Protas EJ. Test-retest strength reliability: hand-held dynamometry in community-dwelling elderly fallers. *Arch Phys Med Rehabil*. 2002;83(6):811–815.
18. Andrews AW, Thomas MW, Bohannon RW. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. *Phys Ther*. 1996;76(3):248–259.
19. Bohannon RW. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. *Arch Phys Med Rehabil*. 1997;78(1):26–32.
20. Visser J, Mans E, de Visser M, et al. Comparison of maximal voluntary isometric contraction and hand-held dynamometry in measuring muscle strength of patients with progressive lower motor neuron syndrome. *Neuromuscul Disord*. 2003;13(9):744–750.
21. Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session of strength assessment. *Phys Ther*. 1986;66(2):206–209.
22. Bohannon RW. Make tests and break tests of elbow flexor muscle strength. *Phys Ther*. 1988;68(2):193–194.
23. Bohannon RW, Andrews AW. Interrater reliability of hand-held dynamometry. *Phys Ther*. 1987;67(6):931–933.
24. Deones VL, Wiley SC, Worrell T. Assessment of quadriceps muscle performance by a hand-held dynamometer and an isokinetic dynamometer. *J Orthop Sports Phys Ther*. 1994;20(6):296–301.
25. Reed RL, Den Hartog R, Yochum K, Pearlmuter L, Ruttinger AC, Mooradian AD. A comparison of hand-held isometric strength measurement with isokinetic muscle strength measurement in the elderly. *J Am Geriatr Soc*. 1993;41(1):53–56.
26. van der Ploeg RJ, Oosterhuis HJ. The “make/break test” as a diagnostic tool in functional weakness. *J Neurol Neurosurg Psychiatry*. 1991;54(3):248–251.
27. Bohannon RW. Hand-held compared with isokinetic dynamometry for measurement of static knee extension torque (parallel reliability of dynamometers). *Clin Phys Physiol Meas*. 1990;11(3):217–222.
28. Bohannon RW. Hand-held dynamometer measurements obtained in a home environment are reliable but not correlated strongly with function. *Int J Rehabil Res*. 1996;19(4):345–347.
29. Brinkmann JR. Comparison of a hand-held and fixed dynamometer in measuring strength of patients with neuromuscular disease. *J Orthop Sports Phys Ther*. 1994;19(2):100–104.
30. 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(4):192–199.
31. Harlaar J, Roebroek ME, Lankhorst GJ. Computer-assisted hand-held dynamometer: low-cost instrument for muscle function assessment in rehabilitation medicine. *Med Biol Eng Comput*. 1996;34(5):329–335.
32. van der Ploeg RJ, Fidler V, Oosterhuis HJ. Hand-held myometry: reference values. *J Neurol Neurosurg Psychiatry*. 1991;54(3):244–247.