Interdevice Reliability and Validity Assessment of the Nicholas Hand-Held Dynamometer

Elaine Trudelle-Jackson, MS, LPT[1]
Allen W. Jackson, EdD[2]
Carolyn M. Frankowski, MS, LPT[3]
Kara M. Long, MS, LPT[4]
Neil B. Meske, MS, LPT[5]

The Nicholas Hand-Held Dynamometer (HHD) has been shown to have excellent interday and intraday reliability when using the same HHD. Since clinics may have more than one HHD with which to evaluate patients, it would be of value to know if two identical HHDs measure the same variable consistently. The purpose of this investigation was to assess interdevice reliability of the Nicholas HHD as well as to determine its validity. Thirty healthy female subjects between the ages of 20 and 56 years ($\bar{X}$ age = 28.4) were tested for hamstring strength. Three measurements of maximum hamstring contractions were obtained using the first HHD (Device A). The average of these three measurements was compared with the average of three measurements obtained after a brief rest using a second HHD (Device B). Measurements from the two HHDs were also compared with measurements obtained from a Kin-Com isokinetic dynamometer. The Kin-Com measurements were used as criteria to determine validity of the HHD. An intraclass correlation coefficient (ICC) calculated to determine reliability between the two HHDs was low (ICC = .58). Pearson product-moment correlation coefficients were calculated between the Kin-Com and each of the two HHDs. These values were .85 and .83 for Device A and B, respectively. Analysis of variance showed no significant difference between the Kin-Com and Device A but a significant difference between the Kin-Com and Device B ($p < .001$). Measurements obtained from two identical HHDs may be significantly different and should not be compared.

Key Words: *muscle strength, validity, Kin-Com*

[1] *Assistant Professor, School of Physical Therapy, Texas Woman's University, 8194 Walnut Hill Lane, Dallas, TX 75231*
[2] *Professor, Department of Kinesiology, Health Promotion and Recreation, University of North Texas, Denton, TX*
[3] *Physical Therapist, Irving Health Care Systems, Irving, TX*
[4] *Physical Therapist, Group I, Grand Rapids, MI*
[5] *Coordinator, Human Performance Center, St. Joseph Medical Center, Wichita, KS*

Measurement of muscle strength is a basic but integral part of the physical assessment of patients. The evaluation of muscle strength is accomplished using a number of methods, ranging from manual muscle testing (MMT) to the use of sophisticated isokinetic dynamometry. Although MMT is the oldest and most commonly used technique to assess muscle strength, it is based on a system of grading movement against the examiner's resistance or resistance provided by gravity. Several studies have been done to examine the reliability of the MMT testing technique and grading system (2,12,15,26). The authors of these studies of MMT conclude that reliable muscle grades are difficult to obtain, particularly for grades greater than fair because these grades require the examiner's subjective judgment concerning the amount of resistance applied during the test (12,26). In addition, several studies have shown that experience of the tester in performing MMT techniques can affect the consistency of muscle testing results (7,15). Bohannon and Andrews urge clinicians to seek more objective and precise methods than MMT to evaluate muscle strength in the clinical setting (6).

For more than 20 years, isokinetic dynamometers have been used in the clinic to evaluate muscle strength (19). Isokinetic dynamometers were first shown to be reliable and valid when their measurements were compared with known weights (11,18). Numerous studies have subsequently reported the consistency of isokinetic dynamometers in the evaluation of human joints, but reliability has been demonstrated primarily at the knee joint (14,25,27). Studies of the consistency of isokinetic dynamometers in the evaluation of other joints have been less conclusive (13,22). In addition, isokinetic dynamometers are not portable and, thus, are not able to evaluate bedridden patients or patients treated in the home setting.

Several investigators report the

usefulness of hand-held dynamometers (HHD) in the clinic (5,10,23). Hand-held dynamometer measurements have been shown to be significantly correlated ($p < .001$) to MMT scores, indicating that the two procedures measure the same variable of strength (4). Unlike MMT, however, determination of HHD scores does not require subjective judgment by the examiner. In addition, the HHD is portable, noninvasive, and inexpensive. The Nicholas HHD is an example of a hand-held dynamometer currently being used in the clinic to evaluate muscle strength. A study comparing Nicholas HHD measurements to isokinetic measurements was performed by Magnusson et al (16) in 1990. The authors compared measurements of shoulder abductor strength obtained using the Nicholas HHD to those obtained with a Cybex isokinetic dynamometer (Cybex, Division of Lumex, Inc., Ronkonkoma, NY). Since the HHD scores were obtained by performing a break test of the shoulder abductors, the strength measurement obtained was that of an eccentric contraction. These HHD values were then compared to concentric shoulder abductor strength values obtained on the opposite extremity measured on the Cybex at 60°/sec. Therefore, a concentric contraction measured on the Cybex at a preset constant speed was compared with an eccentric contraction of the opposite extremity measured by the Nicholas HHD (16). In the same study, Magnusson et al demonstrated that the Nicholas HHD had excellent interday and intraday reliability (16), but interdevice reliability was not investigated. If clinicians are to continue to utilize the Nicholas HHD to measure and document changes in strength in their patients, this measurement tool must be shown to be valid. In addition, since a patient may be evaluated using different Nicholas HHDs on subsequent evaluation sessions, it is critical that interdevice reliability be established. The purpose of this study was to determine the validity and interdevice reliability of the Nicholas HHD. In addition, the test-retest reliability of each of the two Nicholas HHDs used in this study was determined.

MATERIALS AND METHODS

Subjects

The subjects were 30 asymptomatic female volunteers between the ages of 20 and 56 years ($\bar{X}$ age = 28.4 years, $\bar{X}$ weight = 61.1 ± 9.6 kg) from the School of Physical Therapy at Texas Woman's University. The study was described and procedures were explained to potential subjects prior to participation. If they wished to participate after hearing the description, an informed consent form previously approved by Texas Woman's University Human Subjects Review Committee was signed. The subjects were also screened with a brief written questionnaire designed to identify current or past knee and hip pathology that would exclude the volunteers from participation in the study.

Instrumentation

Measurements of hamstring strength were obtained using two Nicholas Manual Muscle Testers (model 01160, Lafayette Instruments, Lafayette, IN). The Nicholas HHD is a digital force gauge capable of measuring forces from 0.0 to 199.99 kg. The dynamometer is calibrated at the factory, but the clinician can reset the zero value on site if necessary. The manufacturer's recommended procedure to assure that the force gauges are properly "zeroed" was performed each time prior to testing in this study.

The Kin-Com (Chattecx, Chattanooga, TN), a computerized isokinetic dynamometer, was used as a criterion measure to which the Nicholas HHD measurements were compared in order to obtain validity estimates.

Procedures

Each day, prior to testing, subjects participated in a light warm-up session on a stationary cycle followed by a hamstring stretching exercise. The subjects then performed four practice contractions of the right hamstring muscles against the HHD. These practice repetitions progressed from the subjects' perception of 25% of a maximal effort for the first repetition to 50% on the second, 75% on the third, and 100% on the final repetition. All practice contractions were performed in the prone position.

The prone testing position is the standard manual muscle testing position for the hamstrings used by Daniels and Worthingham (9). All testing with the HHD and Kin-Com was done in the standard position of prone with the knee flexed to 90°. Once the subjects were positioned, a "make" test was performed as described by Rheault et al (23), ie., a static contraction was performed as the subject exerted force on the external resistance provided by the HHD or Kin-Com force pad. In each case, the force pad was placed 2 inches proximal to the lateral malleolus on the posterior aspect of the lower leg, and the subject was instructed to gradually increase force until a maximal contraction was achieved. The muscle contraction was held for approximately 5 seconds to allow for peak force production as the examiner held the testing device and stabilized the pelvis. Each subject received the following verbal commands as the test was performed: "Pull against the device, pull as hard as you can, pull harder, relax." The force value was then read and recorded to the nearest .1 kg by a second party other than the tester.

The testing sequence took place over a 5-day period for each subject. On day 1, 3 minutes after complet-

ing the warm-up procedure, the subject performed three maximal contractions against the first HHD (Device A). These contractions were separated by 1 minute of rest. Following the third contraction against Device A, the subject rested for 5 minutes. After this rest period, the sequence was repeated, minus the warm-up, using a second HHD (Device B). The subject was then scheduled for a second day of testing 24–48 hours later.

Day 2 of testing replicated day 1 with one exception. Testing with Device B preceded Device A. The same tester did all the testing with the HHDs on both days. The subject was then scheduled for a third and final day of testing 24–48 hours later.

Day 3 again started with the same warm-up procedure. The subject was tested on the Kin-Com utilizing the same test position, device placement, and verbal coaching as when testing with the HHDs. The isometric mode of the Kin-Com software package was utilized as each subject performed three maximal contractions with 1 minute of rest between trials. The order in which the Kin-Com test and HHD test days were conducted was counterbalanced, with 15 subjects performing the Kin-Com test followed by the HHD tests and 15 subjects performing the HHD tests followed by the Kin-Com test.

Data Analysis

Means and standard deviations were calculated for all measured variables. The HHD data were analyzed with a 2 (device) × 2 (days) × 3 (trials) analysis of variance (ANOVA) with repeated measurements. Reliability estimates were determined from application of appropriate subjects by repeated measures ANOVA to average force values representative of the desired reliability comparison. Intraclass correlation coefficients (ICC) were then calculated according to Baumgartner and Jackson (1). Product moment correlations were used to provide validity coefficients for the HHDs. Fisher z transformations were used to develop .95 confidence intervals (CI) for the validity coefficients. Finally, a one-factor ANOVA with repeated measurements was used to contrast the HHD and Kin-Com force data.

RESULTS

Table 1 provides the means and standard deviations for the force (kg) data collected in the study. The 2 (device) × 2 (days) × 3 (trials) ANOVA revealed no significant interactions but a significant device main effect ($F(1,29) = 587.72$, $p < .001$) and a significant days main effect ($F(1,29) = 12.88$, $p < .001$). In examining Table 1, HHD A force values were significantly higher than the force values of HHD B. The overall effect size for this difference is a large one of 1.2 standard deviations. Day 2 produced significantly higher values for both HHDs, with low to moderate effect sizes of .27 to .45 standard deviations.

The intraclass correlations used as reliability estimates are provided in Table 2. The coefficients indicate acceptable reliability for all force values for day-to-day and trial-to-trial consistency. The reliability estimate for the HHD related to device-to-device consistency was low and provided a large standard error of measurement, based on the total observed differences between device force values of 5.12 kg. However, the correlation between the average of all values for HHD A and HHD B was .96 ($p < .001$, .95 CI = .98–.92). Thus, the measurement error was almost totally associated with the mean difference between device force values demonstrated in Table 1.

Device	Trial	Day 1		Day 2	
		$\bar{X}$	SD	$\bar{X}$	SD
HHD A	1	11.9	4.3	13.3	4.4
	2	11.7	4.3	12.9	3.9
	3	12.1	4.6	13.0	4.1
HHD B	1	6.5	4.2	8.1	4.5
	2	6.5	4.4	8.6	4.9
	3	6.7	4.5	8.6	4.2
Kin-Com	1	13.0	3.5		
	2	13.1	3.4		
	3	13.1	3.6		

HHD = Hand-held dynamometer.

TABLE 1. *Force measurements (in kg) for each testing device ($\bar{X}$ and SD).*

Condition	r_{kk}*	.95 CI†
Device	.59	.77–.32
Days HHD A	.87	.94–.74
Days HHD B	.85	.92–.71
Trials HHD A	.98	.99–.96
Trials HHD B	.97	.98–.92
Trials Kin-Com	.94	.97–.88

** The interaction, between subjects by between measures, and the between measures variation are pooled to form the error variance.*
† Confidence intervals.
HHD = Hand-held dynamometer.

TABLE 2. *Reliability estimates presented as intraclass correlation coefficients for each condition.*

The correlations between the Kin-Com force values (average across three trials) were .85 ($p < .001$, .95 CI = .93–.71) for HHD A (average across two days and three trials) and .83 ($p < .001$, .95 CI = .92–.68) for HHD B (average across two days and three trials). This indicates acceptable concurrent validity coefficients for both HHDs. The one-factor ANOVA indicated a significant ($F(2,58) = 159.31$, $p < .001$) difference in the force values between the HHDs and the Kin-Com. Post hoc contrasts within the ANOVA indicated that force values for HHD A were not significantly different from Kin-Com values, and HHD B values were significantly ($F(1,29) = 371.26$, $p < .001$) different from force values of the Kin-Com and HHD A. The Figure illustrates this difference.

DISCUSSION

The purpose of this study was to determine the test-retest reliability

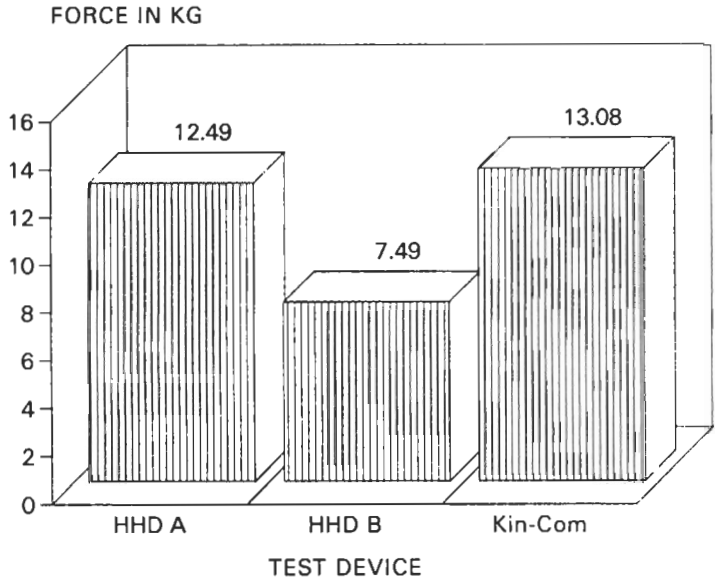

FIGURE. *A comparison of the mean force values for the three test devices.*

and validity of the Nicholas HHD as well as the interdevice reliability. Since the purpose was to evaluate the measurement tool itself, only one male tester was used to administer all tests using the HHD, and all subjects tested were female. The male tester was selected because males generally possess greater strength (20) and, thus, should be able to stabilize the dynamometer more effectively. Adequate dynamometer stabilization is particularly a consideration when testing large muscle groups such as the hamstrings (17,28). The reliability estimates obtained for day-to-day and trial-to-trial consistency for each of the Nicholas HHDs indicates good to high reliability as defined by the rating scale developed by Currier (8). These results are consistent with the findings of previous intrarater reliability studies of HHDs (5,6,21,24).

The calculated interdevice reliability estimate for the Nicholas HHD was poor as defined by the rating scale developed by Currier (8). The correlation between the values obtained for HHD A and HHD B was high, but Device A consistently obtained higher peak force readings. Therefore, although there was a strong linear relationship between the devices, the differences in force readings between the devices resulted in a low reliability estimate.

Potential sources of error between the two HHDs include differences in force output due to familiarity of the subject to the testing device and procedure. However, counterbalancing the order in which the subjects were tested with each device should have minimized this source of error. In addition, only one tester was used in order to eliminate differences due to the tester rather than the device.

It is important to note that although inconsistency between the two HHDs studied was found, the inconsistency may not represent other Nicholas HHDs or other currently available HHDs. Since there are no known published studies investigating the interdevice reliability of HHDs, the results of this study could not be compared to others. There is a possibility that the measurement error between the two Nicholas HHDs evaluated in this study is unique to these two units. However, until further studies of interdevice reliability are conducted using additional Nicholas HHDs and other HHDs, these devices should be used carefully when evaluating strength in patients. When clinicians do utilize HHDs to assess a patient's strength, the same device should be used consistently with that patient.

A comparison of the HHD force values to values obtained on the Kin-Com revealed that the Nicholas HHD is a valid measurement tool for assessing hamstring strength when compared to a criterion device, such as the Kin-Com. Thus, the high correlation obtained between each of the HHDs and the Kin-Com indicates that both Devices A and B were capable of detecting differences in muscle strength. This finding is consistent with a previous study conducted by Bohannon which demonstrated that an Ametek Acuforce II HHD compared favorably with a Cybex II dynamometer in the measurement of isometric knee extensor force (3).

When the means for each of the Nicholas HHDs were compared individually to Kin-Com force data, a significant difference was found between values from HHD B and the Kin-Com, but no significant difference existed between HHD A and the Kin-Com.

CONCLUSION

The results of this study show the Nicholas Manual Muscle Tester to be valid and highly reliable for testing between trials and days. Interdevice reliability between two Nicholas HHDs, however, was shown to be poor. Until further studies are conducted to determine whether the results of this study are applicable to other Nicholas HHDs and other HHDs in general, clinicians should be diligent in assuring that their HHDs are clearly marked so that the same device is used consistently to evaluate a given patient.

JOSPT

REFERENCES

1. *Baumgartner T, Jackson A: Measurement for Evaluation in Physical Edu-*

cation and Exercise Science (4th Ed), Dubuque, IA: W.C. Brown, 1991
2. Beasley WC: Quantitative muscle testing: Principles and applications to research and clinical services. Arch Phys Med Rehabil 42:398–425, 1961
3. Bohannon RW: Hand-held compared with isokinetic dynamometry for measurement of static knee extension torque (parallel reliability of dynamometry). Clin Phys Physiol Meas 11(3):217–222, 1990
4. Bohannon RW: Manual muscle test scores and dynamometer test scores of knee extension strength. Arch Phys Med Rehabil 67:390–392, 1986
5. Bohannon RW: Test-retest reliability of hand-held dynamometry during a single session of strength assessment. Phys Ther 66(2):206–209, 1986
6. Bohannon RW, Andrews AW: Accuracy of spring and strain gauge hand-held dynamometers. J Orthop Sports Phys Ther 10(8):323–325, 1989
7. Bohannon RW, Andrews AW: Interrater reliability of hand-held dynamometry. Phys Ther 67(6):931–933, 1987
8. Currier DP: Elements of Research in Physiotherapy (3rd Ed), Baltimore: Williams & Wilkins, 1990
9. Daniels L, Worthingham C: Muscle Testing: Techniques of Manual Examination (5th Ed), Philadelphia: W.B. Saunders Company, 1986
10. Edwards RH, McDonnell M: Hand-held dynamometry for evaluating voluntary muscle function. Lancet 757–758, 1974
11. Farrel M, Richards JG: Analysis of the reliability and validity of the Kinetic Communicator exercise device. Med Sci Sports Exerc 16(1):44–49, 1986
12. Frese E, Brown M, Norton BJ: Clinical reliability of manual muscle testing. Phys Ther 67(7):1072–1076, 1987
13. Griffin JW: Differences in elbow flexion torque measured concentrically, eccentrically and isometrically. Phys Ther 67:1205–1208, 1987
14. Johnson J, Siegel D: Reliability of an isokinetic movement of the knee extensors. Res Q 49:88–90, 1978
15. Lilienfield AM, Jacobs M, Willis M: A study of the reproducibility of muscle testing and certain other aspects of muscle scoring. Phys Ther Rev 34(6):279–289, 1954
16. Magnusson SP, Gleim GW, Nicholas JA: Subject variability of shoulder abduction strength testing. Am J Sports Med 18(4):349–352, 1990
17. Marino M, Nicholas JA, Gleim GW, Rosenthal P, Nicholas SJ: The efficacy of annual assessment of muscle strength using a new device. Am J Sports Med 10(6):360–364, 1982
18. Moffroid M, Whipple R, Hofkosh J, Lowman E, Thistle H: A study in isokinetic exercise. Phys Ther 49:735–746, 1969
19. Montgomery LC, Douglass LW, Deuster PA: Reliability of an isokinetic test of muscle strength and endurance. J Orthop Sports Phys Ther 10(8):315–322, 1989
20. Morrow JR, Hosler WW: Strength comparisons in untrained men and trained women athletes. Med Sci Sports Exerc 13:194–198, 1981
21. Nies Byl N, Richards 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 9(12):399–405, 1988
22. Perrin DH: Reliability of isokinetic measures. Athl Train 21:319–321, 1986
23. Rheault JR, Beal JL, Kubik KR, Shepley JA: Intertester reliability of the hand-held dynamometer for wrist flexion and extension. Arch Phys Med Rehabil 70:907–910, 1989
24. Riddle DL, Finucane SD, Rothstein JM, Walker ML: Intrasession and intersession reliability of hand-held dynamometer measurements taken on brain-damaged patients. Phys Ther 69(3):182–189, 1989
25. Tredinnick TJ, Duncan PW: Reliability of measurements of concentric and eccentric isokinetic loading. Phys Ther 68(5):656–659, 1988
26. Wadsworth CT, Krishnan R, Sear M: Intrarater reliability of manual muscle testing and hand-held dynametric muscle testing. Phys Ther 67(9):1342–1347, 1987
27. Wessel J, Gray G, Luongo F, Isherwood L: Reliability of work measurements recorded during concentric and eccentric contractions of the knee extensors in healthy subjects. Physiother Can 41:250–253, 1989
28. Wikhom JB, Bohannon RW: Hand-held dynamometer measurements: Tester strength makes a difference. J Orthop Sports Phys Ther 13(4):191–198, 1991