Interrater Reliability of Hand-Held Dynamometry

RICHARD W. BOHANNON
and A. WILLIAMS ANDREWS

Two raters performed hand-held dynamometer testing of six muscle groups of 30 patients to determine the interrater reliability of the procedure. The six muscle groups tested were the shoulder external rotators, elbow flexors, wrist extensors, hip flexors, knee extensors, and ankle dorsiflexors. The patients were of eight different diagnostic groups, with most having hemiparesis secondary to cerebrovascular accidents. We used Pearson product-moment correlations t tests to compare the two raters' scores. The correlation coefficients ranged from .84 to .94 (p < .001). Only two muscle groups (ie, shoulder external rotators and wrist extensors) had mean scores that differed (p < .05) between raters. Although the interrater reliability of the procedure was found to be good to high in the six muscle groups tested, the results of the t tests indicate that hand-held dynamometry should undergo further evaluation.

Key Words: Muscle contraction, Muscles, Physical therapy.

For more than 40 years, manual muscle testing has been an integral part of the physical therapy assessment of patients with neurological impairments.\(^1\) Even before manual muscle testing was practiced widely, however, Lovett and Martin expressed dissatisfaction with the imprecision of subjective estimates of muscle strength.\(^2\) By 1956, Beasley had demonstrated objectively that manual muscle testing is imprecise and that hand-held force gauges and tensiometers can be used to gain a more exact indication of muscle group strength.\(^3\)

Since Beasley's classic work,\(^4\) other studies of hand-held dynamometry,\(^5\)\(^6\) including one involving patients with hemiplegia, have been published. Nevertheless, the reliability of hand-held dynamometry has not been investigated thoroughly. Clarke reported the correlation between muscle strength scores obtained by different examiners using a hand-held Newman myometer to be .82 for finger flexion and .79 for wrist dorsal flexion.\(^8\) Hosking and associates described the reproducibility of myometer muscle strength measurements in young children as "good."\(^9\) They reported that the variation of repeated strength measurements of six muscle groups at one-month intervals rarely exceeded 15% of the initial values. Wiles and Karni reported coefficients of variation of 3.6% to 27.3% for five examiners' myometer measurements of five patients over a 24-hour period.\(^6\) They wrote further that a single examiner measured the same muscle group of three patients on five occasions and obtained a mean coefficient of variation of 8.9%.

Because all of these studies of reliability incorporated "break" tests in which the examiner pushes against the subject's maximal attempt to "hold" a position by contracting a muscle group, their results may not be applicable to the interpretation of "make" tests. In the make test, the dynamometer is held stationary while the subject exerts a maximal force against it with a muscle group. Bohannon reported test-retest correlations for 18 different muscle groups that he tested with a hand-held dynamometer using make tests.\(^9\) The mean of these correlations was .97. To the best of our knowledge, however, no full-length report of the interrater reliability of hand-held dynamometry has been published in which make tests were used.

The purpose of this investigation was to determine the interrater reliability of hand-held dynamometry when make tests are performed on three upper extremity and three lower extremity muscle groups. We anticipated that interrater reliability would be good, but lower than the test-retest reliability reported by Bohannon.\(^9\)

METHOD

Subjects

Thirty patients (18 women and 12 men), who provided informed consent, participated as subjects. The only criteria used in the subject selection were that the patients were assigned to one of the two authors, that the subjects could follow the testing instructions accurately, and that they had measurable strength in the muscle groups being tested. Twenty-one of the patients had suffered cerebrovascular accidents (CVAs), and 3 subjects had spinal cord involvement. One of each of the remaining patients had an amputation, fracture, multiple traumas, Guillain-Barré syndrome, muscular dystrophy, or generalized weakness.

Procedure

The muscle groups of the right side of the body were tested in 17 patients. The muscle groups of the left side of the body were tested in 13 patients. Seven of the patients with CVAs were tested on the paretic side. The therapist to whom a patient was assigned for treatment tested the subject first. Thus, one therapist (R.W.B.) tested 20 subjects first, and the other therapist (A.W.A.) tested 10 subjects first. Each tester performed a single test of each muscle group, but retested the subject if he believed the results to be unrepresentative of a "best" effort. Each tester performed the tests independently and was blind to the results of the other tester. Each rater's tests were

Mr. Bohannon is Chief, Department of Physical Therapy, Southeastern Regional Rehabilitation Center, Cape Fear Valley Medical Center, PO Box 2000, Fayetteville, NC 28302 (USA).

Mr. Andrews is Staff Physical Therapist, Southeastern Regional Rehabilitation Center.

This article was submitted March 3, 1986; was with the authors for revision two weeks; and was accepted May 21, 1986. Potential Conflict of Interest: 4.
performed within five minutes of the other's.

We tested six muscle groups in a gravity-eliminated position, using a calibrated digital Chatillon® force gauge. This gauge has a range of 0 to 115 lb and measures to the nearest 10th of a pound. Only the knee extensor muscles were tested with the subjects seated on the edge of a padded mat table. We tested the other muscle groups, (shoulder external rotators, elbow flexors, wrist extensors, hip flexors, and ankle dorsiflexors) with the patients positioned supine on the mat table. In general, muscle groups were tested in the middle half of their range of motion. Pre-established techniques were used to ensure consistent dynamometer placements, limb positions, joint angles, and stabilization.5,9 Before testing, we demonstrated the appropriate muscle movements to the subjects. Patients were instructed to increase their effort to maximum over a period of 1 to 2 seconds. During each of the 4- to 5-second make tests, consistent verbal encouragement was provided to the subjects.

**Data Analysis**

The dynamometer test score, or the amount of force exerted on the dynamometer that was held by the examiners, was recorded and converted to kilograms. Descriptive statistics (means and standard deviations) were calculated for each rater's scores from all subjects for each muscle group. The rater's scores for each muscle group were compared using Pearson product-moment correlations.11 Thereafter, t tests for correlated samples were calculated to ensure that correlations did not provide, because of systematic variation, a false indication of interrater reliability.

**RESULTS**

The means and standard deviations of the dynamometer scores obtained by the two raters are reported in the Table. The correlations between the two raters' scores ranged from .84 to .94 (p < .001). The largest absolute difference between the raters' mean scores was 1.2 kg. The largest percentage difference between means (expressed in relation to the lower mean) was 11.0%. The means of two raters differed significantly for the shoulder external rotator muscles (p = .03) and wrist extensor muscles (p < .01) only.

**DISCUSSION**

As expected, the correlations between the raters' measurements in this study are consistent with good to high reliability.12 Moreover, the correlations between raters were higher than those reported for the finger flexion and wrist extension movements by Clarke.8 Possible explanations for such a difference might be the use of different test types (make vs break), different testing instruments, a potential difference between the levels of experience of the testers, and differences among the individuals tested. We used a make test and an electronic dynamometer rather than a break test and a spring dynamometer. The raters in our study had considerable experience, one having performed hundreds and the other thousands of dynamometer tests. Furthermore, each rater in this study already had demonstrated previously (unpublished data) a capacity to use the hand-held dynamometer reliably (Pearson product-moment correlations between test-retest scores of ≥ .94). We tested patients with a range of disorders and muscle strengths. Most of the patient types tested in this study have been tested with dynamometers by others.3-7 Most of the patients tested probably had less muscle strength than healthy volunteers. Thus, stabilization and holding against the patient's maximal effort may have been easier than holding against that of a healthy volunteer. This factor could result in higher reliability correlations.

The relationships between raters' scores in this study were lower than the test-retest reliability measurements reported for a single rater.9 Using another rater for a second testing introduces another source of variance that can be expected to reduce the correlation between tests.

Given the many potential sources of error and variance in dynamometer testing, we believe the correlations obtained between raters in this study to be encouraging as to the clinical potential of the procedure. Because significant differences were found between raters' mean scores for the shoulder external rotator and wrist extensor muscle groups, further analysis of our testing procedure is indicated. Perhaps the averaging of several test scores13 or further test standardization will improve inter-rater reliability. Further reliability testing that will involve more than two testers,14 different instruments, and additional muscle groups is recommended.

**CONCLUSION**

Correlations ranging from good to high were found between two testers' dynamometer scores of six muscle groups of 30 patients. Differences in the testers' means for two muscle groups indicate a need for further evaluation of hand-held dynamometry as a clinical procedure.

---

**TABLE**

Comparison of Hand-Held Dynamometer Scores Obtained from Six Muscle Groups of 30 Patients By Two Raters

<table>
<thead>
<tr>
<th>Muscle Group</th>
<th>Dynamometer Scores (kg)</th>
<th>Pearson Product-Moment Correlation*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tester R.W.B.</td>
<td>Tester A.W.A.</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>s</td>
</tr>
<tr>
<td>Shoulder external rotators</td>
<td>7.3</td>
<td>3.9</td>
</tr>
<tr>
<td>Elbow flexors</td>
<td>12.6</td>
<td>6.9</td>
</tr>
<tr>
<td>Wrist extensors</td>
<td>8.4</td>
<td>3.9</td>
</tr>
<tr>
<td>Hip flexors</td>
<td>6.6</td>
<td>3.8</td>
</tr>
<tr>
<td>Knee flexors</td>
<td>18.4</td>
<td>9.4</td>
</tr>
<tr>
<td>Ankle dorsiflexors</td>
<td>13.9</td>
<td>7.9</td>
</tr>
</tbody>
</table>

* For all correlations, p < .001.

---

**Footnotes:**

† 1 lb = 0.4536 kg.
REFERENCES


