Reliability of a New Device Used to Measure Shoulder Subluxation

The purposes of the study were to determine the reliability of a new device used to quantify shoulder subluxation and to estimate its standard error of measurement. *The device is an L-shaped thermoplastic jig with a metric tape measure embedded* in it. A sliding beak-like marker, which can be anchored with a thumbscrew, is used to identify landmarks and to measure the amount of subluxation. Eight male and two female hemiplegic subjects, 40 to 80 years old, consented to be measured for subluxation. Three standardized subluxation measurements were taken by one investigator to determine the reliability with a single rater. One measurement was taken by a second investigator and compared with the first measurement obtained by the first investigator to determine the reliability using more than one rater. Both investigators were experienced physical therapists. Each measurement was read by the other investigator, who disassembled the jig and cleaned the marks from the patient between measurements. For both analyses, an analysis of variance for repeated measures reflected no differences between measurements attributable to raters. The unbiased intraclass correlation coefficent for a single measurement by a single rater was .89 (p < .01) and for more than one rater was .74 (p < .01). The standard error of measurement was \pm 0.77 mm for a single rater and \pm 1.20 mm for more than one rater. We recommend the jig as a tool to measure shoulder subluxation in patients. [Hayes KW, Sullivan JE: Reliability of a new device used to measure shoulder subluxation. Phys Ther 69:762-767, 1989]

Key Words: *Hemiplegia, evaluation; Tests and measurements, general; Upper extremity, shoulder.*

Shoulder subluxation occurs in 33% to 75% of patients with hemiplegia.^{1–5} Usually the humerus displaces inferiorly and anteriorly⁶ during a period when shoulder muscles are flaccid.³ The problem has been associated with pain, brachial plexus and soft

tissue injuries, limited mobility, and impairment of the use of the hand. These problems have significant impact on the physical, functional, and psychosocial rehabilitation of the patient with hemiplegia. Many interventions have been suggested to

K Hayes, MA, PT, is Assistant Professor, Programs in Physical Therapy, Northwestern University Medical School, 345 E Superior St, Rm 1323, Chicago, IL 60611 (USA).

J Sullivan, MS, PT, is Instructor in Clinical Physical Therapy, Programs in Physical Therapy, Northwestern University Medical School. She was Assistant Director, Physical Therapy Department, Rehabilitation Institute of Chicago, 345 E Superior St, 12th Floor, Chicago, IL 60611, when this study was conducted.

The results of this study were presented in poster format at the Joint Congress of the American Physical Therapy Association-Canadian Physiotherapy Association, Las Vegas, NV, June 12–16, 1988.

This article was submitted September 12, 1988; was with the authors for revision for 10 weeks; and was accepted March 21, 1989.

Karen W Hayes Jane E Sullivan

decrease the subluxation, but until recently, the efficacy of those interventions has been difficult to determine because adequate instrumentation to measure the subluxation has not been available.

Clinically, subluxation has been measured by gross descriptive means, such as finger breadths between the acromion and the head of the humerus. Because of variations in investigator finger positioning and size and the crude measurement interval, this method is too subjective for use in clinical evaluation research.

Ritt et al developed a jig to quantify the amount of subluxation at the

shoulder.⁷ They validated the jig using roentgenographic measurements and reported a Pearson product-moment correlation coefficient of .826 as a measure of validity. Interrater agreement as a measure of reliability was also determined using a Pearson product-moment correlation coefficient and ranged from .796 to .995, depending on trial and extremity tested (B Ritt Myers, unpublished data, 1989).

The Pearson product-moment correlation coefficient is hardly an appropriate statistic to use to determine reliability. It represents the degree by which two measures covary rather than agree and is insensitive to systematic covariation. In addition, the Pearson product-moment correlation coefficient is not a measure of reliability because it is not a comparison of variances. By definition, reliability of a measure is the proportion of the variance of the true score, free from error, compared with the variance of the total score. Therefore, the intraclass correlation coefficient (ICC), derived from an analysis of variance (ANOVA), has been used as a more appropriate statistic.8,9

A valid, reliable, interval-level measure is needed to determine the efficacy of therapy in clinical practice and research. The jig developed by Ritt et al⁷ is valid, inexpensive, and easy to use. However, its reliability has not been adequately determined. The purposes of this study were 1) to determine the reliability of the jig with a single rater and with more than one rater and 2) to estimate its error of measurement.

Method

Subjects

Eight male and two female patients with hemiplegia, between 40 and 80 years of age, consented to be measured for subluxation. Nine subjects were hemiplegic secondary to a cere-



Fig. 1. Thermoplastic jig developed by Ritt et al⁷ to measure shoulder subluxation.

brovascular accident (CVA), and one subject was hemiplegic secondary to a brain tumor. All subjects had anteriorly displaced shoulder subluxation, with no isolated movement in the involved upper extremity. All subjects were cognitively intact and cooperated with the procedure.

Instrumentation

The jig* is an L-shaped device constructed of thermoplastic material with a 21-cm tape measure, visible from only one side, embedded in it. A sliding beak-like marker, which can be anchored with a thumbscrew, is used to identify landmarks and to compute measurements (Fig. 1).

Procedure

Both evaluators were graduate physical therapists with more than seven years of clinical experience. The principal investigator (KWH) was unfamiliar with the instrument and practiced with it prior to this study. The second investigator (JES) had used the jig many times previously.

All shoulder subluxation measurements were made in the Physical Therapy Department, Rehabilitation Institute of Chicago (Chicago, Ill). All subjects were seated erect in their wheelchairs with the armrest on the involved side removed. Subjects were instructed to relax during the procedure. Prior to taking any measurements, two landmarks were located and marked. First, the investigator marked the location of the subject's acromion with a pen. Then, after passively flexing the subject's elbow, the investigator placed the short leg of the jig under the subject's elbow and made a dot through a hole in the long arm of the jig 20 cm above the subject's olecranon (Fig. 2). This dot



Fig. 2. Acromion identified with an ink mark; jig used to locate a mark 20 cm above olecranon.

Physical Therapy/Volume 69, Number 9/September 1989



Fig. 3. Short leg of jig placed on acromion; beak moved to dot on arm.

and the mark at the acromion served as landmarks for all future measurements. To take the actual measurements, the subject's elbow was first extended and allowed to hang freely at his or her side. The investigator then placed the short leg of the jig on the acromion mark, moved the beak to the dot on the arm, and fixed the beak (Fig. 3). The point at which the slide portion of the beak rested on the tape measure was read by the other investigator and recorded in millimeters. The first investigator then flexed the subject's elbow, and the other investigator manually reduced the subluxation to the point at which no further reduction was palpable. We were careful not to shift the skin over the acromion during the reduction. The measurement of the distance between the acromion and the dot on the arm was then repeated (Fig. 4). The amount of subluxation was the difference, in millimeters, between the two measurements. Four measurements were taken for each



Fig. 4. Subluxation manually reduced; measurement in Figure 3 is repeated.

subject, one by the second investigator and three by the principal investigator.

The design of the jig is such that the tape measure is facing away from the investigator during measurement, thus eliminating error attributable to knowledge of the previous measurement. One investigator dissassembled the jig before each measurement and handed it to the other investigator, who then performed the measurement and handed it back to the first investigator for reading and recording. The second investigator also removed the acromion and arm markings before another measurement was made. The jig is calibrated in millimeters, and we read each measurement to the nearest one-half millimeter.

Data Analysis

To test the hypothesis that there were no differences among the measurements when a single rater used the jig, we performed an ANOVA for repeated measures on the principal investigator's three measurements using a Statview 512+ software program[†] on a Macintosh SE computer.[‡] We calculated the unbiased ICC as a measure of the reliability of the instrument. The unbiased estimate of reliability of a single measurement is determined by the following formula¹⁰:

Reliability = $\frac{BSMS - resMS}{BSMS + (m-1)(resMS)}$

where BSMS is the between-subjects mean square, resMS is the residual mean square, and m is the number of measurements. The ICC can be calculated to reflect the reliability of the aggregate of the measurements or of a single measurement and can either include or exclude the therapist as a source of error. If the reliability of the aggregate of the measurements were of interest, the total number of measurements would be included. However, in physical therapy, individual therapists perform single measurements of patients to document a patient's condition; therefore, the reliability of a single measurement is of interest. To determine the single measurement reliability, the average reliability is calculated. If the same therapist performs all measurements, the residual mean square is used as the error term.

To test the hypothesis that there were no differences between measurements obtained by different raters, the single measurement obtained by the second investigator and the first measurement obtained by the primary investigator were compared using an ANOVA for repeated measures. The first measurement obtained by the primary investigator was selected to minimize the effect of handling the subject and thus affecting his or her muscle tone. We used the formula for a single measurement in calculating the ICC but included the rater in the error because different raters were using

[†]BrainPower Inc, 24009 Ventura Blvd, #250, Calabasas, CA 91302.

[‡]Apple Computer, Inc, 20525 Mariani Ave, Cupertino, CA 95014.

 Table 1.
 Raw Data for Amount of

 Subluxation (in Millimeters) Recorded by

 Each Investigator^a

Subject	JES	KWH1	KWH2	КЖНЗ
1	3.0	3.0	3.0	3.0
2	3.0	6.0	10.0	5.0
3	0.0	0.0	0.0	0.0
4	3.5	3.0	3.0	1.0
5	7.5	10.0	8.5	12.0
6	2.0	7.0	5.5	6.0
7	7.0	7.5	7.0	9.0
8	-1.0	-1.0	-1.0	-0.5
9	4.0	3.0	2.5	3.0
10	5.0	9.0	9.0	12.0

"JES = single measurement obtained by second investigator; KWH1 = first measurement obtained by primary investigator; KWH2 = second measurement obtained by primary investigator; KWH3 = third measurement obtained by primary investigator.

the jig. To include the rater in the error, the within-subjects mean square (WSMS) is used as the error term. In both analyses, we tested the ICC for significance using an alpha level of .05.^{11,12}

As an indicator of the error of the instrument, we calculated the standard error of measurement (SEM) and established the 95% confidence interval for a single measurement by a single rater and by more than one rater.

Results

The raw data are shown in Table 1. The results of the ANOVAs are shown in Tables 2 and 3. Table 2 demonstrates a significant difference between the measurements of the patients attributable to the variability in the subjects' conditions but no significant difference between the measurements of a single rater. Table 3 indicates a significant difference between measurements attributable to the subjects' conditions but no significant difference between the two raters' measurements. The reliability of the instrument for a single measurement by a single rater was .89 (p < .01) and by more than one rater was .74 (p < .01). The SEM for a single mesaurement by a single rater was ± 0.77 mm. The 95% confidence interval was ± 1.51 mm; therefore, the true measurement will be within ± 1.51 mm of the observed measurement 95% of the time. The SEM for a single measurement by more than one rater was ± 1.20 mm; the 95% confidence interval was ± 2.36 mm.

Discussion

The purposes of this study were to determine the reliability of the jig and to estimate its measurement error. To arrive at a true reliability estimate for an instrument, the error estimate must be reduced optimally. We controlled the following potential sources of error in the design of this study: 1) We standardized the method of measurement and physically marked the landmarks, 2) we were conscious of end-digit preference8 and read the scale to the nearest one-half millimeter, and 3) we blinded the measuring investigator by having the other investigator read the measurement.¹³ The same instrument was used for all measurements.

Potential sources of error still remain in this study and include biological

Table 2. Analysis-of-Variance Summary for Repeated Measures for ThreeMeasurements Obtained by First Investigator (KWH)⁴

Source	df	SS	MS	F
Between subjects	9	419.08	46.56	28.66 ^b
Within subjects	20	32.50	1.62	
Treatments	2	0.60	0.30	0.17
Residual	18	31.90	1.77	
TOTAL	29	451.58		

^{*a*}Reliability estimate for single treatment = .89 (p < .01).

^{*b*}*p* < .0001.

Table 3. Analysis-of-Variance Summary for Repeated Measures for Measurement Obtained by Second Investigator(JES) and for First MeasurementObtained by First Investigator (KWH)

Source	df	SS	MS	F
Between subjects	9	170.76	18.97	6.57ª
Within subjects	10	28.88	2.89	
Treatments	1	9.11	9.11	4.15
Residual	9	19.76	2.20	
TOTAL	19	199.64		

 ${}^{a}p < .01$; reliability estimate for single treatment = .74 (p < .01).

variation among and within the subjects, failure to control head position, shifting of the skin during measurement, variation in amount of force applied during reduction of the subluxation, distractions occurring during measurement, and instability in the jig. Despite these sources of error, however, the reliability coefficient was adequately high to allow reasonable confidence that measurements taken with the jig can be accurate, reliable, and clinically relevant, especially if measured serially by a single therapist.

Although repeated-measures analysis can remove variability between subjects, biological variation within subjects is difficult to control. All subjects were hemiplegic as a result of CVA except one who had a brain tumor (Subject 8). This subject appeared to have subluxation but seemed to sublux further during a manual reduction, producing a negative value. We are unable to explain this phenomenon. Subject 3 also appeared to have subluxation, but reduction was not measurable. We are uncertain whether reduction was prevented by connective tissue stiffness or some other factor. In either of these subjects, if the shoulder subluxed primarily in an anterior direction with little inferior component, reduction

would not be measurable with this instrument.

One subject's arm was too long for the jig, and we could not use the dot through the jig as the distal landmark. We, therefore, simply made a dot more proximally on his arm. Because the distal mark is arbitrary and the measurement used in the data analysis is a difference score using the marked distal landmark both times, altering the landmark for this subject did not introduce substantial error.

Changes in the subject as a result of the measurement itself could not be controlled. Handling the subject during the process of marking landmarks and taking measurements may have caused increases or decreases in muscle tone that could affect the amount of subluxation. Allowing a longer rest period between successive measurements might have decreased the variation between successive measurements, but we could not avoid any alteration of tone during any single measurement. This effect may have occurred with Subjects 2 and 5. In addition, subject motivation may have altered tone. The subjects displayed no isolated motion but did have the ability to contract the proximal upper extremity musculature. Even though the subjects were instructed to relax, any intentional or unintentional contraction of these muscles could have affected the measurement.

Different head positions also produce changes of muscle tone. Head position was uncontrolled in this study but probably varied little. Most subjects had a tendency to watch the procedure, which caused them to rotate their heads toward the side of involvement.

Even though we standardized the method of measurement, the skin shifted slightly during the manual reduction of the subluxation, which moved the landmark at the acromion. Although we made every effort to avoid this shift, it could have contributed some error. Marking the acromion location again after reduction might have reduced the error introduced from this cause.

To decrease error from the application of varying amounts of force in manually reducing the subluxation, we reduced all shoulder subluxations to the point at which we could feel no further reduction. However, each investigator reduced the subluxation for the other. Forces could have varied as well as the amount of skin shift. In our study, a third investigator should have reduced the subluxation in all subjects for both of the measuring investigators.

Environmental variations during the session may have affected the subjects' muscle tone and thus the amount of subluxation. Temperature and lighting remained constant during a session but varied between subjects. Because subjects were measured in the Physical Therapy Department, distractions could not be avoided.

The jig prototype we used had some instability in the slide. We made every effort to keep the slide level, but error could have been introduced. This design error has been corrected in later models of the jig. In addition, this jig is capable of measuring only the inferior component of the subluxation. An instrument capable of measuring the anterior component would provide additional useful information.

The 0.77- and 1.20-mm SEMs probably are overestimates. They may well have been distorted by the small sample size in this study. In addition, accurate measurement of subluxation requires that the sources of error discussed be controlled. Researchers should ensure an adequate number of subjects to diminish the effects of a less-thanperfect reliability. Clinicians should recognize the error when interpreting a patient's condition. However, the fairly small SEM for a single rater offers reassurance of accurate measurements. Clinicians can feel confident using the jig to document patient progress if a single rater uses the instrument.

Several investigators have studied reliability of goniometric measurements and have challenged the previously held position that the average of several measurements is necessary for accuracy.9,14 They found that the reliability of single measurements was adequately high. That finding is supported by this study. The ICCs of the measurement obtained by the second investigator against each measurement obtained by the first investigator (.74, .64, and .66) were compared with the ICC calculated for the measurement obtained by the second investigator and the mean of measurements obtained by the first investigator (.71). Using the mean improves the ICC by a maximum of .07 and reduces it in one instance, by .03. The mean of several measurements does not change the reliability estimate substantially.

Some investigators have also suggested that because intrarater agreement is higher than interrater agreement, a single evaluator should be used in studies involving repeated measures.^{9,14} This finding is strongly supported by the present study. The reliability using one rater was .89; that using more than one rater was .74. Therefore, for clinical intervention studies, we suggest a single evaluator.

Future studies of this nature would be more meaningful if designed as generalizability studies, which allow partitioning of all the sources of variability including subjects, raters, occasions, and all interactions to determine their relative influences.^{15,16} Each rater should measure the subject the same number of times, and the order of raters should be randomized. A study designed in this way can provide an estimate of interrater and intrarater agreement in addition to an overall G coefficient, which is an ICC describing the reliability of the data.

Conclusions

We recommend the jig for use as an evaluation tool to provide accurate and reliable measurements of shoulder subluxation in patients. Physical therapists who use the jig should recognize that the jig is useful only in the measurement of subluxed shoulders that can be reduced manually. In addition, practitioners who use the jig should ensure a standardized procedure and should attempt to control additional sources of error such as head position, skin shift, and force. Measurements obtained by a single rater provide a more accurate estimate than those obtained by more than one rater. For both clinical decision making and clinical efficacy studies related to shoulder subluxation, the jig provides a measure that allows documentation of the change in subluxation.

Acknowledgments

We would like to thank Judith Falconer, PhD, for her careful review of the manuscript and Ellen Perlow, PT, for her assistance with the photography.

References

1 Meredith J, Taft G, Kaplan PL: Diagnosis and treatment of the hemiplegic with brachial plexus injury. Am J Occup Ther 35:656–660, 1981

2 Miglietta O, Lewitan A, Rogoff JB: Subluxation of the shoulder in hemiplegic patients. NY State J Med 59:457–460, 1959

3 Moskovitz H, Goodman CR, Smith E, et al: Hemiplegic shoulder. NY State J Med 69:548– 550, 1969

4 Najenson T, Pikielni SS: Malalignment of the glenohumeral joint following hemiplegia. Annals of Physical Medicine 8:96–99, 1965

5 Taketomi Y: Observations on subluxation of the shoulder joint in hemiplegia: A clinical report. Phys Ther 55:39–40, 1975

6 Hoppenfeld S: Physical Examination of the Spine and Extremities. New York, NY, Appleton-Century-Crofts, 1976

7 Ritt B, Belkin J, Lal S, et al: Comparative Study of Sling Supports for the Subluxed Hemiplegic Shoulder (RT-20, Project R-109). Chicago, IL, Rehabilitation Institute of Chicago, 1980

8 Stratford P, Agostino V, Brazeau C, et al: Reliability of joint angle measurement: A dis-

cussion of methodology issues. Physiotherapy Canada 36:5-9, 1984

9 Rothstein JM, Miller PJ, Roettger RF: Goniometric reliability in a clinical setting: Elbow and knee measurements. Phys Ther 63:1611– 1615, 1983

10 Thorndike RL: Applied Psychometrics. Boston, MA, Houghton Mifflin Co, 1982

11 Shrout PE, Fleiss JL: Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 86:420–428, 1979

12 Lahey MA, Downey RG, Saal FE: Intraclass correlations: There's more there than meets the eye. Psychol Bull 93:586–595, 1983

13 Fish DR, Wingate L: Sources of goniometric error at the elbow. Phys Ther 65:1666–1670, 1985

14 Boone DC, Azen SP, Lin C-M, et al: Reliability of goniometric measurements. Phys Ther 58:1355–1360, 1978

15 Cronbach LJ, Gleser GC, Nanda H, et al: The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles. New York, NY, John Wiley & Sons Inc, 1972

16 Shavelson RJ, Webb NM, Burstein L: Measurement of teaching. In Wittrock MC (ed): Handbook of Research on Teaching, ed 3. New York, NY, Macmillan Publishing Co, 1986, pp 50–91