Reliability of five methods for assessing shoulder range of motion

Kimberley Hayes, Judie R Walton, Zoltan L Szomor and George AC Murrell

University of New South Wales

In symptomatic subjects, the reliability of tests for shoulder joint range of motion has yet to be determined. For this reason, inter-rater and intra-rater agreement trials were undertaken to ascertain the reliability of visual estimation, goniometry, still photography, "stand and reach" and hand behind back reach for six different shoulder movements. Intra-class correlation coefficients (Rho) were derived by using a random effects model. For flexion, abduction and external rotation fair to good reliability was demonstrated for both trials using visual estimation (inter-rater Rho = 0.57-0.70; intra-rater Rho = 0.59-0.67), goniometry (inter-rater Rho = 0.64-0.69; intra-rater Rho = 0.53–0.65) and still photography (inter-rater Rho = 0.62-0.73; intra-rater Rho = 0.56–0.61). The tests had standard errors of measurement of between 14 and 25 degrees (inter-rater trial) and 11 and 23 degrees (intra-rater trial). **[Hayes K, Walton JR, Szomor ZL and Murrell GAC (2001): Reliability of five methods for assessing shoulder range of motion. *Australian Journal of Physiotherapy* 47: 289-294]**

Key words: Clinical Protocols; Range of Motion, Articular; Reproducibility of Results; Shoulder

Introduction

Of all the joints, the shoulder has the greatest range of motion. The assessment of shoulder range of motion is important in the diagnosis of disorders of the shoulder and for the evaluation of the strategies that may alter shoulder function.

There have been a number of tools designed to measure joint range of motion varying from simple visual estimation to high speed cinematography (Clapper and Wolf 1988, Fish and Wingate 1985, Hellebrandt et al 1949, Low 1976, Moore 1949, Moore 1949, Youdas et al 1994). Several of these have been trialled for measurement reliability (Boone et al 1978, Clapper and Wolf 1988, Fish and Wingate 1985, Hellebrandt et al 1949, Low 1976, Mayerson and Milano 1984, Riddle et al 1987, Williams and Callaghan 1990, Youdas et al 1994). However, notably lacking is information specific to the shoulder joint, in symptomatic subjects, under conditions that reflect how measurements are taken in the clinical environment. Given that range of motion reliability varies from one patient population to the next (Ashton et al 1978, Bartlett et al 1985, Harris et al 1985), from one joint to the next (Boone et al 1978, Clapper and Wolf 1988, Hellebrandt et al 1949, Low 1976) and from one joint movement to the next (Hellebrandt et al 1949, Riddle et al 1987), there is a clear need to establish patient specific, joint specific and movement specific reliability indices for clinical practice. For this purpose, we determined the inter-rater and intra-rater reliability, and the standard error of the measurement for five easily applied methods for assessing active and passive shoulder range of motion for six movements, in patients with a spectrum of shoulder dysfunctions (rotator cuff repair, adhesive capsulitis and scapulothoracic fusion).

Method

Subjects Two groups of subjects gave informed consent for trials which investigated the inter-rater and intra-rater reliability of five joint range of motion assessment tests. All patients with shoulder pathology that warranted consultation with the participating orthopaedic surgeon between January 1996 and October 1998 were considered for inclusion to these trials. From this patient list, subjects were randomly selected and contacted by telephone for the purpose of recruitment. This process was repeated until the desired sample size had been obtained.

Inter-rater reliability trial The inter-rater reliability trial consisted of eight volunteers, three males and five females, ranging in age between 57 and 72 years (mean age = 66 years, SD = 5.7). All subjects had a current shoulder complaint. Six patients had undergone rotator cuff repair surgery within the past 24 months, one patient was 17 months post scapulothoracic fusion, and one patient had adhesive capsulitis.

Intra-rater reliability trial The intra-rater reliability trial consisted of nine volunteers, five males and four females, ranging in age between 29 and 74 years (mean age = 64 years, SD = 14.7). One of these subjects had two symptomatic shoulders. Of the 10 shoulders, eight were symptomatic, being within 36 months of rotator cuff repair

surgery. The remaining two shoulders were asymptomatic at the time of testing.

Raters

Inter-rater reliability trial Four raters, identified here as A to D were used for the inter-rater reliability trial. Rater A was an orthopaedic surgeon, Rater B was a sports physician trainee and Raters C and D were qualified physiotherapists. All raters tested all eight subjects with the five range of motion assessment tests. All raters were blinded to the results of each other's assessments.

Intra-rater reliability trial Only Rater A was used for the intra-rater reliability trial. This rater tested all nine subjects (including the one subject being tested bilaterally) on three separate occasions within a 48 hour period, using the same five range of motion assessment tests. The operating/treating surgeon was not involved in assessments for the inter-rater or intra-rater reliability trials.

Measurement tests

The five tests used for measuring range of motion for both the inter-rater and intra-rater trials were:

1. Visual estimation of passive range of motion (measured in degrees) for the movements of flexion, abduction, external rotation and hand behind back (measured to the vertebral level). The subject was seated upright on the edge of a treatment table with feet supported on a footstool. The rater moved the affected extremity (thumb pointing upwards) to the end of passive range of shoulder flexion, abduction and external rotation (0 degrees glenohumeral joint abduction, 90 degrees elbow flexion, neutral supination/pronation forearm position) and made a visual estimation of each resultant angle. For maximum passive hand behind back position, the rater moved the affected extremity upwards and towards the midline to the highest vertebral level reached by the tip of the subject's extended thumb.

2. Goniometry (measured in degrees), using a plastic, 41cm universal goniometer, for the active movements of flexion, abduction and external rotation. The subject was positioned as for visual estimation tests. The subject moved the affected extremity (thumb pointing upwards) to the end of active range of shoulder flexion, abduction and external rotation (0 degrees glenohumeral joint abduction, 90 degrees elbow flexion, neutral supination/pronation forearm position). The flexion angle was formed by aligning the goniometer with the lateral epicondyle of the humerus, the middle of the glenoid fossa, and a vertical line in the coronal plane. The abduction angle was formed by aligning the goniometer with the lateral epicondyle of the humerus, the middle of the posterior glenohumeral joint line, and a vertical line in the sagittal plane. The external rotation angle was formed by aligning the goniometer with the ulna styloid process, the olecranon process of the ulna, and a horizontal line in the transverse plane.

3. Still photography, using a standard polaroid camera, for the active movements of flexion, abduction and external rotation. Each rater labelled eight reference points on each subject prior to the commencement of photographic testing using the following bony landmarks: T1 spinous process, T7 spinous process, postero-lateral acromion process of the scapula, lateral epicondyle of the humerus, a point 6cm inferior to the postero-lateral aspect of the acromion process of the scapula, end of rib 12, olecranon process of the ulna, and ulna styloid process.

For flexion and abduction, the subject stood with feet shoulder width apart. For external rotation, the subject was supine and the affected extremity was supported on the treatment table (0 degrees glenohumeral joint abduction, 90 degrees elbow flexion, neutral supination/pronation position). The subject moved the affected extremity (thumb pointing upwards) to the end of active range of shoulder flexion, abduction and external rotation. The rater took a polaroid photograph of each end of range position from a perspective aligned with the axis of joint motion. Using a plastic 15cm goniometer and the developed photograph, the rater manually calculated the resultant joint angle in degrees. The flexion angle was formed by joining the lateral epicondyle, the point 6cm below the postero-lateral acromion process, and the end of rib 12. The abduction angle was formed by the intersection of a line that ran through the lateral epicondyle and the postero-lateral acromion process and another line that ran through T1 and T7 spinous processes. The external rotation angle was formed by joining the ulna styloid process, the olecranon process of the ulna, and a vertical line that passed through the olecranon process of the ulna. Each rater removed the labels prior to directing the patient to the next rater.

4. Stand and reach test of maximum overhead reach (measured in centimetres), using a metric vertical wall scale. The subject's standing head height (measured in centimetres) was measured against a metric vertical wall scale. The subject turned to face the wall and reached the affected extremity to a maximum overhead position. The stand and reach value was recorded as the position of maximum reach minus the patient's standing head height.

5. Tape measured distance of hand behind back reach in centimetres. The subject was standing with feet shoulder width apart. With thumb extended, the subject reached the affected extremity upwards and towards the midline to a maximum hand behind back position. The rater used a tape measure to record the distance in centimetres from the thumb tip to T1 spinous process.

Procedures For the inter-rater trial, all raters were briefed on the study protocol and a five hour training session was conducted to ensure familiarisation and standardisation of the five range of motion tests. The intra-rater trial was conducted several weeks after the inter-rater trial and involved a different group of subjects. Both trials followed the procedures described.

For each test, raters were asked to correct for compensatory

Table 1. Inter-rater reliability and standard error of the measurement for five methods of assessing shoulder range of motion.

	Mean± SD, (range)		ICC (Rho)	95% CI of ICC	SEM	95% CI
Visual estimation						
Flexion	142±34,	(45 – 180)º	0.70	0.42 – 0.92	19º	±38º
Abduction	121±33,	(45 – 180)º	0.66	0.37 – 0.90	19º	±38º
External rotation	47±21,	(13 – 80)º	0.57	0.26 – 0.87	14º	±28º
Goniometry						
Flexion	132±45,	(0 – 170)º	0.69	0.40 – 0.91	25º	±50º
Abduction	118±39,	(30 – 170)º	0.69	0.37 – 0.92	21º	±42º
External rotation	46±23,	(-15 – 80)º	0.64	0.31 – 0.91	14º	±28º
Still photography						
Flexion	120±45,	(5 – 170)º	0.73	0.46 – 0.93	23º	±46º
Abduction	118±44,	(15 – 160)º	0.73	0.45 – 0.93	23º	±46º
External rotation	50±25,	(-10 – 90)º	0.62	0.32 – 0.89	15º	±30º
Stand and reach	43±7,	(-64 – 54)cm	0.74	0.45 – 0.94	3cm	±6cm
Hand behind back	T10±3,	(T7 – S4)	0.26	-0.01 – 0.69	2	±4
(vertebral level) (cm)	26±8,	(16 – 48)cm	0.39	0.09 – 0.77	6cm	±12cm

Inter-rater reliability and standard error of the measurement for five methods of assessing shoulder range of motion. Each test was performed on eight subjects by four raters. Intra-class correlation coefficients (Rho) were calculated using a two-way mixed effect model. Mean, standard deviation and range of scores were calculated from the raw data. (SD = standard deviation, ICC = intra-class correlation coefficient, CI = confidence interval, SEM = standard error of the measurement).

scapulothoracic or upper body movement and to ensure that each movement occurred in the appropriate cardinal plane (sagittal plane, coronal plane and transverse plane for flexion, abduction and external rotation, respectively). All measurements were taken at the pain limit to range of motion. Each rater performed all five tests on a given subject before commencing assessment of the next subject. This procedure was repeated until all eight subjects had been measured by all four raters.

The order of the raters was randomised for each subject in the inter-rater trial. The order of the subjects was randomised for the intra-rater trial. With the exception of visual estimation, which was performed first, the order of the other four tests was random. This was done to prevent inadvertent biasing of the estimated angle with known cardinal data such as that provided by the goniometer.

Statistical analysis

For the inter-rater and intra-rater reliability trials, two-way random effect intra-class correlation coefficients (2,1) (Rho), together with their confidence intervals, were calculated with SPSS statistical software (Yaffee 1998). In accordance with the suggestions of Fleiss (1986), a Rho value of < 0.4 was deemed as representing poor reliability; 0.4-0.75 as fair to good reliability; and > 0.75 as excellent reliability.

In addition, the standard error of the measurement (SEM) and 95% confidence intervals (±2 SEM) were calculated for each range of motion test and expressed in units of the measure. The standard error of the measurement was calculated according to the equation: $SEM = SD \times \sqrt{(1 - ICC)}$, where SD is the standard deviation and ICC is the Rho score (Portney and Watkins 1993).

Results

Intra-class correlation coefficient The intra-class correlation coefficient (Rho score) is an index of reliability for measurements of the same material. Two types of measurement reliability can be distinguished according to whether each of several raters independently measures the same material (inter-rater reliability), or whether one rater makes two or more measurements of the same material (intra-rater reliability). An instrument with a Rho score of 1 will produce the same measurement each time it is used. In contrast, an instrument with a Rho score of 0 will produce markedly different measurements. The confidence interval of the intra-class correlation coefficient represents a range of possible Rho scores, on the basis of the sample data, for the population score (95% probability).

Inter-rater reliability trial The inter-rater intra-class correlation coefficients (Rho) for each movement tested with each of the range of motion tests are displayed in Table 1.

Table 2. Intra-rater reliability and standard error of the measurement for five methods of assessing shoulder range of motion.

	Mean± SD, (range)	ICC (Rho)	95% CI of ICC	SEM	95% CI
Visual estimation					
Flexion	143±21, (100 – 170)°	0.59	0.28 – 0.85	13°	±26°
Abduction	135±33, (70 – 170)°	0.60	0.30 – 0.86	21°	±42°
External rotation	47±19, (10 – 80)°	0.67	0.38 – 0.89	11°	±22°
Goniometry					
Flexion	135±26, (77 – 165)°	0.53	0.21 – 0.82	17°	±34°
Abduction	129±35, (56 – 170)°	0.58	0.27 – 0.85	23°	±46°
External rotation	41±23, (0 – 80)°	0.65	0.36 – 0.88	14°	±28°
Still photography					
Flexion	128±28, (63 – 168)°	0.56	0.24 – 0.84	19°	±38°
Abduction	131±35, (57 – 170)°	0.61	0.31 – 0.86	22°	±44°
External rotation	48±21, (7 – 79)°	0.60	0.30 – 0.86	13°	±26°
Stand and reach	39±19, (-22 – 59)cm	0.49	0.18 – 0.80	13cm	±26cm
Hand behind back					
(vertebral level)(cm)	T9±2, (T6 – S1)	0.14	-0.11 – 0.55	2	±4
	25±7, (15 – 39)cm	0.39	0.08 – 0.75	6cm	±12cm

Intra-rater reliability and standard error of the measurement for five methods of assessing shoulder range of motion. Each test was performed on nine subjects, on three occasions by a single trained rater. Intra-class correlation coefficients (Rho) were calculated using a two-way mixed effect model. Mean, standard deviation and range of scores were calculated from the raw data. (SD = standard deviation, ICC = intra-class correlation coefficient, CI = confidence interval, SEM = standard error of the measurement).

For the movements of flexion, abduction and external rotation, agreement between the four raters was fair-good as assessed by visual estimation (Rho = 0.70, Rho = 0.66, Rho = 0.57), goniometry (Rho = 0.69, Rho = 0.69, Rho = 0.64) and still photography (Rho = 0.73, Rho = 0.73, Rho = 0.62). Fair-good reliability was also demonstrated for the stand and reach test (Rho = 0.74). For the movement of hand behind back, measured passively to the vertebral level, and as an active reach, agreement between the four raters was poor (Rho = 0.26, Rho = 0.39).

Intra-rater reliability trial The intra-rater intra-class correlation coefficients (Rho) for each movement tested with each of the range of motion tests are displayed in Table 2.

For the movements of flexion, abduction and external rotation, agreement between the three ratings was fair-good as assessed by visual estimation (Rho = 0.59, Rho = 0.60, Rho = 0.67), goniometry (Rho = 0.53, Rho = 0.58, Rho = 0.65) and still photography (Rho = 0.56, Rho = 0.61, Rho = 0.60). Fair-good intra-rater reliability was also demonstrated for the stand and reach test (Rho = 0.49). For the movement of hand behind back, measured passively to the vertebral level, and as an active reach, intra-rater agreement was poor (Rho = 0.14, Rho = 0.39).

Standard error of the measurement The standard error of the measurement is a measure of discrepancy (error) between repeated scores. For the inter-rater trial, the standard error of the measurement represents the extent of error associated with different raters' scores. For the intra-rater trial, the standard error of the measurement represents the extent of error associated with retesting. The confidence interval of the measurement represents the smallest difference attributable to status change, as distinct from measurement error (95% probability).

Inter-rater reliability trial The standard error of the measurement (SEM) together with the mean, standard deviation and range of scores for each movement tested with each of the range of motion tests, are displayed in Table 1.

For the movements of flexion, abduction and external rotation, the standard error of the measurement ranged from 14-19 degrees for visual estimation, 14-25 degrees for goniometry, and 15-23 degrees for still photography. For the stand and reach test, the standard error of the measurement was three centimetres. For the movement of hand behind back, measured passively to the vertebral level, and as an active reach, the standard error of measurement was two vertebral levels, and six centimetres respectively.

Intra-rater reliability trial For the movements of flexion, abduction and external rotation, the standard error of the measurement ranged from 11-21 degrees for visual estimation, 14-23 degrees for goniometry, and 13-22 degrees for still photography. For the stand and reach test, the standard error of the measurement was 13 centimetres. For the movement of hand behind back, measured passively to the vertebral level, and as an active reach, the standard error of measurement was two vertebral levels, and six centimetres, respectively.

Discussion

The results of this study demonstrated fair-good reliability for visual estimation, goniometry, still photography and the stand and reach test for the movements of shoulder flexion, abduction, external rotation and overhead reach in patients with a variety of orthopaedic shoulder disorders (rotator cuff repair, adhesive capsulitis and scapulothoracic fusion) when assessed by either an individual trained measurer or by a number of trained measurers. Hand behind back, whether tested passively or as an active reach, was found to be the least reliable movement of this study.

Williams and Callaghan (1990) have shown that in the normal shoulder, visual estimation when employed by an experienced clinician permits measurements that are as reliable as those obtained with a mechanical goniometer. Using one asymptomatic subject, Low (1976) showed that examiners from various clinical backgrounds were more consistent in measuring full elbow flexion and full wrist extension positions with a goniometer than by estimating the same angles by eye. While there is some evidence to suggest that the goniometer can be used reliably for measuring certain shoulder joint movements (Boone et al 1978), studies of the shoulder in symptomatic patients are limited. In a comprehensive evaluation of passive shoulder range of motion in symptomatic subjects, Riddle (1987) reported variable results (Rho = 0.26-0.90 inter-rater; Rho = 0.87-0.99 intra-rater). The results of this study showed comparable reliability for visual estimation and goniometry for both the inter-rater (visual estimation Rho = 0.57–0.70; goniometry Rho = 0.64–0.69) and the intra-rater (visual estimation Rho = 0.59–0.67; goniometry Rho = 0.53– 0.65) trials.

Photographic procedures have been shown by Fish and Wingate (1985) to be more reliable than goniometry for measurements of elbow joint position in an asymptomatic subject. In the present study, all movements measured by still photography (active range of motion) demonstrated fair-good reliability for both the inter-rater trial (flexion Rho = 0.73, abduction Rho = 0.73, external rotation Rho = 0.62) and the intra-rater trial (flexion Rho = 0.56, abduction Rho = 0.61, external rotation Rho = 0.60). In general, comparable reliability and measurement error was demonstrated for visual estimation, goniometry, and still photography. Of note, large confidence intervals were associated with these same tests for the inter-rater trial (visual estimation = ± 28-38 degrees; goniometry =

± 28-50 degrees; still photography = ± 30- 6 degrees) and the intra-rater trial (visual estimation = ± 22-42 degrees, goniometry = ± 28-46 degrees; still photography = ± 26-44 degrees). On the basis of this data, large differences in repeated scores are therefore prerequisite when using these tests to assess actual change in clinical status.

We evaluated the stand and reach test as it is a functional movement of practical importance. The results of this study demonstrated fair to good reliability for the stand and reach test (inter-rater Rho = 0.74; intra-rater Rho = 0.49). Interestingly, this test was associated with a smaller standard measurement error for the inter-rater trial (3cm) than for the intra-rater trial (13cm).

Poor reliability was demonstrated for both hand behind back tests for both the inter-rater trial (Rho = 0.26 for passive hand behind back; Rho = 0.39 for active reach) and the intra-rater trial (Rho = 0.14 for passive hand behind back; Rho = 0.39 for active reach). These findings may be a reflection of the complexity of the movement itself. To effect the manoeuvre, contributions to range of motion are required from the shoulder complex, elbow, forearm, wrist and hand. Alternatively, the lower reliability associated with this movement may reflect an attenuated range of available scores. In the presence of certain shoulder joint pathologies, hand behind back range of motion is frequently diminished (Murrell and Walton 2001). There was a standard measurement error of two vertebral levels for passive hand behind back and six centimetres for active hand behind back reach for both the inter-rater and intra-rater trials. Hand behind back range of motion is of functional importance and one that warrants further reliability testing.

The intention of this study was not to compare the reliability of active and passive range of motion tests for the shoulder joint, in symptomatic subjects. Rather, this study established the reliability and measurement error of five range of motion tests currently performed in clinical practice.

Both the inter-rater and intra-rater trials were conducted over a relatively short timeframe. The reliability of these methods for longer reassessment timeframes requires that the raters do not change their technique. This study did not establish the reliability of shoulder range of motion assessments over time.

Estimates of measurement reliability are applicable to the range of test scores and to the conditions of assessment. The mean values obtained for flexion, abduction and external rotation, for the inter-rater and intra-rater trials (Table 1, 2) show that most subjects obtained scores for these movements in the upper ranges of motion. We did not determine the reliability of these tests for scores measured primarily in the lower ranges of motion.

In summary, fair to good reliability was demonstrated for three shoulder movements assessed by visual estimation,

goniometry and still photography for both the inter-rater and intra-rater trials. Standard measurement errors in the range of 14–25 degrees for the inter-rater trial, and 11–23 degrees for the inter-rater trial for these three tests highlighted considerable variation in the precision associated with range of motion testing of the shoulder joint in symptomatic subjects.

Authors George Murrell, Department of Orthopaedic Surgery, St George Hospital, Kogarah, New South Wales 2217. E-mail: admin@ori.org.au (for correspondence). Kimberley Hayes, Judie Walton, Zoltan Szomor, Department of Orthopaedic Surgery, St George Hospital, Kogarah, New South Wales 2217.

References

Ashton BB, Pickles B and Roll JW (1978): Reliability of goniometric measurements of hip motion in spastic cerebral palsy. *Developmental Medicine in Child Neurology* 20: 87-94.

Bartlett MD, Wolf LS, Mot DB, Shurtleff MD and Stahell LT (1985): Hip flexion contractures: A comparison of measurement methods. *Archives of Physical Medicine and Rehabilitation* 66: 620-625.

Boone DC, Azen SP, Lin C, Spence C, Baron C and Lee L (1978): Reliability of goniometric measurements. *Physical Therapy* 58: 1355-1360.

Clapper MP and Wolf SL (1988): Comparison of the reliability of the orthoranger and the standard goniometer for assessing active lower extremity range of motion. *Physical Therapy* 68: 214-218.

Fish DR and Wingate L (1985): Sources of goniometric error at the elbow. *Physical Therapy* 65: 1666-1670.

Fleiss JL (1986), cited in Delitto A and Strube MJ (1991): Reliability in the clinical setting. *American Physical Therapy Association Research Section Newsletter* 24: 2-8.

Harris SR, Smith LH and Krukowski L (1985): Goniometric reliability for a child with spastic quadriplegia. *Journal of Paediatric Orthopaedics* 5: 348-351.

Hellebrandt FA, Duvall EN and Moore ML (1949): The measurement of joint motion. Part 3. *Physical Therapy Review* 29: 302-307.

Low JL (1976): The reliability of joint measurement. *Physiotherapy* 62: 227-229.

Mayerson NH and Milano RA (1984): Goniometric measurement reliability in physical medicine. *Archives of Physical Medicine and Rehabilitation* 65: 92-94.

Moore ML (1949): The measurement of joint motion. Part 1. *Physical Therapy Review* 29: 195-205.

Moore ML (1949): The measurement of joint motion. Part 2. *Physical Therapy Review* 29: 256-264.

Murrell GAC and Walton J (2001): Diagnosis of rotator cuff tears. *Lancet* 357:769-770.

Portney LG and Watkins MP (1993): Foundations of Clinical Research. Applications to Practice. Connecticut: Appleton & Lange.

Riddle DL, Rothstein JM and Lamb RL (1987): Goniometric reliability in a clinical setting. *Physical Therapy* 67: 668-673.

Williams JG and Callaghan M (1990): Comparison of visual estimation and goniometry in determination of a shoulder joint angle. *Physiotherapy* 76: 655-657.

Yaffee RA (1998): Enhancement of reliability analysis: application of intraclass correlations with SPSS/Windows v.8, http://www.nyu.edu/acf/socsci/Docs/intracls.html.

Youdas JW, Carey JR, Garrett TR and Suman VJ (1994): Reliability of goniometric measurements of active arm elevation in the scapular plane obtained in a clinical setting. *Archives of Physical Medicine and Rehabilitation* 75:1137-1144.