
ORIGINAL RESEARCH
EVALUATING CHANGE IN CLINICAL STATUS:
RELIABILITY AND MEASURES OF AGREEMENT
FOR THE ASSESSMENT OF GLENOHUMERAL
RANGE OF MOTION

Susan W. Muir, BScPT PhD¹
Charlene Luciak Corea, BScPT²
Lauren Beaupre, BScPT MScPT PhD³

ABSTRACT

Objectives: To evaluate intra-rater and inter-rater reliability and measurement error in glenohumeral range of motion (ROM) measurements using a standard goniometer.

Study design: 17 adult subjects with and without shoulder pathology were evaluated for active and passive range of motion. Fifteen shoulder motions were assessed by two raters to determine reliability. The intra-class correlation coefficients (ICC) were calculated and examined to determine if reliability of ICC ≥ 0.70 existed. The standard error of measurement (SEM) and the minimal clinical difference (MCD) were also calculated.

Results: The criterion reliability was achieved in both groups for intra-rater reliability of standing AROM abduction; supine AROM and PROM abduction, flexion, external rotation at 0° abduction; and for inter-rater reliability of supine AROM and PROM abduction, external rotation at 0° abduction. The SEM ranged from 4°-7° for intra-rater and 6°-9° for inter-rater agreement on movements that achieved the criterion reliability. The MCD ranged from 11°-16° for a single evaluator and 14°-24° for two evaluators.

Conclusions: Assessment of AROM and PROM in supine achieves superior reliability. The use of either a single or multiple raters affects the number of movements that achieved clinically meaningful reliability. Some movements consistently did not achieve the criterion and may not be the best movements to monitor treatment outcome.

Key Words: Reliability, shoulder goniometric measurement

CORRESPONDENCE

Dr. Susan Muir, Department of Geriatric Medicine, Parkwood Hospital, 801 Commissioners Road East, London, Ontario, Canada N6C 5J1, Email: susan.muir@uwo.ca, Phone number: 519-435-1252
Author disclosure: none
Previous presentation: Ontario Physiotherapy Association Annual Conference, March 2007. Received 2nd place award in poster presentations.

¹ Department of Geriatric Medicine, University of Western Ontario, London, Ontario, Canada

² Grey Nuns Community Hospital, Edmonton, Alberta, Canada

³ Department of Physical Therapy, Faculty of Rehabilitation Medicine, Department of Physical Therapy, Faculty of Rehabilitation Medicine; Department of Surgery (Division of Orthopaedic Surgery), Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada.

INTRODUCTION

Clinicians and researchers routinely evaluate change in patients' status over time. The assessment of range of motion (ROM) is important in the 1) diagnosis of glenohumeral disorders, 2) evaluation of treatment progression and effectiveness, and 3) quantifying the amount of change in movement that occurs. It is, therefore, important for clinicians and researchers to have complete and relevant information on the reliability and accuracy of ROM measurement.

Common scenarios that arise in daily clinical practice highlight the need to reliably determine 'real' change in a patient's condition. Clinicians want to know how to determine patients' progress over time with the same clinician performing assessments or how to relate measurements between clinicians when patient care is transferred from one person to another. These same issues also impact clinical research from both a research design and data analysis perspective particularly when multiple evaluators are used to evaluate progress over time.

Clinical assessment of the extremities usually involves obtaining information on an affected and an unaffected side for comparison. The presence of glenohumeral pathology can contribute to variation in measurement due to pain, weakness, fatigue and apprehension in addition to the variation in performing the measurement technique alone. Thus, it is important to ensure that measurements used in a clinical setting are reliable in both the presence and absence of shoulder pathology.

The intra-class correlation coefficient (ICC) quantifies reliability or consistency in a measurement; the closer the value is to 1.0, the better the reliability. However, the ICC value does not provide a quantification of the magnitude of the error. Evaluating the smallest detectable change has also been advocated as an important aspect of a reliability study.¹⁻⁴ The standard error of measurement (SEM) expresses agreement in the same units as the original measurement and indicates the amount of change needed to exceed the error of the measurement itself.¹ Knowledge of the error in the measurement technique allows for the determination of when an observed change equates to a minimum detectable change that is greater than measurement error itself.

Wide variability for both intra and inter-rater reliability for shoulder motion evaluation has previously been reported.^{2,5-11} Limitations in the previous literature that have reported the accuracy of assessing shoulder range of motion with a goniometer should be considered as important omissions. Specifically, these limitations include presentation of estimates without confidence intervals, inadequately powered sample sizes, no sample size calculations, failure to present SEM values and a limited number of possible shoulder movements assessed. (Table 1)

There is a threshold below which the consistency and precision of a measure is considered compromised and ceases to be clinically useful and informative. It is recommended that ICC values be greater than or equal to 0.70 to be considered acceptable as a clinically meaningful measurement tool.¹² In the shoulder reliability literature, this form of statistical analysis has not been previously performed, thus we do not know if goniometric assessment of glenohumeral ROM meets an acceptable standard.

The purpose of this study was to calculate 1) intra and inter-rater reliability ICC values for shoulder range of motion, 2) intra and inter-rater standard error of measurement (SEM) for each movement and 3) the minimal clinical difference (MCD) in a group of people with and without shoulder pathology for each movement assessed by a single evaluator and two evaluators.

MATERIALS AND METHODS

Subjects

A convenience sample of subjects with and without shoulder pathology was recruited from staff and patients attending the outpatient Department of Rehabilitation Medicine at Grey Nuns Community Hospital, Edmonton, Alberta. The study was approved by the University of Alberta Health Research Ethics Board (Biomedical Panel) and Caritas Research Steering Committee, Edmonton, Alberta and informed consent was obtained from all participants.

People were eligible for the study if they were between 18 and 75 years of age, able to easily move between supine and standing positions, and able to actively move their shoulder into 90° of glenohumeral abduction. Exclusion criteria for both groups

Table 1. Summary of Reliability Studies Evaluating Intra-rater and Inter-rater Reliability of Shoulder Range of Motion Using a Standard Goniometer.

Author	Sample size (N), Number of assessors (n)	Study Sample	Measurement	Movements	ICC values	Limits of Agreement	SEM
Riddle et al 1987 ²	N=100 (two groups of 50) n=16	Pathology	PROM	<u>Intra-rater reliability:</u> F E Abd HAbd HAdd ER <u>Inter-rater reliability:</u> F E Abd HAbd HAdd ER	Point estimates only	No	No
Bovens et al 1990 ⁵	N=8 n=3	Normal	PROM	<u>Intra-rater reliability:</u> ER <u>Inter-rater reliability:</u> ER	Point estimates only	No	No
Sabari et al 1998 ¹¹	N=30 n=1	Normal and pathology	PROM and AROM - supine and sitting	<u>Intra-rater reliability:</u> F Abd	Point estimates only	No	No
MacDermid et al 1999 ¹⁰	N=34 n=2	Pathology	PROM	<u>Intra-rater reliability:</u> IR <u>Inter-rater reliability:</u> IR	Point estimates with 95% confidence intervals	No	Yes
Hayes et al 2001 ⁸	<u>Inter-rater:</u> N=8 n=4 <u>Intra-rater:</u> N=9 n=1	Pathology	AROM	<u>Inter-rater reliability:</u> F Abd ER <u>Intra-rater reliability:</u> F Abd ER	Point estimates with 95% confidence intervals	No	Yes
Nadeau et al 2007 ⁶	N=30 n=2	Normal and pathology	AROM	<u>Intra-rater reliability:</u> EL RET PRO <u>Inter-rater reliability:</u>	Point estimates only	No	Yes
				EL RET PRO			

F=flexion, E=extension, Abd= abduction, HAbd=horizontal abduction, HAdd=horizontal adduction, ER=external rotation, IR=internal rotation, EL=elevation, RET=retraction, PRO=protraction

were acute pain/injury of either shoulder, previous fracture of the scapula or proximal humerus, active joint or systemic infection, neurological conditions (e.g. stroke, Parkinson's Disease, brachial plexus injury, etc), significant muscle paralysis of the rotator cuff, deltoid or shoulder girdle musculature, inability to speak or read English, psychiatric illness that precluded providing informed consent or the ability to consistently perform the testing protocol. The self-report of no previous or current shoulder problems placed the participant in the without shoulder pathology group. The shoulder pathology group included participants who reported chronic and stable musculoskeletal injuries of the shoulder.

A sample size for ICC parameter estimation was based on an α value of 0.05, a β value of 0.20, an expected ICC value (intra and inter-rater reliability) of 0.90 with the minimum value in the one-sided 95% confidence interval of 0.70, using 2 replicates of each measurement and 2 evaluators.¹³ Using these parameters and defining the unit of assessment being a shoulder, the estimated sample size required 19 shoulders to be assessed in each group.

The evaluators were two registered physical therapists with 16 and 12 years of experience in the assessment and treatment of orthopedic conditions, and as evaluators in orthopedic surgical trials of shoulder conditions. A study assistant was used for the recording of the measurement data during the test sessions. Data collection began in January 2005 and ended April 2005.

Design

The evaluators and study assistant participated in a one-hour formal training session. Study participants performed a set of warm-up exercises to reduce the risk of a mobilization effect from the repeated movements performed during the assessment. The warm-up routine included 10 repetitions of each exercise of shoulder pendular exercises, and active assisted shoulder extension, flexion, internal and external rotation exercises in standing.

Fifteen movements were assessed; four active range of motion (AROM) movements in standing and eleven movements of both AROM and passive range of motion (PROM), in supine, see Table 2. Scapular

stabilization was used during the evaluation of internal rotation and horizontal adduction in supine. Details of the test positions, manual stabilization and goniometer placement are found in Appendix 1.

Each participant presented on one occasion for approximately one hour and was assessed successively by the two evaluators. A single shoulder was considered the unit of study. Each evaluator independently measured one or both shoulders of each participant twice during the test session, providing a total of four ROM assessments per study shoulder. The evaluator order, the order of the shoulder to be assessed first, the two assessment positions (supine and standing), and the order of the movements in each position were randomly assigned for each participant at the start of the test session by the study assistant.

To prevent measurement bias, the goniometer dial was covered with white paper, as described by Riddle et al.² This method obscured the numerical values on the goniometer to the evaluators, but allowed the study assistant to view the reverse side of the goniometer to record the values.² The recorded values of test measurements were not made available to the evaluators until study recruitment was completed and the last test session was finished.

Joint Measurements

All goniometer measurements were maximal joint motions measured with the JAMAR E-Z Read goniometer, a standard 12 inch, double-armed 360° goniometer, constructed of clear plastic. For testing, the subject was placed in the appropriate starting position, which was with the arm by the side, except where specified otherwise. Goniometer placement was done after the movement was performed and maximal range of motion achieved. Active range of motion was determined by the participant's self-report of reaching maximal amount of motion, while passive range of motion was determined by the assessing physiotherapist's report of reaching maximal passive end feel. No participants were limited by pain in either active or passive range of motion.

Data Analysis

The following analyses were performed for the normal and pathological groups separately.

Table 2. *Shoulder Movements Assessed in Study.*

Test Position	Type of Movement	Specific Movement
Standing	AROM	Abduction Extension Flexion Scaption
Supine	AROM	Abduction Flexion ER at 0° abduction ER at 90° abduction IR at 90° abduction Horizontal adduction
Supine	PROM	Abduction Flexion ER at 0° abduction ER at 90° abduction Horizontal adduction
Abbreviations: AROM, active range of motion; PROM, passive range of motion; ER, external rotation; IR, internal rotation.		

Comparisons were made between these analyses of the number and type of movements that achieved the criterion level of reliability. The intra and inter-rater ICC values were calculated by performing two-way analysis of variance (ANOVA) for each movement using the random effects statistical methodology described by Eliasziw et al.¹⁴ Point estimates and 95% one-sided lower-limit confidence intervals for the ICC values were calculated. In this study, an ICC value with a confidence interval that had a lower limit greater than or equal to 0.70 would

indicate that it achieved the criterion level of reliability deemed necessary for clinical utility. A lower confidence interval bound of below 0.70 would indicate the measure did not achieve the criterion level, regardless of the point estimate value.

The calculation of intra-rater and inter-rater reliability ICC values and standard error of the measurement were performed using the approach described by Eliasziw et al.¹⁴ instead of conventional calculations used to determine SEM. Data analyses were

performed using SAS Statistical Software, version 8.2 (SAS Inc, Cary, NC).

RESULTS

Data from 17 individuals, representing 23 normal shoulders and 11 abnormal shoulders were collected. The average age of subjects in the sample was 45.1 years (range 23-83 years) and there were 14 females and 3 males. The pathologies of study shoulders included: osteoarthritis (1), history of strain with ongoing symptoms due to a motor vehicle accident (1), rotator cuff tendinopathy (5), distal humerus fracture (1), dislocation (1) and instability without history of dislocation (2).

The movements that achieved the criterion level of reliability in the normal and pathology groups had similar point estimates and 95% confidence intervals, SEM, and MCD.

As expected, intra-rater reliability values achieved more criterion levels of reliability than inter-rater values for both PROM and AROM movements in standing and supine for both normal and pathological shoulders, demonstrating that there was less variability when the same evaluator was used (Tables 3 and 4). The values for standing AROM scaption, (See Appendix 1 for definition/description), and supine AROM horizontal adduction did not meet the criterion level in the pathology group and may be a result of low power in this sample. The intra-rater reliability values for standing AROM abduction; supine AROM abduction, flexion, and external rotation (ER) at 0° abduction; and supine PROM abduction, flexion and ER at 0° abduction met or surpassed the pre-specified criterion value in both groups.

Inter-rater reliability values were typically of lower magnitude than intra-rater reliability indicating greater variation, as expected, when two evaluators were used in both the normal and pathology groups. The inter-rater reliability values of 4 movements; supine AROM and PROM of abduction and external rotation at 0° abduction met the criterion for reliability in both groups, suggesting that these movements can be reliably measured and provide clinically useful information.

Importantly, there are movements that consistently did not achieve the criterion value in either group.

The intra-rater reliability measure for standing AROM extension did not achieve the criterion. Several additional movements failed to achieve the criterion value for inter-rater reliability including standing AROM abduction, flexion, and extension; supine AROM internal rotation, horizontal adduction; and supine PROM horizontal adduction.

Both the SEM and the MCD values for intra-rater agreement were smaller than for inter-rater agreement consistent with less measurement variation that is typical when the same evaluator is used. (Tables 3 and 4) Movements that met the criterion level had comparable values in both the normal and pathology group. However, large values were still present for some movements that achieved the criterion value for reliability; in particular supine AROM abduction had a MCD for two raters of 20° in normal shoulders and 24° in pathologic shoulders.

DISCUSSION

This study gives a comprehensive presentation of reliability and minimal clinical difference (MCD) values for 15 movements of the shoulder commonly used in clinical practice and research. The results provide valuable information on the limits of assessment for reliability and enables clinicians to make knowledgeable decisions regarding whether a clinically meaningful change has occurred between testing sessions, or whether the change could primarily be due to variability from measurement error.

This evaluation of shoulder range of motion reliability includes the largest selection of movements compared to previous reports and addresses the limitations present in the current published literature on goniometric measurement. While the authors acknowledge the importance of assessing glenohumeral rotations in 90° abduction, they were only assessed during active movement in supine in the current study. Performance of these movements in standing presents measurement challenges of isolating glenohumeral range from movement due to scapular movement and associated thoracic spine motions of rotation and extension that accompany the glenohumeral movements associated with throwing. Awan et al evaluated three techniques for measuring shoulder internal rotation using an inclinometer and reported that the use of scapular

Table 3. Reliability with One-sided 95% Confidence Intervals, Standard Error of the Measurement (SEM) and Minimum Clinical Difference (MCD) Values are shown for a single rater and two raters when examining patients with normal shoulders. (N = 23)

Movement	Average Range of Motion (SD)	Intra-rater reliability	Intra-rater SEM	MCD for single rater	Inter-rater Reliability	Inter-rater SEM	MCD for two raters
Standing AROM:							
Abduction	170 (14)	0.91 (0.82, 1)*	4	11	0.67 (0.53, 1)	8	22
Extension	53 (8)	0.76 (0.58, 1)	4	11	0.77 (0.66, 1)	4	11
Flexion	160 (12)	0.86 (0.74, 1)*	5	12	0.76 (0.65, 1)	5	14
Scaption	169 (12)	0.91 (0.83, 1)*	4	10	0.82 (0.74, 1)*	5	14
Supine AROM:							
Abduction	167 (16)	0.87 (0.76, 1)*	6	16	0.80 (0.70, 1)*	7	20
Flexion	170 (8)	0.92 (0.83, 1)*	2	7	0.74 (0.63, 1)	4	12
External rotation at 0° abduction	55 (16)	0.91 (0.84, 1)*	5	13	0.91 (0.86, 1)*	5	14
External rotation at 90° abduction	90 (12)	0.81 (0.65, 1)	5	14	0.72 (0.60, 1)	6	18
Internal rotation at 90° abduction	51 (10)	0.87 (0.73, 1)*	4	11	0.62 (0.47, 1)	6	18
Horizontal adduction	36 (12)	0.89 (0.70, 1)*	4	12	0.47 (0.32, 1)	9	26
Supine PROM:							
Abduction	176 (14)	0.91 (0.83, 1)*	4	12	0.88 (0.82, 1)*	5	14
Flexion	177 (6)	0.85 (0.72, 1)*	3	7	0.78 (0.68, 1)	3	9
External rotation at 0° abduction	68 (16)	0.94 (0.88, 1)*	4	11	0.85 (0.77, 1)*	7	18
External rotation at 90° abduction	103 (11)	0.86 (0.64, 1)	5	13	0.49 (0.35, 1)	9	24
Horizontal adduction	43 (10)	0.85 (0.59, 1)	4	12	0.36 (0.21, 1)	9	25
* Movements that achieved criterion value of an ICC with a lower band of the one-sided 95% confidence interval ≥ 0.70 . Shaded areas are the ICC values that achieved the criterion value of reliability and the corresponding SEM and MCD values.							

Table 4. Reliability with One-sided 95% Confidence Intervals, Standard Error of the Measurement (SEM), and Minimum Clinical Difference (MCD). Values are shown for a single rater and two raters when examining patients with shoulder pathology. (N = 11)

Movement	Average Range of Motion (SD)	Intra-rater reliability	Intra-rater SEM	MCD for single rater	Inter-rater reliability	Inter-rater SEM	MCD for two raters
Standing AROM:							
Abduction	160 (15)	0.93 (0.85, 1)*	4	11	0.57 (0.33, 1)	10	28
Extension	44 (9)	0.65 (0.35, 1)	5	14	0.47 (0.31, 1)	6	18
Flexion	151 (8)	0.63 (0.41, 1)	5	14	0.55 (0.29, 1)	6	16
Scaption	159 (13)	0.78 (0.48, 1)	7	18	0.47 (0.33, 1)	10	28
Supine AROM:							
Abduction	153 (27)	0.95 (0.91, 1)*	6	16	0.91 (0.86, 1)*	9	24
Flexion	163 (13)	0.90 (0.84, 1)*	4	11	0.89 (0.82, 1)*	4	12
External rotation at 0° abduction	45 (15)	0.89 (0.78, 1)*	5	15	0.76 (0.65, 1)	8	22
External rotation at 90° abduction	83 (19)	0.93 (0.88, 1)*	5	14	0.89 (0.82, 1)*	7	18
Internal rotation at 90° abduction	42 (9)	0.69 (0.32, 1)	5	14	0.39 (0.24, 1)	7	20
Horizontal adduction	29 (8)	0.82 (0.61, 1)	4	10	0.59 (0.44, 1)	6	15
Supine PROM:							
Abduction	161 (26)	0.94 (0.89, 1)*	7	18	0.92 (0.88, 1)*	8	21
Flexion	171 (12)	0.92 (0.85, 1)*	3	9	0.88 (0.81, 1)*	4	11
External rotation at 0° abduction	59 (15)	0.94 (0.88, 1)*	4	11	0.86 (0.78, 1)*	6	16
External rotation at 90° abduction	94 (19)	0.95 (0.90, 1)*	4	12	0.89 (0.82, 1)*	7	18
Horizontal adduction	39 (8)	0.89 (0.78, 1)*	3	8	0.70 (0.57, 1)	5	14
* Movements that achieved criterion value of an ICC with a lower band of the one-sided 95% confidence interval ≥ 0.70 . Shaded areas are the ICC values that achieved the criterion value of reliability and the corresponding SEM and MCD values.							

stabilization techniques to control for accessory scapulothoracic motion was superior for measurement.¹⁵ Measurement of glenohumeral internal rotation used during the current study used stabilization of the scapula in order to identify the point where scapular motion commenced, at which point the end of range of motion was deemed present and goniometric measurement was taken. Scapular stabilization was not possible to maintain by a single person during goniometric measurement of passive range of motion measurement, so this movement was not included. The current authors found that glenohumeral internal rotation could not be reliably measured when performed as it would be in a typical clinical setting by a single evaluator. This is in contrast to Awan et al.¹⁵ whose protocol used two people in the measurement process, one for positioning and one to perform the measurement.

Comparison of our results to previous studies that used the goniometer as the measurement device is limited by the lack of reported confidence intervals in other studies.^{2,5,6,11} The present study's findings are in accordance with the results from Hayes et al.⁸ in that the movements of intra-rater reliability for standing AROM flexion and inter-rater reliability of standing AROM flexion and abduction fall below the threshold ICC of 0.70 in shoulders with pathology.

This study also highlights the importance of including confidence intervals along with the ICC point estimates when evaluating reliability. Only 7 of 15 common shoulder ROM measurements met the predetermined level of reliability in both groups for a single evaluator. Reliable measurements were achieved using multiple raters for only 4 ROM measurements in both groups, AROM and PROM abduction and external rotation at 0° abduction. The confidence intervals provide a measure of the precision of the estimate and the majority of previous studies have not presented them. The point estimate alone is not sufficient to determine if the measurement exceeds the ICC threshold of 0.70. This study used an a priori specification to determine an adequate sample size and power to evaluate if ICC values meet the ICC threshold of 0.70.

Variation in reliability values can arise from several sources including the inaccurate or inconsistent

land-marking during goniometer placement and lack of stabilization of the shoulder girdle to prevent compensatory scapulothoracic movements during rotations. Movements in supine allow for support of the trunk permitting greater relaxation of the participant and stabilization of the shoulder girdle especially in the evaluation of PROM. The assessment of AROM in positions of sitting or standing while providing an evaluation in a functional position also introduces muscular strength as a potential limiting factor to the maximal attained range of motion. Limited range of motion in a gravity dependent position, such as standing, then needs to be further differentiated between strength and range of motion as limiting factors to assist in devising a treatment program. The assessment of ROM in supine creates different gravity effects and may be complementary to the assessment in standing due to the alterations of muscle strength requirements. The use of a second person to provide the stabilization on the assessment of rotation movements especially in 90° abduction, as found by Awan et al.¹⁵, may need to be encouraged, particularly for research, where it may be important to detect small, but important differences between patient groups. The present study also incorporated measures to limit error due to a warm-up effect, but may not have removed all effects. This finding supports the practice of providing a sufficient warm-up to the area to be assessed before evaluation. Greater variability in PROM than AROM may result from variation in the amount of force used to attain full range, especially for inter-rater reliability, and therefore active movements may be preferable to passive movements in order to evaluate change.

It is uncertain if the movements with the lowest reliability values can be improved with greater training, but it does highlight the importance of training sessions for evaluators and the reporting of this information when performing clinical studies. Therefore, it seems reasonable for research investigators to include a reliability sub-study to confirm the consistency of evaluators and to establish the minimal clinical difference for a given study population. As the values obtained in this study are from a combined population with normal shoulders and shoulders with chronic stable pathology, any study

evaluating shoulders with acute conditions may have greater variation in values.

The setting of a minimally acceptable level for both intra and inter-rater reliability and testing to determine whether those levels could be achieved was a unique aspect of this study. Hypothesis testing in the absence of a criterion value only tests if values are different than zero.¹² We specifically evaluated if goniometric shoulder assessment can achieve clinically useful reliability values using a pre-determined criterion value of $ICC \geq 0.70$.

In this study, reliability with a goniometer was difficult to achieve using multiple raters, a trend consistent with other studies.^{2,5,6,8} The movements that did not meet the criterion value may be unable to accurately reflect change over time and therefore should be used with caution as primary outcomes in research and clinical practice for this purpose. As standard goniometers are not the only measurement method available for range of motion evaluation, there is still room to refine reliability further through evaluation of other apparatus such as electro-goniometers or inclinometers.

It is important therefore to highlight and demonstrate how the information from this study can be used practically in the clinical and research setting. The intra-rater SEM provides the range of values that can be expected on re-testing for a single evaluator.¹⁵ For example; assume a single rater assesses active standing abduction, a movement that achieved the reliability threshold, obtains a value of 135° . The intra-rater SEM, four degrees, suggests that if the same rater repeated that measurement, and there was no expectation that the subject's AROM had truly changed, the range of possible values could be 131° to 139° . This range of values could impact outcome measure scoring systems for functional assessment of the shoulder where points are assigned to the actual range of motion value, as in the Constant and UCLA Shoulder Scores.

The inter-rater SEM gives the range of potential error in different raters' measurements.¹⁶ This value has practical implications on the reporting and comparison of results from independent assessments, as can occur in worker compensation or insurance claims. In either scenario, an independent assessment,

concurrent with community rehabilitation, is not uncommon as part of the routine practice of case management. Active supine abduction, a measurement that met the reliability criterion, has a potential variability in measurement between two raters on a measured value of 135° of 111° to 159° . Passive supine horizontal adduction, a movement that did not meet the criterion threshold, has a range of values of 121° to 149° on a measured value of 135° . If the extremes of possible values were obtained in this scenario, the variability could be misattributed as a lack of sincerity of effort or irritability of the underlying tissue or injury when in fact it is just the inherent error in the measurement process and not a reflection of the capability of the person being assessed.

In clinical research, the MCD for two raters has implications for evaluating the superiority of one treatment regimen over another. For example, using a hypothetical randomized controlled trial of two post-operative rehabilitation protocols after mini-open rotator cuff surgery, treatment regimen 1 produces a statistically significant gain in range of motion for supine active abduction, forward flexion, and external rotation in adduction. The MCD can be used to determine if the statistically significant difference in treatment is in excess of the measurement error and therefore, also clinically meaningful.

There are several limitations in the current study that need to be addressed. The sample size was achieved in the normal shoulder group, but unfortunately, sample size could not be achieved in the shoulder pathology group within the time frame available to complete the study. A lack of movements achieving the a priori established ICC value in the shoulder pathology group could be due in part to insufficient power to find a statistical significant association. A full evaluation of all AROM and PROM glenohumeral rotations at 90° abduction in the two patient test positions limits comprehensive knowledge translation to clinical practice. While there is limited reliability using a standard goniometer, these limitations cannot be translated to other methods of measuring range of motion; therefore evaluation with other measurement tools is recommended considering the prominence that measurement plays in clinical practice and research.

CONCLUSIONS

Intra-rater evaluation can achieve acceptable reliability in the greatest number of movements: standing AROM abduction; supine AROM abduction, flexion, and external rotation (ER) at 0° abduction; and supine PROM abduction, flexion and ER at 0° abduction. Across the groups with normal and shoulder pathology, inter-rater evaluation met the criterion level for reliability for four movements performed in supine: AROM and PROM of abduction and external rotation at 0° abduction. These movements should be considered as acceptable for measuring and quantifying change over time and as primary outcomes in research.

REFERENCES

1. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician*. 1983;32:307-317.
2. Riddle DL, Rothstein JM, Lamb RL. Goniometric reliability in a clinical setting: Shoulder measurements. *Phys Ther*. 1987;67(5):668-673.
3. Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: An illustration using isometric force measurement. *Phys Ther*. 1993;73(6):386-401.
4. Stratford PW, Goldsmith CH. Use of the standard error as a reliability Index of interest: an applied example using elbow flexor strength data. *Phys Ther*. 1997;77(7):745-750.
5. Bovens AMPM, van Baak MA, Vrencken JGPM, Wijnen JAG, Verstappen FTJ. Variability and reliability of joint measurements. *Am J Sports Med*. 1990;18(1):58-63.
6. Nadeaus, Kovacs S, Gravel D, Pottie F, Moffet H, Gagnon D et al. Active movement measurement of the shoulder girdle in healthy subjects with goniometer and tape measure techniques: a study on reliability and validity. *Physiother Theory Pract*. 2007; 23(3): 179-187.
7. Green S, Buchbinder R, Forbes A, Bellamy N. A standardized protocol for measurement of range of movement of the shoulder using the Plurimeter-V inclinometer and assessment of its intrarater and interrater reliability. *Arthritis Care Res*. 1998; 11(1):43-52.
8. Hayes K, Walton J, Szomor Z, Murrell GAC. Reliability of five methods for assessing shoulder range of motion. *Aust J Physiother*. 2001; 47:289-294.
9. Hoving JL, Buchbinder R, Green S et al. How reliably do rheumatologists measure shoulder movement? *Ann Rheum. Dis* 2002; 61:612-616.
10. MacDermid JC, Chesworth BM, Patterson S, Roth JH. Intratester and intertester reliability of goniometric measurement of passive lateral shoulder rotation. *J Hand Ther*. 1999;12:187-192.
11. Sabari JS, Maltzev I, Lubarsky D, Liskay E, Homel P. Goniometric assessment of shoulder range of motion: comparison of testing in supine and sitting positions. *Arch Phys Med Rehabil*. 1998;79:647-651.
12. Streiner DL, Norman Gr. *Health measurement scales: a practical guide to their development and use*. New York: Oxford University Press, 2003.
13. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stats Med*. 1998;17:101-110.
14. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther*. 1994; 74(8):777-788.
15. Awan R, Smith J, Boon AJ. Measuring shoulder internal rotation range of motion: a comparison of 3 techniques. *Arch Phys Med Rehabil*. 2002;83:1229-1234.
16. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. 2nd ed. New Jersey: Prentice Hall Health, 2000.

Appendix 1: Details of Testing Positions and Goniometer Placement.

Test Positions for AROM in Standing

Test Movement	Description
Forward Flexion	The movement included scapular motion with measurement made with full elbow extension and motion leading with the thumb along a path in the sagittal plane. The goniometer was centered in the middle of the glenoid fossa with the arms aligned with the lateral epicondyle and the vertical line in the coronal plane.
Abduction	This movement included scapular motion with the measurement made with full elbow extension and motion leading with the thumb in the coronal plane. The goniometer was centered in the middle of the posterior glenohumeral joint line and the arms were aligned with the lateral epicondyle and the vertical line of the sagittal plane.
Scaption	Defined as elevation through abduction in the plane of the scapula defined to be 30-45° anterior to the coronal plane. This movement was performed with full elbow extension and movement leading with the thumb. The goniometer placement was the same as used when measuring abduction in standing.
Extension	Shoulder extension was measured with full elbow extension and forearm in neutral pronation/supination. Subject moved both arms at the same time to minimize trunk compensation movements. The goniometer placement was the same as used when measuring flexion in standing.

Test Positions for AROM and PROM in Supine

PROM and AROM were measured for each of the movements in supine except for internal rotation where PROM was not performed. For all PROM measurements, overpressure was applied by the evaluator to achieve full allowable motion.

Test Movement	Description
Forward Flexion	Movement in this position was measured with full elbow extension and leading with the thumb. The goniometer was centered in the middle of the glenoid fossa with the arms aligned with the lateral epicondyle of the humerus and horizontal along the midline of the trunk.

Appendix 1: Continued.

Abduction	This movement was measured in the coronal plane with full elbow extension and leading with the thumb. The goniometer was centered in the middle of the glenoid fossa on the anterior aspect of the shoulder joint. The stationary arm was aligned parallel to the spine and the active arm along the shaft of the humerus.
External Rotation in 0° Abduction	This movement was measured with the humerus positioned parallel to the spine or maximum adduction if the person could not achieve 0°. Towels were positioned as needed under the upper arm to bring the humerus in line with the coronal plane. The elbow was flexed to 90° with the forearm in neutral supination/pronation and the person was instructed to maintain adduction. The goniometer was centered at the olecranon with the arms aligned with the shaft of the ulna and the vertical axis of the movement plane.
External Rotation at 90° Abduction	The arm was initially passively positioned in 90° of pure abduction and towels were used, as needed, under the upper arm to bring the humerus level with the coronal plane as the starting position. The elbow was flexed to 90° with the forearm in neutral supination/pronation. The goniometer placement was the same as that used during measurement of external rotation at 0°.
Internal Rotation at 90° Abduction	The initial arm placement was as per external rotation at 90°. Maximal movement was defined at the point when scapular ante-tilt could not be corrected with verbal cueing under manual stabilization by therapist. This was observed when contact between the posterior aspect of the shoulder and the assessment table could not be maintained. The goniometer placement was the same as that used during measurement of external rotation at 0°.
Horizontal Adduction	Starting position of the arm was passive placement in 90° flexion and internal rotation so the forearm aligns horizontally across the body at the level of the shoulders. The elbow was flexed to 90° and the forearm pronated. Maximal movement was defined at the commencement of scapular protraction observed when contact between the posterior aspect of the shoulder and the assessment table could no longer be maintained. Manual stabilization was used on the scapula. The goniometer was positioned over the acromioclavicular joint from the superior aspect and the arms were aligned with the shaft of the humerus and the vertical axis.