Inter-rater reliability of ultrasound measurements of acromion–greater tuberosity distance between experienced and novice raters in healthy people

Praveen Kumar | Matthew Attwood

Department of Allied Health Professions, University of the West of England, Bristol, UK

Correspondence
Dr Praveen Kumar, Faculty of Health and Life Sciences, University of the West of England, Blackberry Hill, Stapleton, Bristol, BS16 1DD, UK
Email: praveen.kumar@uwe.ac.uk

1 | INTRODUCTION

A reduction in the acromion–humeral distance (AHD) is used as a predictive marker in identifying rotator cuff tears (Cholewinski, Kusz, Wojciechowski, Cielinski, & Zoladz, 2008). By contrast, an increase in AHD is suggestive of inferior shoulder subluxation following stroke, which is a common secondary musculoskeletal problem reported in up to 81% of patients (Ada & Foongchomcheay, 2002). Several studies have reported the usefulness of diagnostic ultrasound in the measurement of AHD, which is defined as the nearest distance between the lateral margin of the acromion and the head of the humerus (Azzoni, Cabitza, & Parrini, 2004; Desmeules, Minville, Riederer, Côté, & Frémont, 2004). Recently, the acromion–greater tuberosity (AGT) distance (measured between the acromion process of the scapula and the greater tuberosity of the humerus) has also been used in the diagnoses of rotator cuff tears (Cholewinski et al., 2008) and shoulder subluxation in post-stroke hemiplegia (Kumar, Bradley, Gray, & Swinkels, 2011a; Park, Kim, Sohn, Shin, & Lee, 2007).

To evaluate the effectiveness of treatment interventions, accurate, reliable and valid outcome measures are required. Several studies have reported the reliability of ultrasonographic measurements of AHD both in healthy participants (Desmeules et al., 2004) and in patients with rotator cuff injuries and supraspinatus impingement syndrome (SIS) (Azzoni et al., 2004; Desmeules et al., 2004; Kalra, Seitz, Boardman, & Michener, 2010; Pijs, Kok, Penning, Guldemond, & Arens, 2010). Kalra et al. (2010) reported excellent test–retest reliability (intraclass correlation coefficient [ICC] 0.76–0.92) of AHD measurements in patients with rotator cuff tears. Pijs et al. (2010) reported excellent intrarater reliability for both an experienced rater (musculoskeletal radiologist) (ICC 0.94) and a novice rater (orthopedic resident) (ICC 0.92). That study also reported good inter-rater reliability (ICC 0.70) between experienced and novice raters for AHD measurements in the neutral shoulder position.

Similarly, several other studies have reported the reliability of ultrasonographic measurements of AGT distance both in healthy participants (Kumar, Bradley, & Swinkels, 2010) and in patients with post-stroke hemiplegia (Kumar et al., 2011a; Park et al., 2007). Kumar et al. (2010) reported excellent within-day (ICC 0.98–0.99) and between-day (ICC 0.96–0.97) intrarater reliability in healthy participants when assessed by an experienced rater (a physiotherapist). Similarly, another study reported excellent intrarater (ICC 0.84–0.91) and good inter-rater (ICC 0.79) reliability when assessed by three novice raters (physiotherapy students) (Kumar et al., 2011b). High reliability coefficients reported from these two studies suggest that ultrasonographic measurements are reliable when measured by novice (Kumar et al., 2011b) and experienced (Kumar et al., 2010) raters.

To the best of the authors’ knowledge, no previous studies have investigated inter-rater reliability between experienced and novice raters for AGT measurements. The aim of the present pilot study was to provide a short course of training and assess the inter-rater reliability of ultrasonographic measurements of AGT distance between experienced and novice raters (physiotherapy students) in healthy individuals prior to testing on patient populations. The results of this study should inform future research studies and the clinical application of ultrasonographic measurements in the ongoing assessment of specific shoulder-related pathologies.

2 | MATERIALS AND METHODS

Healthy individuals aged over 45 years were eligible to participate and a convenient sample of n = 11 was recruited from the authors’ academic institution. People with a previous history of injury to the neck or shoulder with ongoing symptoms were excluded from the study. The study received approval from the Faculty of Health and Applied Sciences Research Ethics Sub-committee of the University of
the West of England, Bristol, and each participant gave informed written consent to take part.

2.1 Procedure

Baseline demographic data related to arm dominance (as stated by participants), gender, age and history of previous shoulder injury were recorded. Each participant was asked to perform a few simple arm movements to establish that the range of movement of both shoulders was equal, pain free and within normal parameters (Petty & Moore, 2004).

Ultrasound measurements of AGT distance were undertaken by a physiotherapist (P.K.), who acted as the experienced rater. The experienced rater had been involved in previous reliability studies. Three physiotherapy students acted as the novice raters and underwent shoulder ultrasound training specific to the present study. Each novice rater received 1 h of formal training on the ultrasound technique by the experienced physiotherapist (P.K.) and then practised on one another, unsupervised, for an additional hour, to become familiar with the protocol and measurement procedure.

A portable diagnostic ultrasound, (TITAN model, M-Mode, Depth 3.9, L38/10-5 MHz broadband 38 mm linear array transducer, Sonosite Limited, Hitchin, UK) was used for scanning the shoulder and for recording the AGT distance. The equipment was tested and calibrated according to the manufacturer’s guidelines prior to commencement of the data collection process. The precision of the linear measures based on manufacturer specifications was ±2%.

A standardized position was used for ultrasound scanning and for recording the AGT measurements. Participants were scanned while seated upright on an armless chair with their hips and knees flexed to 90° and feet resting flat on the ground. After adoption of standardized position, three ultrasound images of the first shoulder were obtained and AGT distance was measured on each frozen image by the first rater. AGT distance was defined as the relative distance between the lateral edge of the acromion process of the scapula and the nearest margin of the superior part of the greater tuberosity of the humerus. This was repeated on the opposite shoulder. The participants were then encouraged to move out of the standardized position. The same procedure was then repeated by the other three raters, who ensured that participants were in the standardized position for ultrasound imaging. Therefore, a total of three measurements per shoulder were taken by each rater. All raters were given a number and the order of measurements recorded was randomized. All four raters were blind to their own measurements (values were obscured by placing a sticker on the ultrasound screen) and to each other’s measurements.

2.2 Data analysis

Data were analysed using the Statistical Package for Social Sciences (SPSS v23.0; IBM UK, Business Analytics, Middlesex, UK). Descriptive statistics were used to calculate the mean and standard deviation (SD) of the three AGT distance measurements for both shoulders undertaken by each rater.

To assess the inter-rater reliability of ultrasonographic measurements of AGT distance, ICC of 3.3 with 95% confidence intervals were used. For calculation purposes, the mean of three measurements recorded by the experienced rater were compared with the mean of three measurements recorded by the three novice raters for both the right and left shoulders. Reliability was considered excellent if the ICC value was ≥0.75, fair to good if the value was 0.40–0.74 and poor if the value was ≤0.39 (Shrout & Fleiss, 1979).

The Standard Error of Measurement (SEM) was used to define 95% confidence limits around individual measurements. The minimum detectable change (MDC) was used to quantify the magnitude of change that was not likely to be a result of measurement error (Haley & Fragala-Pinkham, 2006). For MDC, a confidence interval of 90% (MDC90) is commonly recommended (Kolber, Saltzman, Beekhuizen, & Cheng, 2009).

Repeated measures analysis of variance (ANOVA) was used to analyse the variability of repeated ultrasonographic measurements of AGT distance on each shoulder between raters. Significance was set at the $p = 0.05$ level and post-hoc testing was performed using pairwise Bonferroni correction, with significance set at $p = 0.08$.

3 RESULTS

Eleven healthy individuals (nine female, two male) with a mean age of 54 years (SD 5) were recruited into the study. All participants were right-hand dominant. A summary of descriptive data for AGT distance measurements for all four raters is provided in Table 1. The ICC, standard error of measurement and MDC90 for inter-rater reliability for both right and left shoulders are presented in Table 2.

Repeated measures ANOVA indicated significant differences between AGT distance measurements when comparing novice raters with the experienced rater for the left shoulders ($F_{3, 30} = 10.147; p = 0.001$) and right shoulders ($F_{3, 30} = 3.394; p = 0.031$). Post-hoc analysis with pairwise Bonferroni correction showed a significant difference between novice rater 3 ($p < 0.01$) and the experienced rater but not between the latter and the other two raters.

| TABLE 1 | The mean, standard deviation (SD) and 95% confidence interval (CI) for acromion–greater tuberosity distance measurements recorded by four raters (n = 11) |
|----------|---------------------------------|
| LEFT SHOULDER |                               |
| Experienced rater | 1.2 | 2.4 | 1.8 ± 0.4 | 1.6–2.0 |
| Novice 1 | 1.0 | 2.4 | 1.8 ± 0.4 | 1.6–2.1 |
| Novice 2 | 1.2 | 2.4 | 1.7 ± 0.3 | 1.5–1.9 |
| Novice 3 | 0.9 | 2.0 | 1.5 ± 0.3 | 1.3–1.7 |
| RIGHT SHOULDER |                             |
| Experienced rater | 1.1 | 2.1 | 1.7 ± 0.3 | 1.4–1.9 |
| Novice 1 | 1.1 | 2.1 | 1.7 ± 0.3 | 1.5–1.9 |
| Novice 2 | 1.3 | 2.0 | 1.6 ± 0.2 | 1.4–1.7 |
| Novice 3 | 0.9 | 2.0 | 1.5 ± 0.3 | 1.3–1.7 |

The ICC value was compared with the mean of three measurements recorded by the experienced rater for both the right and left shoulders. Reliability was considered excellent if the ICC value was ≥0.75, fair to good if the value was 0.40–0.74 and poor if the value was ≤0.39 (Shrout & Fleiss, 1979).
TABLE 2  Intraclass Correlation coefficient (ICC), standard error of the mean (SEM) and minimum detectable change (MDC) at the 90% confidence interval (MDC90) for acromion–greater tuberosity distance measurements between experienced (Exp) and novice raters

| Right Shoulder | Left Shoulder |
|----------------|--------------|
| ICC (95% CI) | SEM (cm) | MDC90 (cm) | ICC (95% CI) | SEM (cm) | MDC90 (cm) |
| All raters | 0.75 (0.5–0.91) | 0.2 | 0.3 | 0.70 (0.40–0.90) | 0.2 | 0.4 |
| Exp versus rater 1 | 0.85 (0.70–0.95) | 0.1 | 0.2 | 0.83 (0.63–0.94) | 0.1 | 0.2 |
| Exp versus rater 2 | 0.75 (0.60–0.91) | 0.1 | 0.2 | 0.87 (0.73–0.96) | 0.1 | 0.2 |
| Exp versus rater 3 | 0.62 (0.38–0.85) | 0.2 | 0.4 | 0.61 (0.37–0.85) | 0.2 | 0.4 |

SEM = SD √1−ICC, where SD is the baseline standard deviation of the whole sample (Keating and Matyas, 1998).

4 | DISCUSSION

The primary aim of the present study was to assess the inter-rater reliability of ultrasonographic measurements of AGT distance between experienced and novice raters in healthy individuals. One experienced physiotherapist and three physiotherapy students acted as raters and recorded AGT distance measurements using portable ultrasound equipment. The study found good (ICC 0.61) to excellent (0.87) inter-rater reliability.

The findings were in agreement with previous studies in healthy individuals (Kumar et al., 2011a, 2011b) and in patients with shoulder-related problems such as SIS (Cholewinski et al., 2008). Kumar et al. (2011a) tested inter-rater reliability within-day and between-day reliability, and reported excellent between-day intrarater reliability (ICC 0.97–0.98) for AGT measurements in older healthy adults (mean age 64 years [SD 11]). In that study, however, only one rater, an experienced physiotherapist with modest training in shoulder ultrasound, was involved with the recording of measurements.

Another study, involving three physiotherapy students, reported excellent inter-rater reliability (ICC 0.79) of AGT measurements for the right shoulder in a relatively young age group (mean age 21 years [SD 2]) (Kumar et al., 2011b). However, the reliability was not compared with that of an experienced rater. By contrast, the current study involved three novice raters and assessed the inter-rater reliability of the AGT distance measurements by comparison with an experienced rater. To the authors’ knowledge, this is the first report of the inter-rater reliability of AGT distance measurements taken by novice raters and compared with measurements taken by an experienced rater, using portable ultrasound equipment in healthy people. Good reliability of measurements suggests that novice raters with limited training in ultrasound are capable of undertaking reliable ultrasonographic measurements of AGT distance. These results with relatively inexperienced raters are encouraging, suggesting that this technique can be easily learned by clinical physiotherapists with no previous experience in ultrasound. This is because physiotherapists are generally considered to have a good basic knowledge of anatomy and therefore are able to produce reliable results with minimal training. This should be tested in a future study in a patient population.

In conjunction with ICC, this study used SEM, which provides an estimation of how repeated measures on a person are most likely to be distributed around the “true” value (Wyrwich, 2004). On successive testing, there is a 95% probability that repeated measurements on an individual would fall within a mean of ±2 (SEM) cm (Keating & Matyas, 1998). The standard error of measurement for both shoulders, across all raters was ≤0.2 cm, which indicates that, for between-rater measurements, there is a 95% probability that the true measurement would lie within 0.2 cm of the obtained value. These findings were in agreement with those of previous studies which reported a low SEM (≤0.15 cm) when ultrasound measurements were undertaken by experienced (Kumar et al., 2010) and novice (Kumar et al., 2011b) raters.

In the present study, inter-rater reliability was good (ICC 0.61–0.62) for rater 3 but excellent (ICC 0.75–0.87) for the other two raters when compared with the experienced rater. The low reliability coefficients noted for rater 3 could have been due to individual variation in the identification of the bony point for measurement purposes. The mean AGT measurements recorded by rater 3 were generally on the low side when compared with those of the other three raters, suggesting that rater 3 potentially selected different points for measurement purposes. For the purpose of standardization, it is critical that all raters measure the AGT distance using the same bony reference points.

The study had several limitations. First, a small convenience sample from a healthy population was selected. A larger number of healthy participants, recruiting the same number of males and females, would have been helpful. Second, the inter-rater reliability of AGT distance measurements was not tested in a patient population and involving physiotherapists. This would warrant further research in a population with shoulder dysfunction or pain.

In conclusion, inter-rater ultrasonographic measurements of AGT distance were observed to be reliable when assessed by three novice raters (physiotherapy students) when compared with an experienced rater in healthy individuals. Further work to establish the inter-rater reliability of AGT distance measurements in patient populations, involving physiotherapists inexperienced in the ultrasound technique, is required.

ACKNOWLEDGEMENTS

This project was undertaken as part of an undergraduate research study in the BSc(Hons) Physiotherapy programme at the University of the West of England. The authors would like to thank Claire Browett and Debbie Poulston for help with data collection; Nicola Dowdeswel and Serena Clarke for help with recruitment and the randomization process; Professor Shea Palmer and Dr Sue Barnett for their critical comments; and the volunteers for their participation.
REFERENCES
Ada, L., & Foongchomcheay, A. (2002). Efficacy of electrical stimulation in preventing or reducing subluxation of the shoulder after stroke: A meta-analysis. The Australian Journal of Physiotherapy, 48(4), 257–267.

Azzoni, R., Cabitza, P., & Parrini, M. (2004). Sonographic evaluation of subacromial space. Ultrasonics, 42(1-9), 683–687.

Cholewinski, J. J., Kusz, D. J., Wojciechowski, P., Cielinski, L. S., & Zoladz, M. P. (2008). Ultrasound measurement of rotator cuff thickness and acromio-humeral distance in the diagnosis of subacromial impingement syndrome of the shoulder. Knee Surgery, Sports Traumatology, Arthroscopy, 16(4), 408–414.

Desmeules, F., Minville, L., Riederer, B., Côté, C. H., & Frémont, P. (2004). Acromio-humeral distance variation measured by ultrasonography and its association with the outcome of rehabilitation for shoulder impingement syndrome. Clinical Journal of Sport Medicine, 14(4), 197–205.

Haley, S. M., & Fragala-Pinkham, M. (2006). Interpreting change scores of tests and measures used in physical therapy. Physical Therapy, 86(5), 735–743.

Kalra, N., Seitz, A. L., Boardman, N. D., & Michener, L. A. (2010). Effect of posture on acromiohumeral distance with arm elevation in subjects with and without rotator cuff disease using ultrasonography. Journal of Orthopaedics and Sports Physical Therapy, 40(10), 633–640.

Keating, J., & Matyas, T. (1998). Unreliable inferences from reliable measurements. The Australian Journal of Physiotherapy, 44(1), 5–10.

Kolber, M. J., Saltzman, S. B., Beehuizen, K. S., & Cheng, M. S. (2009). Reliability and minimal detectable change of inclinometric shoulder mobility measurements. Physiotherapy Theory and Practice, 25(8), 572–581.

Kumar, P., Bradley, M., Gray, S., & Swinkels, A. (2011a). Reliability and validity of ultrasonographic measurements of acromion-greater tuberosity distance in post-stroke hemiplegia. Archives of Physical Medicine and Rehabilitation, 92(5), 731–736.

Kumar, P., Bradley, M., & Swinkels, A. (2010). Within-day and day-to-day intrarater reliability of ultrasonographic measurements of acromion-greater tuberosity distance in healthy people. Physiotherapy Theory and Practice, 26, 347–351.

Kumar, P., Chetwynd, J., Evans, A., Wardle, G., Crick, C., & Richardson, B. (2011b). Intrarater and intrarater reliability of ultrasonographic measurements of acromion-greater tuberosity distance in healthy people. Physiotherapy Theory and Practice, 27(2), 172–175.

Park, G. Y., Kim, J. M., Sohn, S. I., Shin, I. H., & Lee, M. Y. (2007). Ultrasonographic measurement of shoulder subluxation in patients with post-stroke hemiplegia. Journal of Rehabilitation Medicine, 39(7), 526–530.

Petty, N. J., & Moore, A. P. (2004). Neumusculoskeletal Examination and Assessment: A Handbook for Therapists (3rd edn). (pp. 253–270) Edinburgh: Churchill Livingstone.

Pijls, B. G., Kok, F. P., Penning, L. I., Guldemond, N. A., & Arens, H. J. (2010). Reliability study of the sonographic measurement of the acromiohumeral distance in symptomatic patients. Journal of Clinical Ultrasound, 38(3), 128–134.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86(2), 420–428.

Wyrwich, K. W. (2004). Minimal important difference thresholds and the standard error of measurement: Is there a connection? Journal of Biopharmaceutical Statistics, 14(1), 97–110.

How to cite this article: Kumar P, Attwood M. Inter-rater reliability of ultrasound measurements of acromion–greater tuberosity distance between experienced and novice raters in healthy people. Musculoskeletal Care. 2018;16:163-166. https://doi.org/10.1002/msc.1206