Title: Interrater reliability of ultrasound measurements of acromion-greater tuberosity distance between experienced and novice raters in healthy people

Short title: Reliability of AGT distance between experienced and novice raters

Praveen Kumar*, Matthew Attwood

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

Bristol, United Kingdom

*Author for correspondence

Dr Praveen Kumar

PhD, MSc, PG Cert (HE), MCSP, MIAP, MSPA

Senior Lecturer in Physiotherapy

Faculty of Health and Life Sciences

University of the West of England

Blackberry Hill, Stapleton

Bristol, BS16 1DD, United Kingdom

Email: Praveen.Kumar@uwe.ac.uk

Word count: 1990

INTRODUCTION

Reduction in the acromion-humeral distance (AHD) is used as a predictive marker in identifying rotator cuff tears (Cholewinski et al, 2008). In 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 and 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 et al, 2004; Desmeules et al, 2004). Recently, 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 (Park et al, 2007; Kumar et al, 2011a).

To evaluate the effectiveness of treatment interventions, accurate, reliable and valid outcome measures are required. Several studies have reported reliability of ultrasonographic measurements of AHD in both 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 et al, 2010; Pijls et al 2010). Kalra et al (2010) reported excellent test-retest reliability (ICC 0.76-0.92) of AHD measurements in patients with rotator cuff tears. Pijls 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 interrater reliability (ICC 0.70) between experienced and novice raters for AHD measurements in the neutral shoulder position.

Similarly, several other studies have reported reliability of ultrasonographic measurements of AGT distance on both healthy participants (Kumar et al, 2010) and in patients with post-stroke hemiplegia (Park et al, 2007; Kumar et al, 2011a). Kumar et al (2010) reported excellent within-day (ICC 0.98-0.99) and between-day (ICC 0.96-0.97) intrarater reliability on healthy participants when assessed by an experienced rater (a physiotherapist). Similarly, another study reported excellent intrarater (ICC 0.84-0.91) and good interrater reliability (ICC 0.79) 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 this pilot study was, to provide short course of training and assess 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.

MATERIALS AND METHODS

Healthy individuals aged over 45 years of age were eligible to participate in this pilot study and a convenient sample of n=11 were 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.

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 and Moore, 2011).

Ultrasound measurements of AGT distance were undertaken by a physiotherapist (PK) who acted as the experienced rater. The experienced rater was involved in previous reliability studies. Three physiotherapy students acted as the novice raters and underwent shoulder ultrasound training specific to this study. Each novice rater received 1 hour of formal training on the ultrasound technique by experienced physiotherapist (PK) and then practiced 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, L38/10-5MHz 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 is ± 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 standardised position, 3 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 standardised position for ultrasound imaging. Therefore, a total of 3 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

Data Analysis

Data were analysed using the Statistical Package for Social Sciences (SPSS v23.0). Descriptive statistics were used to calculate the mean and standard deviation (SD) of the 3 AGT distance measurements for both shoulders undertaken by each rater.

To assess interrater reliability of ultrasonographic measurements of AGT distance, intraclass correlation coefficients (ICC 3, 3) with 95% confidence intervals were used. For calculation purpose, the mean of three measurements recorded by experienced rater were compared with the mean of three measurements recorded by the three novice raters for both right and left shoulders. Reliability was considered excellent if the ICC value was ≥ 0.75, fair to good if the value was 0.40 to 0.74, and poor if the ICC value was ≤ 0.39 (Shrout and Fleiss, 1979).

The standard error of measurement (SEM) was used to define 95% confidence limits around individual measurements. Minimum detectable change (MDC) was used to quantify the magnitude of change that was not likely to be a result of measurement error (Haley and Fragala-Pinkham, 2006). For MDC, a confidence interval of 90% (MDC90) is commonly recommended (Kolber et al, 2009).

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

RESULTS

Eleven healthy individuals (9 female, 2 male) with a mean age of 54 SD 5 years 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. ICC, standard error of measurement, and MDC90 for interrater reliability for both right and left shoulders are presented in table 2.

Repeated measures ANOVAs indicated significant differences between AGT distance measurements when comparing novice raters to 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 experienced rater but not between experienced and other 2 raters.

DISCUSSION

The primary aim of this study was to assess interrater 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. This study found good (ICC 0.61) to excellent (0.87) inter-rater reliability.

The findings are in agreement with previous studies on healthy individuals (Kumar et al, 2011a, b) and in patients with shoulder-related problems such as SIS (Cholewinski et al, 2008). Kumar et al, (2011a) tested inter-rater reliability both 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 SD 11 years). 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 interrater reliability (ICC 0.79) of AGT measurements for the right shoulder in a relatively younger age group (mean age of 21 years SD 2) (Kumar et al, 2011b). However, the reliability was not compared with an experienced rater. On the contrary, the current study involved three novice raters and assessed inter-rater reliability of the AGT distance measurements by comparing with an experienced rater. To our knowledge, this is the first report of inter-rater reliability of AGT distance measurements taken by novice raters and compared with experienced rater using portable ultrasound 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 very 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 having a good basic knowledge of anatomy and therefore with minimal training are able to produce reliable results. This needs to be tested in future study on patient population.

In conjunction with ICC, this study used SEM that 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 and Matyas, 1998). The standard error of measurement for both shoulders across all raters was ≤0.2 cm, which indicates that, for between raters measurements, there is a 95% probability that the true measurement would lie within 0.2cm of the obtained value. These findings are in agreement with previous studies which report a low SEM (≤ 0.15cm) when ultrasound measurements were undertaken by experienced (Kumar et al, 2010) and novice raters (Kumar et al, 2011b).

In this study, interrater reliability was good (ICC 0.61-0.62) for rater 3 but excellent (ICC 0.75-0.87) for other 2 raters when compared with the experienced rater. Low reliability coefficients noted for rater 3 could be due to some individual variation in identification of bony point for measurement purposes. The mean AGT measurements recorded by rater 3 were generally on a lower side when compared to other 3 raters suggesting that rater 3 potentially selected different points for measurement purposes. For the purpose of standardization, however, it is critical that all raters measure the AGT distance using the same bony reference points.

This study has several limitations. First, a small convenience sample from a healthy population was selected for the study. Adequate number of healthy subjects recruiting the same number of males and females would have been helpful. Second, the interrater reliability of AGT distance measurements were not tested in patient population and involving physiotherapists. This would warrant further research in a population with shoulder dysfunction or pain.

In conclusion, interrater 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 interrater reliability of AGT distance measurements on patient populations involving physiotherapists inexperienced in ultrasound technique is required.

ACKNOWLEDGEMENTS

This project has been undertaken as part of an undergraduate research study on 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 randomisation process. The authors would like to thank Professor Shea Palmer and Dr Sue Barnett for their critical comments and 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. Australian Journal of Physiotherapy 48(4): 257-67. DOI: https://doi.org/10.1016/S0004-9514(14)60165-3

Azzoni R, Cabitza P, & Parrini M (2004) Sonographic evaluation of subacromial space. Ultrasonics 42(1-9): 683-87. DOI: https://doi.org/10.1016/j.ultras.2003.11.015

Cholewinski JJ, Kusz DJ, Wojciechowski P, Cielinski LS, & Zoladz MP (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): 408–14. DOI: https://doi.org/10.1007/s00167-007-0443-4

Desmeules F, Minville L, Riederer B, Côté CH and & 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 SM, & Fragala-Pinkham M (2006) Interpreting change scores of tests and measures used in physical therapy. Physical Therapy 86(5):735-743

Kalra N, Seitz AL, Boardman ND, & Michener LA (2010) Effect of posture on acromiohumeral distance with arm elevation in subjects with and without rotator cuff disease using ultrasonography. The Journal of Orthopaedics and Sports Physical Therapy 40(10): 633-40. DOI: https://doi.org/10.2519/jospt.2010.3155.

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

Kolber MJ, Saltzman SB, Beekhuizen KS, & Cheng MS (2009) Reliability and minimal detectable change of inclinometric shoulder mobility measurements. Physiotherapy Theory and Practice 25(8):572-81. DOI: https://doi.org/10.3109/09593980802667995.

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(5): 347-351. DOI: https://doi.org/10.3109/09593980903059522.

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): 731-36. DOI: https://doi.org/ 10.1016/j.apmr.2010.10.018.

Kumar P, Chetwynd J, Evans A, Wardle G, Crick C, & Richardson B (2011b) Interrater and intrarater reliability of ultrasonographic measurements of acromion-greater tuberosity distance in healthy people. Physiotherapy Theory and Practice 27(2):172-175. DOI: https://doi.org/10.3109/09593985.2010.481012

Park GY, Kim JM, Sohn SI, Shin IH, & Lee MY (2007) Ultrasonographic measurement of shoulder subluxation in patients with post-stroke hemiplegia. Journal of Rehabilitation Medicine 39(7): 526-30. DOI: https://doi.org/ 10.2340/16501977-0099

Petty NJ, & Moore AP (2011) Neuromusculoskeletal examination and assessment: A handbook for therapists, 4th edn. Edinburgh, Churchill Livingstone. pp 253-270.

Pijls BG, Kok FP, Penning LI, Guldemond NA, & Arens HJ (2010) Reliability study of the sonographic measurement of the acromiohumeral distance in symptomatic patients. Journal of Clinical Ultrasound 38(3): 128-134. DOI: https://doi.org/10.1002/jcu.20674.

Shrout PE, & Fleiss JL (1979) Intra-class correlations: uses in assessing rater reliability. Psychological Bulletin 86(2): 420-8.

Wyrwich KW (2004) Minimal important difference thresholds and the standard error of measurement: is there a connection? Journal of Biopharmaceutical Statistics 14(1):97-110.

Table 1: The mean, standard deviation (SD) and 95% confidence interval (CI) for AGT distance measurements recorded by three raters (n=11)

	Min	Max	Mean±SD	95% CI
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

Table 2: Inter-rater reliability coefficients, SEM and MDC90 for AGT distance measurements between experienced and novice raters

	RIGHT SHOULDER			LEFT SHOULDER		
	ICC (95% CI)	SEM (cm)	MDC (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 vs rater 1	0.85 (0.70-0.95)	0.1	0.2	0.83 (0.63-0.94)	0.1	0.2
Exp vs rater 2	0.75 (0.60-0.91)	0.1	0.2	0.87 (0.73-0.96)	0.1	0.2
Exp vs rater 3	0.62 (0.38-0.85)	0.2	0.4	0.61 (0.37-0.85)	0.2	0.4