Background

Range of motion (ROM) impairments of the overhead athletes’ shoulder are commonly addressed through mobility-based treatments, however, adaptations from humeral torsion (HT) are not amenable to such interventions. A clinical measurement to quantify HT has been proposed, however, the validity is not conclusive.

Purpose

The primary aim of this study is to determine the intrarater reliability and standard error of measurement (SEM) of the biceps forearm angle (BFA) measurement. The secondary aim of this study is to investigate the convergent validity of the BFA compared to diagnostic ultrasound.

Study Design

Cross Sectional Reliability and Validity Study

Methods

HT measurements, utilizing diagnostic ultrasound, were compared to BFA in 74 shoulders (37 subjects) over two sessions. Each measurement was performed three times and a third investigator recorded measures to ensure blinding. Reliability was investigated using utilizing an intraclass correlation coefficient (ICC 3,k)

Results

Intrarater reliability values were 0.923 and 0.849 for diagnostic ultrasound and BFA methods respectively. Convergent validity was $r = 0.566$. The standard error of measurement for diagnostic ultrasound and BFA was $3^\circ$ and $5^\circ$, respectively. The 95% limits of agreement between the two measurement methods were $-24.80^\circ$ and $19.80^\circ$ with a mean difference of $-2.50^\circ$ indicating that on average the diagnostic ultrasound measurement was lower than that of the BFA method.

Conclusion

The BFA is a reliable clinical method for quantifying HT, however, demonstrates moderate to poor convergent validity when compared to diagnostic ultrasound.

Level of Evidence

2b
INTRODUCTION

Shoulder pain affects up to 67% of the adult population throughout their lifespan.1,2 The etiology of shoulder pain is multifactorial and inclusive of numerous impairments, including but not limited to, restricted mobility. Posterior shoulder tightness (PST), in particular, has been associated with more common diagnoses such as labral tears, rotator cuff related pain syndrome, and post-operative arthrofibrosis among both the general and athletic populations, with a predilection towards overhead athletes.3-10 PST has been defined as a limitation of the extensibility within the posterior soft tissue structures of the shoulder including both contractile and non-contractile elements as well as osseous changes seen in the form of humeral torsion (HT) within the overhead athlete through training adaptations.11 Moreover, PST has been associated with restricted internal rotation (IR), horizontal adduction (HA), and flexion range of motion in conjunction with increased external rotation (ER) in the dominant arm of the overhead athlete. Adaptive changes occurring within the throwing athletes shoulder complex have been studied extensively.12-15 Range of motion (ROM) adaptations within the throwing shoulder of baseball pitchers has gained significant attention in the literature, as these anatomical changes are relevant for both diagnosis and program design.12,14 These adaptations within the dominant arm of the overhead athlete are necessary to a certain extent in order to perform at a particular level. However, excessive adaptations in ROM, or lack thereof, have been proposed as possible risk factors for both shoulder and elbow injuries among baseball pitchers resulting in lost playing time.16,17

Given the association between PST and shoulder pain in overhead athletes, sports medicine professionals often seek to quantify, and when necessary, decrease PST as a key element of their interventions.6,18-32

Within the overhead athlete it is necessary to determine which training adaptations surrounding the shoulder complex are contributing to differing degrees of mobility when compared to the non-throwing shoulder. There have been a number of proposed methods utilized to quantify PST; however, a recent systematic review35 provides evidence to support the use of a comprehensive approach including measures of HA, IR, and HT among the overhead athlete population. Once these elements have been identified, a thorough intervention strategy can be implemented in order to address any limitations. While soft tissue limitations may be addressed through select interventions, excessive adaptations in ROM resulting from HT are not amenable to mobility-based treatments. Identifying the contribution of HT to mobility impairments may serve to influence interventions and guide ROM expectations. HT can be measured utilizing computed tomography (CT) and diagnostic ultrasound by determining the angle created by the forearm and the proximal humerus. However, these methods require costly imaging investments and training, thus are not readily available in many clinical settings. Dashottar and Borstad34 proposed a clinical method for quantifying HT, which is referred to as the biceps forearm angle (BFA) and relies on palpation of the bicipital tuberosities and does not require the use of imaging modalities. Conflicting evidence surrounding the degree of validity found within this measurement technique exists, lending uncertainty to its use as a practical alternative to ultrasound.35 Finally, a recent study by Yaaari et al. utilized both the palpation and ultrasound methods for measuring HT among baseball pitchers concluding that the palpation method significantly underestimated HT compared to diagnostic US.36 The current investigation utilized a general population with a majority of individuals having participated in various overhead sports in the past. The authors hypothesize that BFA method, utilizing palpation, for quantifying HT will demonstrate good validity when compared to diagnostic ultrasound as well as good intrarater reliability. The primary aim of this study is to determine the intrarater reliability and standard error of measurement (SEM) of the BFA measurement. The secondary aim of this study is to investigate the convergent validity of the BFA compared to diagnostic ultrasound.

METHODS

PARTICIPANTS

A convenience sample of 37 adults, 15 males and 22 females, mean age 24 (± 2.4) for a total of 74 shoulders were screened for eligibility in this study between February 2019 - May 2019. Of the 37 participants, 26 had previously participated in overhead sports. The inclusion criteria for this study consisted of individuals enrolled at the University of Indianapolis, graduate or undergraduate courses, age 18 – 40 years old, and able to assume all testing positions. Exclusion criteria for this study consisted of a history of or current shoulder pain or injury that had required the attention of a healthcare provider within the past year. Demographic data were collected from each participant including gender, age, height, body mass, and handedness. The mean ± standard deviation (SD) age, body mass index, and height for subjects was 24 ± 3 years, 24.93 ± 3.16 kg/m², and 171.76 ± 11.30 cm respectively with 35 individuals being right-handed and two individuals being left-handed. All participants that met the inclusion criteria and agreed to participate in the study were provided with and signed an informed consent form approved by the Institutional Review Board at the University of Indianapolis, Study #0906. All 37 participants that agreed to participate in the study were available for follow up.

INSTRUMENTS

A standard plinth and Baseline® digital inclinometer (Fabrication Enterprises, White Plains, NY) were utilized to quantify all ROM measurements. The digital inclinometer was set at a zero point along a level vertical surface before measurements were taken on each participant and after any handling of the inclinometer to ensure an accurate zero starting point. A General Electric LOGIQ™ diagnostic ultrasound (GE Healthcare, Wauwatosa, WI) with 12 MHz linear - multi frequency linear array probe was utilized to perform all measurements of HT. The following parameters were utilized during all ultrasound procedures; frame rate of 35 cycles per second, coded harmonic imaging to reduce noise and enhance true signal, gain of 45, speckle reduction algorithm/frame averaging (S/A) of 3/2, B-mode, depth of
field was 3.5 cm (this was adjusted on two participants due to increased thickness of soft tissue), and a dynamic range of 72.

PROCEDURES

Study participants were seen twice, once during an initial data collection period and again three to five days later, with an average of four days, to have reliability components repeated for HT and BFA. Although there would be no activities in between the initial and follow up data collection period that would influence HT, all participants were asked to avoid any overhead upper extremity resistance training as well as participation in overhead sport activities. This was done to eliminate any potential for increased shoulder musculature tightness or soreness that, although unlikely, may influence the measurements.

All ROM measurement procedures were performed by the primary researcher with a clinical background in musculoskeletal orthopedics and over 10 years of experience. The protocols for each ROM measurements were adapted from published measurement protocols demonstrating good reliability when utilized by the primary researcher of this study. During all ROM measurement procedures, CS placed the digital inclinometer in the appropriate locations when instructed to by the primary researcher. Both the primary researcher and the CS were blinded to the measurement recording by having a cover placed over the digital read out on the inclinometer. Following each measurement, CS would hand the digital inclinometer to a third investigator that removed the cover, recorded the measurement and then zeroed out the inclinometer before replacing the cover and returning to CS.

Measurements of HT were performed by the primary researcher who additionally has over four years of experience utilizing diagnostic ultrasound for research purposes surrounding the shoulder. The primary researcher was blinded to the angle of rotation of the forearm utilizing a visual barrier and blinded to the measurement taken via the digital inclinometer by CS in the same manner as described above to ensure both were blinded.

All BFA measurements were performed by CS who was blinded to the angle of rotation of the forearm utilizing a visual barrier. CS was also blinded to the measurement taken via the digital inclinometer by the primary researcher in the same manner as described initially in order to ensure that both CS and the primary researcher were blinded to the results. All measurements were performed on both upper extremities.

HUMERAL TORSION MEASUREMENT WITH DIAGNOSTIC ULTRASOUND

The participant was positioned in supine with the arm abducted to 90 degrees, neutral rotation of the glenohumeral joint, elbow flexed to 90 degrees, and the forearm and wrist in a neutral position. A folded towel was placed under the humerus until it was visually level with the acromion to create a neutral position along the horizontal plane. Ultrasound gel was utilized as a conductor between the skin and the linear probe. A horizontal line was placed on the viewing screen diagnostic ultrasound, utilizing the measurement function, prior to initializing each imaging study. The probe was then placed at the proximal humerus perpendicular to the shaft of the humerus in order to visualize the greater and lesser tubercles. The probe was moved proximally and distally until the apex of the greater and lesser tubercles were thought to be visualized. Additionally, a small bubble level was placed on the probe to allow for a more consistent angle to be maintained during the procedure. The forearm of the participant was utilized to move the glenohumeral joint into IR and ER passively until the apexes of the tubercles were in line with the horizontal line created on the viewing screen (Figure 1). At this time CS placed the digital inclinometer along the distal anterior forearm to record the measurement (Figure 2).
Table 1. Intrarater reliability measurements

<table>
<thead>
<tr>
<th>Method</th>
<th>Measurement 1 mean angle°(SD)</th>
<th>Measurement 2 mean angle°(SD)</th>
<th>ICC 3,k (95% CI)</th>
<th>SEM°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic ultrasound</td>
<td>19.10 (11.74)</td>
<td>18.89 (12.46)</td>
<td>0.923(0.88-0.95)</td>
<td>3</td>
</tr>
<tr>
<td>Biceps forearm angle</td>
<td>20.28 (12.03)</td>
<td>21.40 (11.94)</td>
<td>0.849(0.76-0.91)</td>
<td>5</td>
</tr>
</tbody>
</table>

**BICEPS FOREARM ANGLE WITH PALPATION**

The participant was positioned in supine with the arm abducted to 45 degrees, CS palpated the proximal humerus to place the greater and lesser tubercles underneath one thumb. Maintaining the position of the greater and lesser tubercles under CS's thumb, the participant's arm was passively abducted to 90 degrees, with neutral rotation of the glenohumeral joint, the elbow flexed to 90 degrees, and the forearm and wrist in a neutral position. A folded towel was placed under the humerus until it was visually level with the acromion to create a neutral position parallel to the floor. The forearm of the participant was utilized to move the glenohumeral joint into IR and ER passively until the apexes of the tubercles were equally palpable under the thumb and thought to be along the horizontal plane. At this time, the primary researcher placed the digital inclinometer along the distal anterior forearm to record the measurement (Figure 3).

**DATA ANALYSIS**

Collected data were transferred to the Macintosh version of SPSS Statistics Version 23.0 for analysis. Descriptive data including mean measurement angles with SD were calculated for each series of measurements. The intrasession reliability of HT was calculated utilizing the intraclass correlation coefficient (ICC) model 3, k. The mean value of each series of measurements was utilized for the analysis. Model 3, k was used for the intrarater analysis to determine if this particular instrument can be used repeatedly with confidence by the same clinician. Our interpretation of the ICC values were based on guidelines offered by Portney and Watkins, whereby a value of above 0.75 was classified as good and a value of 0.50 to 0.75 would be considered to have moderate to poor reliability. The standard error of measurement (SEM) is not affected by intersubject variability and is important for clinical utilization of a measurement procedure; therefore, it will be reported in conjunction with the ICC’s using the formula: SEM = SD with a 68% confidence interval and the ICC (3,k) = r. Pearson product-moment coefficient of correlation (r) using a significance level of p = 0.01 was used for the analysis for the construct validity component of the investigation. Finally, a Bland Altman plot was utilized to calculate the mean difference between measurements as well as evaluate the 95% limits of agreement.

**RESULTS**

**RELIABILITY**

The data analysis of measurements revealed good intrarater reliability of both the HT measurement via diagnostic ultrasound (ICC = 0.923) and the measurements of BFA via palpation (ICC = 0.849). **Table 1** contains the mean angular measurements, SD, ICC (95% CI), and SEM.

**VALIDITY**

Convergent validity between HT, as measured via diagnostic ultrasound, and BFA, measured via palpation, was supported by a statistically significant moderate correlation (r = 0.566), (p = <0.001). **Table 2** contains the mean, SD, and average values for both HT and BFA.

The 95% limits of agreement between the two measurement methods were -24.80° and 19.80° with a mean difference of -2.50° (SD 11.38), the negative values are present because HT be either have either a positive or negative value (Figure 4).

**DISCUSSION**

The purpose of this investigation was to evaluate intrarater reliability and validity of a proposed method for quantifying
HT that does not require costly imaging modalities and extensive training. Prior to the proposed method put forward by Dashottar and Borstad, methods utilized to quantify humeral torsion involved CT, magnetic resonance imaging (MRI), and most recently, diagnostic ultrasound. These methods, although extremely accurate, require equipment that is not readily available to a majority of clinicians looking to quantify HT as part of a comprehensive examination of an overhead athlete. Therefore, the proposed method for quantifying HT via palpation seeks to allow an alternative in the absence of the aforementioned equipment. Conceptually, the palpation method is very similar in nature to that of the method utilizing diagnostic ultrasound. In both instances it is necessary to line up the apex of the greater and lesser tubercles on the horizon and then measure the angle of the forearm in this position in order to determine the amount of torsion present in the humerus. The precision to which the diagnostic ultrasound would be able to identify the greater and lesser tubercles would appear to be much improved, in comparison to palpation, given the ability to penetrate varying depths of the overlying soft tissue and very clearly visualize these osseous landmarks. Furthermore, utilizing the measurement feature on the diagnostic ultrasound machine, it is possible to create a true horizontal line on the screen in order to line up the apex of the tubercles once identified.

Despite the aforementioned differences, the validity results from the original study by Dashottar and Borstad demonstrated a correlation \( r = 0.85 \) between diagnostic ultrasound measurements and measurements through palpation among 49 shoulders. However, clinical measurements based on palpating the arm in 45 and 90 degrees of abduction as well as horizontal adduction have been proposed in another study that reportedly lack validity when compared to diagnostic ultrasound \( r \leq 0.326 \).

The results of the correlation analysis from previous studies differ from those reported in the current investigation \( r = 0.566 \), however, given the contrasting reports from previous studies, further investigation was warranted. There are several possible reasons for these differences which require consideration prior to proposing the method of palpation as a reasonable alternative for quantifying HT as opposed to current diagnostic ultrasound or other imaging studies. First, the palpation methods utilized in each study differed. The original study by Dashottar and Borstad performed the palpation method with the arm directly at the side in order to better palpate the tubercles as there is less soft tissue directly over the tubercles in this position. In the study by Feuerherd et al the palpation was performed and measured with the arm directly positioned in 45 degrees of abduction as well as in 90 degrees of abduction. The method utilized in the current study placed the arm in 45 degrees of abduction to palpate the tubercles and then moved the arm into 90 degrees of abduction while maintaining the palpated position. This method was performed to initially accommodate for less soft tissue being over the tubercles and then moved into the position in which the diagnostic ultrasound measurement was performed in order to remain consistent with both measurement positions. It is unclear from the methods section in the original study what position the arm was in when the ultrasound measurement was taken. Secondly, the experience and training of the individuals taking the measurements, both utilizing the diagnostic ultrasound and palpation method, may have also been different. From the original study it is noted that the individual performing the ultrasound measurements spent three months familiarizing themselves with the ultrasound unit and identifying anatomical structures of the shoulder which was further refined by a pilot study of 20 shoulders. In the current study, the researcher performing the ultrasound measurements had four years of experience utilizing diagnostic ultrasound for research regarding the shoulder while the researcher performing the palpation method had very little experience. The previous study indicates that the same individual performed both the diagnostic ultrasound measurements as well as the palpation measurements. It is possible that experience with palpation, particularly the palpation method proposed in both studies, may influence the results, much like is seen with experience surrounding diagnostic ultrasound methods.

---

**Table 2. Convergent validity between humeral torsion measurements via diagnostic ultrasound and biceps forearm angle**

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean angle°(SD)</th>
<th>Range°</th>
<th>Minimum°</th>
<th>Maximum°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic ultrasound</td>
<td>18.89 (12.46)</td>
<td>75.97</td>
<td>1.00</td>
<td>76.97</td>
</tr>
<tr>
<td>Biceps forearm angle</td>
<td>21.40 (11.94)</td>
<td>56.50</td>
<td>1.00</td>
<td>57.50</td>
</tr>
</tbody>
</table>

**Figure 4. Bland-Altman plot indicating differences between palpation and ultrasound measurements of humeral torsion**
The results of the current study demonstrate that both the method utilizing diagnostic ultrasound and palpation for quantifying HT exhibit good intrarater reliability, ICC=0.923 and ICC=0.849 respectively. However, the SEM differs slightly between the diagnostic ultrasound and palpation methods, 3° and 5° respectively meaning that the measurement obtained via diagnostic ultrasound may be 3° more or less than the true measurement and slightly greater, 5° when considering the palpation method. The previous study reported an SEM value that were slightly lower than the current study, 3° for the palpation method, but are very close to the value calculated for the diagnostic ultrasound method. A higher SEM for the palpation methods in this study could indicate a greater variability in measurements due to a larger standard deviation, leading to a greater probability of measurement error. Furthermore, a recent study by Yaari et al. that quantified HT in baseball pitchers concluded that the palpation method significantly underestimated the amount of humeral torsion when compared to diagnostic US and that these methods should not be utilized interchangeably.

Lastly, the overall mean difference between measurement methods in the current study was -2.50° (SD 11.38) which appears to be relatively low and indicates that HT was on average 2.5° lower for the diagnostic ultrasound. However, the range between the 95% limits of agreement (-24.80° and 19.80°) is relatively large. These results indicate that it is possible to either overestimate or underestimate the HT angle by a range of at least 19.8°, which suggests the need for caution if using the two measurement methods interchangeably. These results differ greatly from the study by Dashottar and Borstad, which reported 95% limits of agreement between -8.3° and 7.9°, which indicate that the palpation method could either overestimate or underestimate the measurement by approximately 8°. The mean difference in the referenced study was -0.2° (SD 4.1°). Dashottar and Borstad proposed the reason for their values with greater error may be a result of the anatomy of the lesser tubercle, more specifically when its angle relative to the bicipital groove is lower, not being as sensitive to palpation. This reason may also be speculated to have some influence on the results of the current study but do not account for such a wide range. The authors of the current study believe the possible explanations for this wide range of error may lie in both the palpation method and experience of the one performing it, as mentioned earlier in the discussion.

Although both methods demonstrate the ability to be reproduced among the same rater with some consistency, the question surrounding the validity of the palpation method when compared to the diagnostic ultrasound method still remains. It is possible that overall experience with palpation and, particularly, experience with the BFA may contribute to the overall ability of an individual to quantify HT via palpation. This variable, along with intrarater reliability, will be important to investigate prior to determining if this method is a practical alternative to measuring HT via diagnostic ultrasound as student clinicians and residents are likely to employ this method.

LIMITATIONS

There are several possible limitations that influence the overall results and interpretation of the findings of this current study. First, the convenience sample utilized for this study was a healthy college aged population and although it consisted of many overhead athletes or former athletes, it may not be representative of the overall population of interest. The authors did not include an investigation of intrarater reliability for either measurement and therefore the overall clinical utility of the methods investigated is not fully understood. Furthermore, the training of the individuals for both measurements may also be seen as a limitation of this study. Although the student was trained on the use of the BFA measurement by the principal investigator, she had limited experience with this technique prior to beginning this study. Future studies should seek to determine if training experience with the method utilizing palpation influences convergent validity as well as investigating intrarater reliability.

CONCLUSION

The palpation method for quantifying HT appears to demonstrate a good degree of intrarater reliability among an asymptomatic population. Of concern is the lack of concurrent validity when compared to diagnostic ultrasound. Although the palpation method for quantifying HT lacks validity when compared to diagnostic ultrasound, it may be a plausible alternative to quantifying HT when other methods are not available. Specifically, in cases where costly imaging modalities are not available, the palpation method may be useful for side-to-side comparisons of HT. Having some understanding of side-to-side differences allows the clinician to have an understanding of the underlying etiology of shoulder stiffness in both non-operative and post-operative populations. Further investigation is warranted to determine the influence of experience among those performing the measurement and the position of the glenohumeral joint during palpation. Although the use of asymptomatic participants may limit generalization, the use of symptomatic participants is unlikely to confound the results as most are able to achieve the needed passive range required for testing.

CONFLICTS OF INTERESTS

The authors report no conflicts of interest.

FUNDING

The authors received no funding for any portion of this study.

ETHICAL APPROVAL

Institutional Review Board Approval through the University of Indianapolis #906
Submitted: March 17, 2021 CST, Accepted: August 05, 2021 CST
REFERENCES


