Interobserver and intraobserver comparison of imaging glenoid morphology, glenoid version and humeral head subluxation

Paul J Cagle Jr¹ , Birgit Werner¹ , Dave R Shukla²,
Daniel A London³ , Bradford O Parsons³ and Neal L Millar⁴

Abstract

Background: Glenoid morphology, glenoid version and humeral head subluxation represent important parameters for the treating physician. The most common method of assessing glenoid morphology is the Walch classification which has only been validated with computed tomography (CT).

Methods: CT images and magnetic resonance imaging (MRI) images of 25 patients were de-identified and randomized. Three reviewers assessed the images for each parameter twice. The Walch classification was assessed with a weighted kappa value. Glenoid version and humeral head subluxation were compared with a reproducibility coefficient.

Results: The Walch classification demonstrated almost perfect intraobserver agreement for MRI and CT images ($k = 0.87$). Weighted interobserver agreement values for the Walch classification were fair for CT and MRI ($k = 0.34$). The weighted reproducibility coefficient for glenoid version measured 9.13 (CI 7.16–12.60) degrees for CT and 13.44 (CI 10.54–18.55) degrees for MRI images. The weighted reproducibility coefficient for percentage of humeral head subluxation was 17.43% (CI 13.67–24.06) for CT and 18.49% (CI 14.5–25.52) for MRI images.

Discussion: CT and MRI images demonstrated similar efficacy in classifying glenoid morphology, measuring glenoid version and measuring posterior humeral head subluxation. MRI can be used as an alternative to CT for measuring these parameters.

Keywords

glenoid morphology, glenoid version, Walch classification, humeral head subluxation

Date received: 11th May 2017; revised: 4th February 2018; accepted: 7th February 2018

Introduction

The ability to classify glenoid bony morphology and quantify glenoid version and humeral head subluxation represents important diagnostic information for practicing clinicians. The most commonly used classification system for the assessment of glenoid morphology, in the context of primary osteoarthritis, was described by Walch et al.¹ This classification system is made up of three categories. Category A represents central erosion of the glenoid graded as a 1 or 2 depending on the severity. Category B represents posterior wear of the glenoid again graded as a 1 or 2 based on severity. Category C is used to describe a dysplastic glenoid. In 2016, the Walch classification underwent a modification to add the category of a B3 and D.² This classification

system has been shown to be useful as a prognostic indicator for outcomes following total shoulder arthroplasty, with recent data demonstrating that a significant change in glenoid version can correlate with inferior

¹Rhön-Klinikum AG, Bad Neustadt an der Saale, Germany

²Newport Orthopaedic Institute (affiliated with the Mayo Clinic), Newport Beach, USA

³Icahn School of Medicine at Mount Sinai Hospital, New York, USA

⁴University of Glasgow, Glasgow, UK

Corresponding author:

Paul J Cagle, Leni & Peter W. May Department of Orthopaedic Surgery, Icahn School of Medicine, Mount Sinai Medical Center, Mount Sinai West Hospital, 1000 Tenth Avenue, Suite 3A-35, New York, NY 10019, USA.
Email: paul.cagle@mountsinai.org

results following anatomic total shoulder arthroplasty. Consequently, there is an evolving recommendation to consider reverse total shoulder arthroplasty in certain patients depending on the glenoid morphology.³

The Walch classification was based on and has been validated using computed tomography (CT) scans, but has never been evaluated with magnetic resonance imaging (MRI) of the shoulder.^{4,5} Scalise et al.⁴ demonstrated a fair intrarater and moderate interrater reliability of the classification system when applied to CT images of 24 patients. Nowak et al.⁵ went on to examine 26 shoulders with CT imaging and demonstrated moderate interobserver reliability and substantial intraobserver reliability for the Walch classification.

Glenoid version and the percentage of humeral head subluxation are important parameters to consider prior to shoulder arthroplasty as well. Glenoid version represents the angle of the glenoid surface in relation to the scapula. Friedman et al.⁶ reported a method to accurately measure glenoid version utilizing CT images and reported normal values ranging from -2 to 11 degrees, with negative values representing retroversion. This method was later validated by Kidder et al.⁷

To our knowledge, no study has examined shoulder MRI images in reference to the reliability of the Walch classification, glenoid version measurement and humeral head subluxation measurements, although MRI has been utilized to assess posterior glenohumeral subluxation of the shoulder. One study had examined pediatric patients following an obstetric brachial plexus injury, and the authors did not seek to validate the method.⁸ Additionally, MRI offers the opportunity for increased visualization of the soft tissue, assessment of rotator cuff atrophy and avoidance of radiation exposure, and it is not uncommon for patients to arrive to a clinician's office having already obtained an MRI scan. Thus, a study validating the Walch classification and these measurements with MRI would prove useful. The purpose of this study is to compare CT and MRI images to determine if MRI can be used reliably to assess glenoid morphology, measure glenoid version and measure humeral head subluxation.

Materials and methods

Shoulder arthroplasty procedures occurring between September 2013 and July 2014 were examined. Twenty-five patients planned to undergo a total shoulder arthroplasty were identified. Only patients who had CT and MRI images within a six-month time frame were included. All images were obtained by one of the non-reviewing investigators. The images were deidentified and placed in a randomized order. The complete set of randomized images was sent to three reviewers. Each reviewer was a board certified

orthopaedic surgeon who also had completed a fellowship in shoulder surgery. Each investigator reviewed the images and classified the CT and MRI series for the Walch classification (A1, A2, B1, B2 or C), glenoid version measurement in degrees and posterior humeral head subluxation measured as a percentage. The images were then placed in a different random order and each reviewer repeated the analysis three months later in order to minimize the chance of recall bias.

Glenoid version measurements were performed with the method described by Friedman.⁶ The glenoid was assessed just inferior to the coracoid on axial images. A line is drawn down the long axis of the scapula, and a subsequent line is drawn perpendicular at the glenoid articular surface. A line is then drawn parallel with the glenoid surface. The angle difference between the perpendicular line and the line parallel to the glenoid surface represents the glenoid version. The value was reported in degrees, and retroversion was reported as a negative number. Humeral head subluxation was measured as a percentage using the scapula referencing method, and posterior humeral head subluxation was reported as a negative number.⁷ This was measured by a similar line drawn down the long axis of the scapula. This line is carried lateral through the humeral head. The distance from the line to the anterior humeral head and the posterior humeral head is measured. The anterior distance is then divided by the anterior distance added to the posterior distance and reported as a percentage.

Interobserver and intraobserver reliability for the Walch classification was determined utilizing a weighted kappa coefficient (k). The level of agreement was described by Landis and Koch.⁹ For the continuous variables (glenoid version angle and humeral head subluxation), the repeatability and reproducibility coefficients were chosen to determine intraobserver and interobserver agreements, respectively. These metrics were chosen to help facilitate comparisons between studies as they are consistent regardless of changing between subject (rater) variability. Additionally, this measure allows for the assignment of units, in our calculations degrees and percentage, to provide context and easy clinical interpretation to the calculated values.^{10,11} Confidence intervals (CIs) were calculated with each reliability coefficient to provide further support for each comparison. The sample size was also equal to the standards used to validate the Walch classification with CT scans.^{5,6}

Results

The weighted intraobserver agreements of the Walch classification was $k=0.87$ (almost perfect agreement) for CT and $k=0.87$ (almost perfect agreement) for

MRI images. Intraobserver kappa values for each reviewer are demonstrated with the weighted values and standard error calculations in Table 1. The weighted interobserver agreement values for the Walch classification were $k=0.34$ (fair agreement) for CT and $k=0.34$ (fair agreement) for MRI. This is demonstrated in Table 2 with standard error calculations.

The weighted reproducibility coefficient for glenoid version demonstrated a value expressed as a degree measurement with an associated CI. The values were 9.13 (CI 7.16–12.60) for CT images and 13.44 (CI 10.54–18.55) for MRI. The reproducibility coefficient measurement for each individual reviewer and the weighted value is demonstrated in Table 3 with associated CIs. The weighted reproducibility coefficient and associated CIs for percentage humeral head subluxation demonstrated a value of 17.43 (CI 13.67–24.06) for CT images and 18.49 (CI 14.5–25.52) for MRI images. The reproducibility coefficient measurement

for each individual reviewer and the weighted value is demonstrated in Table 4 with associated CIs.

Discussion

Defining glenoid morphology has important implications for clinicians planning for shoulder arthroplasty, and the national frequency of total shoulder arthroplasty continues to increase.¹² A CT scan to assess glenoid morphology is a common practice, and understanding the glenoid is of the utmost importance as multiple studies have demonstrated the glenoid component to be the most common site of shoulder arthroplasty component failure.^{13–16} An MRI offers the opportunity to avoid radiation and improved visualization of the soft tissue structures of the shoulder. Although MRI imaging often costs more than CT imaging, many patients present for consultation after already having had recent MRI images. The ability to use those images for preoperative planning avoids the extra cost of an additional CT, and it avoids exposing the patient to unnecessary radiation.

Walch et al.¹ defined a classification of glenoid morphology with a described high rate of intraobserver ($k=0.65$) and interobserver ($k=0.70$) reliability. Scalise et al.⁴ then went on to demonstrate fair intraobserver ($k=0.34$) and moderate interobserver ($k=0.44$) reliability of the Walch classification system when they examined the CT scans of 24 shoulders. Nowak et al.⁵ utilized CT images of 26 shoulder to demonstrate substantial intraobserver reliability ($k=0.611$) and moderate interobserver reliability ($k=0.508$) for the Walch classification. Our data demonstrated an even higher degree of intraobserver agreement with Walch classification having an almost perfect reliability ($k=0.87$) for MRI and CT images. Our interobserver agreement was similar to prior studies with fair agreement ($k=0.34$) for both MRI and CT images. These data demonstrate that both MRI and CT have reliability similar to other published rates when used to classify glenoid

Table 1. Intraobserver kappa values for Walch classification with standard error measurement (SEM).

	CT (SEM)	MRI (SEM)
Reviewer 1	0.94 (0.05)	0.9 (0.06)
Reviewer 2	0.83 (0.09)	0.7 (0.1)
Reviewer 3	0.84 (0.11)	1.0 (–)
Weighted kappa value	0.87	0.87

Table 2. Weighted interobserver kappa values for Walch classification with standard error measurement (SEM).

	CT	MRI
Weighted kappa	0.34 (0.08)	0.34 (0.08)

Table 3. Reproducibility coefficient for glenoid version measured in degrees.

	CT	MRI
Reviewer 1	9.68 (7.59–13.37)	8.91 (6.98–12.29)
Reviewer 2	8.4 (6.59–11.6)	10.21 (8.01–14.09)
Reviewer 3	5.11 (4.01–7.05)	4.05 (3.18–5.59)
Weighted value	9.13 (7.16–12.60)	13.44 (10.54–18.55)

Table 4. Reproducibility coefficient for humeral head subluxation measured in percentage.

	CT	MRI
Reviewer 1	6.97 (5.47–9.63)	7.61 (5.97–10.5)
Reviewer 2	14.38 (11.28–19.86)	13.46 (10.56–18.58)
Reviewer 3	17.51 (13.74–24.18)	13.28 (10.41–18.33)
Weighted value	17.43 (13.67–24.06)	18.49 (14.5–25.52)

Figure 1. CT and MRI images demonstrate the capacity to appreciate similar glenoid morphology.

Figure 2. CT and MRI images demonstrate the capacity to measure humeral head subluxation.

morphology utilizing the Walch classification system. Though in comparison, the intraobserver reliability in our study was higher than the interobserver reliability. A selection of images from one patient is demonstrated in Figure 1.

Glenoid retroversion and percentage of humeral head subluxation are important parameters to evaluate prior to shoulder arthroplasty and can even potentially indicate that a surgeon should use a different type of prosthesis. A recent report by Denard and Walch³ suggests considering 27° or more of glenoid retroversion or 80% or greater posterior humeral head subluxation as indications for a reverse total shoulder arthroplasty. Although the assessment of humeral head subluxation using MRI for patients with obstetric brachial plexus injury (OBPI) has been reported, no study has compared the reliability of measurements of humeral head subluxation and glenoid version using CT and MRI images.⁹

These data demonstrate excellent reproducibility of glenoid version measurement for CT and MRI images. The measurement difference for glenoid version between CT and MRI images was only 4.31° and the CIs overlapped for the data sets. This supports the hypothesis that an MRI or a CT can be utilized to reproducibly measure the degree of glenoid version. A similar trend was noted in the measurements of posterior humeral head subluxation with a measurement difference of only 1.06% with overlapping CIs between CT and MRI images. This supports the utilization of MRI measurements for humeral head subluxation. Figure 2 demonstrates a series of images from the same patient and this illustrates a similar percentage humeral head subluxation measurement for the CT and MRI images.

There are limitations to our study. The sample size was small at 25 and the time frame between reviews was three months. Although the small sample size and the limited time between reads could have led to bias, this sample size and the time frame has been used in the literature on two occasions to validate the Walch classification for intraobserver and interobserver reliability.^{4,5} Furthermore, all three reviewers were board certified orthopaedic surgeons with specific fellowship training in shoulder surgery. It is possible with less experience that the results would be less reliable. Finally, as there is no accepted standard for these measurements, this paper references a comparison of common methods.

Overall, this study demonstrated that MRI can be reliably utilized to classify glenoid morphology with the Walch classification, to measure glenoid version and to assess the percentage of humeral head subluxation. Based on these data, we believe a treating clinician may choose to utilize either an MRI or a CT to assess these parameters.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: BOP is a consultant for Arthrex and receives royalties from Arthrex. BOP is an editor for Journal of Bone and Joint Surgery Reviews.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical Review and Patient Consent

This study was approved by the institution's Internal Review Board. All images were presented in a de-identified manner.

ORCID iDs

Paul J Cagle Jr <http://orcid.org/0000-0002-9617-8369>

Birgit Werner <http://orcid.org/0000-0003-0324-2286>

Daniel London <http://orcid.org/0000-0002-1711-0762>

References

- Walch G, Badet R, Boulahia A, et al. Morphologic study of the glenoid in primary glenohumeral osteoarthritis. *J Arthroplasty* 1999; 6: 756–760.
- Bercik MJ, Kruse K, Yalozis M, et al. A modification to the Walch classification of the glenoid in primary glenohumeral osteoarthritis using three-dimensional imaging. *J Shoulder Elbow Surg* 2016; 25: 1601–1606.
- Denard PJ and Walch G. Current concepts in the surgical management of primary glenohumeral arthritis with a biconcave glenoid. *J Shoulder Elbow Surg* 2013; 22: 1589–1598.
- Scalise JJ, Codsi MJ, Brems JJ, et al. Inter-rater reliability of an arthritic glenoid morphology classification system. *J Shoulder Elbow Surg* 2008; 17: 575–577.
- Nowak DD, Gardner TR, Bigliani LU, et al. Interobserver and intraobserver reliability of the Walch classification in primary glenohumeral arthritis. *J Shoulder Elbow Surg* 2010; 19: 180–183.
- Friedman RJ, Hawthorne KB and Genev BM. The use of computerized tomography in the measurement of glenoid version. *J Bone Joint Surg Am* 1992; 74: 1032–1037.
- Kidder JF, Rouleau DM, Pons-Villanueva J, et al. Humeral head posterior subluxation on CT scan: validation and comparison of 2 methods of measurement. *Tech Should Surg* 2010; 11: 72–76.
- van Geleijn Vittinga VM, van Kooten EO, Margriet G, et al. An MRI study on the relations between muscle atrophy, shoulder function and glenohumeral deformity in shoulders of children with obstetric brachial plexus injury. *J Brachial Plex Peripher Nerve Inj* 2009; 18: 5.
- Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
- Vaz S, Falkmer T, Passmore AE, et al. The case for using the repeatability coefficient when calculating test-retest reliability. *PLOS One* 2013; 8: e73990.

11. Bartlett JW and Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol* 2008; 31: 466–475.
12. Kim SH, Wise BL, Zhang Y, et al. Increasing incidence of shoulder arthroplasty in the United States. *J Bone Joint Surg Am* 2011; 93: 2249–2254.
13. Torchia ME, Cofeld RH and Settergren CR. Total shoulder arthroplasty with the Neer prosthesis: long-term results. *J Shoulder Elbow Surg* 1997; 6: 495–505.
14. Kasten P, Pape G, Raiss P, et al. Mid-term survivorship analysis of a shoulder replacement with a keeled glenoid and a modern cementing technique. *J Bone Joint Surg Br* 2010; 92: 387–392.
15. Neer CS 2nd, Watson KC and Stanton FJ. Recent experience in total shoulder replacement. *J Bone Joint Surg Am* 1982; 64: 319–337.
16. Bohsali KI, Wirth MA and Rockwood CA Jr. Complications of total shoulder arthroplasty. *J Bone Joint Surg Am* 2006; 88: 2279–2292.