



Reliability and discriminative accuracy of 5 measures for craniocaudal humeral position: an assessment on conventional radiographs

Arjen Kolk, MD ^{a,b}, Celeste L. Overbeek, MD ^{a,b}, Jurriaan H. de Groot, PhD ^b,
Rob G.H.H. Nelissen, MD, PhD ^{a,b}, Jochem Nagels, MD ^{a,b,*}

^a Department of Orthopaedics, Leiden University Medical Center, Leiden, The Netherlands

^b Laboratory for Kinematics and Neuromechanics, Department of Orthopaedics and Rehabilitation, Leiden University Medical Center, Leiden, The Netherlands

ARTICLE INFO

Keywords:

Rotator cuff
shoulder
rotator cuff tears
supraspinatus
radiography
reliability (epidemiology)

Level of evidence: Level IV; Case-Control
Design; Diagnostic Study

Hypothesis: This study aimed to examine the reliability and diagnostic discriminative accuracy of 5 different methods that quantify the craniocaudal humeral position with respect to the scapula on conventional radiographs.

Methods: In this retrospective, cross-sectional diagnostic study, 2 observers randomly assessed the conventional anteroposterior shoulder radiographs of 280 subjects with rotator cuff imaging for the (1) acromiohumeral (AH) interval, (2) upward migration index (UMI), (3) glenohumeral center-to-center measurement (GHCC), (4) glenohumeral arc measurement (GHa), and (5) scapular spine–humeral head center method (SHC). Reliability was assessed by means of relative consistency (intraclass correlation coefficient) and absolute consistency. Discriminative accuracy for detecting a rotator cuff tear was calculated.

Results: Relative consistency (intraclass correlation coefficient) for the AH interval, UMI, GHCC, GHa, and SHC was 0.961, 0.913, 0.806, 0.924, and 0.726, respectively. The AH interval had the highest absolute consistency with a random residual measurement error of 0.58 mm compared with 1.0–3.2 mm for the other measurements. The discriminative accuracy of the AH interval did not significantly differ from that of the UMI (−0.010; 95% confidence interval [CI], −0.042 to 0.022; $P = .545$) but was significantly better than that of the GHCC (0.112; 95% CI, 0.043–0.181; $P = .001$), GHa (0.074; 95% CI, 0.009–0.139; $P = .027$), and SHC (0.178; 95% CI, 0.100–0.256; $P < .001$).

Conclusion: Assessment of the craniocaudal humeral position is performed with good to excellent intraobserver and interobserver reliability. The discriminative accuracy for detecting a rotator cuff tear on a single radiograph was highest for the AH interval and UMI. We recommend using the AH interval or UMI as an indirect measure of the presence of a rotator cuff tear on conventional radiographs.

© 2019 The Authors. Published by Elsevier Inc. on behalf of American Shoulder and Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The incidence of shoulder complaints is 11–29 per 1000 person-years, with higher rates in elderly persons.^{12,29} Shoulder complaints are most commonly related to subacromial inflammation or a rotator cuff (RC) tear.^{29,30} RC tears have a prevalence of 13% in the fifth decade of life, and this prevalence further increases up to 50% in elderly persons, resulting in a great social but also economic burden in the United States.^{19,27,31,35} Medical treatment of RC tears coincides with huge preoperative health care costs especially for

diagnostic imaging of the RC by means of magnetic resonance imaging (MRI).³⁶ This social and economic burden underlines the need for more cost-effective ways to diagnose RC disorders in patients.

Conventional radiographs are an available and inexpensive first diagnostic step to differentiate between shoulder pathologies and may help to identify an RC tear through the assessment of the acromiohumeral (AH) interval.^{4,8,10,17,32} Whether AH interval measurements on conventional anteroposterior (AP) radiographs are of added value in the diagnosis of RC tears depends on the random measurement error (ie, absolute consistency), the observed measurement error with respect to the total amount of variability (ie, relative consistency), and the closeness of the measurement to the true subacromial distance (ie, accuracy).^{24,26,33} The latter has previously been evaluated through the calculation of

The Medical Ethical Committee of the Leiden University Medical Center approved this retrospective study (no. P15.044) and waived the requirement for written informed consent.

* Corresponding author: Jochem Nagels, MD, Leiden University Medical Center, Postzone J11-R, PO Box 9600, 2300 RC Leiden, The Netherlands.

E-mail address: j.nagels@lumc.nl (J. Nagels).

<https://doi.org/10.1016/j.jseint.2019.11.005>

2666-6383/© 2019 The Authors. Published by Elsevier Inc. on behalf of American Shoulder and Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

the structural difference in the AH interval between radiographs and MRI—computed tomography (CT).^{28,34} Moreover, magnification and projection have been revealed to introduce measurement error, which subsequently led to the recommendation of methods that should correct for magnification (ie, upward migration index [UMI]) or projection (spino-humeral head center method).^{21,28}

Nowadays, several measurement methods exist to assess the position of the humerus relative to the scapula (ie, further referred to as “proximal migration”): AH interval, UMI, glenohumeral center-to-center measurement (GHCC), glenohumeral arc measurement (GHa), and scapular spine—humeral head center method (SHC).^{1,6,9,18,21,28} Remarkably, the relative consistency has been evaluated in patients only for the AH interval; nonetheless, the absolute consistency was not reported.¹³ Moreover, the amount of random measurement error that is associated with the UMI, GHCC, GHa, and SHC—and thus their reliability—remains unknown. Therefore, it has still to be determined whether the introduction of these methods has resulted in improved reliability and higher discriminative accuracy for detecting an RC tear.

The primary aim of this study was to assess the reliability and diagnostic discriminative accuracy of 5 measures (ie, AH interval, UMI, GHCC, GHa, and SHC) of proximal migration of the humerus in detecting RC tears on a single conventional radiograph. Because the AH interval is an easy applicable technique and is associated with high relative reliability, we hypothesized that the AH interval measurement would have superior reliability and diagnostic accuracy for detecting an RC tear compared with the other measurements.

Materials and methods

Participants

In this retrospective, cross-sectional diagnostic study, we evaluated subjects with (1) subacromial pain syndrome (SAPS), (2) an isolated full-thickness supraspinatus tear, (3) a massive posterolateral RC tear involving the supraspinatus and infraspinatus muscle, (4) a massive anterosuperior RC tear involving the subscapularis and supraspinatus muscle, (5) asymptomatic shoulders (ie, healthy controls), or (6) radiographic osteoarthritis of the glenohumeral joint.

We identified a consecutive series of eligible subjects with shoulder pain who visited the outpatient clinic at the Leiden University Medical Center through screening of the nationwide diagnosis-related financial coding system from August 2005 to October 2015. Radiographs of healthy controls were selected from a prior study.⁵

We applied the following general inclusion criteria: shoulder pain for at least 3 months and the availability of a true AP radiograph. Furthermore, the quality of conventional radiographs was assessed regarding the presence of the scapular spine because two-thirds of the scapular spine was required for some of our outcomes.^{11,15} Finally, because we aimed to assess the discriminative value of our outcome measures in the diagnosis of RC tears, patients in the subgroups with RC-related complaints (subgroups 1–4) were required to have undergone an evaluation of the RC with ultrasonography or MRI. The RC was assessed for signs of tendinitis and tears of the supraspinatus, infraspinatus, or subscapularis muscle for correct classification according to our subgroups by a radiologist. The general exclusion criteria were the presence of or a history of shoulder surgery, tumor, cervical radiculopathy, frozen shoulder, fracture within the shoulder region, muscular dystrophy, or rheumatologic disease.

The diagnosis was obtained by an orthopedic surgeon after evaluation of patient history, physical examination findings,

radiography, and ultrasound or MRI. All patients with SAPS (subgroup 1) experienced movement-related pain and pain at night in combination with positive Hawkins and Neer impingement tests. On imaging, all patients had tendinitis, a partial-thickness RC tear, or bursitis, whereas patients with a full-thickness RC tear, labral tear, and biceps tendinopathy were excluded. Patient with an RC tear (subgroups 2–4) had movement-related shoulder pain, a positive Hawkins test, a positive Neer impingement test, and a full-thickness RC tear on ultrasound or MRI.

Each diagnostic subgroup comprised 60 patients, except for the healthy control subgroup (n = 10) and osteoarthritis subgroup (n = 30). In this way, 280 shoulder radiographs from 280 different subjects with 6 different clinical entities were analyzed.

Outcome measures

All outcome measures are depicted in [Figure 1](#). Five methods were used to quantify the craniocaudal position of the humerus with respect to the scapula (ie, proximal migration): (1) The AH interval was defined as the shortest distance measured from the cortical undersurface of the anterior part of the acromion and the humeral head.^{2,4,10,22,32} (2) The UMI was obtained by dividing the distance between the geometric center of the humeral head and the cortical undersurface of the acromion by the radius of the humeral head.^{18,28} (3) The GHCC was defined as the distance from the geometric center of the humeral head to a perpendicular line running through the middle of the articular surface of the glenoid fossa (positive when pointing superiorly).^{6,9} (4) The GHa—originally designed for the evaluation of shoulder arthroplasty and later modified for the assessment of the native shoulder¹—was defined as the distance from the anatomic neck of the humeral head to the inferior rim of the glenoid (positive when pointing superiorly).¹ (5) The SHC was defined as the perpendicular distance between a line running through the straight part of the supraspinatus fossa floor and the geometric center of the humeral head (positive when pointing inferiorly).^{15,21} Distances were expressed in millimeters.

We also assessed the influence of the cranial-caudal projection on radiographs on proximal migration measurements as described in cadavers by Nagels et al.²¹ To measure the cranial-caudal projection, the projection of the coracoid process with respect to a reference line drawn through the supraspinatus fossa floor was evaluated ([Fig. 2](#)).²¹ The coracoid process projects either below (cranial projection), on (neutral projection), or well above (caudal projection) the reference line.

Radiologic assessments

Standard true AP radiographs (ie, oblique view) with the arm in the neutral position and the hand in the anatomic position (ie, external rotation) were selected and retrieved from the hospital's picture archiving and communication system. Radiographs were made with the patients standing at a film distance of approximately 120 mm and with 15° of craniocaudal tilt. Radiographs were processed with CXDI Control Software NE (Canon Europe, Amstelveen, The Netherlands), stored in DICOM (Digital Imaging and Communications in Medicine) file format, and assessed using Digimizer software (version 4.6.1 [2005–2016]; MedCalc Software, Ostend, Belgium). Each measurement was conducted on a blank radiograph. Two observers (A.K. and C.L.O.) independently assessed the outcome parameters on each of the 280 radiographs in a random sequence. At 1 to 7 days (median, 3 days) after the first session, both observers assessed the outcome parameters in a second session. In total, each outcome measure was rated 4 times, resulting in 1120 observations per method.

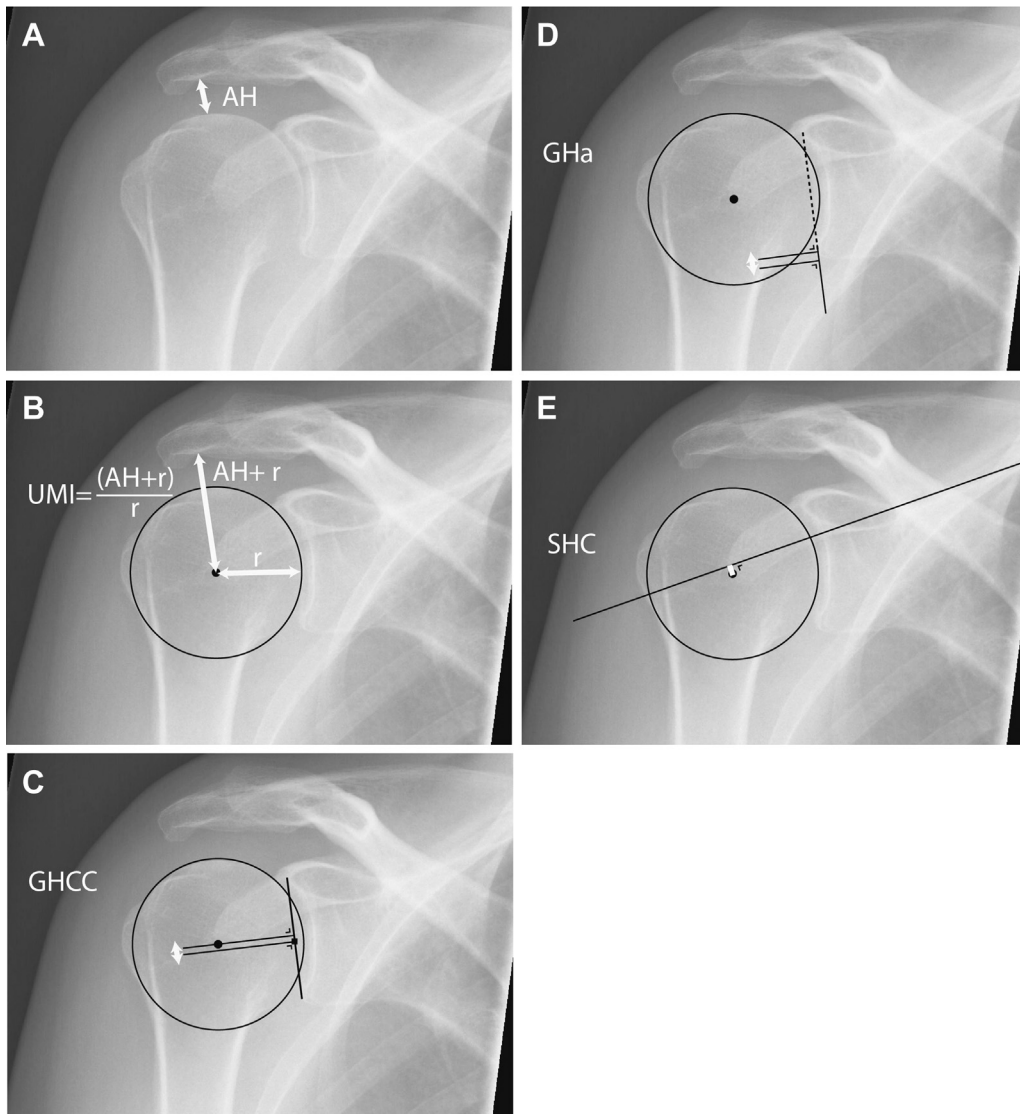


Figure 1 Radiographic assessments of glenohumeral joint. (A) Acromiohumeral interval (AH). (B) Upward migration index (UMI). This measure was calculated as the sum of the AH and radius (r) of the humeral head divided by r , that is, $(AH + r)/r$. ○: a circle fitting the humeral head ●: geometric center of rotation. (C) Glenohumeral center-to-center measurement (GHCC). ○, a circle fitting the humeral head ○: geometric center of the humeral head. ■, middle of the articular surface of the glenoid. —: two perpendicular lines to a line running to the glenoid fossa floor, one line through the center of the humeral head and the second perpendicular line through the middle of the articular surface of the glenoid. (D) Glenohumeral arc measurement (GHa). ○, circle fitting the humeral head ●: center of the humeral head. —, line touching the glenoid articular surface. —, perpendicular lines at the level of the inferior rim of the glenoid and at the level of the anatomical neck of the humerus. (E) Scapular spine–humeral head center method (SHC). ●, Circle fitting the humeral head. ○, geometric center of the humeral head. —, line running through the straight part of the fossa floor.

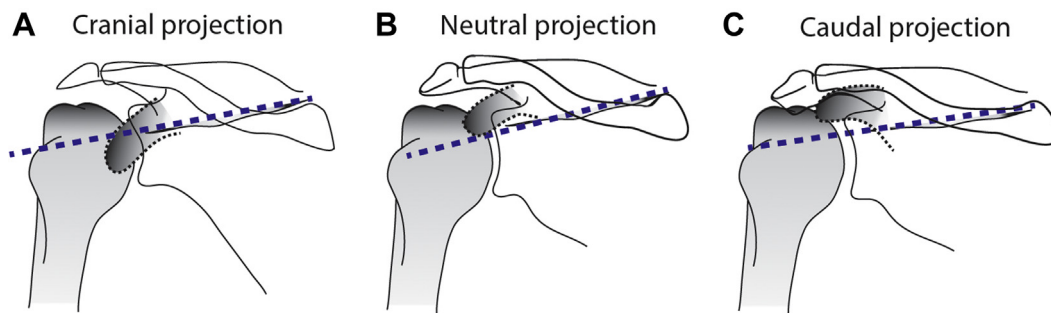


Figure 2 Coracoid projection (—). The projection of the coracoid process was evaluated with respect to a line drawn through the supraspinatus fossa floor (—) according to the method described by Nagels et al.²¹ (A) When the coracoid process was projecting below the (—), this indicated a cranial projection. (B) Neutral projection. (C) When the coracoid process was projecting well above the —, this indicated a caudal projection.

Blinding and randomization of radiographs

Radiographs were anonymized by removing all information that could expose the patient's identity. The order of radiographs was randomized with the purpose to blind the assessor to the clinical diagnosis, as well as to ensure that learning effects could not influence outcomes across the diagnostic subgroups. Randomization was performed using random permutation in MATLAB (version 2013b; The MathWorks, Natick, MA, USA).

Four lists of 280 random unique integers were generated (ie, 1 list per session for both observers). Custom-made software was used for saving the original DICOM files according to these random lists. The order of randomization was revealed following all measurements. Correct randomization was checked at the end of the study ([Supplementary Fig. S1](#)).

Reliability indices

Assessments were evaluated for random errors in light of the total amount of variability within the population (ie, relative consistency) and for absolute random errors among repeated assessments (ie, absolute consistency).³³ Relative consistency is usually expressed with the intraclass correlation coefficient (ICC), whereas absolute consistency can be expressed as the standard error of measurement (SEM) or root-mean-square error (RMSE).^{24,33}

The ICC is a ratio of the true variability in the population to the total variance as observed.^{26,33} This ratio not only assesses the level of agreement between consecutive ratings but also represents the existing real variance between patients (equation 1).^{7,24,26,33} When the real variance among patients increases or when the residual variability decreases, the ICC increases.²⁶ Consequently, the ICC represents a signal-to-noise ratio and will depend on the variance within the population of interest.^{24,26,33} The ICC provides information on the ability to differentiate individuals with a disease within the population.³³

Absolute consistency among measurements is frequently referred to as “precision” and has the advantage that it can be expressed in units of the metric system. Absolute consistency was quantified using the SEM.³³ The SEM is less dependent on the variability within the group and is expressed through the same scale as the outcome measure.^{26,33} When the ICC is used to calculate the SEM (equation 2), structural errors between sessions may have an effect on the SEM.³³ To account for this, the RMSE for the residuals can be used to describe the absolute consistency.³³ The RMSE does not rely on the ICC and contains the random residual error of a measurement.³³ In a similar way, the RMSE for different sessions (ie, $RMS_{\text{within observers}}$) and the RMSE between observers (ie, $RMS_{\text{between observers}}$) were used to describe variation. The SEM and RMSE provide information on whether changes are real. Hence, these properties were used to obtain the smallest change that is not expected to be the result of measurement error and thus can be considered a real change in the outcome measure, also known as the “minimal detectable change” (MDC).^{16,33}

Because variance may differ between diagnostic subgroups, the ICC, SEM (when calculated with an ICC), and MDC were first calculated within each of the diagnostic subgroups ([Supplementary Table S1](#)). We found consistent indices among the 6 subgroups per outcome measure. Therefore, we present overall indices in the “Results” section, whereas all indices per subgroup are presented in [Supplementary Table S1](#).

Statistical analysis

Means and differences among subgroups

Means and 95% confidence intervals (CIs) were calculated per diagnostic subgroup. The differences in measurements were determined with a linear mixed model. Covariance within the residuals of 4 repeated measures was modeled with an unstructured covariance matrix. The dependent variable was 1 of our 5 outcome measures (eg, AH interval). The fixed factors were as follows: session, observer, session \times observer, and subgroup.

Evaluation of reliability indices

Two-way random-effects models with absolute agreement were applied to calculate the intraobserver and interobserver reliability ($ICC_{2,1}$) according to Shrout and Fleiss.²⁴ In addition, we calculated the overall reliability ratio (ICC_{overall}), standard deviation (SD), SEM, RMSE, and MDC. A detailed description and formulas are provided in [Supplementary Appendix S1](#). Although there is no consensus as to the interpretation of the ICC, we used the following criteria to interpret the ICC: 0.00–0.39, poor; 0.40–0.59, fair; 0.60–0.74, good; and 0.75–1.00, excellent.^{3,33}

Evaluation of discriminative accuracy

“Discriminative accuracy” refers to the ability of a measure to distinguish patients with disease from patients without disease.²⁵ It can be assessed with receiver operating characteristics and the area under the curve (AUC).²⁵ Because the presence of osteoarthritis is usually evident on radiographs and because asymptomatic subjects are not exposed to radiography in a clinical setting, we only evaluated the discriminative accuracy of our measurement methods using the subgroups with RC-related complaints (subgroups 1–4). Given that the AH distance is the most common way to evaluate proximal migration, we compared the AUCs of the other measuring methods with the AUC of the AH interval. Because AUCs were derived from the same set of radiographs, we accounted for correlated data.¹⁴ The optimal cutoff value was set at the highest sum of sensitivity and specificity.

Evaluation of cranial-caudal projection variation

Linear regression analysis was performed with cranial-caudal translation of the humerus as the dependent variable to evaluate the association between the cranial-caudal projection and proximal migration measurements in patients with RC-related complaints (subgroups 1–4). The independent variable was cranial, neutral, or caudal projection of the coracoid with respect to the scapular spine. All our outputs were obtained using IBM SPSS Statistics for Windows (version 20.0; IBM, Armonk, NY, USA). A 2-sided $P < .05$ was considered statistically significant.

Results

Means and between-group differences

The patient characteristics of our 6 subgroups are described in [Table I](#). Means for proximal migration measures are described in [Supplementary Table S2](#). The SAPS subgroup had a significantly larger mean AH interval compared with the subgroups with a supraspinatus tear (1 mm; 95% CI, 0.46–2.2 mm) or massive RC tear (3 mm; 95% CI, 1.9–3.7 mm). A comparable pattern was found for the UMI, but differences between the SAPS and supraspinatus tear subgroups were less prominent for the GHCC, GHa, and SHC ([Supplementary Table S3](#)).

Table I
Characteristics of diagnostic subgroups

	Subgroups with RC-related complaints				Healthy controls (n = 10)	Glenohumeral osteoarthritis (n = 30)
	SAPS (n = 60)	Supraspinatus tear (n = 60)	Massive posterolateral RC tear (n = 60)	Massive anterolateral RC tear (n = 60)		
Age (SD), yr	49 (6.5)	62 (11.9)	59 (8.6)	62 (8.4)	50 (6.6)	68 (12.9)
Female, n (%)	29 (48)	27 (45)	27 (45)	25 (42)	5 (50)	20 (67)
Left side affected, n (%)	24 (40)	26 (43)	19 (32)	18 (30)	6 (60)	11 (37)
RC imaging, n (%)						
MRI	15 (25)	38 (63)	56 (93)	57 (95)	NA	NA
Ultrasound	45 (75)	22 (37)	4 (7)	3 (5)	NA	NA

RC, rotator cuff; SAPS, subacromial pain syndrome; SD, standard deviation; MRI, magnetic resonance imaging; NA, not applicable.

Reliability indices

Relative consistency of proximal migration measurements was excellent, with ICCs larger than 0.8 for the AH interval, UMI, GHCC, and GHa (Table II) for all of the diagnostic subgroups (Fig. 3, Supplementary Table S1). The overall ICC for the SHC was good. The highest overall relative consistency (ie, ICC) was found for the AH interval measurement (ICC, 0.961). Absolute consistency was best for the AH interval, with a small random residual measurement error of 0.58 mm. The random residual measurement error was larger for the GHCC (1.0 mm), GHa (1.4 mm), and SHC (3.2 mm).

Discriminative accuracy

Receiver operating characteristic curves of proximal migration measurements are presented in Figure 4. The AUCs of the AH interval measurement and UMI were higher than the AUCs of the GHCC, GHa, and SHC (Table III). The AH interval had a significantly higher discriminative accuracy (Table III) than the GHCC ($P = .001$), GHa ($P = .027$), and SHC ($P < .001$). Although the UMI was designed to improve diagnostic accuracy via the adjustment for the magnification factor, this correction did not lead to a significant change in discriminative accuracy in a clinical group of patients (Table III). The optimal cutoff value for detecting an RC tear was 9.8 mm and 1.32 for the AH interval and UMI, respectively.

Evaluation of cranial-caudal projection variation

We found an association between the projections of the coracoid and the measures of proximal migration in the 4 diagnostic subgroups with RC-related complaints (Supplementary Figure S2). AH interval, UMI, and GHa values were lower in the case of a caudal

projection of the coracoid in each subgroup. In contrast, a cranial projection of the coracoid process was associated with a smaller GHCC and SHC in each subgroup. In the total study sample (ie, combining subgroups 1–4), we found that a caudal projection was significantly related to a smaller AH interval and UMI. A cranial projection was significantly correlated with a smaller SHC (Table IV).

Discussion

This study evaluated the reliability and discriminative accuracy of 5 methods to assess proximal migration of the humerus on radiographs in a clinical setting. Our quantitative evaluation demonstrated good to excellent relative consistency scores for all 5 methods. The AH interval method had the highest absolute consistency as indicated by the SEM and RMSE. We demonstrated superior discriminative accuracy for the AH interval and UMI by using the AUC. Furthermore, we showed that cranial-caudal projection errors of the coracoid are associated with the outcome of the AH interval and UMI measurements on radiographs.

Of the 5 measurements used to assess proximal migration, the AH interval method had the best absolute consistency and highest relative consistency. The interobserver measurement error associated with the AH interval is in line with the measurement error between observers as reported in the literature (range, 0–4 mm).¹³ Because variation between subjects within the population (eg, 6–14 mm) is generally larger than the maximum demonstrated difference between observers, we may deduce high intraobserver and interobserver ICCs, which are consistent with the excellent relative consistency in our study.^{4,10,13} Prior work indicated a measurement error of less than 1 mm between observers and excellent relative consistency for the AH interval method when assessing proximal

Table II
Reliability of outcome measures

	AH interval, mm	UMI ($[\text{AH interval} + r]/r$)	GHCC, mm	GHa, mm	SHC, mm
Intraobserver reliability (95% CI)	0.960 (0.952–0.968)	0.913 (0.897–0.928)	0.794 (0.716–0.848)	0.924 (0.909–0.937)	0.728 (0.663–0.781)
Interobserver reliability (95% CI)	0.962 (0.954–0.969)	0.912 (0.895–0.927)	0.830 (0.800–0.857)	0.924 (0.910–0.937)	0.722 (0.649–0.780)
ICC overall (95% CI)	0.961 (0.953–0.968)	0.913 (0.896–0.927)	0.806 (0.749–0.849)	0.924 (0.909–0.937)	0.726 (0.658–0.780)
SD	3.0	0.111	2.6	5.0	6.5
RMS _{patients}	5.8	0.214	4.8	9.7	12
RMS _{within observers}	2.1	0.0342	9.1	3.9	21
RMS _{between observers}	0.11	0.0738	0.81	3.0	23
RMSE	0.58	0.0327	1.0	1.4	3.2
SEM	0.58	0.0326	1.1	1.4	3.4
MDC	1.6	0.0905	3.1	3.8	9.4

AH, acromioclavicular; UMI, upward migration index; r , radius; GHCC, glenohumeral center-to-center measurement; GHa, glenohumeral arc measurement; SHC, scapular spine–humeral head center method; ICC, intraclass correlation coefficient; CI, confidence interval; SD, standard deviation; RMS, root-mean-square; RMSE, root-mean-square error; SEM, standard error of measurement; MDC, minimal detectable change; RMS_{patients}, root mean square error for between-patient variance; RMS_{within observers}, root mean square error between sessions; RMS_{between observers}, root mean square error between observers.

The properties of the AH interval, UMI, GHCC, GHa, and SHC are presented for the entire group, comprising 280 radiographs.

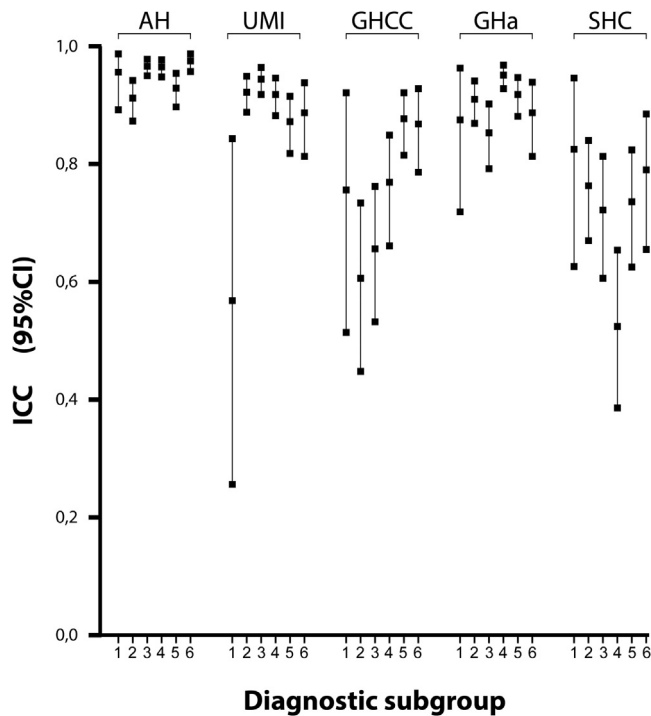


Figure 3 Intra-class correlation coefficients. The plot shows the intra-class correlation coefficient (ICC) with its 95% confidence interval (CI) in each diagnostic subgroup: (1) healthy controls, (2) subacromial pain syndrome, (3) isolated full-thickness supraspinatus tear, (4) massive posterosuperior rotator cuff tear, (5) massive antero-superior rotator cuff tear, and (6) osteoarthritis. AH, acromiohumeral interval; UMI, upward migration index; GHCC, glenohumeral center-to-center measurement; GHa, glenohumeral arc measurement; SHC, scapular spine–humeral head center method.

migration in cadavers.²¹ Nagels et al.²¹ found comparable intra-observer and interobserver relative consistency scores for GHa. In contrast to our results, Nagels et al observed higher values for the GHCC method and SHC. The laboratory-controlled conditions including the application of tantalum markers without over-projection of the soft tissues in patients may have caused this discrepancy.²¹ Identification of anatomic scapular landmarks used in the SHC could be more difficult on conventional radiographs, leading to larger measurement errors in our observations. It is interesting to note that these previous studies predominantly described the relative consistency of measurement for proximal migration in relation to the variability within the population using the ICC_{2,1} value. In this study, we added absolute consistency and RMSE to the literature, both of which do not mask poor consistency due to high variability between subjects.

The AH distance and UMI have significantly higher discriminative accuracy than the GHCC, GHa, and SHC on conventional radiographs. The cranial-caudal projection of the coracoid is associated with the outcome of the AH interval and UMI measurement. This finding supports the use of the coracoid base for the

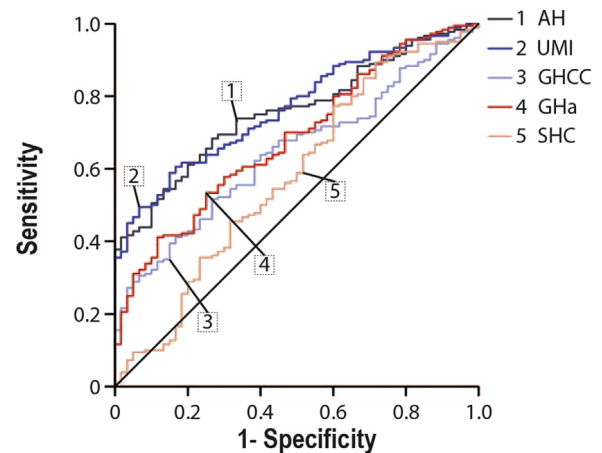


Figure 4 The area under the curve was calculated for the 5 measurement methods to evaluate the discriminative value of these methods to classify patients as having or not having a rotator cuff tear. AH, acromiohumeral interval; UMI, upward migration index; GHCC, glenohumeral center-to-center measurement; GHa, glenohumeral arc measurement; SHC, scapular spine–humeral head center method.

identification of projection errors and emphasizes the importance of projection in the assessment of the subacromial space as advocated by Nagels et al.²¹ Because the influence of the cranial-caudal projection on the measurement results was the lowest for the SHC, this measure has been proposed previously for consecutive measurements in a single patient.²¹ In the population, however, the identification of anatomic scapular landmarks on conventional radiographs can differ among patients.²¹ Lower errors in the identification of anatomic scapular landmarks using the AH interval or UMI may outweigh the impact of projection errors, leading to a higher discriminative accuracy of the AH interval and UMI in this study. Good to excellent inter-method agreement exists when comparing the subacromial distance measurements on radiographs and MRI-CT.^{28,34} Van de Sande and Roziog²⁸ showed that the correlation coefficient of the UMI improved with correction to the size of the humeral head. It should be noted that a reduction in within-subject variance by calibrating the radiograph and CT scan measures via the UMI might lead to a reduction in measurement error within a single patient, but it has no impact on the natural variance in anatomy within the population. In this way, the UMI will improve the correlation between radiographic and CT scanning measurements as demonstrated by van de Sande and Roziog. However, the UMI will not calibrate the measurements between subjects because the size of the humeral head is unknown and is likely to vary within the population. Consequently, the UMI is unable to eliminate all positioning and magnification errors on conventional radiographs, which might explain the higher relative consistency and discriminative accuracy for the AH interval than for the UMI.

In clinical practice, the AH interval and UMI are both reliable tools to assess proximal migration of the humeral head. The normal AH interval is within the range from greater than 6 mm to 14

Table III
Area under curve

	AUC (95% CI)	Mean difference (95% CI)	P value	Cutoff value	Sensitivity	Specificity
AH interval	0.757 (0.695-0.818)	Reference		9.8 mm	0.68	0.73
UMI ((AH interval + r)/r)	0.767 (0.606-0.828)	-0.010 (-0.042 to 0.022)	.545	1.32	0.61	0.83
GHCC	0.645 (0.572-0.718)	0.112 (0.043-0.181)	.001	0.3 mm	0.52	0.73
GHa	0.683 (0.609-0.756)	0.074 (0.009-0.139)	.027	0.2 mm	0.41	0.88
SHC	0.579 (0.492-0.666)	0.178 (0.100-0.256)	<.001	5.2 mm	0.89	0.28

AUC, area under curve; CI, confidence interval; AH, acromiohumeral; UMI, upward migration index; r, radius; GHCC, glenohumeral center-to-center measurement; GHa, glenohumeral arc measurement; SHC, scapular spine–humeral head center method.

Table IV
Projection of coracoid process and its correlation with measurements

Projection	Patients with RC disease		P value
	Mean	95% CI	
AH interval, mm			
Cranial	−0.26	−0.983 to −0.462	.478
Neutral	Reference		
Caudal	−1.9	−3.33 to −0.420	.012*
UMI ([AH interval + r]/r)			
Cranial	−0.0259	−0.0530 to 0.0012	.061
Neutral	Reference		
Caudal	−0.0557	−0.1102 to −0.0011	.045*
GHCC, mm			
Cranial	−0.60	−1.22 to 0.030	.062
Neutral	Reference		
Caudal	0.47	−0.784 to 7.32	.458
GHa, mm			
Cranial	0.24	−1.06 to 1.54	.717
Neutral	Reference		
Caudal	−2.2	−4.84 to 0.39	.096
SHC, mm			
Cranial	−1.7	−3.20 to −0.131	.033*
Neutral	Reference		
Caudal	1.2	−1.85 to 4.33	.429

RC, rotator cuff; CI, confidence interval; AH, acromiohumeral; UMI, upward migration index; r, radius; GHCC, glenohumeral center-to-center measurement; GHa, glenohumeral arc measurement; SHC, scapular spine–humeral head center method.
* Statistically significant.

mm.^{4,10,22,32} An AH interval of less than 6 mm is generally considered indicative of an RC tear.^{4,10,22,32} Although an AH interval of less than 6 mm is associated with an RC tear, this cutoff value is potentially too strict and misclassifies patients with a tear as having a normal RC. In our study, the cutoff value for the AH interval measurements was calculated based on sensitivity and specificity. We found a cutoff value with optimal sensitivity and specificity of 9.8 mm for the AH interval. This is considerably larger than the 6-mm value described in the literature. For clinical purposes, however, the positive predictive value (PPV) and negative predictive value (NPV) are important to determine a threshold. The PPV and NPV are influenced by the prevalence of disease and will vary between hospitals.²³ As a consequence, future studies should identify the optimal diagnostic cutoff value that can be used to confirm the presence of an RC tear (ie, PPV) and to rule out an RC tear (ie, NPV) on a single radiograph, resulting in a more effective use of health care expenditures.

This study has some limitations. First, the presence of an RC tear was determined with either ultrasound or MRI, and this might have resulted in verification bias. However, previous research indicated that ultrasound and MRI have comparable accuracy for diagnosing full-thickness RC tears.²⁰ Second, we assessed the effect of projection errors within different patients; projection errors within a single patient were not recorded. For clinical follow-up of patients, projection errors within a single patient will add variance influencing the psychometric properties (eg, MDC) of the measurement. The impact of this variance may differ among the 5 methods, with consequences for the reliability and consistency of these methods in follow-up studies. Third, we assumed that screening of the RC (MRI or ultrasound to assess for the presence of an RC tear) was performed independently from obtaining the radiographs. However, radiographic results can influence the decision to obtain further imaging, with a subsequent effect on the composition of the subgroups. Because orthopedic surgeons do not (fully) rely on subacromial space measurements in clinical practice, we believe this impact is limited. The selection may lead to a modification of variation within the subgroups between subjects and thus may eventually impact reliability indices.

Conclusion

Our study showed good to excellent relative consistency scores for all tested methods (AH interval, UMI, GHCC, GHa, and SHC). Absolute consistency was best for the AH interval, with a small random residual measurement error. Diagnostic accuracy was highest for the AH interval and UMI. Even though the UMI was designed to correct for magnification in the AH interval score, diagnostic accuracy of the UMI was not significantly superior to that of the AH interval method in the diagnosis of RC tears. Furthermore, we for the first time showed that the projection of the coracoid was associated with the outcome of the AH interval and UMI measurement on conventional radiographs and thus should be accounted for when evaluating shoulder radiographs. As a result of higher consistency and discriminative accuracy for the AH interval and UMI, we recommend the use of these methods when evaluating conventional shoulder radiographs for the presence of humeral cranialization as an indirect measure of a full-thickness RC tear.

Acknowledgments

The authors gratefully acknowledge Kanishk Kaushik for assisting in the patient selection by screening of the diagnosis-related financial codes.

Disclaimer

This study was funded by the Dutch Arthritis Association (grant No. 2013-1-303).

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Supplementary data

Supplementary data to this article can be found online at [10.1016/j.jseint.2019.11.005](https://doi.org/10.1016/j.jseint.2019.11.005).

References

- Aliabadi P, Weissman BN, Thornhill T, Nikpoor N, Sosman JL. Evaluation of a nonconstrained total shoulder prosthesis. *AJR Am J Roentgenol* 1988;151:1169–72.
- Azzoni R, Cabitza P, Parrini M. Sonographic evaluation of subacromial space. *Ultrasonics* 2004;42:683–7. <https://doi.org/10.1016/j.ultras.2003.11.015>.
- Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic* 1981;86:127–37.
- Cotton RE, Rideout DF. Tears of the humeral rotator cuff; a radiological and pathological necropsy survey. *J Bone Joint Surg Br* 1964;46:314–28.
- de Witte PB, Henseler JF, van Zwet EW, Nagels J, Nelissen RG, de Groot JH. Cranial humerus translation, deltoid activation, adductor co-activation and rotator cuff disease—different patterns in rotator cuff tears, subacromial impingement and controls. *Clin Biomech (Bristol, Avon)* 2013;29:26–32. <https://doi.org/10.1016/j.clinbiomech.2013.10.014>.
- Deutsch A, Altchek DW, Schwartz E, Otis JC, Warren RF. Radiologic measurement of superior displacement of the humeral head in the impingement syndrome. *J Shoulder Elbow Surg* 1996;5:186–93.
- Eliaszew 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;74:777–88.
- Flatow EL, Soslowky LJ, Ticker JB, Pawluk RJ, Hepler M, Ark J, et al. Excursion of the rotator cuff under the acromion. Patterns of subacromial contact. *Am J Sports Med* 1994;22:779–88.
- Franklin JL, Barrett WP, Jackins SE, Matsen FA III. Glenoid loosening in total shoulder arthroplasty. Association with rotator cuff deficiency. *J Arthroplasty* 1988;3:39–46.

10. Golding FC. The shoulder—the forgotten joint. *Br J Radiol* 1962;35:149–58.
11. Goud A, Segal D, Hedayati P, Pan JJ, Weissman BN. Radiographic evaluation of the shoulder. *Eur J Radiol* 2008;68:2–15. <https://doi.org/10.1016/j.ejrad.2008.02.023>.
12. Greving K, Dorrestijn O, Winters JC, Groenhof F, van der Meer K, Stevens M, et al. Incidence, prevalence, and consultation rates of shoulder complaints in general practice. *Scand J Rheumatol* 2012;41:150–5. <https://doi.org/10.3109/03009742.2011.605390>.
13. Gruber G, Bernhardt GA, Clar H, Zacherl M, Glehr M, Wurnig C. Measurement of the acromiohumeral interval on standardized anteroposterior radiographs: a prospective study of observer variability. *J Shoulder Elbow Surg* 2010;19:10–3. <https://doi.org/10.1016/j.jse.2009.04.010>.
14. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839–43.
15. Henseler JF, de Witte PB, de Groot JH, van Zwet EW, Nelissen RG, Nagels J. Cranial translation of the humeral head on radiographs in rotator cuff tear patients: the modified active abduction view. *Med Biol Eng Comput* 2014;52:233–40. <https://doi.org/10.1007/s11517-013-1057-2>.
16. Henseler JF, Kolk A, van der Zwaal P, Nagels J, Vliet Vlieland TP, Nelissen RG. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. *J Shoulder Elbow Surg* 2015;24:376–81. <https://doi.org/10.1016/j.jse.2014.07.003>.
17. Henseler JF, Raz Y, Nagels J, van Zwet EW, Raz V, Nelissen RG. Multivariate analyses of rotator cuff pathologies in shoulder disability. *PLoS One* 2015;10:e0118158. <https://doi.org/10.1371/journal.pone.0118158>.
18. Hirooka A, Wakitani S, Yoneda M, Ochi T. Shoulder destruction in rheumatoid arthritis. Classification and prognostic signs in 83 patients followed 5–23 years. *Acta Orthop Scand* 1996;67:258–63.
19. Iyengar JJ, Samagh SP, Schairer W, Singh G, Valone FH III, Feeley BT. Current trends in rotator cuff repair: surgical technique, setting, and cost. *Arthroscopy* 2014;30:284–8. <https://doi.org/10.1016/j.arthro.2013.11.018>.
20. Lenza M, Buchbinder R, Takwoingi Y, Johnston RV, Hanchard NC, Faloppa F. Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. *Cochrane Database Syst Rev* 2013;9:CD009020. <https://doi.org/10.1002/14651858.CD009020.pub2>.
21. Nagels J, Verweij J, Stokdijk M, Rozing PM. Reliability of proximal migration measurements in shoulder arthroplasty. *J Shoulder Elbow Surg* 2008;17:241–7. <https://doi.org/10.1016/j.jse.2007.07.011>.
22. Petersson CJ, Redlund-Johnell I. The subacromial space in normal shoulder radiographs. *Acta Orthop Scand* 1984;55:57–8.
23. Rothman KJ. *Epidemiology in clinical settings*. In: *Epidemiology: an introduction*. New York: Oxford University Press; 2012. p. 235–53.
24. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
25. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38. <https://doi.org/10.1097/EDE.0b013e3181c30fb2>.
26. Streiner DL, Norman GR. “Precision” and “accuracy”: two terms that are neither. *J Clin Epidemiol* 2006;59:327–30. <https://doi.org/10.1016/j.jclinepi.2005.09.005>.
27. Tempelhof S, Rupp S, Seil R. Age-related prevalence of rotator cuff tears in asymptomatic shoulders. *J Shoulder Elbow Surg* 1999;8:296–9.
28. van de Sande MA, Rozing PM. Proximal migration can be measured accurately on standardized anteroposterior shoulder radiographs. *Clin Orthop Relat Res* 2006;443:260–5. <https://doi.org/10.1097/01.blo.0000196043.34789.73>.
29. van der Windt DA, Koes BW, de Jong BA, Bouter LM. Shoulder disorders in general practice: incidence, patient characteristics, and management. *Ann Rheum Dis* 1995;54:959–64.
30. Vecchio P, Kavanagh R, Hazleman BL, King RH. Shoulder pain in a community-based rheumatology clinic. *Br J Rheumatol* 1995;34:440–2.
31. Vitale MA, Vitale MG, Zivin JG, Braman JP, Bigliani LU, Flatow EL. Rotator cuff repair: an analysis of utility scores and cost-effectiveness. *J Shoulder Elbow Surg* 2007;16:181–7. <https://doi.org/10.1016/j.jse.2006.06.013>.
32. Weiner DS, Macnab I. Superior migration of the humeral head. A radiological aid in the diagnosis of tears of the rotator cuff. *J Bone Joint Surg Br* 1970;52:524–7.
33. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19:231–40. <https://doi.org/10.1519/15184.1>.
34. Werner CM, Conrad SJ, Meyer DC, Keller A, Hodler J, Gerber C. Intermethod agreement and interobserver correlation of radiologic acromiohumeral distance measurements. *J Shoulder Elbow Surg* 2008;17:237–40. <https://doi.org/10.1016/j.jse.2007.06.002>.
35. Yamamoto A, Takagishi K, Osawa T, Yanagawa T, Nakajima D, Shitara H, et al. Prevalence and risk factors of a rotator cuff tear in the general population. *J Shoulder Elbow Surg* 2010;19:116–20. <https://doi.org/10.1016/j.jse.2009.04.006>.
36. Yerasiosian MG, Terrell RD, Wang JC, McAllister DR, Petrigliano FA. The costs associated with the evaluation of rotator cuff tears before surgical repair. *J Shoulder Elbow Surg* 2013;22:1662–6. <https://doi.org/10.1016/j.jse.2013.08.003>.