8-2021

Significant variability exists in preoperative planning software measures of glenoid morphology for shoulder arthroplasty

Alex R. Webb

Blake M. Bodendorfer

Nicholas C. Laucis
Henry Ford Health, nlaucis1@hfhs.org

David X. Wang

Daniel M. Dean

See next page for additional authors

Authors

Alex R. Webb, Blake M. Bodendorfer, Nicholas C. Laucis, David X. Wang, Daniel M. Dean, Joseph L. Rabe, Steven B. Soliman, Chad L. Klochko, Evan H. Argintar, David M. Lutton, and Brent B. Wiesel

SEMINARS IN ARTHROPLASTY 000 (2021) 1–11

Available online at www.sciencedirect.com
ScienceDirect
www.jsesarthroplasty.org

Significant variability exists in preoperative planning software measures of glenoid morphology for shoulder arthroplasty

Alex R. Webb, MD[a], Blake M. Bodendorfer, MD[b], Nicholas C. Laucis, MD[c], David X. Wang, MD[d], Daniel M. Dean, MD[b], Joseph L. Rabe, MD[b], Steven B. Soliman, DO[c], Chad L. Klochko, MD[c], Evan H. Argintar, MD[e], David M. Lutton, MD[f], and Brent B. Wiesel, MD[b,*]

[a]Emory University School of Medicine, Department of Orthopaedic Surgery, Atlanta, GA, USA
[b]MedStar Georgetown University Hospital, Department of Orthopaedic Surgery, Washington, DC, USA
[c]Henry Ford Health System, Department of Radiology, Detroit, MI, USA
[d]Allegheny General Hospital, Department of Orthopaedic Surgery, Pittsburgh, PA, USA
[e]MedStar Washington Hospital Center, Department of Orthopaedic Surgery, Washington, DC, USA
[f]Washington Circle Orthopaedic Associates, Washington, DC, USA

ARTICLE INFO

Keywords:
Total shoulder arthroplasty
Reverse shoulder arthroplasty
Glenoid deformity
Patient specific instrumentation
Glenoid version
glenoid inclination
3-dimensional computed tomography

ABSTRACT

Background & Hypothesis: We sought to assess the reliability of 4 different shoulder arthroplasty 3-dimensional preoperative planning programs. Comparison was also made to manual measurements conducted by 2 fellowship-trained musculoskeletal radiologists. We hypothesized that there would be significant variation in measurements of glenoid anatomy affected by glenoid deformity.

Methods: A retrospective review of computed tomography (CT) scans of patients undergoing shoulder arthroplasty was undertaken. A total of 76 computed tomographies were analyzed for glenoid version and inclination by 4 templating software systems (VIP, Blueprint, TrueSight, ExactechGPS). Inter-rater reliability was assessed via intra-class correlation coefficient (ICC). For those shoulders with glenohumeral arthritis (58/76), ICC was also calculated when sub-grouping by modified Walch classification. Lin's concordance correlation coefficient was calculated for each system with 2 musculoskeletal-trained radiologists' measurements.

Results: Measurements of glenoid version and inclination differed between at least 2 programs by 5°-10° in 75% and 92% of glenoids respectively, and by >10° in 18% and 45% respectively. ICC was excellent for version but only moderate for inclination. ICC was highest among Walch A glenoids for both version (near excellent) and inclination (good), and lowest among Walch D for version (near poor) and Walch B for inclination (moderate). When measuring version, VIP had the highest concordance with manual measurement; Blueprint had the lowest. For inclination Blueprint had the highest concordance; ExactechGPS had the lowest.

This study was approved by the Georgetown University Institutional Review Board (IRB# 2018-0004).

*Corresponding author: Brent B. Wiesel, MD, Department of Orthopaedic Surgery, MedStar Georgetown University Hospital, Georgetown University Medical Center, 3800 Reservoir Road Northwest, Ground Floor Pasquerilla Healthcare Center, Washington, DC 20007, USA.
E-mail address: brent.wiesel@gmail.com (B.B. Wiesel).

https://doi.org/10.1053/j.sart.2021.06.008

Discussion & Conclusion: Despite overall high reliability for measures of glenoid version between 4 frequently utilized shoulder arthroplasty templating softwares, this reliability is significantly affected by glenoid deformity. The programs were overall less reliable when measuring inclination, and a similar trend of decreasing reliability with increasing glenoid deformity emerged that was not statistically significant. Concordance with manual measurement is also variable. Further research is needed to understand how this variability should be accounted for during shoulder arthroplasty preoperative planning.

Level of Evidence: Level III; Retrospective Comparative Study

Total shoulder arthroplasty (TSA) and reverse total shoulder arthroplasty (RTSA) are mainstays of definitive management of both degenerative and acute shoulder conditions.[2] Clinical outcomes research in recent years has demonstrated the ability of both of these to produce long-term satisfaction and improved quality of life in patients electing to undergo these procedures.[32] Despite their overall success, both TSA and RTSA are subject to relatively high failure rates compared to other common arthroplasties[2,43] highlighting the need for better approaches to avoid the most common mechanisms of failure. Among the most common characteristics of failed shoulder replacements are stiffness, instability, rotator cuff deficiency, glenoid component loosening and scapular notching all of which are closely associated with glenoid component malposition.[11,14,15] Studies of glenoid component malpositioning in TSA have linked excessive retroversion of the glenoid component with osteolysis and loosening while implantation of the glenoid component with a superior tilt increases risk of secondary rotator cuff dysfunction and worse clinical outcomes.[19,41] Likewise in RTSA component malpositioning has been demonstrated to adversely affect both clinical and radiographic outcomes.[9]

Three-dimensional (3D) imaging has become increasingly relied upon during preoperative planning in an effort to improve component positioning in shoulder arthroplasty.[5,30] This approach involves using 3D computerized tomography (CT) reconstruction to develop preoperative plans using any of the commercially available programs designed for shoulder arthroplasty planning. Although these technologies are not universally employed among shoulder arthroplasty surgeons and currently there are no absolute indications for their use, the new generation of both surgeons and patients living in a more technology-laden environment coupled with the historically higher failure rates of shoulder arthroplasty previously mentioned have led to high interest in these programs. Preliminary evidence examining their potential benefit in improving component positioning has been promising, with limited evidence to the contrary.[5,16,17,21,24,36,38] Conclusive data is limited by small sample size and dearth of in vivo, clinical follow-up data, but as these technologies become more widely available more surgeons particularly at lower volume centers may be able to avoid some of the pitfalls associated with component malpositioning in shoulder arthroplasty.

Manufacturer differences with preoperative planning programs can complicate orthopedic surgeons' understanding of the utility of 3D planning. Recent comparisons have shown high variability between software measurements of glenoid version and inclination,[8] while other studies have suggested that manufacturer differences may contribute to different clinical outcomes[24]. Previous studies only examined 2 of the common preoperative planning programs and did not compare the programs to human measurement. It is additionally unclear how the accuracy of these programs is affected by preoperative scapular deformity. Unfortunately, there is evidence to suggest that the most deformed shoulders may be the most difficult for preoperative planning programs to assess.[16,24]

The purpose of this study is to assess the interrater reliability between different preoperative planning programs and manual measurement for shoulder arthroplasty with the hypothesis that there will be a statistically significant degree of variability between these modes of measurement. Primary outcome measures include the degree of inter-software variability in CT-based measurements of glenoid version and inclination. Secondarily we will examine whether degree of glenoid deformity is an independent predictor of program reliability.

Materials and methods

The method proposed by Walter[26] et al was implemented *a priori* and determined a minimum sample size of 76 patients for the 4 preoperative planning programs as "raters." An estimated underlying Intraclass Correlation Coefficient (ICC) of at least 0.8 and minimal accepted ICC of 0.7 were used to calculate the minimum sample size to provide 80% power at α = 0.05. This minimum sample size also allowed for a 10% attrition rate for subjects whose imaging would not be readily available. Shoulder CT scans for 76 consecutive patients undergoing either TSA or RTSA with 3D preoperative planning were retrospectively obtained. Digital Imaging and Communications in Medicine (DICOM) files from each scan were uploaded to 4 different preoperative planning programs (VIP by Arthrex, Naples, FL, USA; BluePrint by Wright Medical, Memphis, TN, USA; TrueSight Planner by Stryker/Materialise NV, Kalamazoo, MI, USA/Leuven, Belgium; ExactechGPS by Exactech, Gainesville, FL, USA;). None of the CT scans demonstrated truncation of any portion of the scapula which helped to eliminate a potential source of differences between those programs using landmark-based and those using automated measurements of scapular anatomy.[8] Native glenoid version and inclination of each shoulder was then measured separately by each program. A negative value from all modes of measurement corresponds to either retroversion or inferior inclination. Two separate board-certified and fellowship-trained musculoskeletal radiologists (MSK) blindly assessed each glenoid for version, inclination, and modified Walch classification[1] (for glenohumeral osteoarthritis only). Radiologists were blinded from each other, the results of the preoperative planning programs, as well as their own results during second pass measurements taken for intra-rater reliability.

Manual glenoid measurement methodology

All manual measurements and classifications were performed by 2 board-certified and MSK fellowship-trained radiologists. The axial CT images of each glenoid was reformatted into the scapular anatomic plane as described by Zale et al[42] using clinical 3D medical imaging software (Horos; Nimble Co LLC d/b/a Purview in Annapolis, MD, USA). Glenoid version was then measured using Friedman's line as described by Friedman et al.[12,31] Glenoid inclination was measured in the scapular plane measuring angle β, the angle between the floor of the supraspinatus and the glenoid fossa as described by Maurer et al.[28] Glenoid version and inclination were measured on selected axial and coronal slices. Axial slice selection was determined by selecting a middle glenoid slice inferior to the coracoid minimizing artifact and distortion as described by van de Bunt et al.[4] Each glenoid with diagnosed glenohumeral osteoarthritis (OA) was then classified according to the modified Walch classification.[1] When radiologists disagreed on classification, the more reliable radiologist's classification was used.

Three-dimensional preoperative planning software glenoid measurement methodologies

DICOM files for each shoulder CT were also separately uploaded to the 4 different 3D preoperative planning programs and analyzed for native glenoid inclination and version. Methodologies for glenoid analysis by VIP and Blueprint have been previously described[8] as has the Stryker platform.[25,30] In brief, VIP uses a midglenoid approach to digitally measure glenoid inclination and version based on manual identification of scapular landmarks by a trained technician. Blueprint is an automated preoperative planning program that creates a best-fit sphere on the face of the glenoid to 3-dimensionally measure version and inclination without the need for a company technician. TrueSight employs the definitions of scapular architecture outlined by Frankle et al[10] to measure glenoid anatomy in a landmark-based approach similar to VIP. Inclination is measured relative to the scapular neutral inclination axis defined along Friedman's axis using a glenoid center calculated from the smooth surface of the glenoid. This smooth surface is used to create a glenoid face plane which is also used to measure version relative to the scapular plane. ExactechGPS also employs trained technicians and a landmark-based approach to manually segment all CTs and create a scapular coordinate system based on the Friedman axis. The axis is determined manually via the line connecting the glenoid center and the trigonum spinae, each of which is identified using an average of 4 points (anterior, posterior, superior and inferior) for the glenoid face and 3 points for the trigonum. The scapular plane is then made by connecting this axis to the inferior most point on the scapular border. Inclination is measured within the scapular plane as the angle between the Friedman axis and a line connecting the superior and inferior most points on the glenoid. Glenoid version is then calculated as the angle between the Friedman axis and a line connecting the anterior and posterior most points of the glenoid face within a plane orthogonal to the scapular plane.

Statistical analysis

Intra-rater reliability was assessed for version measurements, inclination measurements and modified Walch classifications by using a randomized blinded sample of 19 of the shoulders included in the study. An estimated necessary sample size for intra-rater reliability was calculated by Walter's method[39] for $\alpha = 0.05$ and $\beta = 0.20$, using a minimum acceptable intra-class correlation coefficient (ICC) of 0.7 and an estimated underlying ICC of 0.9. Each MSK radiologist reassessed each of these images for version, inclination and modified Walch classification a minimum of 2 weeks from the first assessment. Intra-rater reliability for each radiologist's reads was calculated using a single rater, absolute agreement, 2-way mixed effects ICC. Radiologist inter-rater reliability for both version and inclination were likewise calculated from measurements for the entire study population. Single rater, consistency, 2-way mixed effects ICC were also calculated for the version and inclination as measured by the 4 templating softwares to assess inter-rater reliability. ICC was calculated across the entire dataset and for subgroups of different Walch classes based on glenoid wear. Overall ICC was classified by the following criteria: <0.50 indicates poor reliability, 0.50-0.75 indicates moderate reliability, 0.75-0.90 indicates good reliability and >0.90 indicates excellent reliability.[22] One-way Analysis of Variance (ANOVA) was used to compare ICC among Walch A, B, and D categories as well as among the further stratified Walch A1, A2, B1, B2, B3, and D cases. Threshold for statistical significance was set at $P < .05$.

The measurements made by the 2 board-certified MSK radiologists were averaged for version and inclination to obtain a single value. Lin's concordance correlation coefficient (CCC)[26,27] using a Fisher Z-transformation was calculated for each templating software against the MSK reads for both inclination and version. All confidence intervals indicate 95% confidence intervals. All statistics were conducted using R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria) and JMP Pro version 14 (SAS Institutes, Cary, NC). All figures were created using JMP.

Results

A total of 76 shoulders were included in this study, with 51 undergoing TSA and 25 undergoing RTSA. Fifty-nine of the shoulders underwent surgery due to glenohumeral OA and the remaining 17 were due to rotator cuff arthropathy (RCA). All 17 RCA cases underwent RTSA and an additional 8 glenohumeral OA cases were indicated for RTSA instead of TSA for reasons including patient preference and severe glenoid deformity with or without severe posterior subluxation of the humeral head. The 51 shoulders undergoing TSA were all due to glenohumeral OA. All 59 shoulders with glenohumeral OA were classified by each MSK radiologist according to the modified Walch classification system.[1] Intra-rater reliability ICC of the 2 MSK radiologists for modified Walch classification were 0.517 (95% CI: 0.081-0.783) and 0.791 (95% CI: 0.541-0.913), indicating moderate and good reliability respectively. Inter-rater reliability was measured by Cohen's Kappa. Between the 2 observers for Walch classification, $\kappa = 0.798$ indicating substantial agreement, with observers disagreeing on 13 out of 76 occasions (82% concordance rate). The more reliable observer's classification was used for these 13 cases. Overall 21 shoulders were classified as Walch A (6 A1, 15 A2), 27 as Walch B (9 B1, 15 B2, 3 B3),1 as Walch C, and 10 as Walch D. Further analysis by Walch classification excludes the single Walch C glenoid.

Intra-rater reliability for the 2 MSK radiologists' measurements of glenoid version were 0.987 (95% CI: 0.966-0.995) and 0.937 (95% CI: 0.843-0.975) indicating excellent reliability. Intra-rater reliability for the 2 MSK radiologists' measurements of glenoid inclination were 0.950 (95% CI: 0.840-0.982) and 0.966 (95% CI: 0.915-0.987) also indicating excellent reliability. Inter-rater reliability for the 2 MSK radiologists' measurements of glenoid version and inclination were 0.915 (95% CI: 0.868-0.945) and 0.902 (95% CI: 0.850-0.937) respectively, similarly indicating excellent reliability. For comparison to other modalities, we averaged the measurements by the 2 radiologists for each glenoid's version and inclination to obtain a single value.

The average (standard deviation) overall version in degrees for the cohort as measured by MSK radiology, VIP, Blueprint, TrueSight and ExactechGPS was -5.3 (7.7), -8.3 (8.1), -10.9 (9.1), -9.8 (8.0) and -10.1 (8.0), respectively. The average overall inclination in degrees was 9.0 (6.7), 9.5 (6.0), 9.3 (8.3), 5.9 (6.0) and 1.1 (5.8), respectively. Average version (Table I) and inclination (Table II) for each modified Walch class[37] (as determined by the more reliable rater) as measured by each mode of measurement are also given. Measures of glenoid version differed between at least 2 modes of measurement by a minimum of 1° in 76 (100%) glenoids, 5° in 57 (75%) glenoids, and 10° in 14 (18%) glenoids. Measures of glenoid inclination differed between at least 2 modes of measurement by a minimum of 1° in 75 (99%) glenoids, 5° in 70 (92%) glenoids, and 10° in 34 (45%) glenoids. The full pairwise comparison of all 5 modes of measurement is also displayed for both version (Table III) and inclination (Table IV).

The overall ICC of the 4 softwares' glenoid version measurements was 0.914 (95% CI: 0.876-.944) indicating excellent inter-rater reliability (Table V). ICC values for comparisons by Walch class and subclass for glenoid version measurements across the 4 templating softwares are also presented in Table V, with *P* values for all pairwise comparisons by Walch class and subclass presented in Table VI. Walch A glenoids (0.885; 95% CI: 0.793-0.946) trended towards greater inter-rater reliability than Walch B (0.739; 95% CI: 0.594-0.855) ($P = .085$) and showed statistically greater reliability than Walch D (0.561; 95% CI: 0.249-0.84) ($P = .002$). The 4 softwares likewise showed statistically greater inter-rater reliability when measuring version of Walch B glenoids than Walch D ($P = .034$). Further stratification of glenoid by modified Walch subclass revealed additional significant differences in the inter-rater reliability of the softwares' measures of glenoid version (Table VI). VIP's measurements of glenoid version demonstrated the highest rate of concordance with the averaged MSK radiologist read as measured by CCC (0.810; 95% CI: 0.712-0.877), while Blueprint's demonstrated the lowest (0.702; 95% CI: 0.5875-0.789) (Table VII, Fig. 1).

In measuring inclination, the overall ICC of 0.705 (95% CI: 0.602-0.795) indicates moderate reliability of the 4 templating softwares (Table V). ICC values for comparisons by Walch class and subclass for glenoid inclination measurements across the 4 templating softwares are also presented in Table V, with *P* values for all pairwise comparisons by Walch class and subclass presented in Table VI. Walch A glenoids (ICC = 0.766; 95% CI: 0.610-0.884) demonstrated a higher ICC than Walch B (ICC = 0.647; 95% CI: 0.476-0.796) and comparable ICC to Walch D (ICC = 0.765; 95% CI: 0.519-0.925), and no statistical significance was found between the 3 analyzed Walch classes ($P = .216$). Further stratification of glenoids by modified Walch subclass showed no statistically significant differences in ICC between the measured inclination of different subclasses. A similar continuous correlation was seen between the Walch A and B subclasses, however, wherein Walch A1 glenoids demonstrated the highest ICC and B3, the

Table I – Average version (standard deviation) in degrees for each mode of measurement by modified Walch class.

Walch Class	MSK	VIP	Blueprint	TrueSight	ExactechGPS
A (n = 21)	-2.9 (3.6)	-4.3 (6.4)	-7 (7.4)	-6.7 (7)	-6.1 (6.8)
A1 (n = 6)	-3.2 (2.6)	-5.3 (5.5)	-6.2 (7.3)	-6.8 (4.3)	-10.1 (5.7)
A2 (n = 15)	-2.7 (4)	-4 (6.9)	-7.3 (7.6)	-6.6 (8)	-4.5 (6.7)
B (n = 27)	-9.5 (5.2)	-14 (5)	-16.9 (5.3)	-14.6 (4.6)	-15 (4.1)
B1 (n = 9)	-7.4 (5.3)	-12.8 (5.4)	-14.9 (4.9)	-13.6 (5.2)	-14.1 (4.1)
B2 (n = 15)	-10.2 (4.3)	-15 (4.2)	-18 (5.1)	-15.7 (3.9)	-15.8 (4)
B3 (n = 3)	-12.4 (8.2)	-12.2 (7.9)	-17.7 (7.4)	-12.3 (6.1)	-13 (5)
C (n = 1)	-33.5	-32.7	-38.0	-33.0	-34.1
D (n = 10)	5.9 (3)	2.1 (3.7)	0.5 (6.5)	0 (5)	0 (4.9)
RCA (n = 17)	-7.6 (7.0)	-8.4 (5.9)	-10.9 (7.0)	-10.3 (7.0)	-12.1 (6.1)
Overall (n = 76)	-5.3 (7.7)	-8.3 (8.1)	-10.9 (9.1)	-9.8 (8)	-10.1 (8)

MSK, musculoskeletal radiologist; RCA, Rotator cuff arthropathy.

Table II – Average inclination (standard deviation) in degrees for each mode of measurement by modified Walch class.

Walch Class	MSK	VIP	Blueprint	TrueSight	ExactechGPS
A (n = 21)	8 (7.3)	7.4 (5.3)	6.5 (9.4)	4.7 (6)	1.2 (5.9)
A1 (n = 6)	10.7 (7.1)	10.6 (3.6)	11.2 (5.9)	7.8 (4.2)	4.5 (5)
A2 (n = 15)	6.9 (7.4)	6.1 (5.5)	4.6 (10.1)	3.4 (6.2)	-0.2 (5.8)
B (n = 27)	8.2 (5.8)	8.9 (4.5)	9.9 (7.2)	6.6 (5)	-0.3 (5.4)
B1 (n = 9)	9.2 (6.9)	11 (4.2)	12.4 (9.2)	8 (4.9)	3.2 (5)
B2 (n = 15)	8.3 (5.1)	7.1 (4.2)	8.8 (6.4)	5.4 (5.2)	-2.6 (3.9)
B3 (n = 3)	4.8 (6.1)	11.4 (3.3)	8 (3.6)	8.3 (4)	0.4 (9.1)
C (n = 1)	0.9	11.4	9.0	4.0	-3.2
D (n = 10)	12 (6.7)	12.3 (5.3)	12.8 (6.4)	7 (6.4)	4.3 (4.6)
RCA (n = 17)	10.1 (8.5)	11.3 (8.5)	9.8 (9.0)	6.2 (7.7)	1.8 (6.6)
Overall (n = 76)	9 (6.7)	9.5 (6)	9.3 (8.3)	5.9 (6.1)	1.1 (5.8)

MSK, musculoskeletal radiologist; RCA, Rotator cuff arthropathy.

lowest (Table V). Of the 4 softwares Blueprint had the highest concordance to MSK radiology as measured by CCC (0.660; 95% CI: 0.499-0.776) and ExactechGPS had the lowest (0.370; 95% CI: 0.236-0.489) (Table VII) with non-overlapping 95% confidence intervals (Fig. 2).

Discussion

Three-dimensional (3D) preoperative planning may enable better shoulder arthroplasty outcomes by facilitating the accurate placement of the glenoid component. This study helps elucidate this possibility by determining the interrater reliability (intraclass correlation coefficient or ICC) among 4 different preoperative planning softwares when measuring glenoid version and inclination as well as the concordance correlation coefficient (CCC) between each of the softwares and the averaged measurements of 2 board-certified, fellowship-trained MSK radiologists. The overall ICC of the programs was determined to be excellent when measuring version but only moderate when measuring inclination. As hypothesized, version ICC however was affected by the Walch class of the studied glenoids, with the highest inter-rater reliability among Walch A glenoids (0.885; good, near excellent), intermediate reliability among Walch B glenoids (0.739; moderate, near good) and lowest among Walch D (0.561; moderate, near poor). While A vs. D and B vs. D were both statistically significant, the A vs. B comparison only trended toward statistical significance (good $P = 0.085$). Additionally, version ICC for the 4 studied programs was significantly higher among certain modified Walch subclasses than others (Table VI). CCC calculation showed that when measuring version, VIP had the highest concordance with the MSK radiologists' measurements, while Blueprint had the lowest. Regarding inclination, no significant differences in ICC were noted between Walch classes or subclasses, which is not surprising given that the classification primarily accounts for glenoid deformity in the axial plane. Finally, CCC for inclination showed Blueprint with the highest concordance to the MSK radiologists' measurements and ExactechGPS with the lowest.

Outcomes in both TSA and RTSA are highly dependent on glenoid component positioning.[9,11,15,41] Achieving component positioning that accounts for patient anatomy and provides satisfactory clinical outcomes similar to natural shoulder function is not straightforward, and is limited by both pre- and intra-operative factors. Preoperative evaluation of scapular anatomy is critical for surgical planning given the high variability of both glenoid version and inclination.[7] The importance of appropriate preoperative evaluation is enhanced in the deformed shoulder as degree of difficulty in glenoid placement correlates with preoperative degree of deformity.[20] Two- and 3-dimensional CT-based imaging modalities have proven useful to this end[13] with recent evidence showing improved accuracy of 3D reconstructed CTs relative to 2D.[3,18,40] Many surgeons therefore employ 3D preoperative planning during shoulder arthroplasty using one of several commercially available programs. Some evidence in recent years has shown increased accuracy of component placement when using these programs,[16,17,21,25,35] although the evidence is not uniformly positive.[8,24] Additionally, questions remain regarding the accuracy of these 3D preoperative planning software measurements of scapular anatomy given their different methodologies.

Our results demonstrate that, despite overall high reliability for measures of glenoid version between the preoperative planning programs when measuring scapular anatomy, a more granular analysis reveals that this reliability is significantly affected by glenoid deformity. The programs were overall less reliable when measuring inclination, and although a similar trend of decreasing reliability with increasing glenoid deformity emerged it was not statistically significant in our sample. These findings support the previous evidence reported by Denard et al[8] which also showed that 2 programs produced different results when measuring glenoid version and inclination. Their study reported variation in version and inclination measures between VIP and Blueprint for a 63-patient cohort. Results from that study demonstrated a difference in either of those measurements by more than 10° in almost 25% of cases and by at least 5° in more than half. Although our study only showed a measured version or inclination difference of more than 10° between VIP and Blueprint in just 5% of glenoids, we similarly observed a difference of at least 5° in one of those measurements in almost 54% of all glenoids. By comparing discordance across twice as many preoperative platforms as Denard et al's previous work, our current study saw much higher overall rates of discordance. For example, greater than 93% of all glenoids had a measured difference in inclination

Table III – Degree of variance in measured version between all modes of measurement.

	VIP vs. BluePrint	VIP vs. TrueSight	VIP vs. ExactechGPS	VIP vs. MSK	Blueprint vs. TrueSight	Blueprint vs. ExactechGPS	Blueprint vs. MSK	TrueSight vs. ExactechGPS	TrueSight vs. MSK	ExactechGPS vs. MSK
1-5°	45 (59%)	46 (61%)	53 (70%)	42 (55%)	45 (59%)	46 (61%)	30 (39%)	54 (71%)	37 (49%)	32 (42%)
5°<x<10°	21 (28%)	11 (14%)	11 (14%)	19 (25%)	12 (16%)	15 (20%)	36 (47%)	9 (12%)	29 (38%)	29 (38%)
>10	0 (0%)	1 (1%)	0 (0%)	3 (4%)	0 (0%)	1 (1%)	7 (9%)	1 (1%)	5 (7%)	5 (7%)

MSK, musculoskeletal radiologist.

Table IV – Degree of variance in measured inclination between all modes of measurement.

	VIP vs. BluePrint	VIP vs. TrueSight	VIP vs. ExactechGPS	VIP vs. MSK	Blueprint vs. TrueSight	Blueprint vs. ExactechGPS	Blueprint vs. MSK	TrueSight vs. ExactechGPS	TrueSight vs. MSK	ExactechGPS vs. MSK
1-5°	46 (61%)	43 (57%)	27 (36%)	34 (45%)	41 (54%)	23 (30%)	38 (50%)	41 (54%)	43 (57%)	23 (30%)
5°<x<10°	16 (21%)	21 (28%)	31 (41%)	30 (39%)	21 (28%)	29 (38%)	29 (38%)	18 (24%)	17 (22%)	33 (43%)
>10	4 (5%)	1 (1%)	13 (17%)	4 (5%)	8 (11%)	17 (22%)	3 (4%)	4 (5%)	7 (9%)	14 (18%)

MSK, musculoskeletal radiologist.

Table V – Intra-class correlation coefficient (ICC) for measurement of glenoid version and inclination among 4 observed PSI preoperative planning softwares.

	Intra-Class Correlation Coefficient (ICC)	
	Version	Inclination
Overall (n = 76)	0.914	0.705
Walch Class		
Walch A (n = 21)	0.885	0.766
Walch A1 (n = 6)	0.828	0.754
Walch A2 (n = 15)	0.940	0.738
Walch B (n = 27)	0.739	0.647
Walch B1 (n = 9)	0.762	0.682
Walch B2 (n = 15)	0.646	0.585
Walch B3 (n = 3)	0.935	0.579
Walch D (n = 10)	0.561	0.765

<0.5 is considered poor reliability; 0.5< moderate <0.75; 0.75< good <0.90; 0.90< excellent.

Table VI – Pairwise comparison of version and inclination inter-rater reliability (ICC) by Walch class and subclass among 4 preoperative planning software programs.

	Version		Inclination	
Walch class	ICC Δ	P value	ICC Δ	P value
A-B	0.146	.0851	0.119	.1147
A-D	0.324	.0006*	0.001	.9998
B-D	0.178	.0345*	-0.118	.1182
Walch Sub-Class	ICC Δ	P value	ICC Δ	P value
A1-A2	0.112	.5686	0.016	1.0000
A1-B1	0.066	.9202	0.072	.9729
A1-B2	0.182	.1122	0.169	.5050
A1-B3	0.107	.6143	0.175	.4677
A1-D	0.267	.0068*	-0.011	1.0000
A2-B1	0.178	.1258	0.056	.9911
A2-B2	0.294	.0026*	0.153	.6074
A2-B3	0.005	1.0000	0.159	.5688
A2-D	0.379	.0001*	-0.027	.9997
B1-B2	0.116	.5322	0.097	.9089
B1-B3	0.173	.1448	0.103	.8861
B1-D	0.201	.0632	-0.083	.9508
B2-B3	0.289	.0031*	0.006	1.0000
B2-D	0.085	.8026	-0.18	.4373
B3-D	0.374	.0001*	-0.186	.4021

ICC, inter-rater reliability.

* Indicates $P < 0.05$.

Table VII – Concordance correlation coefficient (CCC) for 4 preoperative planning softwares measurements of glenoid version and inclination as compared to fellowship-trained musculoskeletal radiologists.

	Concordance Correlation Coefficient (CCC)	
Program	Version	Inclination
VIP	0.810	0.476
Blueprint	0.702	0.660
TrueSight	0.749	0.540
ExactechGPS	0.733	0.370

alone between the 4 platforms of at least 5°. Denard et al did show higher rates (63.6%) of discordant measurements between the 2 studied programs among those CTs assessed to be "unclean" (i.e. with extra bony fragments) by Blueprint reconstruction. The researchers therefore posited that automated vs. manual identification of scapular landmarks may be a primary contributor to cross-platform differences. In the present study, however, differences between those 2 platforms may have been mitigated by complete scans in all 76 cases. None of the studied CTs demonstrated inferior or medial truncation of the scapula which may impact manual and automated scapular measurements differently. Nonetheless we still saw inter-rater reliability varying with glenoid morphology with better concordance to MSK radiologist measurements when measuring inclination with an automated system (Blueprint) vs. better concordance when measuring version with a landmark-based system (VIP).

Shah et al's 2019 study[34] comparing glenoid measurements from both an automated and a manual program to precision measurements on 3D-printed replicas of the scapular CTs similarly showed high rates of variability between the 2 different methods. Authors of this study suggested 5 primary contributors to this variability that largely fall into 2 categories: differences in program technique (e.g. circular inferior glenoid plane vs. maximum circular plane) and differences in glenohumeral pathology (e.g. glenoid morphology). There is well documented significant variability of radiologists' measurements of glenoid version and inclination.[18,23,31,33] Labral calcifications, small osteophytes on the glenoid rim, and CT artifacts can alter a radiologists' choice to draw the line representing the anterior and posterior corners of the glenoid.[4] The decision to include or not include an osteophyte can significantly alter one's measurement of glenoid version, and there is limited discussion in previously published methods on determining if or how much of an osteophyte to include in a measurement. At our radiologists' institution, radiologists attempt to follow a well-defined glenoid cortex to estimate the glenoid corners in the presence of large or bulky osteophytes to minimize measurement variability due to osteophytes. The ability to exclude osteophytes is an advantage of measurements using 2D CT rather than measurements using 3D volumetric models, although this advantage is likely offset due to variability in slice selection on 2D CT. Further study to quantify this variability is needed. While our findings substantiate their suggestion that anatomic elements such as glenoid morphology may contribute to variability between preoperative planning programs, we similarly offer these findings not to endorse a particular methodology, but rather to help inform clinical decision making with an acknowledgement of the benefits these programs provide.[29]

Additionally, our findings suggest that program reliability varies with the degree of glenoid deformity with cross-platform reliability decreasing in a near gradient-like fashion from Walch A to D when measuring version. These findings are in contrast to Denard et al[8] who saw similar rates of 5° or more variation across Walch classes, although they did not report 10°+ variation rates for this analysis. Authors of that study acknowledged however that conclusions regarding this outcome measure are limited by their relatively small sample size. Chalmers et al however previously showed that B2

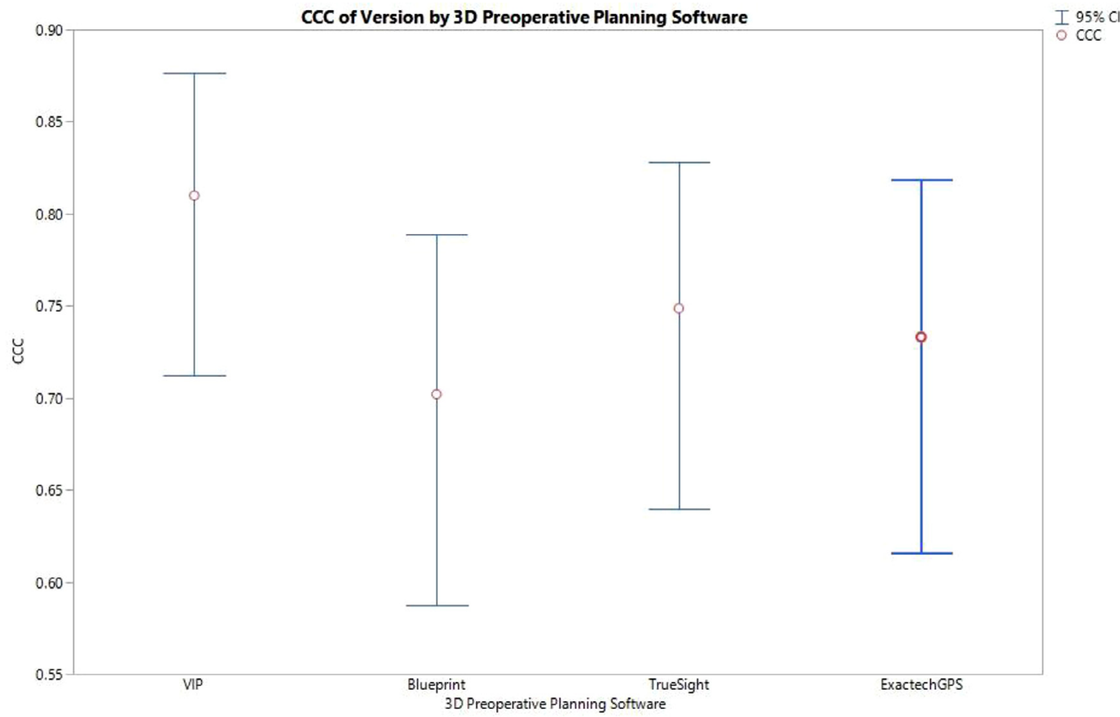

Figure 1 – Concordance correlation coefficients (CCC) with 95% confidence intervals for all 4 programs' measures of version as compared to 2 musculoskeletal-trained radiologists. CCC *(concordance correlation coefficient)*.

glenoids require specific consideration during CT measurement especially of version.[6] Although our study only showed a trend toward significance in version ICC between Walch A and Walch B glenoids at the class level (P = .085), there were significant differences that became apparent when stratifying by subclass, particularly among the Walch A2 vs. Walch B2 which were our largest sample size sub-groups. This finding likely highlights the inherent difficulty in determining which scapular landmarks remain pertinent for planning component positioning in the B2 glenoid. The biconcave deformity of this subclass presents a unique challenge in deciding whether to include landmarks from the neoglenoid,

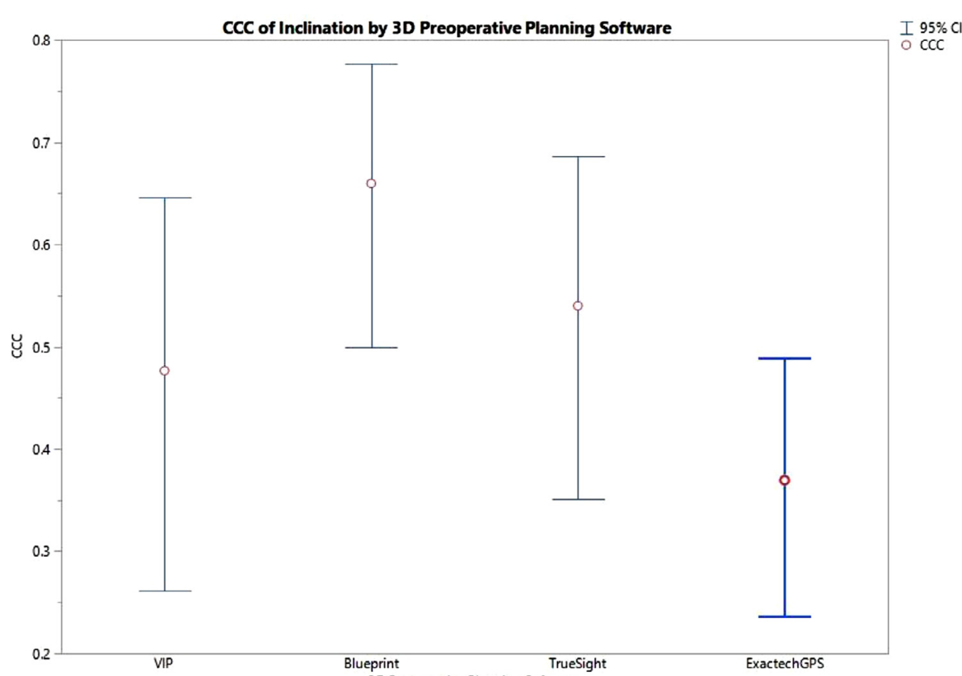

Figure 2 – Concordance correlation coefficients (CCC) with 95% confidence intervals for all 4 programs' measures of inclination as compared to 2 musculoskeletal-trained radiologists. CCC *(concordance correlation coefficient)*.

paleoglenoid, or both when measuring version, and thus may be variably interpreted by all modes of measurement. This same logic may explain the consistently and significantly lower inter-rater reliability when measuring version among Walch D glenoids whereby anterior subluxation and the resultant asymmetric wear pattern leading to an anteverted biconcavity is inconsistently handled by different planning program methodologies. Bercik et al[1] originally considered differentiating D1 and D2 subtypes based on the presence or absence of an anteverted biconcavity, and it stands to reason that the same challenges present in the B2 glenoid may be present in the D glenoid cohort evaluated in this study. This may indicate that the programs and technicians need modification to specifically account for glenoid anteversion and anterior humeral head subluxation.

This study is not without limitations. Although this study was adequately powered to yield several statistically significant findings, conclusions for our secondary outcome measures were limited by subgroup sizes. While our results did show several significant differences between modified Walch sub-classes, perhaps larger sample sizes would have revealed additional meaningful differences between other sub-classes. Additionally, the datasets were reformatted into the scapular plane with selected axial and coronal slices for glenoid and version measurement before being sent to the radiologists for measurements, matching the institution's workflow, which likely contributed to the high inter-rater and intra-rater reliability. If selected slices were not used, radiologists'' measurement variability would be expected to increase. It is also worth noting that while we elected to use radiologist measurement of glenoid anatomy as a comparator for the ICC analysis, this decision was made simply because this is 1 historical standard of measuring glenoid anatomy and not necessarily superior to other methods. Finally, we did not examine surgical or clinical outcomes for any patients and additional research is needed to better understand how the observed program variability affects meaningful outcomes. Despite these limitations there are many strengths to the current study as well. Ours is the largest sample size to date and is appropriately powered for a reliability study based on *a priori* minimum sample size estimation. Additionally, while most previous quantitative research has examined 1-2 preoperative planning programs, this study is the first to evaluate and directly compare 4 frequently used programs. This is also the only study to our knowledge to include a manual measurement comparator with inclusion of both inter- and intra-rater reliability. Finally, the current study significantly expands upon previous findings for both variability in measurements among programs and increasing variability with worsening glenoid deformity.

Conclusions

This study revealed significant differences among 4 CT-based preoperative planning software programs when measuring glenoid anatomy. Overall, the software inter-rater reliability as measured by intra-class correlation coefficient (ICC) was lower when measuring inclination in all shoulders and when measuring both version and inclination in shoulders with more glenoid erosion. Rates of concordance between each program and MSK fellowship-trained radiologists as measured by Lin's concordance correlation coefficient (CCC) also showed variable concordance between programs. Further research is needed to better understand how this variability should be accounted for during preoperative planning for shoulder arthroplasty.

Disclaimers

Funding: The Georgetown Department of Orthopaedic Surgery received non-financial support for this research from Stryker/Materialise in the form of complimentary glenoid measurements of our study population.

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.The authors, their immediate families, and any research foundation with which they are affiliated did not receive any financial payments or other benefits from any commercial entity related to the subject of this article.

REFERENCES

[1] Bercik MJ, Kruse K. 2nd, Yalizis M, Gauci M-O, Chaoui J, Walch G. A modification to the Walch classification of the glenoid in primary glenohumeral osteoarthritis using three-dimensional imaging. J Shoulder Elbow Surg 2016;25:1601–6. https://doi.org/10.1016/j.jse.2016.03.010.

[2] Bohsali KI, Wirth MA, Rockwood C.A. Jr. Complications of total shoulder arthroplasty. J Bone Joint Surg Am. 2006;88:2279–92. https://doi.org/10.2106/JBJS.F.00125.

[3] Budge MD, Lewis GS, Schaefer E, Coquia S, Flemming DJ, Armstrong AD. Comparison of standard two-dimensional and three-dimensional corrected glenoid version measurements. J Shoulder Elbow Surg 2011;20:577–83. https://doi.org/10.1016/j.jse.2010.11.003.

[4] van de Bunt F, Pearl ML, Lee EK, Peng L, Didomenico P. Glenoid version by CT scan: an analysis of clinical measurement error and introduction of a protocol to reduce variability. Skeletal Radiol 2015;44:1627–35. https://doi.org/10.1007/s00256-015-2207-4.

[5] Cancienne JM, Dempsey IJ, Garrigues GE, Cole BJ, Brockmeier SF, Werner BC. Trends and impact of three-dimensional preoperative imaging for anatomic total shoulder arthroplasty. Shoulder Elbow 2020. https://doi.org/10.1177/1758573220908865.

[6] Chalmers PN, Salazar D, Chamberlain A, Keener JD. Radiographic characterization of the B2 glenoid: the effect of computed tomographic axis orientation. J Shoulder Elbow Surg 2017;26:258–64. https://doi.org/10.1016/j.jse.2016.07.021.

[7] Churchill RS, Brems JJ, Kotschi H. Glenoid size, inclination, and version: an anatomic study. J Shoulder Elbow Surg 2001;10:327–32.

[8] Denard PJ, Provencher MT, Lädermann A, Romeo AA, Parsons BO, Dines JS. Version and inclination obtained with 3-dimensional planning in total shoulder arthroplasty: do different programs produce the same results? JSES Open Access 2018;2:200–4. https://doi.org/10.1016/j.jses.2018.06.003.

[9] Feeley BT, Zhang AL, Barry JJ, Shin E, Ho J, Tabaraee E. Decreased scapular notching with lateralization and inferior baseplate placement in reverse shoulder arthroplasty with high humeral

inclination. Int J Shoulder Surg. 2014;8:65–71. https://doi.org/10.4103/0973-6042.140112.

[10] Frankle MA, Teramoto A, Luo Z-P, Levy JC, Pupello D. Glenoid morphology in reverse shoulder arthroplasty: classification and surgical implications. J Shoulder Elbow Surg 2009;18:874–85. https://doi.org/10.1016/j.jse.2009.02.013.

[11] Franta AK, Lenters TR, Mounce D, Neradilek B, Matsen F.A. 3rd. The complex characteristics of 282 unsatisfactory shoulder arthroplasties. J Shoulder Elbow Surg 2007;16:555–62. https://doi.org/10.1016/j.jse.2006.11.004.

[12] Friedman RJ, Hawthorne KB, Genez BM. The use of computerized tomography in the measurement of glenoid version. J Bone Joint Surg Am. 1992;74:1032–7.

[13] Ganapathi A, McCarron JA, Chen X, Iannotti JP. Predicting normal glenoid version from the pathologic scapula: a comparison of 4 methods in 2- and 3-dimensional models. J Shoulder Elbow Surg 2011;20:234–44. https://doi.org/10.1016/j.jse.2010.05.024.

[14] Gomes NS. Patient-specific instrumentation for total shoulder arthroplasty. EFORT Open Rev 2016;1:177–82. https://doi.org/10.1302/2058-5241.1.000033.

[15] Hasan SS, Leith JM, Campbell B, Kapil R, Smith KL, Matsen FA. Characteristics of unsatisfactory shoulder arthroplasties. J Shoulder Elbow Surg 2002;11:431–41. https://doi.org/10.1067/mse.2002.125806.

[16] Hendel MD, Bryan JA, Barsoum WK, Rodriguez EJ, Brems JJ, Evans PJ. Comparison of patient-specific instruments with standard surgical instruments in determining glenoid component position: a randomized prospective clinical trial. J Bone Joint Surg Am. 2012;94:2167–75. https://doi.org/10.2106/JBONE JOINT SURG AM.K.01209.

[17] Heylen S, Van Haver A, Vuylsteke K, Declercq G, Verborgt O. Patient-specific instrument guidance of glenoid component implantation reduces inclination variability in total and reverse shoulder arthroplasty. J Shoulder Elbow Surg 2016;25:186–92. https://doi.org/10.1016/j.jse.2015.07.024.

[18] Hoenecke H.R. Jr, Hermida JC, Flores-Hernandez C, D, Lima DD. Accuracy of CT-based measurements of glenoid version for total shoulder arthroplasty. J Shoulder Elbow Surg 2010;19:166–71. https://doi.org/10.1016/j.jse.2009.08.009.

[19] Ho JC, Sabesan VJ, Iannotti JP. Glenoid component retroversion is associated with osteolysis. J Bone Joint Surg Am. 2013;95. https://doi.org/10.2106/JBJS.L.00336. e82.

[20] Iannotti JP, Greeson C, Downing D, Sabesan V, Bryan JA. Effect of glenoid deformity on glenoid component placement in primary shoulder arthroplasty. J Shoulder Elbow Surg 2012;21:48–55. https://doi.org/10.1016/j.jse.2011.02.011.

[21] Iannotti JP, Weiner S, Rodriguez E, Subhas N, Patterson TE, Jun BJ. Three-dimensional imaging and templating improve glenoid implant positioning. J Bone Joint Surg Am. 2015;97:651–8. https://doi.org/10.2106/jbjs.n.00493.

[22] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

[23] Kwon YW, Powell KA, Yum JK, Brems JJ, Iannotti JP. Use of three-dimensional computed tomography for the analysis of the glenoid anatomy. J Shoulder Elbow Surg 2005;14:85–90. https://doi.org/10.1016/j.jse.2004.04.011.

[24] Lau SC, Keith PPA. Patient-specific instrumentation for total shoulder arthroplasty: not as accurate as it would seem. J Shoulder Elbow Surg 2018;27:90–5. https://doi.org/10.1016/j.jse.2017.07.004.

[25] Levy JC, Everding NG, Frankle MA, Keppler LJ. Accuracy of patient-specific guided glenoid baseplate positioning for reverse shoulder arthroplasty. J Shoulder Elbow Surg 2014;23:1563–7. https://doi.org/10.1016/j.jse.2014.01.051.

[26] Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255–68.

[27] Lin LI-K. A note on the concordance correlation coefficient. Biometrics 2000;56:324–5.

[28] Maurer A, Fucentese SF, Pfirrmann CWA, Wirth SH, Djahangiri A, Jost B. Assessment of glenoid inclination on routine clinical radiographs and computed tomography examinations of the shoulder. J Shoulder Elbow Surg 2012;21:1096–103. https://doi.org/10.1016/j.jse.2011.07.010.

[29] Min KS, Fox HM, Bedi A, Walch G, Warner JJP. Patient-specific planning in shoulder arthroplasty. Bone Joint J 2020;102-B:365–70. https://doi.org/10.1302/0301-620x.102b3.bjj-2019-1153.r1.

[30] Rodríguez JA, Entezari V, Iannotti JP, Ricchetti ET. Pre-operative planning for reverse shoulder replacement: the surgical benefits and their clinical translation. Ann Joint 2019;4:4. https://doi.org/10.21037/aoj.2018.12.09.

[31] Rouleau DM, Kidder JF, Pons-Villanueva J, Dynamidis S, Defranco M, Walch G. Glenoid version: how to measure it? Validity of different methods in two-dimensional computed tomography scans. J Shoulder Elbow Surg 2010;19:1230–7. https://doi.org/10.1016/j.jse.2010.01.027.

[32] Savin DD, Frank RM, Sumner S, Richardson C, Nicholson GP, Romeo AA. Good functional outcomes expected after shoulder arthroplasty irrespective of body mass index. J Shoulder Elbow Surg 2018;27:S43–9. https://doi.org/10.1016/j.jse.2018.03.022.

[33] Scalise JJ, Codsi MJ, Bryan J, Brems JJ, Iannotti JP. The influence of three-dimensional computed tomography images of the shoulder in preoperative planning for total shoulder arthroplasty. J. Bone Joint Surg. Am. 2008;90:2438–45. https://doi.org/10.2106/jbjs.g.01341.

[34] Shah SS, Sahota S, Denard PJ, Provencher MT, Parsons BO, Hartzler RU. Variability in total shoulder arthroplasty planning software compared to a control CT-derived 3D printed scapula. Shoulder Elbow 2019. https://doi.org/10.1177/1758573219888821.

[35] Suero EM, Citak M, Lo D, Krych AJ, Craig EV, Pearle AD. Use of a custom alignment guide to improve glenoid component position in total shoulder arthroplasty. Knee Surg Sports Traumatol. Arthrosc. 2013;21:2860–6. https://doi.org/10.1007/s00167-012-2177-1.

[36] Throckmorton TW, Gulotta LV, Bonnarens FO, Wright SA, Hartzell JL, Rozzi WB. Patient-specific targeting guides compared with traditional instrumentation for glenoid component placement in shoulder arthroplasty: a multi-surgeon study in 70 arthritic cadaver specimens. J Shoulder Elbow Surg 2015;24:965–71. https://doi.org/10.1016/j.jse.2014.10.013.

[37] Walch G, Badet R, Boulahia A, Khoury A. Morphologic study of the glenoid in primary glenohumeral osteoarthritis. J Arthroplasty. 1999;14:756–60.

[38] Walch G, Vezeridis PS, Boileau P, Deransart P, Chaoui J. Three-dimensional planning and use of patient-specific guides improve glenoid component position: an in vitro study. J Shoulder Elbow Surg 2015;24:302–9. https://doi.org/10.1016/j.jse.2014.05.029.

[39] Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;17:101–10.

[40] Werner BS, Hudek R, Burkhart KJ, Gohlke F. The influence of three-dimensional planning on decision-making in total shoulder arthroplasty. J Shoulder Elbow Surg 2017;26:1477–83. https://doi.org/10.1016/j.jse.2017.01.006.

[41] Young AA, Walch G, Pape G, Gohlke F, Favard L. Secondary rotator cuff dysfunction following total shoulder arthroplasty for primary glenohumeral osteoarthritis: results of a multicenter study with more than five years of follow-up. J Bone Joint Surg Am 2012;94:685. https://doi.org/10.2106/JBJS.J.00727.

[42] Zale CL, Pace GI, Lewis GS, Chan J, Kim HM. Interdepartmental imaging protocol for clinically based three-dimensional computed tomography can provide accurate measurement of glenoid version. J Shoulder Elbow Surg 2018;27:1297–305. https://doi.org/10.1016/j.jse.2017.11.020.

[43] Zumstein MA, Pinedo M, Old J, Boileau P. Problems, complications, reoperations, and revisions in reverse total shoulder arthroplasty: a systematic review. J Shoulder Elbow Surg 2011;20:146–57. https://doi.org/10.1016/j.jse.2010.08.001.