Using Computed Tomography to Assess Proximal Humerus Fractures

Arun J. Ramappa, MD, Vaishali Patel, MD, Karan Goswami, MD, David Zurakowski, PhD, Corrie Yablon, MD, E. Ken Rodriguez, MD, PhD, Paul Appleton, MD, and Joseph P. DeAngelis, MD

Abstract

Computed tomography (CT) scans are often used to evaluate proximal humerus fractures.

We conducted a study to determine if use of preoperative CT scans affects surgical decision-making with respect to proximal humerus fractures. Three board-certified orthopedic surgeons interpreted plain radiographs of 40 proximal humerus fractures and then CT scans with reformatted images. Results were assessed for interrater reliability.

Use of CT significantly improved interobserver reliability in fracture characterization and assessment. Surgeons were more likely to identify a displaced fracture ($P < .01$), an impaction ($P < .001$), and involvement of the anatomical neck ($P < .03$). However, CT did not improve agreement with use of AO (Arbeitsgemeinschaft für Osteosynthesefragen) fracture classification and did not significantly alter treatment recommendations.

CT scans provide more detail about the character of proximal humerus fractures (displacement, involved segments) but do not significantly influence surgical treatment recommendations when compared with plain radiographs alone.

Proximal humerus fractures account for 4% to 5% of fractures in adults and have an incidence of 6.6/1000 person-years.[1] In patients older than 65 years, these injuries are the third most common in the United States, after fractures of the proximal femur and distal radius. As the population ages, the incidence of proximal humerus fractures will increase and represent a major cause of morbidity.[2] A concerted approach to evaluating these fractures could translate into substantial patient benefits and cost savings.

Plain radiographs and computed tomography (CT) scans are used routinely by orthopedic surgeons to assess proximal humerus fractures. Because plain radiographs are used to evaluate the character of a proximal humerus fracture, its alignment and displacement, these images inform orthopedic surgeons' decisions regarding best treatment. Although approximately 80% of proximal humerus fractures are appropriate for conservative (ie, nonoperative) treatment,[3] proximal humerus fractures with angulation of more than 45° or displacement of more than 1 cm may benefit from surgical intervention.[4,5]

Several authors have expressed concern about using plain radiographs to classify proximal humerus fractures.[6-9] In these studies, interobserver and intraobserver reliability of proximal humerus fracture classification was unsatisfactory, demonstrating poor reproducibility based on plain radiographs alone. In addition, Brorson and colleagues[10,11] found poor agreement on Neer classification of proximal humerus fractures using anteroposterior and lateral radiographs.

The best evaluation of proximal humerus fractures often requires multiple plain radiographs. This standard assessment takes time and careful attention to patient positioning. The patient may experience considerable discomfort as the fractured arm is poised for imaging. For this reason, CT scans have been used to supplement plain radiographs. In addition, CT allows for better characterization of complex proximal humerus fractures.[12,13] However, the benefit of CT may not justify its use because of the associated radiation exposure and expense.

As 2-dimensional (2-D) CT scans have failed to increase the reliability of CT scans for Neer classification,[6,9,14] the precise role of CT in proximal humerus fractures has not been established; benefits of CT in the treatment of proximal humerus fractures remains unclear.[7,8,13,15]

We conducted a study to evaluate the role of CT scans in preoperative assessment of proximal humerus fractures and to determine if preoperative CT scans significantly affect surgeon treatment recommendations when compared with assessment using plain radiographs alone.

Materials and Methods

After obtaining institutional review board approval, we retrospectively reviewed the cases of patients who presented to a level I academic medical center emergency department between January 2007 and January 2008; we identified patients

Authors' Disclosure Statement: The authors report no actual or potential conflict of interest in relation to this article.

with proximal humerus fractures and patients with a complete set of plain radiographs and a CT scan of the proximal humerus.

A priori analysis using effect size of 0.5, α of 0.05, and β of 0.95 yielded a sample size of 34 patients. For radiographic review, 40 representative cases were selected.

For each patient, 3 board-certified orthopedic surgeons reviewed radiographs and CT scan independently. Each surgeon was asked to classify the fracture in accordance with the AO (Arbeitsgemeinschaft für Osteosynthesefragen) Comprehensive Classification of Fractures (Davos, Switzerland). They were also asked to evaluate each fracture for displacement, dislocation, impaction, number of major fracture fragments, involvement of the anatomical neck, involvement of the surgical neck, fracture of the greater tuberosity, and fracture of the lesser tuberosity. Last, they were asked whether they would recommend surgical treatment. All cases were evaluated in a blinded, randomized fashion, without radiology reports or other supporting documentation.

For the plain radiographs and CT scans, interobserver agreement was calculated using multirater κ (SPSS; IBM, Armonk, New York). Kappa is a chance-corrected measure of agreement that takes into account agreement expected by chance alone. Since its introduction by Cohen in 1960, this statistic has been the one most commonly used to describe agreement in a variety of interobserver studies.[6-8,16-19] A κ of 1.00 indicates perfect agreement, and a κ of .00 indicates no more agreement than expected by chance alone. Our interpretation of κ values was performed using the criteria of Landis and Koch[20]: Less than .2 represents poor or slight agreement; .21 to .40, fair agreement; .41 to .60, moderate agreement; .61 to .80, good agreement; .81 or more, excellent agreement. In this study, we considered a difference between κ values to be significant when the upper and lower boundaries of the respective 95% confidence intervals did not overlap. $P < .05$ was considered statistically significant.

The 3 surgeons' data were aggregated, and results compared using a binomial distribution on all variables, except AO classification. The analysis used a generalized estimating equations (GEE) approach, assuming the data were repeated-measures, to assess whether the treatment recommendation differed between plain radiographs and CT scan. For AO fracture classification, a Wilcoxon paired analysis was used. All statistical analysis was performed using SPSS.

Results

Of the patients enrolled, 77.5% were women. Mean (SD) age was 65.5 (16.9) years (range, 28 to 94 years). The interobserver κ values are presented in **Table I**.

Interobserver Analysis

When evaluating plain radiographs alone, this group demonstrated *moderate* or *good* interobserver reliability in the evaluation of proximal humerus fractures. Involvement of the surgical neck (.43), greater tuberosity (.51), and lesser tuberosity (.46) and presence of 3 or more fracture parts (.53) revealed *moderate* κ values. Presence of a 0- or 1-part fracture (.67) showed *good* interobserver reliability.

Table I. Interobserver Agreement on Proximal Humerus Fracture Characteristics and Surgical Decision-Making Based on Plain Radiographs and CT Scans

	Plain Radiographs		CT Scans	
Characteristic	Multirater κ[a]	*P*	Multirater κ[a]	*P*
AO class	.44	< .001[b]	.43	< .001[b]
Displacement	–.07	.56	.22	.24
Anatomical neck	.02	.92	.57	< .001[b]
Surgical neck	.43	< .001[b]	.65	< .001[b]
Greater tuberosity	.50	< .001[b]	.42	< .001[b]
Lesser tuberosity	.46	< .01[b]	.64	< .001[b]
0 or 1 part	.67	< .001[b]	.68	< .001[b]
2 parts	.37	< .01[b]	.44	< .001[b]
3+ parts	.53	< .001[b]	.57	< .001[b]
Impacted	.22	.18	.50	< .001[b]
Comminution	–.08	.67	–.12	.62
Dislocation	.44	.09	.75	< .001[b]
Plan to operate	.38	< .001[b]	.65	< .001[b]

Abbreviation: AO, Arbeitsgemeinschaft für Osteosynthesefragen.
[a]Evaluating agreement of 3 independent surgeons reviewing the same 40 cases.
[b]Statistically significant.

Surgeon agreement on many fracture characteristics was higher when CT scans were used to characterize fractures. Agreement was higher when assessing involvement of the anatomical neck (.57), surgical neck (.65), and lesser tuberosity (.64) and when determining whether the fracture was impacted (.50), comminuted (-0.12), or dislocated (.75). Furthermore, the reliability of classification of lesser tuberosity involvement (κ = .46) and dislocation (κ = .44) was *moderate* with use of plain radiographs but improved to *good* with use of CT (κ = .64, .68). CT κ values were all statistically significant, except in the assessment of displacement and comminution.

Interobserver reliability for classification according to AO type (A, B, C) was *moderate* when observers used plain radiographs (.44) or CT scans (.43).

Of importance, the multirater κ for operative versus nonoperative decision-making using plain radiographs showed *fair* reliability (κ = .38). However, this agreement on whether to operate improved to *good* (κ = .65) when CT was used.

Analysis of Pooled Data

Pooling the 3 surgeons' data and comparing the results between plain radiographs and CT using a binomial distribution on all variables other than AO class with a GEE approach indicated that CT changed fracture classification or decision

Table II. **Surgeons' Evaluation of Proximal Humerus Fracture Characteristics Using Plain Radiographs and CT Scans (N = 40 Cases)**

	Plain Radiographs			CT Scans		
Characteristic	Surgeon 1	Surgeon 2	Surgeon 3	Surgeon 1	Surgeon 2	Surgeon 3
AO class, A/B/C	15/10/15	12/17/11	12/11/17	10/13/17	16/14/10	16/12/12
Displacement	1	16	19	0	10	10
Anatomical neck	3	12	3	12	12	8
Surgical neck	29	20	32	22	24	27
Greater tuberosity	33	32	29	35	27	31
Lesser tuberosity	6	10	10	13	12	10
0 or 1 part	6	5	7	6	11	11
2 parts	16	12	18	12	9	10
3+ parts	18	23	15	22	20	19
Impacted	5	11	8	12	17	16
Comminution	4	14	2	13	0	0
Dislocation	2	5	5	5	4	4
Operative	15	23	12	15	16	13

Abbreviation: AO, Arbeitsgemeinschaft für Osteosynthesefragen.

to operate in comparison with plain radiographs. Each surgeon's evaluation of fracture characteristics using plain radiographs and CT is given in **Table II**. According to the analysis, CT significantly changed results for classification of 4 fracture characteristics: displacement ($P = .01$), anatomical neck ($P = .03$), 2-part fracture ($P = .02$), and impaction ($P < .001$). Using CT, surgeons were significantly less likely to indicate displacement and more likely to detect impaction (**Figure**).

No significant differences were seen in characterizing the surgical neck ($P = .31$), greater tuberosity ($P = .87$), lesser tuberosity ($P = .16$), 0- or 1-part fractures ($P = .09$), 3-part-or-more fractures ($P = .48$), comminution ($P = .21$), and dislocation ($P = .84$). In addition, the Wilcoxon paired analysis used for AO class showed no significant difference between plain radiographs and CT ($P = .52$), and no significant difference was detected between plain radiographs and CT datasets regarding plan to operate on the fracture ($P = .41$) (**Figure**).

Discussion

In the treatment of proximal humerus fractures, CT scans are often used to supplement plain radiographs. Although CT scans have been found to allow for better characterization of complex fractures, the role of CT in proximal humerus fractures has not been established. In this study, we tried to determine if a preoperative CT scan affected surgeon treatment recommendations when compared with plain radiographs alone.

This study found a similar interobserver reliability for AO classification of proximal humerus fractures with both plain radiographs (.44) and CT scans (.43), consistent with findings reported by Bernstein and colleagues[6] and Sidor and colleagues,[7] who described poor interrater agreement using the AO and Neer classification systems with both radiographs and CT scans.

In a comparison of fracture characteristics, CT scans significantly improved interobserver reliability in assessing involvement of the anatomical neck, surgical neck, and lesser tuberosity and presence of impaction or dislocation. Thus, CT is more reliable than plain radiographs in assessing these fracture parameters.

Surgeon agreement on treatment was also significantly higher using CT scans than using plain radiographs alone. This finding suggests that CT provides additional information that increases agreement among surgeons, even though the CT scan did not significantly change the individual surgeon's operative planning. Based only on the recommendation for treatment, it is difficult to assert whether the differences between plain radiographs and CT scans are of sufficient clinical importance to warrant use of CT.

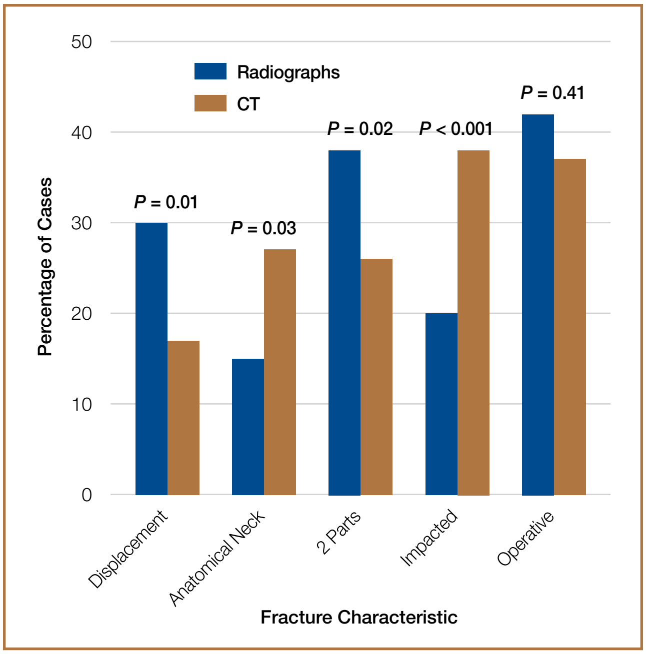

Figure. Comparison of fracture evaluation by radiographs and CT scans.

This investigation found that CT scans demonstrate improved reliability in fracture characterization and operative decision-making in proximal humerus fractures. However, several studies have questioned whether additional imaging improves consensus when assessing proximal humerus fractures. In a study assessing the Neer classification system with use of plain radiographs, Bernstein and colleagues[6] determined that intraobserver reliability was good and interobserver reproducibility was moderate. Adding a CT scan resulted in a slight increase in intraobserver reliability but no increase in interobserver reproducibility. Sjödén and colleagues[9,14] also concluded that adding a CT scan did not improve the reproducibility of proximal humerus fracture classification. They showed that 2-D and 3-D CT did not significantly improve interobserver or intraobserver reliability of AO and Neer classifications made using plain radiographs alone. However, the CT scans in their studies had several technical limitations. Slice thickness was 3 mm, resulting in reformatted images with reduced quality. In addition, the images reviewed were on photographic film, not digital images, and this format did not allow for interactive assessment of fractures.

[Computed tomography] use in managing proximal humerus fractures should be carefully weighed, in light of its limited impact on the plan of care.

Sallay and colleagues[21] divided their observers into 2 groups (experts, nonexperts) based on experience in shoulder surgery. Each group reviewed plain radiographs and 3-D CT scans of 12 patients with a proximal humerus fracture. Both groups demonstrated low reliability in identifying displaced fracture fragments on plain films. Adding a CT scan did not improve reliability or reproducibility. However, the study had a small sample size (n = 12), selection bias, and inconsistent radiographic views.

More recent studies have endorsed use of CT and 3-D reconstructions in proximal humerus fracture assessment. Bahrs and colleagues[22] compared conventional radiographic views with 2-D and 3-D CT to establish indications for CT diagnostic images. They showed that CT diagnostics allowed for significantly better assessment of the fractured region than conventional radiographs did. They concluded that primary assessment with conventional radiographs often, but not always, shows a clear presentation of the relevant bony structures, whereas CT with thin-slice technology always provides a clear presentation of the fractured region. Clinically, the authors suggested that CT should be performed when the proximal humerus and the shoulder joint are not presented with sufficient radiographic quality to establish a treatment plan.

Similarly, Brunner and colleagues[23] suggested that stereo-visualization using 3-D volume-rendering datasets was valuable in analyzing and classifying complex fractures of the proximal humerus. This technique demonstrated improved interobserver reliability of AO and Neer fracture classification when compared with plain radiographs and 2-D CT scans. However, the authors did not determine if this method affected clinical or operative decision-making.

Like all radiographic reviews, the present study has inherent limitations. Patients who did not have both plain radiographs and a CT scan were excluded from the study. As a result, this study may be biased toward more complex fractures, which commonly warrant CT evaluation. However, because the results demonstrated better reliability for the AO classification in less complex fracture patterns,[8,9,14] this selection bias appears not to have compromised the results. In addition, it has been shown that repeated training on the accurate use of a classification system improves reliability.[6] Participants in this investigation did not participate in a structured review. However, because the AO classification system is the primary tool for fracture characterization at the sponsoring institution, the facility and experience of the observers may have compensated for lack of pre-participation review.

Conclusion

In the evaluation of proximal humerus fractures, CT scans are often obtained as an adjunct in determining the plan of care. This investigation compared use of plain radiographs alone with use of a preoperative CT scan to determine if CT significantly affected surgeon decision-making. CT significantly improved interobserver reliability in fracture characterization and assessment, but did not improve agreement in AO fracture classification. In addition, CT did not significantly alter surgeon treatment recommendations. Although CT provides more detail about fracture characteristics, its use in managing proximal humerus fractures should be carefully weighed, in light of its limited impact on the plan of care.

Dr. Ramappa is Chief of Sports Medicine and Shoulder Surgery, Department of Orthopaedics, Beth Israel Deaconess Medical Center, Boston, Massachusetts. Dr. Patel is Resident, Department of Medicine, Duke University Medical Center, Durham, North Carolina. Dr. Goswami is Orthopaedic Resident, University Hospitals Coventry and Warwickshire, Coventry, CV2 2DX, United Kingdom, and Honorary Academic Fellow, Education and Development, Warwick Medical School, University of Warwick, Coventry, CV4 7AL, United Kingdom. Dr. Zurakowski is Director of Biostatistics, Departments of Anesthesiology and Surgery, Boston Children's Hospital, Boston, Massachusetts. Dr. Yablon is Diagnostic Radiologist, Department of Radiology, University of Michigan School of Medicine, Ann Arbor, Michigan. Dr. Rodriguez is Chief of Orthopaedic Trauma, Dr. Appleton is Orthopaedic Traumatologist, and Dr. DeAngelis is Attending Orthopaedic Surgeon and Director of Research, Sports Medicine and Shoulder Surgery, Department of Orthopaedics, Beth Israel Deaconess Medical Center, Boston, Massachusetts.

Address correspondence to: Arun J. Ramappa, MD, Department of Orthopaedics, Beth Israel Deaconess Medical Center, 330 Brookline Ave, ST-10, Boston, MA 02215 (tel, 617-667-3940; fax, 617-667-2155; e-mail, aramappa@bidmc.harvard.edu).

References

1. Baron JA, Barrett JA, Karagas MR. The epidemiology of peripheral fractures. *Bone*. 1996;18(3 suppl):209S-213S.
2. Kannus P, Palvanen M, Niemi S, Parkkari J, Järvinen M, Vuori I. Increasing number and incidence of osteoporotic fractures of the proximal humerus in elderly people. *BMJ*. 1996;313(7064):1051-1052.
3. Neer CS 2nd. Displaced proximal humeral fractures: part I. Classification and evaluation. 1970. *Clin Orthop*. 2006;(442):77-82.
4. Neer CS 2nd. Displaced proximal humeral fractures. Part I. Classification and evaluation. By Charles S. Neer, I, 1970. *Clin Orthop*. 1987;(223):3-10.
5. Neer CS 2nd. Displaced proximal humeral fractures. II. Treatment of three-part and four-part displacement. *J Bone Joint Surg Am*. 1970;52(6):1090-1103.
6. Bernstein J, Adler LM, Blank JE, Dalsey RM, Williams GR, Iannotti JP. Evaluation of the Neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. *J Bone Joint Surg Am*. 1996;78(9):1371-1375.
7. Sidor ML, Zuckerman JD, Lyon T, Koval K, Cuomo F, Schoenberg N. The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. *J Bone Joint Surg Am*. 1993;75(12):1745-1750.
8. Siebenrock KA, Gerber C. The reproducibility of classification of fractures of the proximal end of the humerus. *J Bone Joint Surg Am*. 1993;75(12):1751-1755.
9. Sjödén GO, Movin T, Güntner P, et al. Poor reproducibility of classification of proximal humeral fractures. Additional CT of minor value. *Acta Orthop Scand*. 1997;68(3):239-242.
10. Brorson S, Bagger J, Sylvest A, Hróbjartsson A. Low agreement among 24 doctors using the Neer-classification; only moderate agreement on displacement, even between specialists. *Int Orthop*. 2002;26(5):271-273.
11. Brorson S, Bagger J, Sylvest A, Hrøbjartsson A. Improved interobserver variation after training of doctors in the Neer system. A randomised trial. *J Bone Joint Surg Br*. 2002;84(7):950-954.
12. Borrelli J Jr, Goldfarb C, Catalano L, Evanoff BA. Assessment of articular fragment displacement in acetabular fractures: a comparison of computerized tomography and plain radiographs. *J Orthop Trauma*. 2002;16(7):449-456.
13. Castagno AA, Shuman WP, Kilcoyne RF, Haynor DR, Morris ME, Matsen FA. Complex fractures of the proximal humerus: role of CT in treatment. *Radiology*. 1987;165(3):759-762.
14. Sjödén GO, Movin T, Aspelin P, Güntner P, Shalabi A. 3D-radiographic analysis does not improve the Neer and AO classifications of proximal humeral fractures. *Acta Orthop Scand*. 1999;70(4):325-328.
15. Billet FP, Schmitt WG, Gay B. Computed tomography in traumatology with special regard to the advances of three-dimensional display. *Arch Orthop Trauma Surg*. 1992;111(3):131-137.
16. Andersen DJ, Blair WF, Steyers CM Jr, Adams BD, el-Khouri GY, Brandser EA. Classification of distal radius fractures: an analysis of interobserver reliability and intraobserver reproducibility. *J Hand Surg Am*. 1996;21(4):574-582.
17. Cohen JA. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37-46.
18. Cole RJ, Bindra RR, Evanoff BA, Gilula LA, Yamaguchi K, Gelberman RH. Radiographic evaluation of osseous displacement following intra-articular fractures of the distal radius: reliability of plain radiography versus computed tomography. *J Hand Surg Am*. 1997;22(5):792-800.
19. Wainwright AM, Williams JR, Carr AJ. Interobserver and intraobserver variation in classification systems for fractures of the distal humerus. *J Bone Joint Surg Br*. 2000;82(5):636-642.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
21. Sallay PI, Pedowitz RA, Mallon WJ, Vandemark RM, Dalton JD, Speer KP. Reliability and reproducibility of radiographic interpretation of proximal humeral fracture pathoanatomy. *J Shoulder Elbow Surg*. 1997;6(1):60-69.
22. Bahrs C, Rolauffs B, Südkamp, et al. Indications for computed tomography (CT-) diagnostics in proximal humeral fractures: a comparative study of plain radiography and computed tomography. *BMC Musculoskelet Disord*. 2009;10:33.
23. Brunner A, Honihmann P, Treumann T, Babst R. The impact of stereo-visualisation of three-dimensional CT datasets on the inter- and intraobserver reliability of the AO/OTA and Neer classifications in the assessment of fractures of the proximal humerus. *J Bone Joint Surg Br*. 2009;91(6):766-771.