The Role of Computed Tomography in Evaluating Intra-Articular Distal Humerus Fractures

Betsy M. Nolan, MD, Stephan J. Sweet, MD, MPH, Eric Ferkel, MD, Aniebiet-Abasi Udofia, MD, MBA, and John Itamura, MD

Abstract

Computed tomography (CT) is often used to evaluate intra-articular distal humerus fracture patterns, but it increases radiation exposure and cost.

We conducted a study to determine the effect of adding CT evaluation to plain radiographic evaluation on the classification of, and treatment plans for, intra-articular distal humerus fractures. Nine blinded orthopedic surgeons evaluated 30 consecutive fractures for classification and surgical approach. Evaluations were performed first using plain radiographs and then again using the same radiographs plus CT images. Statistical analysis was performed using the κ correlation coefficient and Cramer V testing. We hypothesized that adding CT images to plain radiographs would change the classification and treatment of these fractures and would improve interobserver agreement on classification and treatment.

Intraobserver reliability (Cramer V) was fair (.393) for classification and moderate (.426) for treatment. Interobserver reliability (Cohen κ) did not improve with CT: For classification, κ was .21 without CT and .20 with CT; for treatment, κ was .28 without CT and .27 with CT. When classifying the fractures, attending surgeons chose the multiplanar fracture pattern 25.6% of the time without CT, and remained consistent at 23.3% with CT. Trainees chose this fracture pattern much less often without CT than with CT. Use of CT changed the treatment for multiplanar fractures (73.7% lateral approach vs 51.9% posterior approach with olecranon osteotomy).

When added to plain radiographic evaluation, CT evaluation changes classification and treatment plans. Interobserver reliability did not improve. Less experienced surgeons were more likely to identify multiplanar fracture patterns with use of CT. We recommend performing CT for all intra-articular distal humerus fractures.

Elbow fractures constitute 7% of all adult fractures, and 30% of these fractures are distal humerus fractures.[1,2] Of these, 96% involve disruption of the articular surface.[3] Intra-articular distal humerus fracture patterns can be difficult to characterize on plain radiographs, and therefore computed tomography (CT) is often used. The surgeon's understanding of the fracture pattern and the deforming forces affects choice of surgical approach. In particular, multiplanar fracture patterns, including coronal shear fractures of the capitellum or trochlea, are often difficult to recognize on plain radiographs. Identification of a multiplanar fracture pattern may require a change in approach or fixation. CT is useful for other intra-articular fractures, such as those of the proximal humerus,[3-6] but involves increased radiation and cost.

We conducted a study to determine the effect of adding CT evaluation to plain radiographic evaluation on the classification of, and treatment plans for, intra-articular distal humerus fractures. We hypothesized that adding CT images to plain radiographs would change the classification and treatment of these fractures and would improve interobserver agreement on classification and treatment.

Materials and Methods

After obtaining University of Southern California Institutional Review Board approval, we retrospectively studied 30 consecutive cases of adult intra-articular distal humerus fractures treated by Dr. Itamura at a level I trauma center between 1995 and 2008. In each case, the injured elbow was imaged with plain radiography and CT. Multiple machines were used for CT, but all according to the radiology department's standard protocol. The images were evaluated by 9 independent observers from the same institution: 3 orthopedic surgeons (1 fellowship-trained shoulder/elbow subspecialist, 1 fellowship-trained upper extremity subspecialist, 1 fellowship-trained orthopedic trauma surgeon), 3 shoulder/elbow fellows, and 3 senior residents pursuing upper extremity fellowships on graduation.

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

No observer was involved in the care of any of the patients. All identifying details were removed from the patient information presented to the observers. For each set of images, the observer was asked to classify the fractures according to the Mehne and Matta classification system,[7,8] which is the predominant system used at our institution.

Diagrams of this classification system were provided, but there was no formal observer training or calibration. Seven treatment options were presented: (1) open reduction and internal fixation (ORIF) using a posterior approach with olecranon osteotomy, (2) ORIF using a posterior approach, (3) ORIF using a lateral approach, (4) ORIF using a medial approach, (5) ORIF using an anterior/anterolateral approach, (6) total elbow arthroplasty, and (7) nonoperative management. The only clinical data provided were patient age and sex.

Images were evaluated in blinded fashion. Two rounds of evaluation were compared. In round 1, plain radiographs were evaluated; in round 2, the same radiographs plus corresponding 2-dimensional (2-D) CT images. A minimum of 1 month was required between viewing rounds.

Statistical Analysis

Statistical analysis was performed by the Statistical Consultation and Research Center at our institution. Cohen κ was calculated to estimate the reliability of the fracture classification and treatment plan made by different observers on the same occasion (interobserver reliability). Cramer V[9] was calculated to estimate the reliability of the fracture classification and treatment plan made by the same observer on separate occasions (intraobserver reliability). It measures the association between the 2 ratings as a percentage of their total variation. The κ value and Cramer V value were also used to evaluate results based on the observers' training levels. Both κ and Cramer V values are interpreted as follows: .00 to .20 indicates slight agreement; .21 to .40, fair agreement; .41-.60, moderate agreement; .61 to .80, substantial agreement; and ≥.81, almost perfect agreement. Zero represents no agreement, and 1.00 represents perfect agreement.

Results

Overall intraobserver reliability for classification was fair (.393). It was moderate for the treatment plan (.426) between viewing rounds. Residents had the highest Cramer V value at .60 (moderate) for classification reliability, and attending surgeons had the highest value at .52 (moderate) for treatment plan. All 3 groups (residents, fellows, attending surgeons) showed moderate intraobserver agreement for treatment plan (**Table 1**).

Interobserver reliability did not improve with the addition of CT in round 2. Reliability was fair at both viewing rounds for classification and for treatment. For classification, the overall κ value was .21 for the first round and .20 for the second round. For treatment plan, the overall κ value was .28 for the first round and .27 for the second round. Attending surgeons decreased in agreement with regard to treatment plan with the addition of CT (.46, moderate, to .32, fair). Fellows had only slight agreement for both rounds with regard to classification as well as treatment (**Table 2**).

ORIF using a posterior approach with an olecranon osteotomy was the most common choice of treatment method overall at both time points (58.1% and 63.7%) and was still the most common choice when each group of observers (residents, fellows, faculty) was considered separately (**Figure 1**).

When classifying the fractures, attending surgeons chose the multiplanar fracture pattern 25.6% of the time when viewing radiographs only, and remained consistent in choosing this pattern 23.3% of the time when CT was added to radiographs. Fellows and residents chose this fracture pattern much less often (8.9% and 7.8%, respectively) when viewing radiographs only. Both fellows and residents increased their choice of the multiplanar fracture pattern by 10% (18.9% for fellows, 17.8% for residents) when CT was added (**Figure 2**).

Overall, the recognition of a multiplanar fracture pattern increased when CT was added. On 30 occasions, an answer

Table 1. Intraobserver Agreement

Raters	Cramer V
All	
Classification	.393 (fair)
Treatment	.426 (moderate)
Attendings	
Classification	.455 (moderate)
Treatment	.520 (moderate)
Fellows	
Classification	.304 (fair)
Treatment	.502 (moderate)
Residents	
Classification	.600 (moderate)
Treatment	.401 (fair)

Table 2. Interobserver Agreement (Cohen κ)

	Overall		Attendings		Fellows		Residents	
Viewing[a]	Classification	Treatment	Classification	Treatment	Classification	Treatment	Classification	Treatment
1	.21 (fair)	.28 (fair)	.29 (fair)	.46 (moderate)	.10 (poor)	.08 (poor)	.22 (fair)	.25 (fair)
2	.20 (poor)	.27 (fair)	.21 (fair)	.32 (fair)	.05 (poor)	.15 (poor)	.32 (fair)	.38 (fair)

[a]1, plain radiographs only; 2, plain radiographs plus computed tomography.

was changed from another classification pattern to the multiplanar pattern when CT was added. Only 6 times did an observer change a multiplanar pattern selection at round 1 to another choice at round 2.

Adding CT in round 2 changed the treatment plan for multiplanar fractures. At round 1, 73.7% chose ORIF using a lateral approach for treatment of the multiplanar fracture versus 10.5% who chose ORIF using a posterior approach with an olecranon osteotomy. The choice of the posterior approach with olecranon osteotomy increased to 51.9% at round 2, using the technique we have previously described.[5,10]

Overall intraobserver reliability for classification was fair (.393). It was moderate for the treatment plan (.426) between viewing rounds. Residents had the highest Cramer V value at .60 (moderate) for classification reliability, and faculty had the highest value at .52 (moderate) for treatment plan. All 3 groups (residents, fellows, attending surgeons) showed moderate intraobserver agreement for treatment plan (Table 1).

Interobserver reliability did not improve with the addition of CT in round 2. Reliability for classification was fair for round 1 and slight for round 2. Reliability was fair at both viewing rounds for treatment. For classification, the overall κ value was .21 for round 1 and .20 for round 2. For treatment plan, the overall κ value was .28 for round 1 and .27 for round 2. Attending surgeons decreased in agreement with regard to treatment plan with the addition of CT (.46, moderate, to .32, fair). Fellows had only slight agreement for both rounds with regard to classification as well as treatment (Table 2).

Discussion

In this study, CT changed both classification and treatment when added to plain radiographs. Interestingly, interobserver

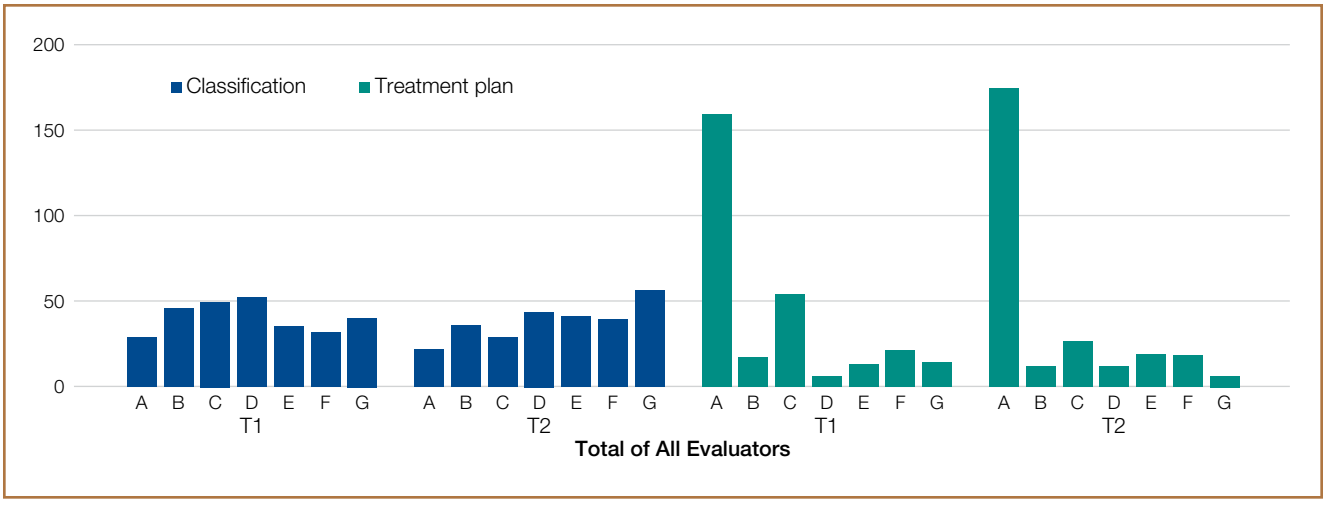

Figure 1. Classification and treatment choice by level of training. A through G are the classification and treatment options chosen based on the handout that was given to evaluators. Abbreviations: T1, time 1 (first viewing); T2, time 2 (second viewing).

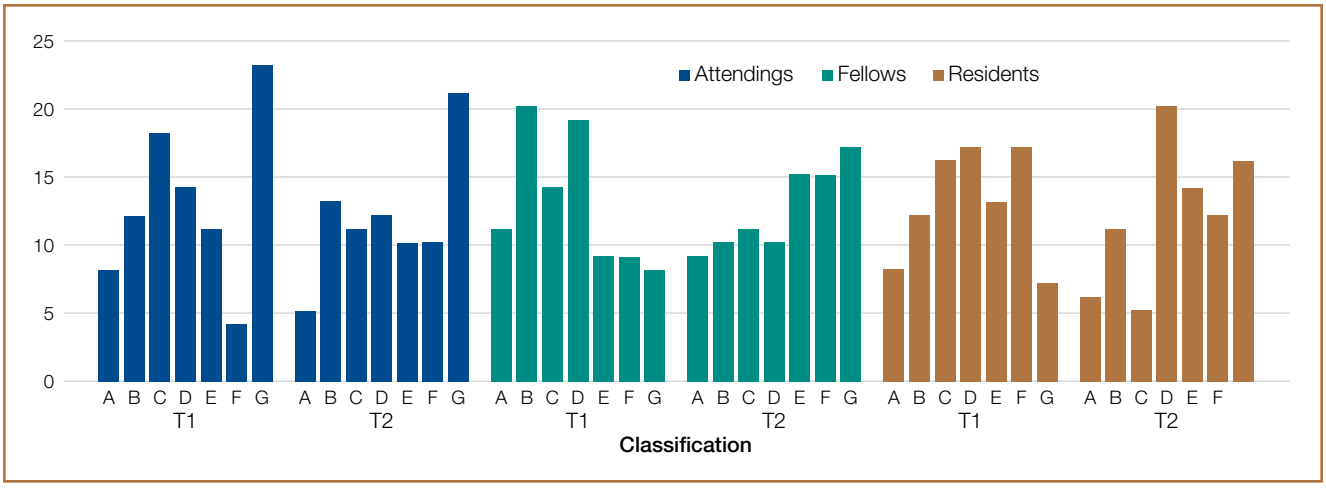

Figure 2. Classification by level of experience. A through G are the classification and treatment options chosen based on the handout that was given to evaluators. Abbreviations: T1, time 1 (first viewing); T2, time 2 (second viewing).

reliability did not improve for classification or treatment with the addition of CT. This finding suggests substantial disagreement among qualified observers that is not resolved with more sophisticated imaging. We propose this disagreement is caused by differences in training and experience with specific fracture patterns and surgical approaches.

Our fair to moderate interobserver reliability using radiographs only is consistent with a study by Wainwright and colleagues,[11] who demonstrated fair to moderate interobserver reliability with radiographs only using 3 different classification systems. CT did not improve interobserver reliability in the present study.

To our knowledge, the effect of adding CT to plain radiographs on classification and treatment plan has not been evaluated. Doornberg and colleagues[2] evaluated the effect of adding 3-dimensional (3-D) CT to a combination of radiographs and 2-D CT. Using the AO (Arbeitsgemeinschaft für Osteosynthesefragen) classification[12] and the classification system of Mehne and Matta, they found that 3-D CT improved intraobserver and interobserver reliability for classification but improved only intraobserver agreement for treatment. Interobserver agreement for treatment plan remained fair. In parallel with their study, fracture classification in our study was more often changed with CT than the treatment plan was. Training level appeared not to affect this finding. We found fair interobserver agreement for treatment choice as well, which was not improved by adding CT. Doornberg and colleagues[2] concluded that the "relatively small added expense of three-dimensional computed tomography scans seems worthwhile."

When evaluating specific fracture patterns in the Mehne and Matta classification system, we observed that less experienced surgeons (residents, fellows) were much more likely to identify multiplanar fracture patterns with the aid of CT. Use of CT did not change attending surgeons' recognition of these multiplanar fractures, suggesting that the faculty were more capable of appreciating these fracture patterns with radiographs only (**Figure 3**). We also observed that adding CT changed the predominant treatment plan for multiplanar fractures from a lateral approach to a posterior approach with an olecranon osteotomy. Failure to appreciate this component of the fracture before surgery could lead to an increased intraoperative difficulty level. Failure to appreciate it during surgery could lead to unexpected postoperative displacement and ultimately poorer outcome.

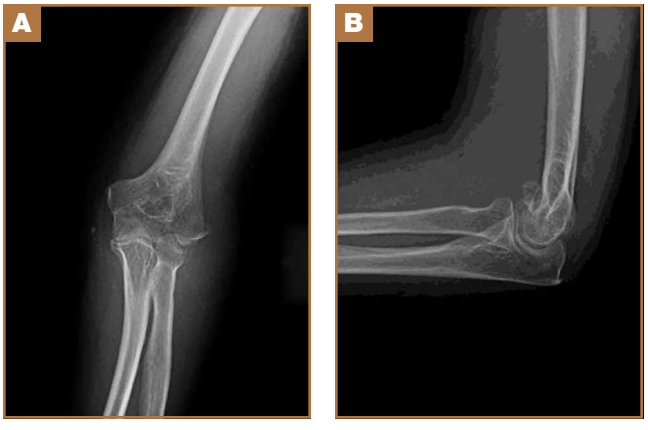

Figure 3. (A, B) Plain radiographs show multiplane fracture.

There are limitations to our study. There is no gold standard for assessing the accuracy of classification decisions. Intraoperative classification could have served as a gold standard, but the fractures were not routinely assigned a classification during surgery. Brouwer and colleagues[13] evaluated the diagnostic accuracy of CT (including 3-D CT) with intraoperative AO classification as a reference point and found improvement in intraobserver agreement but not interobserver agreement when describing fracture characteristics—and no significant effect on classification.

We used a single classification system, the one primarily used at our institution and by Dr. Itamura. There are many systems,[7,12,14] all with their strengths and weaknesses, and no one system is used universally. Adding a system would have allowed us to compare results of more than one system. Our aim, however, was to keep our form simple for the sake of participation and completion of the viewings by each volunteer.

Only 2-D CT was used for this study, as 3-D images were not available for all patients. Although this is a potential weakness, it appears that, based on the study by Doornberg and colleagues,[2] adding 3-D imaging resulted in only modest improvement in the reliability of classification and no significant improvement in agreement on treatment recommendation.

In addition, our results were likely biased by the fact that 8 of the 9 evaluators were trained by Dr. Itamura, who very often uses a posterior approach with an olecranon osteotomy for internal fixation of distal humerus intra-articular fractures, as previously described.[8,10] Therefore, selection of this treatment option may have been overestimated in this study. Nevertheless, after reviewing the literature, Ljungquist and colleagues[15] wrote, "There do not seem to be superior functional results associated with any one surgical approach to the distal humerus."

We did not give the evaluators an indication of patients' activity demands (only age and sex), which may have been relevant when considering total elbow arthroplasty.

Last, performing another round of evaluations with only plain radiographs, before introducing CT, would have provided intraobserver reliability results on plain radiograph evaluation, which could have been compared with intraobserver reliability when CT was added. Again, this was excluded to encourage participation and create the least cumbersome evaluation experience possible, which was thought appropriate, as this information is already in the literature.

Conclusion

Adding CT changed classifications and treatment plans. Raters were more likely to change their classifications than their treatment plans. The addition of CT did not increase agreement between observers. Despite the added radiation and cost, we recommend performing CT for all intra-articular distal humerus fractures because it improves understanding of the

fracture pattern and affects treatment planning, especially for fractures with a coronal shear component, which is often not appreciated on plain radiographs.

Dr. Nolan is Clinical Assistant Professor of Orthopaedic Surgery, University of Oklahoma Health Science Center, Oklahoma City, Oklahoma. Dr. Sweet is Orthopaedic Surgeon, Kaiser Permanente Medical Group, Ventura, California. Dr. Ferkel is Orthopaedic Surgery Resident, University of Southern California, Los Angeles, California. Dr. Udofia is Orthopaedic Surgery Resident, Howard University, Washington, DC. Dr. Itamura is Professor of Orthopaedic Surgery, University of Southern California, and Attending Surgeon, Kerlan-Jobe Orthopaedic Clinic, Los Angeles, California.

Acknowledgment: The authors thank Dr. Christianne Lane, University of Southern California Department of Biostatistics, for her assistance with statistical analysis.

Address correspondence to: Betsy M. Nolan, MD, University of Oklahoma Health Science Center, WP 1380, 920 Stanton L. Young Blvd., Oklahoma City, OK 73104 (tel, 405-456-1000 ext. 5166; email, shoulderandelbowdoc@gmail.com).

Am J Orthop. 2015;44(9):E326-E330. Copyright Frontline Medical Communications Inc. 2015. All rights reserved.

References

1. Anglen J. Distal humerus. *J Am Acad Orthop Surg*. 2005;13(5):291-297.
2. Doornberg J, Lindenhovius A, Kloen P, van Dijk CN, Zurakowski D, Ring D. Two and three-dimensional computed tomography for the classification and management of distal humerus fractures. Evaluation of reliability and diagnostic accuracy. *J Bone Joint Surg Am*. 2006;88(8):1795-1801.
3. Pollock JW, Faber KJ, Athwal GS. Distal humerus fractures. *Orthop Clin North Am*. 2008;39(2):187-200.
4. 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.
5. Palvanen M, Kannus P, Niemi S, Parkkari J. Secular trends in the osteoporotic fractures of the distal humerus in elderly women. *Eur J Epidemiol*. 1998;14(2):159-164.
6. 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.
7. Jupiter JB, Mehne DK. Fractures of the distal humerus. *Orthopedics*. 1992;15(7):825-833.
8. Zalavras CG, McAllister ET, Singh A, Itamura JM. Operative treatment of intra-articular distal humerus fractures. *Am J Orthop*. 2007;36(12 suppl):8-12.
9. Cramer H. *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press; 1946.
10. Panossian V, Zalavras C, Mirzayan R, Itamura JM. Intra-articular distal humerus fractures. In: Mirzayan R, Itamura JM, eds. *Shoulder and Elbow Trauma*. New York, NY: Thieme; 2004:67-78.
11. 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.
12. Müller ME, Nazarian S, Koch P, Schatzker J. *The Comprehensive Classification of Fractures in Long Bones*. Berlin, Germany: Springer-Verlag; 1990.
13. Brouwer KM, Lindenhovius AL, Dyer GS, Zurakowski D, Mudgal C, Ring D. Diagnostic accuracy of 2- and 3-dimensional imaging and modeling of distal humerus fractures. *J Shoulder Elbow Surg*. 2012;21(6):772-776.
14. Riseborough EJ, Radin EL. Intercondylar T fractures of the humerus in the adult. A comparison of operative and non-operative treatment in 29 cases. *J Bone Joint Surg Am*. 1969;51(1):130-141.
15. Ljungquist KL, Beran MC, Awan H. Effects of surgical approach on functional outcomes of open reduction and internal fixation of intra-articular distal humeral fractures: a systematic review. *J Shoulder Elbow Surg*. 2012;21(1):126-135.