

Interobserver reliability of radial head fracture classification : two-dimensional compared with three-dimensional CT

AUTHOR(S)

T Guitton, D Ring, Richard Page, Science of Variation Group

PUBLICATION DATE

02-11-2011

HANDLE

10536/DRO/DU:30047653

Downloaded from Deakin University's Figshare repository

Deakin University CRICOS Provider Code: 00113B

Deakin Research Online

This is the published version:

Guitton, Thierry G., Ring, David, Page, Richard S. and Science of Variation Group 2011, Interobserver reliability of radial head fracture classification : two-dimensional compared with three-dimensional CT, *Journal of bone and joint surgery*, vol. 93, no. 21, pp. 2015-2021.

Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30047653

Reproduced with the kind permission of the copyright owner

Copyright : 2011, The Journal of Bone and Joint Surgery

Copyright © 2011 by The Journal of Bone and Joint Surgery, Incorporated

Interobserver Reliability of Radial Head Fracture Classification: Two-Dimensional Compared with Three-Dimensional CT

Thierry G. Guitton, PhD, and David Ring, MD, PhD, on behalf of the Science of Variation Group*

Investigation performed at Massachusetts General Hospital, Boston, Massachusetts, as part of the Science of Variation Group Collaborative

Background: The Broberg and Morrey modification of the Mason classification of radial head fractures has substantial interobserver variation. This study used a large web-based collaborative of experienced orthopaedic surgeons to test the hypothesis that three-dimensional reconstructions of computed tomography (CT) scans improve the interobserver reliability of the classification of radial head fractures according to the Broberg and Morrey modification of the Mason classification.

Methods: Eighty-five orthopaedic surgeons evaluated twelve radial head fractures. They were randomly assigned to review either radiographs and two-dimensional CT scans or radiographs and three-dimensional CT images to determine the fracture classification, fracture characteristics, and treatment recommendations. The kappa multirater measure (κ) was calculated to estimate agreement between observers.

Results: Three-dimensional CT had moderate agreement and two-dimensional CT had fair agreement among observers for the Broberg and Morrey modification of the Mason classification, a difference that was significant. Observers assessed seven fracture characteristics, including fracture line, comminution, articular surface involvement, articular step or gap of ≥2 mm, central impaction, recognition of more than three fracture fragments, and fracture fragments too small to repair. There was a significant difference in kappa values between three-dimensional CT and two-dimensional CT for fracture fragments too small to repair, recognition of three fracture fragments, and central impaction. The difference between the other four fracture characteristics was not significant. Among treatment recommendations, there was fair agreement for both three-dimensional CT and two-dimensional CT.

Conclusions: Although three-dimensional CT led to some small but significant decreases in interobserver variation, there is still considerable disagreement regarding classification and characterization of radial head fractures. Three-dimensional CT may be insufficient to optimize interobserver agreement.

Disclosure: None of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of any aspect of this work. One or more of the authors, or his or her institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. No author has had any other relationships, or has engaged in any other activities, that could be perceived to influence or have the potential to influence what is written in this work. The complete **Disclosures of Potential Conflicts of Interest** submitted by authors are always provided with the online version of the article.

*The Science of Variation Group: Brandon E. Earp, Amy L. Ladd, Peter J. Evans, Christina E. Kuo, Jan Biert, C.N. van Dijk, Phani K. Dantuluri, David E. Ruchelsman, K.J. Ponsen, Maximillian Soong, Tim R. Davis, Ashok K. Shyam, Laura S. Phieffer, C. Michael LeCroy, Martin Richardson, Andrew H. Schmidt, Peter L. Jebson, Paul E. Levin, Gregory J. Della Rocca, Charles A. Goldfarb, Kyle J. Jeray, David M. Kalainov, George S.M. Dyer, Neal T. Chen, A. Lee Osterman, George S. Athwal, Luke P. Leenen, Thomas W. Wright, Marc F. Swiontkowski, David J. Slutsky, Frede Frihagen, Scott F. Duncan, Rick F. Papandrea, Kevin C. Chung, Philip Blazar, Robert J. Feibel, Robert D. Zura, Huub J. van der Heide, Robert Z. Tashjian, Leon Elmans, John A. Jiuliano, Marco Rizzo, Samir Sodha, John A. McAuliffe, Randall W. Culp, Jorge Orbay, Charles Cassidy, Robert G.H. Albers, Leonid I. Katolik, Reid A. Abrams, Mark E. Baratz, Kenneth A. Egol, Joseph M. Conflitti, Doug P. Hanel, Jose M. Nolla, Michael Hausman, Andrew E. Caputo, Rudolf W. Poolman, Terry S. Axelrod, Michael D. McKee, J.C. Goslings, Parag K. Sancheti, Peter R.G. Brink, Carrie R. Swigart, Thomas B. Hughes, Keith A. Segalman, P.V. van Eerten, Brett D. Crist, Edward Diao, Richard S. Page, Lisa L. Lattanza, George Thomas, Jason C. Fanuele, Peter Kloen, Taco Gosens, Charalampos Zalavras, John S. Taras, Jeffrey A. Greenberg, Eric M. Hammerberg, Louis W. Catalano 3rd, Rodrigo F. Pesantez, Arie B. van Vugt, Steve C. Kronlage, Michael A. Baskies, Martin I. Boyer, Peter V. Giannoudis, Michael J. Prayson, and Elena Grosso.

The classification of radial head fractures according to the Broberg and Morrey modification of the Mason classification[1] has substantial interobserver variation with interpretation of radiographs[2,3]. As with classification and characterization of most fractures, the interobserver variation is greater than the intraobserver variation. Evidence suggests that more sophisticated imaging modalities, such as three-dimensional computed tomography (CT), improve intraobserver reliability more than interobserver reliability[4,5]. A major limitation of most studies of observer variation is the use of only a few observers, frequently relatively junior surgeons.

A new collaboration motivated to better understand interobserver variation[6] consists of observers who have completed all training and are independently treating patients. This collaboration provides an opportunity to further investigate interobserver variability and how to reduce it.

Treatment decisions for radial head fractures are often based on radiographic criteria and measurements according to the Broberg and Morrey modification of the Mason classification[7,8]. This investigation tested the hypothesis that three-dimensional CT images improve the interobserver reliability of the classification and characterization of radial head fractures compared with the reliability associated with use of two-dimensional CT and radiographs.

Materials and Methods

Study Design

Independent observers (all orthopaedic surgeons) from several countries were invited to evaluate twelve radial head fractures from a convenience sample. Fractures were selected to represent a full spectrum of radial head morphologies and overall injury patterns and were viewed in an online survey. Observers were randomly assigned to review either radiographs and two-dimensional CT or radiographs and three-dimensional CT and then to determine the fracture classification, the fracture characteristics, and treatment recommendations. The randomization sequence was determined with use of a random number generator in Microsoft Excel for Windows (Microsoft, Redmond, Washington). The study was performed under a protocol approved by the Institutional Research Board at the principal investigator's (D.R.'s) hospital.

This was the inaugural study from a nascent collaborative called the Science of Variation Group. The objectives of the collaborative are to study variation in the definition, interpretation, and classification of injury and disease. The Science of Variation group has created a web-based platform that facilitates large international interobserver studies. With multiple fully trained surgeons from diverse countries and institutions participating, this approach has the potential to provide a powerful forum for studying, understanding, and ultimately reducing interobserver variation during patient care.

Observers

A total of 206 surgeons were invited via e-mail to join the Science of Variation Group. We used lists of various professional organizations as well as friends and acquaintances to identify surgeons to invite for participation. We welcome any interested surgeon who wishes to join. Other than an acknowledgment as part of the author collaborative in the paper, no incentives were provided. One hundred surgeons were interested in participation and logged on to the web site. Forty-eight surgeons were randomized to two-dimensional CT scans and radiographs and fifty-two to three-dimensional CT scans and radiographs. Four weekly reminders to complete the online survey were e-mailed. Eighty-eight surgeons completed the study. Three observers were excluded because of inability to view the online study due to hospital restriction. This study presents an analysis of the eighty-five observers who completed the study: thirty-nine in the two-dimensional CT group and forty-six in the three-dimensional CT group.

Fractures

Radiographs and computed tomography scans of radial head fractures were identified from a list of all cases treated by the senior investigator from the beginning of the year 2000 until the end of the year 2006 at a Level-1 trauma center. The scanning technique was evaluated to determine suitability for three-dimensional reconstructions (slice thickness between 0.62 and 1.25 mm, no metal implants). Inclusion criteria were (1) radial head fracture, (2) CT scan appropriate for three-dimensional reconstruction, and (3) patient age of eighteen years or older. Inadequate quality of the CT scan prompted exclusion of the associated case from the study. Radiographs and CT scans of radial head fractures from thirty patients were blinded by an independent research fellow for use in this study. Seven fractures were part of an injury that included either an elbow dislocation (six patients, four with associated fracture of the coronoid process) or fracture of the proximal portion of the ulna (one patient). Five fractures were isolated injuries. Two of the authors (one subspecialty-trained upper-extremity surgeon [D.R.] and one research fellow in upper-extremity trauma [T.G.G.]) selected twelve cases among which the radial head fractures were of different size, morphology, and location, representing most of the different patterns of traumatic elbow instability with radial head fracture. Radiographs, two-dimensional CT scans, and three-dimensional CT reconstructions were uploaded to the research group's web site. The three-dimensional CT reconstructions were created with use of Vitrea imaging software (Vital Images, Minnetonka, Minnesota). For each case, videos with two-dimensional CT images and three-dimensional CT images along the sagittal, coronal, and axial planes were created. The three-dimensional CT videos included a reconstruction of the entire elbow and a reconstruction with the distal part of the humerus subtracted. Observers could scroll through the videos or play them automatically.

Evaluation

Observers logged in independently on the web site. After logging on to the web site, they were asked to provide the following demographic and professional information: (1) location of practice, (2) years in independent practice, (3) participation in resident or fellow education, (4) number of radial head fractures treated per year, and (5) clinical specialty. Subsequently, observers were asked to classify the fractures according to the Broberg and Morrey modification of the Mason classification[7,8]. Type-4 fractures (radial head fracture associated with an elbow dislocation) were excluded. Observers were provided with the original description and corresponding images of the classification system.

The observers were also asked seven questions regarding fracture characteristics: (1) Does the fracture line separate the entire articular surface from the radial neck? (2) Is there any comminution of the radial neck? (3) Does the fracture involve the articular surface? (4) Is there an articular step or gap of ≥2 mm? (5) Is there any central impaction of the articular surface? (6) Are there more than three fracture fragments? (7) Are any of the fragments too small to repair? They were also asked which of the following was their preferred treatment recommendation: (1) nonoperative management; (2) open reduction and internal fixation with screws, wires, or pins; (3) open reduction and internal fixation with plate and screws; (4) radial head excision; or (5) radial head replacement (arthroplasty). Observers were blinded to clinical information. Observers could comment on each case, and all questions had to be completed in order to continue with the next case. The observers completed the study at their own time and pace.

Statistical Analysis

The kappa multirater measure (κ) was used to estimate agreement among surgeons with respect to fracture classification, fracture characteristics, and treatment approach. Kappa values are commonly used to describe chance-corrected agreement in a variety of intraobserver and interobserver studies[9-11]. Agreement among observers was calculated with use of the kappa multirater

TABLE I Post Hoc Power Analysis for Comparing Two-Dimensional and Three-Dimensional CT Images

Variable	Power (1 – β)*	Minimum No. per Group for 80% Power*
Broberg and Morrey modification of the Mason classification	0.99	20
Fracture characteristics		
Fracture line	0.29	179
Comminution radial neck	0.19	289
Articular surface involvement	0.05	500
Articular step or gap ≥2 mm	0.25	209
Central impaction	0.72	57
Three fracture fragments	0.99	5
Too small to repair	0.99	8
Proposed treatment	0.99	10

*Power is based on a comparison of kappa interobserver agreement with use of a two-sample independent Z test (thirty-nine surgeons viewed two-dimensional CT and forty-six surgeons viewed three-dimensional CT).

measure described by Siegel and Castellan[12]. Kappa values were interpreted with use of the guidelines proposed by Landis and Koch[10]: values of 0.01 to 0.20 indicate slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 or more, almost perfect agreement. Zero indicates no agreement beyond that expected due to chance alone, –1.00 means total disagreement, and +1.00 represents perfect agreement[10,11]. Two-sample independent Z-tests were performed for each variable to compare the kappa for two-dimensional CT with that of three-dimensional CT. Since the samples compared in this study were not independent (the same set of fractures were rated by the two-dimensional CT and three-dimensional CT group), this method produced conservative estimates of the p values. A post hoc power analysis was performed with use of nQuery Advisor (version 7.0, nQuery Advisor; Statistical Solutions, Saugus, Massachusetts) to identify the power of each comparison and the sample size necessary to achieve a power of 80% given that both the effect size and rater ratio remain constant at each iteration (Table I).

Sources of Funding
No funding was received in direct support of this study.

Results

Observer Demographics

A total of eighty-five observers participated in this investigation. The observer demographics are summarized in a table in the Appendix. Included in this group of orthopaedic surgeons were three general orthopaedic surgeons, twenty-five orthopaedic trauma surgeons, eleven shoulder and elbow surgeons, thirty-eight hand and wrist surgeons, and eight other surgeons. Among the surgeons that were classified as "other," there were three hand surgeons, two trauma surgeons, and three upper-extremity surgeons (hand, wrist, elbow, and shoulder).

Interobserver Reliability (Table II)

Classification

When fractures were classified according to the Broberg and Morrey modification of the Mason classification system[1], the use of two-dimensional CT scans was associated with fair agreement and the use of three-dimensional CT reconstructions was associated with moderate agreement (the kappa multirater measure and the standard error of the mean were 0.37 and 0.010, respectively, for two-dimensional CT and 0.49 and 0.023, respectively, for three-dimensional CT; $p < 0.001$) (Table II).

Fracture Characteristics

Agreement regarding central impaction of the articular surface was fair with use of two-dimensional CT scans and slight with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.22 and 0.027, respectively, for two-dimensional CT and 0.15 and 0.010, respectively, for three-dimensional CT; $p = 0.006$). Interobserver agreement regarding the presence of more than three fracture fragments was fair with use of two-dimensional CT scans and substantial with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.38 and 0.011, respectively, for two-dimensional CT and 0.61 and 0.010, respectively, for three-dimensional CT; $p < 0.001$). Agreement on presence of fragments too small to repair was moderate with use of two-dimensional CT scans and substantial with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.47 and 0.013, respectively, for two-dimensional CT and 0.61 and 0.010, respectively, for three-dimensional CT; $p < 0.001$) (Table II).

Treatment

Interobserver agreement on treatment was fair with both two-dimensional CT scans and three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.26 and 0.012, respectively, for two-dimensional CT and 0.40 and 0.013, respectively, for three-dimensional CT; $p < 0.001$) (Table II).

TABLE II Interobserver Reliability

Variable	Two-Dimensional CT			Three-Dimensional CT			P Value
	Category	κ	95% CI	Category	κ	95% CI	
Broberg and Morrey modification of the Mason classification	Fair	0.37	(0.35, 0.39)	Moderate	0.49	(0.45, 0.54)	<0.001*
Fracture characteristics							
Fracture line	Moderate	0.41	(0.31, 0.51)	Moderate	0.50	(0.43, 0.58)	0.139
Comminution radial neck	Fair	0.33	(0.25, 0.41)	Fair	0.38	(0.33, 0.42)	0.328
Articular surface involvement	Slight	0.02	(–0.46, 0.49)	Poor	0.00	(–0.81, 0.80)	0.961
Articular step or gap ≥2 mm	Substantial	0.65	(0.58, 0.71)	Substantial	0.73	(0.63, 0.82)	0.152
Central impaction	Fair	0.22	(0.17, 0.28)	Slight	0.15	(0.13, 0.16)	0.006*
Three fracture fragments	Fair	0.38	(0.35, 0.40)	Substantial	0.61	(0.59, 0.63)	<0.001*
Too small to repair	Moderate	0.47	(0.44, 0.49)	Substantial	0.61	(0.59, 0.63)	<0.001*
Proposed treatment	Fair	0.26	(0.24, 0.29)	Fair	0.40	(0.37, 0.43)	<0.001*

*Significant value. κ = kappa multirater measure. CI = confidence interval.

Observer Demographics and the Broberg and Morrey Modification of the Mason Classification

When classifying fractures according to the Broberg and Morrey modification of the Mason classification, agreement among United States observers was fair with use of two-dimensional CT scans and moderate with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.32 and 0.01, respectively, for two-dimensional CT and 0.52 and 0.03, respectively, for three-dimensional CT; $p < 0.001$) (Table III).

Agreement among observers who were in practice five or fewer years was moderate with use of two-dimensional CT scans and substantial with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.44 and 0.03, respectively, for two-dimensional CT and 0.62 and 0.08, respectively, for three-dimensional CT; $p = 0.039$). Agreement among observers who were in practice from six to ten years was fair with use of two-dimensional CT scans and moderate with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.32 and 0.05, respectively, for two-dimensional CT and 0.53 and 0.05, respectively, for three-dimensional CT; $p = 0.002$). Agreement among observers who were in practice from eleven to twenty years was fair with use of two-dimensional CT scans and moderate with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.35 and 0.02, respectively, for two-dimensional CT and 0.45 and 0.04, respectively, for three-dimensional CT; $p = 0.011$).

Agreement among observers who treated five or fewer radial head fractures per year was fair with use of either two-dimensional CT scans or three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.27 and 0.03, respectively, for two-dimensional CT and 0.32 and 0.14, respectively, for three-dimensional CT; $p = 0.76$). Agreement among observers who treated six to ten radial head fractures per year was fair with use of two-dimensional CT scans and moderate with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.39 and 0.04, respectively, for two-dimensional CT and 0.48 and 0.04, respectively, for three-dimensional CT; $p = 0.063$). Agreement among observers who treated eleven to twenty radial head fractures per year was moderate with use of either two-dimensional CT scans or three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.44 and 0.03, respectively, for two-dimensional CT and 0.46 and 0.05, respectively, for three-dimensional CT; $p = 0.66$). Agreement among observers who treated more than twenty radial head fractures per year was moderate with use of either two-dimensional CT scans or three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.46 and 0.06, respectively, for two-dimensional CT and 0.52 and 0.05, respectively, for three-dimensional CT; $p = 0.41$).

Agreement among orthopaedic traumatology specialist observers was fair with use of two-dimensional CT scans and moderate with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.37 and 0.03, respectively, for two-dimensional CT and 0.47 and 0.04, respectively, for three-dimensional CT; $p < 0.05$). Agreement among hand and wrist specialist observers was fair with use of two-dimensional CT scans and moderate with use of three-dimensional CT reconstructions (the kappa multirater measure and the standard error of the mean were 0.31 and 0.03, respectively, for two-dimensional CT and 0.54 and 0.03, respectively, for three-dimensional CT; $p < 0.001$).

TABLE III Differences in Agreement on the Broberg and Morrey Modification of the Mason Classification Between Observer Demographics and Imaging Modality*

	Two-Dimensional CT (N = 39 Observers)				Three-Dimensional CT (N = 46 Observers)				
Variable	N	Category	κ	SE	N	Category	κ	SE	P Value†
Practice									
Europe	6	Moderate	0.50	0.05	9	Moderate	0.43	0.04	0.275
United States	28	Fair	0.32	0.01	31	Moderate	0.52	0.03	<0.001
Other	5	Moderate	0.44	0.07	6	Moderate	0.51	0.10	0.561
Years in practice									
0-5	11	Moderate	0.44	0.03	7	Substantial	0.62	0.08	0.039
6-10	6	Fair	0.32	0.05	15	Moderate	0.53	0.05	0.002
11-20	18	Fair	0.35	0.02	17	Moderate	0.45	0.04	0.011
21-30	4	Moderate	0.47	0.10	7	Fair	0.38	0.08	0.449
Fractures per year									
0-5	10	Fair	0.27	0.03	3	Fair	0.32	0.14	0.757
6-10	10	Fair	0.39	0.04	15	Moderate	0.48	0.04	0.063
11-20	13	Moderate	0.44	0.03	11	Moderate	0.46	0.05	0.663
>20	6	Moderate	0.46	0.06	17	Moderate	0.52	0.05	0.411
Specialization									
General orthopaedics	2	Moderate	0.49	0.21	1	‡	‡	‡	‡
Orthopaedic traumatology	12	Fair	0.37	0.03	13	Moderate	0.47	0.04	<0.05
Shoulder and elbow	7	Moderate	0.50	0.05	4	Fair	0.30	0.11	0.107
Hand and wrist	14	Fair	0.31	0.03	24	Moderate	0.54	0.03	<0.001
Other	4	Moderate	0.47	0.09	4	Fair	0.35	0.11	0.396

*N = Number of observers, κ = kappa multirater measure, and SE = standard error of the mean. †P value of <0.05 was significant. ‡There was only one general orthopaedic surgeon in this category, so no statistics are possible.

When classifying fractures according to the Broberg and Morrey modification of the Mason classification with use of two-dimensional CT, there were significant differences in agreement between European surgeons and U.S. surgeons (the kappa multirater measure was 0.50 vs. 0.30, respectively; p = 0.001), surgeons who treat as many as five fractures per year and surgeons who treat more than twenty fractures per year (the kappa multirater measure was 0.27 vs. 0.46, respectively; p = 0.006), and orthopaedic traumatology specialists and shoulder and elbow specialists (the kappa multirater measure was 0.37 vs. 0.50, respectively; p = 0.017). When three-dimensional CT was used, the only differences were between surgeons who had been in practice for five years or less and surgeons who had been in practice for twenty-one to thirty years (p = 0.037) (Table IV).

TABLE IV Differences in Observer Agreement for the Broberg and Morrey Modification of the Mason Classification by Imaging Modality

Variable	Two-Dimensional CT	Three-Dimensional CT
Practice		
Europe vs. United States	0.001*	0.06
Years in practice		
0-5 vs. 21-30	0.715	0.037*
Fractures per year		
0-5 vs. >20	0.006*	0.175
Specialization		
Orthopaedic traumatology vs. shoulder and elbow	0.017*	0.175

*Asterisks indicate significant value (p < 0.05).

Discussion

The collaborative, Internet-based approach has facilitated large, international studies of interrater variation[13,14]. Only fully trained surgeons participated, many of whom had substantial clinical experience. Inclusion of surgeons from multiple countries and continents should increase the generalizability of the results. Using high-speed Internet connections and improved compression techniques, we were able to provide high-quality reproduction images and movies via the Internet.

The use of three-dimensional CT images led to small but significant decreases in variation between observers for fracture classification and some fracture characteristics as compared with the differences that resulted from use of two-dimensional CT, but a notable amount of variation remains even with this more sophisticated imaging. Our previous belief that three-dimensional CT images are easier for surgeons to interpret is open to debate since only three to seven fracture characteristics were more reliably identified with three-dimensional CT. On the plus side, three-dimensional CT produced a higher agreement with regard to the Broberg and Morrey modification of the Mason classification than that previously reported in the literature[2,3] and, in comparison with two-dimensional CT, three-dimensional CT was associated with less disagreement in classification across various cultures, training, subspecialty, and levels of experience. Nonetheless, even with use of three-dimensional CT, agreement was only fair or moderate at best. Furthermore, some might interpret these data as showing much less influence on interobserver variation than might be expected.

We speculate that the better agreement of younger surgeons—particularly with three-dimensional CT—is related to greater familiarity with this imaging technique, or perhaps to greater reliance on the precise definitions rather than experience or intuition. We speculate that the very poor agreement regarding articular surface involvement might reflect misunderstanding of the question—that is, based on comments received as part of the survey, some observers may have thought we were referring to involvement of the part of the radial head that articulates with the lesser sigmoid notch of the ulna. The poor agreement regarding central impaction likely reflects the lack of a precise or consistent definition of this term. The findings of this study are otherwise consistent with those of prior studies on the distal part of the humerus, distal part of the radius, and the coronoid[4,5,14].

Three-dimensional reconstructions are made from CT scans and therefore do not require additional scanning or expose the patient to additional radiation. It has been calculated at the investigators' institution that the costs for additional three-dimensional reconstructions are an additional 20% of the costs of a CT scan. The availability of free software such as OsiriX[15] makes it possible for every orthopaedic surgeon to quickly and easily create three-dimensional reconstructions himself or herself, with minimal training.

Other potential sources of interobserver variation include unfamiliar or unclear definitions as well as differences in culture, training, and exposure. In our opinion, the fact that well-trained, experienced observers disagree indicates that there are variations in these factors that lead different experts to see different things in sophisticated images. In other words, reducing interobserver variation seems to depend on something more than better imaging. Additional research to identify and reduce sources of observer variation in the interpretation of diagnostic images is merited.

There are several weaknesses in this study. First, the quality of the radiographs was limited to what had been obtained at the time of injury, which reflects usual practice but not what might be achieved with specific protocols. In addition, we provided limited information about the patient and the injury. There was also a spectrum bias by selecting cases to represent the known variety of injuries, with the result that less common complex fractures were overrepresented compared with the more common minimally or slightly displaced fractures. Our study reflects what would be expected with relatively complex fractures of the radial head—the reliability would be expected to be higher if we included more of the nondisplaced or minimally displaced fractures that include the majority of radial head fractures. Another shortcoming is that a small number of observers either uncommonly or never treat radial head fractures, but we did not plan for exclusions on this basis and therefore did not do so after the fact, so as to avoid introducing bias. The power is based on the total number of observations, allowing us to use a smaller number of cases and thereby decrease the burden on and increase the participation of observers, but for small differences there may not have been sufficient power. Specifically, while our primary study question (reliability of the Broberg and Morrey modification of the Mason classification) was adequately powered, five of the secondary study questions or comparisons were underpowered, and thus our findings should be interpreted with caution. Finally, this is an artificial research situation, given that, in clinical practice, both two-dimensional CT and three-dimensional CT reconstructions would be available for patients.

In conclusion, there is considerable disagreement regarding classification and characterization of radial head fractures, even with use of three-dimensional CT. The use of three-dimensional CT may not be sufficient to optimize interobserver agreement.

Appendix

(eA) A table showing demographic data regarding the observers is available with the online version of this article as a data supplement at jbjs.org. ■

Thierry G. Guitton, PhD
David Ring, MD, PhD
Orthopaedic Hand and Upper Extremity Service,
Massachusetts General Hospital,
Yawkey Center, Suite 2100, 55 Fruit Street,
Boston, MA 02114.
E-mail address for D. Ring: dring@partners.org

References

1. Broberg MA, Morrey BF. Results of treatment of fracture-dislocations of the elbow. Clin Orthop Relat Res. 1987;216:109-19.

2. Morgan SJ, Groshen SL, Itamura JM, Shankwiler J, Brien WW, Kuschner SH. Reliability evaluation of classifying radial head fractures by the system of Mason. Bull Hosp Jt Dis. 1997;56:95-8.

3. Matsunaga FT, Tamaoki MJ, Cordeiro EF, Uehara A, Ikawa MH, Matsumoto MH, dos Santos JB, Belloti JC. Are classifications of proximal radius fractures reproducible? BMC Musculoskelet Disord. 2009;10:120.

4. 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 humeral fractures. Evaluation of reliability and diagnostic accuracy. J Bone Joint Surg Am. 2006;88:1795-801.

5. Harness NG, Ring D, Zurakowski D, Harris GJ, Jupiter JB. The influence of three-dimensional computed tomography reconstructions on the characterization and treatment of distal radial fractures. J Bone Joint Surg Am. 2006;88:1315-23.

6. The Science of Variation Group. Science of Variation. http://www.scienceofvariationgroup.nl. Accessed 2010 Oct 11.

7. Mason ML. Some observations on fractures of the head of the radius with a review of one hundred cases. Br J Surg. 1954;42:123-32.

8. Johnston GW. A follow-up of one hundred cases of fracture of the head of the radius with a review of the literature. Ulster Med J. 1962;31:51-6.

9. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37-46.

10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74.

11. Posner KL, Sampson PD, Caplan RA, Ward RJ, Cheney FW. Measuring interrater reliability among multiple raters: an example of methods for nominal data. Stat Med. 1990;9:1103-15.

12. Siegel S, Castellan NJ. Nonparametric statistics for the behavioral sciences. 2nd ed. New York: McGraw-Hill; 1988.

13. Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, Norman GR, Guyatt GH; Collaboration for Outcome Assessment in Surgical Trials (COAST) Musculoskeletal Group. Evaluating agreement: conducting a reliability study. J Bone Joint Surg Am. 2009;91 Suppl 3:99-106.

14. Lindenhovius A, Karanicolas PJ, Bhandari M, van Dijk N, Ring D; Collaboration for Outcome Assessment in Surgical Trials. Interobserver reliability of coronoid fracture classification: two-dimensional versus three-dimensional computed tomography. J Hand Surg Am. 2009;34:1640-6.

15. Rosset A, Spadola L, Ratib O. OsiriX: an open-source software for navigating in multidimensional DICOM images. J Digit Imaging. 2004;17:205-16.