

Inter- and intraobserver reliability in the assessment of glenoid fracture classifications

Dirk P.H. VAN OOSTVEEN, Olivier P.P. TEMMERMAN, Steven J. VERBERNE, Marco J.M. HOOZEMANS, Bart J. BURGER, Harm C.A. GRAAT, Daniël HOORNENBORG, Michiel H.J.M. VAN DE REST, Didi GUBLER, Arthur VAN NOORT

From the Spaarne Gasthuis Hoofddorp, Noordwest Ziekenhuisgroep Alkmaar, VUMC Amsterdam, UMC Utrecht and AMC Amsterdam.

Throughout literature a number of glenoid classification systems have been described but lack clear correlation with the fracture patterns found in clinical cases. This study aimed to evaluate the intra- and interobserver agreement for fracture classifications of the glenoid, using either plain radiographs, computed tomography (CT) scans.

The study was retrospective, using images with a variety of fracture types. Six observers classified the fracture patterns. Classifications of Ideberg, OTA-AO, Mayo and Euler/Ruedi were used. Agreement was determined using kappa coefficients.

Currently used glenoid fracture classification systems have a fair to moderate intraobserver reliability. Combining plain radiographs and CT scans led to a better observer agreement. For interoberserver reliability, the system of Euler scored slightly better than other systems.

Although Ideberg's classification is the most widely used system, this study does not support superiority of it. Based on this study there is need for a more reliable glenoid classification system.

Keywords : glenoid fracture ; Ideberg classification ; OTA/AO classification ; Euler/Ruedi classification ; CT scan ; plain radiographs.

INTRODUCTION

Fractures of the scapula comprise 0.4 -1 per cent of all fractures, whereas the glenoid is involved in up to 10 percent of these patients (*3*,*10*,*24*,*26*,*27*). Little

No benefits or funds were received in support of this study. The authors report no conflict of interests. has been published regarding these fractures, but these intra-articular fractures can result in persistent pain, instability and osteoarthritis (*3*, *10*, *26*). Clinical diagnosis and prognosis is aided by classifying the fractures (*8*). Ideally, fracture classification systems

- Dirk P.H. Van Oostveen M.D.¹
- Olivier P.P. Temmerman M.D. Ph.D.^{2,3}
- Steven J. Verberne M.D.²
- Marco J.M. Hoozemans Ph.D.^{3,4}
- Bart J. Burger M.D., Ph.D.^{2,3}
- Harm C.A. Graat M.D., Ph.D.²
- Daniël Hoornenborg M.D.⁵
- Michiel H.J.M. Van De Rest M.D.⁶
- Didi Gubler M.D.⁶
- Arthur Van Noort M.D., Ph.D.⁷

¹Department of Orthopaedic Surgery, Jeroen Bosch Hospital, 's-Hertogenbosch, The Netherlands. ²Department of Orthopaedic Surgery, Noordwest Ziekenhuisgroep, Alkmaar, the Netherlands. ³CORAL Centre for Orthopaedic Research Alkmaar,

Department of Orthopaedic Surgery, Noordwest Ziekenhuisgroep, The Netherlands.

⁴Research Institute MOVE, Faculty of Human Movement Sciences, VU University Amsterdam, The Netherlands.

^sDepartment of Orthopaedic Surgery, Slotervaart Hospital, Amsterdam, The Netherlands.

⁶Department of Radiology, Spaarne Gasthuis, Hoofddorp, The Netherlands.

⁷Department of Orthopaedic Surgery, Spaarne Gasthuis, Hoofddorp, The Netherlands.

Corresponding author: D.P.H. van Oostveen -tel: 073-5532000

E-mail: D.v.oostveen@jbz.nl

© 2017 Acta Orthopaedica Belgica.

should be simple, a guide for treatment options and predict clinical outcome.

The first well described classification system was published by Ideberg in 1984(14). This classification was based on plain radiographs, trauma mechanisms and treatment options of 200 glenoid fractures that where retrospectively analyzed. The original Ideberg classification was modified by Goss into 6 types and subtypes (3,10,12,14,24,25). He excluded the bony bankart lesions from the system and introduced new (sub) types.

In an attempt to provide a glenoid fracture classification including the scapular body, Mayo et al. published his classification in 1998 (20). Based on 31 patients that underwent CT and open fixation of a glenoid fracture, he argued that the scapular component in glenoid fractures is underestimated. Outcome was not related to the classification, but was solely dependent on concomitant injury, especially neurologic pathology (20).

An overall fracture classification system, with a separate part for glenoid fractures was developed by the Orthopaedic Trauma Association in collaboration with the AO foundation(19). It is based on the anatomical location and morphology of the fracture. It has an alphanumeric code for each type of fracture, the glenoid is type 14B or 14C. The hierarchy is from less severe to more severe and from less to more detailed in injury description. In contrast to the other classification systems, the AO/OTA focuses separately on glenoid fractures, excluding combinations with coracoid or acromion fractures.

Recently this system has been changed and validated (13,15), but this was published after our study had been carried out.

The scapula fracture classification by Euler and Habermeyer as modification to Ruedi has a subset of glenoid fractures (8,21). This classification is utilised only in German literature (9,23,27). In this classification system fractures can be classified as a combination of different types, as was described by Schofer et al.

The variety in characteristics of these four commonly applied classification systems and the applied imaging modalities may cause debate regarding the optimum diagnostic algorithm. This may be augmented by the increased radiation levels to which the patient is exposed in routine application of CT imaging. Therefore, for a physician, knowledge regarding imaging techniques and their reliability and validity in different fracture classifications is a prerequisite. Unfortunately, apart from the new OTA classification, the reliability and validity of glenoid fracture classification systems based on CT imaging have not been described yet, but are essential clinimetric characteristics of these prognostic tests (15). Therefore, the purpose of this study was to determine the inter- and intra-observer reliability of the use of both plain radiography and CT scans in the assessment of glenoid fractures using the four most commonly used glenoid fracture classification systems.

MATERIALS AND METHODS

A cohort study was performed. To collect a variety of glenoid fracture types, we screened all patients with glenoid fractures that presented themselves at the departments of orthopedic surgery of three university hospitals and two general teaching hospitals in The Netherlands between 2005-2012, using the ICPC code (International Classification of Primary Care) and operation codes. Only cases in which both plain radiographs and CT scans were available in a digital format were included in the current study. Patients were excluded from this study if the imaging data were incomplete, such that retrospective analysis could not be performed.

The true AP shoulder and scapular Y shoulder radiographs were used. For CT imaging, the axial, coronal and sagital images were included for image interpretation. Since patients presented at different hospitals, minor variation (eg scanning sequence, slice thickness) in imaging protocols were accepted.

The diagnostic images were independently interpreted by two experienced musculoskeletal radiologists, two senior orthopedic shoulder surgeons and two orthopaedic registrars in the second half of their training, all of whom were blinded to clinical pre-test data and outcome. Images of the 4 classification systems were distributed to the observers prior to the start of this study (fig 1-4). An oral explanation of the classification systems was given at an instructional meeting by the first author (author 1). The principal investigators (Author 1 and 2) did not participate in the image interpretation. To overcome a learning curve within the study, 2 sets of X-rays and CT scans of patients with glenoid fractures that were not included in this study were used in the plenary session for a collective scoring session by the observers. Only in these two cases the observers could debate regarding the fracture patterns and subsequent classifications. The observers were instructed to assess independently to the other raters without discussion or external help in scoring. All plain radiographs of fractures of the actual study population were first randomly scored in all 4 classification systems. Subsequently scoring for CT took place, where patients were presented in a different order, without any link to the conventional images. All observers reviewed the same patient series in a second session (3 months later) to assess intraobserver reliability.

We assessed all four above described, commonly used classification systems for glenoid fractures: Ideberg's classification (modified by Goss), the Mayo classification, the OTA/AO classification and the Euler/Ruedi classification.

The kappa (k) coefficient was calculated in order to evaluate inter- and intraobserver reliability. Calculation was performed using computercalculated kappa statistics (Microsoft Office Excel 2007, Redmon, Washington USA) and the webbased statistical tool http://www.statstodo. com/CohenKappa Pgm.php. Cohen's kappa (7) was used for intraobserver and Fleiss' kappa for interobserver agreement. Kappa's coefficient is a measure that adjusts for the agreement that is expected by chance and ranges between -1.0 and 1.0. Absolute agreement (100%) is represented with k = 1.0, k = 0.0 is considered as random agreement and k = -1.0 represents agreement less than which would occur by random chance. The interpretation of kappa coefficients was performed using the criteria of Landis and Koch and implies k = > 0.8 as almost perfect correlation, k = 0.6-0.8 as substantial, k = 0.4 - 0.6 as moderate, k = 0.2-0.4 as fair, and k = < 0.2 as slight (17). The results of the first observer session were used to determine interTable I. - interobserver reliability of CT and X-ray imaging techniques

	Kappa Coefficients for the Interobserver Reliability When Classified By 6 Observers				
	Interobserver Reliability				
Classification	2D CT Scan (SE)*	X-ray (SE)*	p Value		
Ideberg OTA	0.239 (0.02) 0.189 (0.02)	0.284 (0.03) 0.169 (0.02)	0.226 0.488		
Mayo Euler	0.300 (0.02) 0.333 (0.03)	0.317 (0.03) 0.350 (0.03)	0.642 0.693		

* standard error

observer reliability. The first and second observer session three months later were used to calculate the intra-observer reliability. To determine whether kappa values differed between observers, imaging modalities and fracture classifications, paired t-tests were performed with IBM SPSS Statistics version 20 (IBM Corporation, Amon, NY, USA) and a p-value below 0.05 was accepted as statistically significant.

RESULTS

Using the criteria described we retrospectively enrolled 72 patients that presented with a glenoid fracture at the participating orthopedic surgery departments during 2005-2009. In order to create a case file containing a comprehensive spectrum of glenoid fractures, the two principal investigators selected a total of 24 patients to cover the various fracture types as described in the 4 classification systems assessed. Exclusion of the 48 cases was on basis of an identical anterior rim fracture as already were selected within the 24 patients. After completion of the observational sessions, the study comprised a total of 192 data sets (4 classification systems, 4 different imaging methods, 2 scoring episodes, 6 observers). In one patient the plain radiographs were scored as uninterpretable. In order to compare intra-observer reliability data we calculated means, to compare the four classification systems.

Interobserver reliability of the Ideberg classification was fair ($\kappa = 0.239$) when classified on basis of CT scans (Table I) and showed a

	Mean Kappa Coefficients for the Intraobserver Reliability When Classified With 2D CT Scans and With X-ray				
	Mean intraobserver reliability				
Classification	2D CT Scan	X-ray	p Value		
Ideberg	0.642	0.601	0.57		
ΟΤΑ	0.600	0.544	0.41		
Mayo	0.629	0.528	0.14		
Euler	0.669	0.584	0.27		

Table II. — Mean Kappa intraobserver reliability of both CT and X-ray

non-significant difference in reliability when compared to an assessment using plain radiography ($\kappa = 0.284$). Intraobserver reliability (Table II) was substantial ($\kappa = 0.642$) with CT scans and stayed substantial ($\kappa = 0.601$) when classified on the basis of plain radiography. Also, the means of the kappas of the 6 observers were not statistically different between CT scans and plain radiographs.

Overall, the OTA classification showed slight interobserver reliability (Table I) with plain radiographs (κ =0.169) that improved when assessed with 2D CT scans (κ = 0.189). Intraobserver reliability was good with CT scans (κ = 0.600) and was slightly lower, but not significant, when classified on the basis of plain radiographs (κ = 0.544) (Table II). No statistically significant differences regarding single kappa values could be found between the observers with regard to intra –and inter-observer variation.

The Mayo classification showed fair interobserver reliability when classified with plain radiography ($\kappa = 0.317$). After the additional use of CT scans ($\kappa = 0.300$), interobserver reliability did not improve (Table I). Intraobserver reliability was substantial ($\kappa = 0.629$) with CT scans and plain radiographs revealed a kappa of $\kappa = 0.528$, which was not significantly different (table II).

Interobserver reliability revealed fair agreement between all observers when CT scans (k = 0.333) were used (Table I). A similar result was found for plain radiography (k = 0.350). Intraobserver variability revealed a substantial agreement for CT scans (k = 0.669) and moderate for plain radiographs (k = 0.584). We found no statistical difference between observers (Table II).

Mean Interobserver Reliability					
Classification	Senior Orthopaedics Observers 1&2 (SE)*	Radiologists Observers 3&4 (SE)*	Orthopaedics Registrars Observers 5&6 (SE)*	p Value	
Ideberg	0.224 (0.12)	0.372 (0.13)	-	0.41	
Ideberg	0.224 (0.12)	-	0.568 (0.13)	0.07	
Ideberg	-	0.372 (0.13)	0.568 (0.13)	0.31	
ΟΤΑ	0.160 (0.10)	0.405 (0.14)	-	0.17	
ΟΤΑ	0.160 (0.10)	-	0.229 (0.13)	0.68	
OTA	-	0.375 (0.13)	0.212 (0.13)	0.39	
Mayo	0.222 (0.14)	0.229 (0.15)	-	0.97	
Mayo	0.215 (0.13)	-	0.632 (0.14)	0.04**	
Mayo	-	0.220 (0.13)	0.541 (0.14)	0.12	
Euler	0.236 (0.14)	0.394 (0.14)	-	0.43	
Euler	0.227 (0.13)	-	0.664 (0.13)	0.03**	
Fuler	-	0.374 (0.13)	0.612 (0.14)	0.23	

Table III. — Mean Kappa interobserver reliability for X-ray

Comparing interobserver reliability of plain radiographs between the three observer groups, revealed no significant differences in kappa values between the consultant radiologists and orthopedic surgeons for any of the classification methods (table III).

Surprisingly however, evaluation of the Mayo classification revealed a significantly higher kappa value for orthopedic registrars compared to radiologists. A similar result was found when evaluating the Euler classification. Similarly, using CT-scans in the evaluation of glenoid fractures, kappa values in the OTA classification revealed a

Table IV. — Mean Kappa in subgroup interobserver reliability for CT

	Mean Kappa Coefficien	With 2D CT-scan				
Classification	Senior Orthopaedics Observers 1&2 (SE)*	Radiologists Observers 3&4 (SE)*	Orthopaedics Registrars Observers 5&6 (SE)*	p Value		
Ideberg	0.392 (0.12)	0.258 (0.14)	-	0.47		
Ideberg	0.374 (0.12)	-	0.422 (0.12)	0.78		
Ideberg	-	0.305 (0.14)	0.450 (0.12)	0.44		
OTA	0.152 (0.10)	0.271 (0.14)	-	0.50		
OTA	0.114 (0.09)	-	0.457 (0.13)	0.04**		
OTA	-	0.292 (0.15)	0.504 (0.13)	0.30		
Mayo	0.398 (0.13)	0.137 (0.11)	-	0.14		
Mayo	0.398 (0.13)	-	0.487 (0.13)	0.63		
Mayo	-	0.127 (0.11)	0.464 (0.13)	0.06		
Euler	0.509 (0.13)	0.178 (0.15)	-	0.12		
Euler	0.401 (0.13)	-	0.483 (0.13)	0.66		
Fuler	-	0.195 (0.15)	0.494 (0.14)	0.16		

significantly higher kappa value for the orthopedic registrars compared to orthopedic surgeons (Table IV).

DISCUSSION

In the present study we evaluated the clinimetric characteristics of four fracture classification systems and only found fair interobserver reliability for both plain radiography and CT scans in the assessment of glenoid fractures. In clinical practice, plain radiography is often the first imaging modality in patients with suspected glenoid fractures, which is mostly complemented with a CT scan for a more detailed insight in its fracture pattern. In an attempt to further standardize treatment options, communication and clinical outcome, various fracture classification systems were introduced. However, for a fracture classification system to be applicable, interobserver and intraobserver reliability should ideally be perfect.

A possible explanation for these "fair interobserver reliability" results found in the present study may lie in the multifaceted character of these systems. The negative effect of complexor multiple options classification systems on clinimetrics of a classification system has been described in earlier studies concerning various complex radiographic classifications (5,11,22). These effects may be further augmented by the scapular geometry and its orientation and relation to the humeral head in the field of view when assessed using plain radiographs and CT scans. In a further attempt to identify factors influencing observer variation, Humphrey et al. assessed the effect of limited imaging data of a fracture type on interobserver reliability. Surprisingly however, they concluded that a carefully defined CT image was no better than the results reported based on the full CT data, thus not providing an explanation for our reported results.

Intraobserver variation was found to be moderate to substantial. Since the observers consequently scored the fracture types in a different sequence compared to their fellow observers, variation found between subgroups may lie in subgroup characteristic of the observers (eg radiologist versus orthopedic surgeons)(δ). In our study we found the opposite: subgroup analyses revealed that orthopaedic registrars showed significantly higher kappa values as compared to orthopaedic surgeons for the Mayo, Euler and the OTA classifications, suggesting other factors than experience explaining the disagreement. We do not have a good explanation for this fact. It is in contrast to a study by Lindenhovius et al. who concluded that differences between observers was found in experience in the specific fracture area with observers with more experience scoring higher values of reliability (18). One could argue that, since glenoid fractures are rare, experience is lacking and this might diminish the possible advantage of consultants over registrars.

In a recent study by Brorson et al. it was illustrated that repetitive training of fracture classification systems limits learning effects (4,16,19). It may therefore be assumed that the intraobserver reliability may be reduced due to the learning effects. In the present study we attenuated this effect by organizing a "run in" session for all observers. The significantly better kappa values for the orthopedic registrars compared to the shoulder surgeons is therefore not clarified by the learning effects.

Although Ideberg's classification is the most widely used system, this study does not support superiority of it. The revision of the original classification underlines the difficulty to fit all different fractures into one uniform classification. This was also supported by Armitage et al. who performed a study on 90 scapular fracture cases illustrating a large variety of articular fractures. The author concluded that articular fractures of the glenoid described an almost random pattern(1). The recently proposed new OTA-classification system might overcome this problem, since it has higher observer reliability among experts assessing CT scans(15). The higher observer reliability may be explained by both the applied imaging modality used for assessment of the glenoid fracture as well as the anatomically based orientation -rather than the fracture pattern itself that is described. Another contributing factor for the high reliability is the high level of experience among the observers. Future studies focusing on this topic should compare the new classification to other commonly used classification systems.

Some strengths of our study include the "runin" viewing session to account for a possible learning effect. Plain radiographs and CT scans were assessed separately in random order. Due to the large number of pure glenoid fractures, the 24 selected cases covered all glenoid fracture types that were described in the 4 classification systems. Furthermore, three different observer subgroups that are involved in classifying glenoid fractures in clinical practice, participated in this study. Limitations of this study regard the spectrum bias introduced by selection of the 24 glenoid fractures for review rather than a series of consecutively treated patients that include the different types and amount of imaging techniques, for each individual patient(2). Furthermore there was some difference in imaging protocols. This was the result of different hospital origins which was necessary in order to include all fracture types of which some have a low prevalence.

CONCLUSION

Currently used glenoid fracture classification systems have a moderate to substantial intraobserver reliability. The interobserver reliability of all systems was slight to fair. The system of Euler scored slightly better and the AO/OTA system slightly less favourable than the other systems, but without any significant differences. Trained orthopaedic registrars scored well and in some series their score was significantly better than orthopaedic surgeons and radiologists. Results of this study show a lower intra- and interobserver reliability compared to other fracture classification systems for upper extremity fractures. This indicates the need for a more reliable glenoid classification system.

Acknowledgements

We thank all the hospitals that participated in this study. We greatly appreciate the contribution of Nicky Isenia for image processing, Ron Glandorf for initial statistical analysis, Bas Fransen for his contribution to the writing and Peter Struijs for cooperation in extracting raw data.

REFERENCES

1. Armitage B M, Wijdicks C A, Tarkin I S et al. Mapping of scapular fractures with three-dimensional computed tomography. J Bone Joint Surg Am 2009; 91: 2222-2228.

oups injuries and clavicular, glenoid, and scapular fractures. J Bone Joint Surg Am 2009; 91: 2492-2510.
4. Brorson S. Fractures of the proximal humerus. Acta Orthop Suppl 2013; 84: 1-32.

75:184-194.

 Churchill R S, Brems J J, Kotschi H. Glenoid size, inclination, and version: an anatomic study. J Shoulder Elbow Surg 2001; 10: 327-332.

2. Audige L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic

3. Bahk M S, Kuhn J E, Galatz L M, Connor P M,

review of their methodologies. Acta Orthop Scand 2004 ;

Williams G R. Jr. Acromioclavicular and sternoclavicular

- Churchill R S, Brems J J, Kotschi H. Glenoid size, inclination, and version: an anatomic study. *J Shoulder Elbow Surg* 2001; 10: 327-332.
- 7. Cohen J A. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960 ; 20 : 37-46.
- **8. Euler E, Habermeyer P, Kohler W, Schweiberer L.** Scapula fractures--classification and differential therapy. *Orthopade* 1992; 21: 158-162.
- **9. Goebel M, Seebauer L.** [Open operative treatment of acute glenoid fractures following anterior and posterior shoulder dislocation]. *Oper Orthop Traumatol* 2008 ; 20 : 228-238.
- **10.** Goss T P. Fractures of the glenoid cavity. *J Bone Joint Surg Am* 1992; 74: 299-305.
- 11. Goss T P. Double disruptions of the superior shoulder suspensory complex. *J Orthop Trauma* 1993; 7: 99-106.
- 12. Gramstad G, Marra G. Treatment of glenoid fractures. techniques in shoulder & elbow surgery 2002; 3:102-110.
- Harvey E, Audige L, Herscovici D et al. Development and validation of the new international classification for scapula fractures. J Orthop Trauma 2012; 26: 364-369.
- **14. Ideberg R.** Fractures of the scapula involving the glenoid fossa. In: Surgery of the shoulder. (Eds. Bateman JE, Welsh RP). *Philadelphia*, 1984 ; 63-66.
- **15. Jaeger M, Lambert S, Sudkamp N P et al.** The AO Foundation and Orthopaedic Trauma Association (AO/ OTA) scapula fracture classification system: focus on glenoid fossa involvement. *J Shoulder Elbow Surg* 2013 ; 22 : 512-520.
- **16. Jeray K J, Cole P A.** Clavicle and scapula fracture problems: functional assessment and current treatment strategies. *Instr Course Lect* 2011; 60: 51-71.
- **17. Landis J R, Koch G G.** An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977; 33: 363-374.
- Lindenhovius A, Karanicolas P J, Bhandari M, van D N, Ring D. Interobserver reliability of coronoid fracture classification: two-dimensional versus three-dimensional computed tomography. *J Hand Surg Am* 2009; 34: 1640-1646.
- 19. Marsh J L, Slongo T F, Agel J et al. Fracture and dislocation classification compendium 2007: Orthopaedic

Trauma Association classification, database and outcomes committee. *J Orthop Trauma* 2007; 21: S1-133.

- **20. Mayo K A, Benirschke S K, Mast J W.** Displaced fractures of the glenoid fossa. Results of open reduction and internal fixation. *Clin Orthop Relat Res* 1998; 347: 122-130.
- **21. Ruedi Th, Euler E, Habermeyer P.** Skapulafrakturen. In: Schulterchirurgie. (Eds. Habermeyer P, Krueger P, Schweiberer L). Urban&Schwarzenberg, München Wien Baltimore, 1990.
- **22.** Scheibel M, Magosch P, Lichtenberg S, Habermeyer P. Open reconstruction of anterior glenoid rim fractures. *Knee Surg Sports Traumatol Arthrosc* 2004 ; 12 : 568-573.
- 23. Schofer M D, Sehrt A C, Timmesfeld N, Stormer S, Kortmann H R. Fractures of the scapula: long-term results

after conservative treatment. Arch Orthop Trauma Surg 2009; 129: 1511-1519.

- 24. van Noort A. Scapular fractures. In: Rockwood and Green's Fractures in Adults. (Eds. Bucholz RW, Court-Brown CM, Heckman JD, Tornetta P). Wolters Kluwer, Philadelphia, 2010; 7: 1144-1164.
- 25. van Oostveen D P, Temmerman O P, Burger B J, van N A, Robinson M. Glenoid fractures: a review of pathology, classification, treatment and results. *Acta Orthop Belg* 2014; 80: 88-98.
- 26. Wening J V, von Fritschen U, Lechert B. Therapie von Glenoid- und Skapulafrakturen. *Trauma und Berufskarankheit* 2001; 4 : 544-549.
- 27. Wiedemann E. Fractures of the scapula. Unfallchirurg 2004; 107 : 1124-1133.