DOI: http://dx.doi.org/10.1590/1413-785220172504169249

Original Article

INTRA- AND INTER-OBSERVER AGREEMENT IN THE AO AND GARNAVOS SYSTEMS FOR DIAPHYSEAL HUMERUS FRACTURE

CONCORDÂNCIA INTRA E INTEROBSERVADOR DOS SISTEMAS AO E GARNAVOS NA FRATURA DA DIÁFISE DO ÚMERO

Roberto Meriqui Neto[1], Rodrigo Yuzo Masuda[1], Artur Yudi Utino[1], Rafael Pierami[1], Fábio Teruo Matsunaga[1], Marcel Jun Sugawara Tamaoki[1]

1. Universidade Federal de São Paulo, Escola Paulista de Medicina, Department of Orthopedics and Traumatology, Shoulder and Elbow Orthopedics Division, São Paulo, SP, Brazil.

ABSTRACT

Objective: The objective of this study was to compare inter- and intra-observer agreement using the Garnavos and AO/ASIF systems for classifying humeral diaphysis fractures. Methods: Eighty X-ray images taken of humeral diaphysis fractures in adult patients (age≥18 years) between January 2013 and September 2015 in the Radiology Department of Hospital São Paulo were selected for subsequent classification by five orthopedic surgeons with differing levels of experience. The images were examined at two different times and reproducibility analysis was evaluated using Fleiss' kappa to verify intra- and inter-observer agreement. Results: High-level agreement was observed for both classification systems, but particularly for the AO/ASIF classification. Inter-observer evaluation yielded excellent levels of agreement for both classifications, but principally for the Garnavos classification. Conclusions: Good or excellent inter- and intra-observer agreement was seen for both the AO/ASIF and Garnavos classification systems. However, intra-observer agreement was higher for the AO/ASIF system and inter-observer agreement was higher for the Garnavos classification. ***Level of Evidence II, Diagnostic Studies – Investigating a Diagnostic Examination.***

Keywords: Humeral fractures/classification. Adult. Orthopedics.

RESUMO

Objetivo: Averiguar a superioridade da concordância inter e intraobservadores do sistema de classificação Garnavos com relação ao sistema AO/ASIF para as fraturas diafisárias do úmero. Métodos: Foram selecionadas 80 radiografias com fraturas da diáfise do úmero de pacientes adultos (idade ≥ 18 anos) no período de janeiro/2013 a setembro/2015, no Departamento de Radiologia do Hospital São Paulo. Essas radiografias foram classificadas por cinco ortopedistas com diferentes níveis de experiência. Foram examinadas em dois momentos distintos e a análise da reprodutibilidade foi avaliada pelo índice Kappa de Fleiss para verificar a concordância intra e interobservadores. Resultados: Foram obtidas concordâncias intraobservadores de alto nível, tanto para a classificação AO/ASIF quanto para a Garnavos, especialmente para a classificação AO/ASIF. A avaliação interobservadores apresentou níveis de concordância excelentes para ambas as classificações, principalmente para a classificação Garnavos. Conclusões: Observamos concordância intra e interobservadores boa ou excelente tanto para o sistema de classificação AO/ASIF e quanto para o sistema de Garnavos. No entanto, houve maior concordância intraobservador na classificação AO/ASIF e concordância elevada interobservador na classificação de Garnavos. ***Nível de Evidencia II, Estudos Diagnósticos - Investigação de um Exame para Diagnóstico.***

Descritores: *Fraturas do úmero/classificação. Adulto. Ortopedia.*

Citation: Meriqui Neto R, Masuda RY, Utino AY, Pierami R, Matsunaga FT, Tamaoki MJ S. Intra- and inter-observer agreement in the ao and garnavos systems for diaphyseal humerus fracture. Acta Ortop Bras. [online]. 2017;25(4):159-61. Available from URL: http://www.scielo.br/aob.

INTRODUCTION

Fractures of the humeral diaphysis account for approximately 3% of fractures in adults[1-7] and 20% of humerus fractures.[7] Recent studies have estimated that these fractures are increasing in number, and incidence is expected to double by 2030.[6] Non-surgical treatment is still the gold standard for this type of fracture.[1,5]

Fracture classification is essential to determine epidemiology, guarantee communication between orthopedists, and to define treatment algorithms.[8] Multiple classification systems have been developed based on the location and morphology of injuries to categorize each type of long bone injuries; these must be clinically relevant, simple, reliable, reproducible and valid,[9,10] and ideally should also establish the method of treatment, complications and outcome.[10]

Fractures of the humeral diaphysis are predominantly classified according to the AO/ASIF system.[10-12] This classification has low inter- and intra-observer agreement and low reliability.[13-18]

A new classification proposed by Garnavos et al.[10] was proved to have greater inter- and intra-observer agreement, be easier to remember, and to yield more rapid classification in comparison with the AO/ASIF classification. Furthermore, this new classification

All authors declare no potential conflict of interest related to this article.

Study conducted at Universidade Federal de São Paulo, Escola Paulista de Medicina, Department of Orthopedics and Traumatology, Shoulder and Elbow Orthopedics Division, São Paulo, SP, Brazil.

Correspondence: Rua Borges Lagoa, 783, 5º andar, Vila Clementino, São Paulo, SP, Brazil. 04038-032. betoepm@gmail.com

Article received in 09/13/2016, approved in 03/24/2017.

has a predictive value for selecting treatment method, complication rate, and the functional outcome of the injury.[10]
This effectiveness of this new classification system led us to conduct this study to test the inter- and intra-observer agreement of the AO/ASIF classification and the Garnavos system.

MATERIALS AND METHODS

This research project was submitted to and approved by the institutional review boards of Plataforma Brasil, Hospital São Paulo, and UNIFESP (CAAE no.: 58279916.0.0000.5505) and the institution waived the need for an informed consent form for this type of study.
Radiographs were selected consecutively from January 2013 to September 2015 from the Department of Diagnostic Imaging at Hospital São Paulo, SP (quaternary care). We included X-rays from adults aged >18 years who presented humeral diaphyseal fractures. The images were selected by two orthopedic surgeons who did not participate in the fracture classification process to include X-rays with two orthogonal planes and good image quality.
The radiographs were classified by five examiners with different levels of experience. Two examiners were considered expert level (≥4 years of experience as an orthopedist specialized in the shoulder and elbow), one examiner was considered advanced (≥1year experience as an orthopedist specialized in the shoulder and elbow), and two were considered basic level (second- and third-year resident orthopedists).
The examiners received an explanation of the classification systems prior to classification in order to minimize bias from difficulties in interpreting and inexperience with the new classification. Moreover, during the classification process the examiners had access to the brochure fully describing the AO and Garnavos classifications for humerus shaft fractures.
The classifications were done by the five examiners on two different occasions with an interval of 15 days between the first and second assessments. During the first session, the X-rays were arranged in chronological order, and during the second session, the X-rays were randomized. In both cases, closed digital files were organized. Each of the examiners independently evaluated the radiographs. They were given all the time they need for assessment, and were instructed to not discuss the classification systems until the end of the classification stage. They also did not have access to the patient's history or any clinical information.
No correct response was established, but rather we looked for intra- and inter-observer agreement.

Statistical analysis

The statistical analysis was performed by a specialist in medical statistics. Fleiss' kappa was used to evaluate the intra- and inter-observer agreement for each classification. Use of Fleiss' kappa coefficient is considered most appropriate when faced with a situation where multiple examiners or assessments are involved and when the scale under evaluation presents many categories.[19]
The test was interpreted according to Altman[20] as "proportional agreement with correction of chance". Kappa is the coefficient of agreement that has a value ranging from +1 (representing perfect agreement) through 0 (representing agreement the same as chance) up to -1 (representing complete disagreement). There are no definitions for accepted levels of agreement, but some studies suggest that results between 0 and 0.2 show minimal agreement, 0.21–0.40 is poor agreement, 0.41–0.60 is moderate agreement, and 0.61–0.80 is good agreement. A value exceeding 0.80 is considered optimal agreement.[4,7,8,11]
Humeral diaphyseal fractures can be divided, according to the Arbeitsgemeinschaft für Osteosynthesefragen classification (AO), into long bone fracture in bone 1 (humerus) and segment 2 (shaft). Depending on the type of fracture, it can be classified as A (simple), B (wedge), and C (complex). Group A can be further divided into types A1 (spiral), A2 (oblique) and A3 (transverse), respectively. Spiral, flexion, and comminuted wedge fractures are classified as B1, B2, and B3, respectively, and C1 comprises complex spiral fractures, C2 complex segmental fractures, and C3 complex comminuted fractures.[21]
In the Garnavos classification for long bones, the humerus shaft is the bone segment between two parallel lines perpendicular to the long axis of the humerus, which pass through the surgical neck, and the line that passes 1 cm above the apex of the olecranon fossa. First, the fracture is classified according to its location. To do so, the segment is divided into three equal parts which are labeled P (proximal segment), M (middle segment), and D (distal segment); if the fracture line affects more than one segment, it receives more than one letter, so for example a fracture affecting the proximal and medial segments is labeled PM. A fracture can also be labeled J if it extends to the joint. Next, the fractures were classified according to their morphology into three patterns: simple fractures were divided into transverse or oblique (labeled as T) or spiral (S), intermediate fractures (with 1 or 2 significant fragments) were labeled I, and complex fractures (≥3 fragments or large comminution) were labeled C. If a fracture was segmented, each of the fractures was classified independently, with the most proximal segment classified first.[10]

RESULTS

The five examiners evaluated the radiographs separately. Examiners 1 and 3 were basic level, examiner 2 was advanced level, and examiners 4 and 5 were expert level.
A high degree of intra-observer agreement was seen. Optimal agreement was seen between four examiners (kappa >0.8) for the AO classification, and one examiner showed good agreement (0.6<kappa≤0.8). For the Garnavos classification intra-observer agreement was also high, but two examiners showed optimal agreement and three showed good agreement. (Table 1 and Figure 1)
A high degree of inter-observer agreement was also evident; for both AO classification as well as the Garnavos system agreement was optimal. We also observed that agreement increased between the first and second evaluations. The greatest inter-observer agreement was seen for the Garnavos classification. (Table 2 and Figure 2)

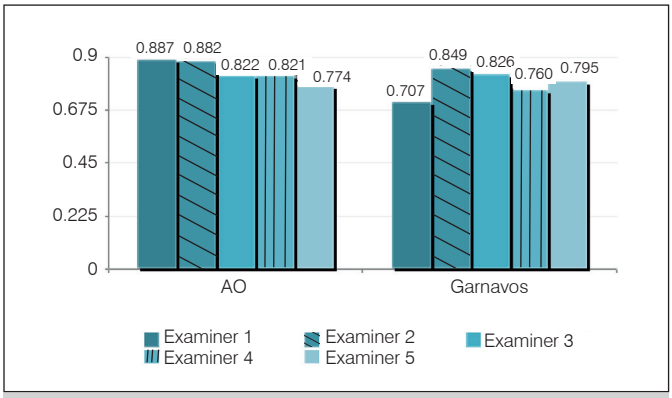

Figure 1. Intra-observer agreement for AO classification and Garnavos system..

Table 1. Intra-observer agreement for AO classification and Garnavos system.

	AO		Garnavos	
	Kappa	P-value	Kappa	P-value
Examiner 1	0.887	<0.001	0.707	<0.001
Examiner 2	0.882	<0.001	0.849	<0.001
Examiner 3	0.822	<0.001	0.826	<0.001
Examiner 4	0.821	<0.001	0.760	<0.001
Examiner 5	0.774	<0.001	0.795	<0.001

Table 2. Inter-observer agreement for AO classification and Garnavos system.

		Kappa	P-value
AO	Series 1	0.836	<0.001
	Series 2	0.844	<0.001
Garnavos	Series 1	0.878	<0.001
	Series 2	0.890	<0.001

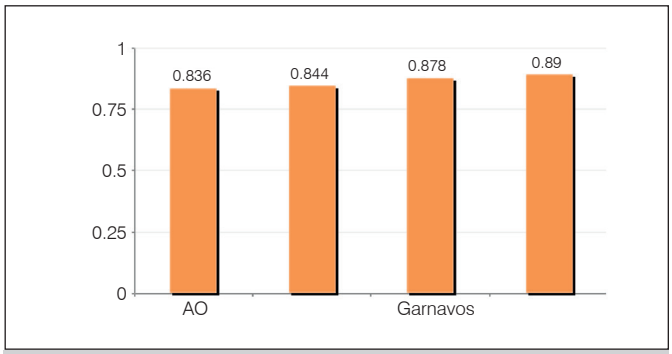

Figure 2. Inter-observer agreement for AO classification and Garnavos system.

DISCUSSION

The AO/ASIF classification is a well-established method, which is most commonly used to describe humeral diaphyseal fractures. A new classification system, the Garnavos system, has recently been introduced for diaphyseal fractures of the long bones. Few studies were seen in the literature addressing this new classification. In his article, Garnavos noted poor agreement (0.2<kappa≤0.4) for the AO classification for both intra- and inter-observer evaluation. For the Garnavos classification, this author observed good inter-observer agreement (0.6<kappa≤0.8) and moderate intra-observer agreement (0.4<kappa≤0.6).[10]

In our study, we observed good to optimal inter- and intra-observer agreement. Furthermore, we also observed that the AO classification system obtained a higher rate of intra-observer agreement than the Garnavos system. This fact was already expected, since we are more accustomed to this system of classification. Nevertheless, in the inter-observer comparison, we observed higher agreement for the Garnavos system. This may be explained by the simplification of the Garnavos classification grouping transverse and oblique fractures. We also observed that inter-observer agreement increased in the second evaluation period, showing that familiarity and practice in classifying fractures increased agreement.

In contrast with the literature, our data showed high agreement, perhaps because we selected only X-rays of the humerus shaft with two orthogonal planes and good image quality.

The difficulties related to the new classification system involved adaptation, because this classification system was unknown to the evaluators until they were involved in this study, and because this system does not make a clear division between the regions of the humerus. For example, fractures can be classified as P, M, or PM, since the classification does not provide parameters to make such a distinction. We also observed that the examiners took slightly longer to classify the fractures under the new system during the first session, but this time was not measured.

The strengths of our study included the use of five examiners with different levels of experience, a reasonable number of radiographs evaluated (80), and selection of radiographs with two good-quality views. Weaknesses included the fact that we did not measure the time needed to classify the fractures and did not compare the classification with each patient's clinical data, which made it impossible to assess any prognosis associated with the established treatment. Interestingly, this study found greater inter-observer agreement for the Garnavos classification, which could facilitate communication between orthopedists and epidemiological studies. Further studies are needed with more institutions to evaluate the prognosis and complications of this new classification, since in our opinion an ideal classification system has not yet been established.

CONCLUSIONS

We observed good or excellent intra- and inter-observer agreement for both the AO/ASIF classification and the Garnavos system. However, there was greater intra-observer agreement for the AO/ASIF classification and high inter-observer reliability for the Garnavos classification.

AUTHORS' CONTRIBUTIONS: Each author made significant individual contributions to this manuscript. RMN (0000-0002-5023-0627)*, RYM (0000-0003-0414-5752)*, and AYU (0000-0002-6320-9659)* were the main contributors to writing the manuscript. RMN and FTM (0000-0001-7328-1446)* were responsible for collecting the imaging tests for evaluation and collecting the clinical data. RMN, RYM, and AYU evaluated the data for the statistical analysis and revised the manuscript. RMN, RYM, AYU, RP (0000-0002-1745-4362)*, FTM, and MJST (0000-0002-9539-4545)* evaluated and classified the X-ray images, conducted the bibliographic research, and contributed to the intellectual concept of the study. *ORCID (Open Researcher and Contributor ID).

REFERENCES

1. Carroll EA, Schweppe M, Langfitt M, Miller AN, Halvorson JJ. Management of humeral shaft fractures. J Am Acad Orthop Surg. 2012;20(7):423-33.
2. Tytherleigh-Strong G, Walls N, McQueen MM. The epidemiology of humeral shaft fractures. J Bone Joint Surg Br. 1998;80(2):249-53.
3. Court-Brown CM. The epidemiology of fractures and dislocations. In: Court-Brown CM, Heckman JD, McQueen MM, Ricci WM, Tornetta P 3rd, editors. Rockwood and Green's fractures in adults. 8th ed. Philadelphia Wolters Kluwer Health; 2015. p. 59-108.
4. Ekholm R, Adami J, Tidermark J, Hansson K, Törnkvist H, Ponzer S. Fractures of the shaft of the humerus. An epidemiological study of 401 fractures. J Bone Joint Surg Br. 2006;88(11):1469-73.
5. Garnavos C. Humeral shaft fractures In: Court-Brown CM, Heckman JD, McQueen MM, Ricci WM, Tornetta P 3rd. editors. Rockwood and Green's Fractures in adults. 8th ed. Philadelphia: Wolters Kluwer Health; 2015. p. 1287-336.
6. Kim SH, Szabo RM, Marder RA. Epidemiology of humerus fractures in the United States: nationwide emergency department sample, 2008. Arthritis Care Res (Hoboken). 2012;64(3):407-14.
7. Matsunaga FT, Tamaoki MJ, Matsumoto MH, dos Santos JB, Faloppa F, Belloti JC. Treatment of the humeral shaft fractures--minimally invasive osteosynthesis with bridge plate versus conservative treatment with functional brace: study protocol for a randomised controlled trial. Trials. 2013;14:246.
8. Utino AY, de Alencar DR, Fernadez Maringolo L, Negrão JM, Blumetti FC, Dobashi ET. Intra and interobserver concordance of the AO classification system for fractures of the long bones in the pediatric population. Rev Bras Ortop. 2015;50(5):501-8.
9. Audigé L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. J Orthop Trauma. 2005;19(6):401-6.
10. Garnavos C, Kanakaris NK, Lasanianos NG, Tzortzi P, West RM. New classification system for long-bone fractures supplementing the AO/OTA classification. Orthopedics. 2012;35(5):e709-19.
11. Fracture and dislocation compendium. Orthopaedic Trauma Association Committee for Coding and Classification. J Orthop Trauma. 1996;10(Suppl 1):v-ix, 1-154.
12. Pignataro GS, Junqueira AE, Matsunaga FT, Matsumoto MH, Belloti JC, Tamaoki MJ. Evaluation of the reproducibility of the AO/ASIF classification for humeral shaft fractures. Rev Bras Ortop. 2015;50(4):378-82.
13. Johnstone DJ, Radford WJ, Parnell EJ. Interobserver variation using the AO/ASIF classification of long bone fractures. Injury. 1993;24(3):163-5.
14. Newey ML, Ricketts D, Roberts L. The AO classification of long bone fractures: an early study of its use in clinical practice. Injury. 1993;24(5):309-12.
15. Taylor JK. AO fracture classification logos, as evocative signposts. J Orthop Trauma. 1996;10(2):146.
16. Swiontkowski MF, Agel J, McAndrew MP, Burgess AR, MacKenzie EJ. Outcome validation of the AO/OTA fracture classification system. J Orthop Trauma. 2000;14(8):534-41.
17. Martin JS, Marsh JL. Current classification of fractures. Rationale and utility. Radiol Clin North Am. 1997;35(3):491-506.
18. Garbuz DS, Masri BA, Esdaile J, Duncan CP. Classification systems in orthopaedics. J Am Acad Orthop Surg. 2002;10(4):290-7.
19. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360-3.
20. Altman DG. Practical statistic for medical research. 3rd ed. Oxford: Oxford Medical Publication; 1995.
21. Müller M, Narzarian S, Koch P, Koch P, Schatzker J. The comprehensive classification for fractures of long bones. Berlin: Springer-Verlag; 1990.