

# INTER/INTRA-OBSERVER EVALUATION BETWEEN RADIOGRAPHS AND TOMOGRAPHIES FOR PROXIMAL HUMERUS FRACTURE

## AVALIAÇÃO INTER E INTRA-OBSERVADOR ENTRE RADIOGRAFIAS E TOMOGRAFIAS PARA FRATURA DE ÚMERO PROXIMAL

GUILHERME AUGUSTO STIRMA<sup>1</sup> , ARMANDO ROMANI SECUNDINO<sup>1</sup> , GUILHERME FERREIRA GARCIA GONZALEZ<sup>1</sup> , WILSON CARLOS SOLA JUNIOR<sup>1</sup> , GEOVANNA ANDRADE LABRES DE SOUZA<sup>1</sup> , LEONARDO DAU<sup>1</sup> 

1. Universidade Federal do Paraná, Hospital do Trabalhador, Department of Orthopedic Surgery, Curitiba, PR, Brazil.

### ABSTRACT

**Objective:** The use of images in 3D reconstruction is an instrument that facilitates the interpretation of the fracture, observations of deviations, rotations and articular surface. **Objective:** To evaluate the inter-observer and intra-observer reliability of the Neer x AO proximal humerus fracture classification on radiographs versus computed tomography with three-dimensional reconstruction (3D). **Methods:** We evaluated the digital radiographs (anteroposterior and profile) and computerized tomography with 3D reconstruction of patients presenting with a proximal humerus fracture, surgically treated at an Orthopedics and Traumatology Service. All radiographs and computed tomography were classified (Neer and AO) by eight (8) orthopedic surgeons, specialists in the upper limb and sent, following the pre-established numeration by the author, in a spreadsheet to the author of the study. **Results:** The Neer and AO scores were more reproducible when determined by computed tomography with 3D reconstruction, mainly in fractures of greater complexity (Neer 4 parts and AO group C). However, in absolute values, inter and intra-observer reproducibility and concordance still remain low. **Conclusion:** Computed tomography with 3D reconstruction allows a better analysis of fractures of group C and Neer 4 parts. However, the inter and intra-observer agreement does not increase significantly in comparison to the radiographs. **Level of evidence III, Study of non-consecutive patients, without gold standard, applied uniformly.**

**Keywords:** Tomography. Proximal Humeral Fracture. Inter and intra-observer.

### RESUMO

**Objetivo:** O uso de imagens em reconstrução 3D são um instrumento facilitador na interpretação da fratura, observações dos desvios, rotações e superfície articular. **Objetivo:** Avaliar a confiabilidade inter-observador e intra-observador da classificação da fratura de úmero proximal, descrita por Neer x AO, em radiografias versus tomografias computadorizadas com reconstrução tridimensional (3D). **Métodos:** Avaliamos as radiografias digitais (anteroposterior e perfil) e tomografias computadorizadas com reconstrução 3D de pacientes que apresentavam fratura de úmero proximal, tratados cirurgicamente em um Serviço de Ortopedia e Traumatologia. Todas as radiografias e tomografias computadorizadas foram classificadas (Neer e AO) por oito (8) cirurgiões ortopédicos especialistas em membro superior e enviadas, seguindo a numeração pré-estabelecida pelo autor, em uma planilha para o autor do trabalho. **Resultados:** A classificação de Neer e AO foram mais reprodutíveis quando determinadas pela tomografia computadorizada com reconstrução 3D, principalmente em fraturas de maior complexidade (Neer 4 partes e AO grupo C). Porém, em valores absolutos, a reprodutibilidade e concordância inter e intraobservador ainda permanecem baixas. **Conclusão:** A tomografia com reconstrução 3D, permite uma melhor análise das fraturas do grupo C e Neer 4 partes. Entretanto, não aumenta significativamente a concordância global inter e intraobservador em comparação as radiografias. **Nível de Evidência III, Estudo de pacientes não consecutivos, sem padrão ouro, aplicados uniformemente.**

**Descritores:** Tomografia. Fraturas do Úmero Proximal. Inter e Intraobservador.

**Citation:** Stirma GA, Secundino AR, Gonzalez GFG, Sola Junior WC, Souza GAL, Dau L. Inter/intra-observer evaluation between radiographs and tomographies for proximal humerus fracture. *Acta Ortop Bras.* [online]. 2020;28(1):36-9. Available from URL: <http://www.scielo.br/aob>.

### INTRODUCTION

Proximal humerus fracture corresponds to 5% of fractures, and it is the third most common fracture, only behind distal radius fractures, femur in its proximal portion, and it corresponds to 80% of humerus fractures.<sup>1</sup> The most frequent mechanism of trauma is the fall on the same level. Approximately 80% of cases present or not small deviations and can

be treated conservatively.<sup>2</sup> However, understanding the most complex fractures can be a challenge to the orthopedic surgeon. Inadequate and poorly performed radiographs may alter or even hinder analysis.<sup>3</sup> In 1970, Charles Neer created the classification of four segments for humerus fracture in his proximal portion, namely greater tuberosity, lesser tuberosity, humeral head and humeral shaft.

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

Study developed at the Universidade Federal do Paraná (UFPR), Department of Orthopedics and Traumatology, Hospital do Trabalhador, SESA, Curitiba, PR, Brazil. Correspondence: Guilherme Augusto Stirma. Rua São Mateus, 370, Portão, Curitiba, PR, Brazil, 81070080. [guilhermeaugusto89@hotmail.com](mailto:guilhermeaugusto89@hotmail.com)

Article received on 10/05/2018, approved on 04/17/2019.



After 46 years, it continues to be used due to its usability, guidance in the treatment and explanation of pathological characteristics of the injury.<sup>4-6</sup> However, its reliability is increasingly contested due to the low inter-observer agreement,<sup>4</sup> explained by the poor image quality and poor positioning of patients.<sup>7</sup> Charles Neer claims this low agreement occurs due surgeons' inexperience, in the case of 4-part fracture.<sup>8</sup> The AO classification (Arbeitsgemeinschaft für Osteosynthesefragen) values the vascularization of the humeral head.<sup>1</sup> Created in 1986 and revised in 1990, it uses an A-to-C system related to the fracture pattern. A subdivision into 3 subgroups (1, 2 and 3) is added based on the degree of fragmentation and complexity of the fracture, obtaining 27 fractures with different patterns.<sup>1,9</sup> Conventional radiography has an important role in the initial evaluation. However, computed tomography and 3D reconstruction have stood out in observations of deviations, rotations and joint surface due to technology improvement. The AO and Neer classifications have shown low reproducibility during conventional radiographic and tomographic evaluation. Images in 3D reconstruction facilitates the interpretation of the fracture. Neer emphasizes that better understanding of the fracture pattern is essential to recommend a treatment.<sup>3</sup> Our study sought to evaluate the inter- and intra-observer reliability of the classification of proximal humerus fracture described by Neer compared with AO classification on radiographs, versus computed tomographies with three-dimensional reconstruction (3D).

## MATERIAL AND METHODS

This project was submitted to the ethics committee in human research and was approved on 11/02/2016 by code 59901816.0.0000.5225. Based on the codes of procedures and surgery records, we identified all patients that underwent initial digital radiographs and computed tomographies with 3D reconstruction for proximal humerus fracture. All patients were treated surgically in the orthopedics and traumatology service of a large hospital and signed an informed consent form. All radiographs and computed tomographies were classified by 8 orthopedic surgeons specialized in the upper limb. The tests were previously edited by one of the authors (who did not participate in the evaluation) to remove the identification and randomization of the sequence of patients. Radiographies were first sent digitally to each orthopedist and, about one month after, tomographies. Each orthopedist classified each fracture using Neer (number and fractured segments), and using AO (with subgroups) and classified in tables, following the pre-established numbering, in a spreadsheet to the author responsible for randomization of the images. After data collection, radiographic and tomographic classifications were compared by inter- and intra-observer analysis. A statistical study of the data, values found and a discussion on the basis of the current literature in already published data were conducted. Patients without initial radiographs and computed tomographies for proximal humerus fractures and pathological fractures were excluded from the study.

We used the Kappa coefficient of agreement for statistical analysis between the inter- and intra-observer agreement. The coefficient values found in this test can be classified as follows: 0-0.19 as unsatisfactory, 0.20-0.39 low agreement, 0.40-0.59 moderate agreement, 0.60-0.79 satisfactory agreement and 0.80-1.00 as almost perfect.

## RESULTS

### Inter-observer

In total, 54 patients were included in the sample. The tomographies and radiographs of the 54 cases were evaluated by eight orthopedists specialized in the upper limb.

Regarding the radiographs for the Neer classification, kappa agreement values were 0.275 (2 parts), 0.083 (3 parts), 0.204 (4 parts), 0.178 general kappa,  $p < 0.001$ . (Table 1). In tomographies, the kappa values were 0.229 (2 parts), 0.147 (3 parts), and 0.32 (4 parts), 0.22 mean kappa, with  $p < 0.001$ . (Table 2)

The results regarding the radiographs classified according to AO showed kappa values of 0.232 to A1, 0.194 to A2, 0.266 to A3, 0.15 to B1, 0.21 to B2, 0.078 to B3, 0.045 to C1, 0.133 to C2 and 0.419 to C3, with 0.201 general kappa. (Table 3)

Regarding the tomographies, the results showed kappa 0.535 to A1, 0.273 to A2, 0.28 to A3, 0.242 to B1, 0.221 to B2, 0.236 to B3, 0.114 to C1, 0.479 to C2 and 0.311 to C3, with 0.277 general mean. (Table 4)

**Table 1.** Concordance Table with radiographs by Neer.

	0	2 parts	3 parts	4 parts
Kappa of the category		0.275	0.083	0.204
P-value of Kappa of the category		< 0.001	0.001	< 0.001
General Kappa		0.178		
P-value		< 0.001		
95% CI Kappa		upper: 0.213 lower: 0.144		

**Table 2.** Concordance Table with tomographies by Neer.

	2 parts	3 parts	4 parts
Kappa of the category	0.229	0.147	0.32
P-value of Kappa of the category	< 0.001	< 0.001	< 0.001
General Kappa		0.22	
P-value		< 0.001	
95% CI Kappa		upper: 0.256 lower: 0.184	

**Table 3.** Concordance Table with AO radiographs.

	A1	A2	A3	B1	B2	B3	C1	C2	C3
Kappa of the category	0.232	0.194	0.266	0.15	0.21	0.078	0.045	0.133	0.419
P-value of Kappa of the category	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.002	0.081	< 0.001	< 0.001
<b>General Kappa</b>									
General Kappa		0.201							
P-value		< 0.001							
95% CI Kappa		upper: 0.221 lower: 0.18							

**Table 4.** Concordance Table with AO tomographies.

	A1	A2	A3	B1	B2	B3	C1	C2	C3
Kappa of the category	0.535	0.273	0.28	0.242	0.221	0.236	0.114	0.479	0.311
P-value of Kappa of the category	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
<b>General Kappa</b>									
General Kappa	0.277								
P-value	< 0.001								
95% CI Kappa	upper: 0.298 lower: 0.256								

On radiographs, according to Neer classification, the mean agreement between the classification was 4.71 physicians for each case, while by the AO classification there was an agreement between 4 or more physicians, totaling 36 cases.

In the tomographies, according to Neer classification, the agreement between the classification was 5.06 physicians for each patient, while in AO there was an agreement between 4 or more physicians, totaling 42.

#### Intra-observer

Regarding intra-observer evaluations, there was agreement in the classification on radiographs with tomographies on average of 26.92 cases, ranging from 15 to 31 according to Neer classification in the 54 patients and 17.125, ranging from 12 to 22 correct answers, according to AO classification.

#### DISCUSSION

Radiography is the standard method for evaluation, diagnosis and classification. However, computed tomography is expected to facilitate and improve the reproducibility of the analyzed fractures, providing a greater intra-observer agreement, enabling a better choice of treatment and a more reliable and reproducible classification system.<sup>7,10</sup> Despite numerous complaints regarding its reproducibility, Neer classification is widely accepted and commonly used to guide treatment and anticipate prognosis; it is pedagogically useful, of easy learning and separates fractures into broad categories, being easy to understand.<sup>5,7</sup> AO classification (Arbeitsgemeinschaft für Osteosynthesefragen) divides fractures according to their complexity and facilitates choice of treatment and prognosis. The AO is one of the most complete classification system, however, its intra- and inter-observer reproducibility has reduced.<sup>1,9,11</sup>

In our study, we evaluated 54 patients with proximal humerus fracture, whose initial evaluation was performed by radiography and tomography with 3D reconstruction.

In the evaluation of the inter-observer results regarding radiographs according to Neer classification, we observed that kappa agreement ranged between 0.083 (analysis with fractures classified into 3 parts), 0.204 (4 parts) and 0.275 (2 parts), with 0.178 general kappa, with  $p < 0.001$ . (Table 1) These data are lower than those of Papakonstantinou et al.<sup>12</sup>, which showed a 0.40-0.58 global kappa, Bernstein et al.<sup>13</sup>, a 0.52 kappa, Siebenrock and Gerber<sup>14</sup>, a 0.40 kappa, and Sidor et al.<sup>15</sup>, a 0.48 kappa. Brorson and Hróbjartsson<sup>16</sup> conducted a systematic review, finding 11 studies with kappa ranging from 0.17 to 0.52. However, of the revised studies, the higher the number of evaluations and the larger the group that classified them, less is the kappa agreement. Among the studies mentioned, Schwartz and Cuny<sup>17</sup> used 11 orthopedists to evaluate the radiographs of 21 patients, obtaining a 0.17 kappa value; Kristiansen<sup>18</sup> studied 100 patients, obtaining a 0.07-0.48 kappa value. The best result was found in the study by Bernstein et al.<sup>13</sup>, with 20 cases analyzed by 2 orthopedists and 2 orthopedic residents, which obtained a 0.52 kappa value.

In the evaluation of the results of the tomographies, we found a 0.22 mean kappa, with  $p < 0.001$ , ranging between 0.147 for fractures classified in 3 parts, 0.229 in 2 parts and 0.32 in 4 parts, as shown in Table 2. Brorson and Hróbjartsson<sup>16</sup> had a 0.34-0.72 mean. We can justify the low agreement in our study by the evaluation of tomography with 3D reconstruction being conducted without radiographic analysis. Sjöden et al.<sup>19</sup>, on the other hand, showed that the addition of tomography did not improve the Neer classification reproducibility. However, our study showed a small improvement in reproducibility in computed tomography (5.06), obtaining a better agreement in the classification of computed tomography versus 4.71 on radiographs. When analyzing the results of the radiographs classified according to AO, a greater agreement was obtained when classifying fractures in C3, 0.419 kappa, and A3, 0.266 kappa, with a 0.201 general mean. Tomographies showed higher agreement when classified according to AO and compared with radiographs, A1 0.535 kappa, C2 0.479, C3 0.311 with a 0.277 general mean, as shown in Tables 3 and 4. The values found were similar to the results of Matsushigue et al.<sup>9</sup>, in which a 0.25 kappa value was obtained for radiographs and a 0.36 kappa for tomographies. The values were higher than in the analysis by Majed et al.<sup>10</sup>, which showed weak inter-observer reliability, with a 0.11 kappa. Values below Sjöden et al.<sup>19</sup>, a 0.31 kappa, Siebenrock and Gerber et al.<sup>14</sup>, a 0.42 kappa and Papakonstantinou et al.<sup>12</sup> with a 0.31-0.54 kappa were observed in our analysis. The high complexity of the classification system and the high number of categories and subcategories explains the low inter-observer agreement.<sup>12,14,15,19</sup> In our study, we showed the Neer and AO classification were more reproducible and presented better results when performed through tomography with 3D reconstruction, especially in fractures of greater complexity (Neer 4 parts and AO group C). However, inter- and intra-observer reproducibility and agreement (26.92 cases, ranging from 15 to 31 according to Neer and 17.125, ranging from 12 to 22 correct answers, according to AO in the 54 cases analyzed) still remain low in absolute values.

The statistical method used in our study was kappa agreement analysis. This measure of agreement presents values between 1 (one), representing total agreement, and values near 0 (zero), representing no agreement. Although this form of calculation is planned for two observers, Kappa was used with more than 2 observers in our study and in the other studies we analyzed. Thus, the Kappa values obtained are below the real, since the rate of chance is calculated for each observer. However, Kappa is still the most assertive statistical method for this type of analysis.<sup>20</sup>

One of the limitations of our study was its retrospective nature. All radiographs were performed in the emergency room, in emergency situations, some with limited quality. This is the reason why we could not repeat radiographs or request new ones so that they would improve quality.

Eight orthopedists specialized in the upper limb participated in our study to level the agreement indexes and to obtain professionals with the same experience level. The classification was not repeatedly

applied at different times because, according to studies, it would not change the reproducibility.

## CONCLUSION

The 3D resection tomography did not significantly improve inter- and intra-observer global agreement for Neer and AO

classifications compared with radiographs. We found a low agreement for the evaluation of proximal humerus fracture, except in group C and Neer fracture 4 parts. Despite being applied to 8 specialists in the upper limb, this supports previous studies on the difficulty of achieving good reliability and reproducibility of classifications.

**AUTHORS' CONTRIBUTIONS:** Each author contributed individually and significantly to the development of the manuscript. GAS: writing, review, intellectual concept of the article, preparation of the entire study project Research. ARS: writing, review, intellectual concept of the article. GFGG: article review. WCSJ: writing, review, intellectual concept of the article. GALS: data collection, article review. LD: writing, review, intellectual concept of the article.

## REFERENCES

1. Robinson BC, Athwal GS, Sanchez-Sotelo J, Rispoli DM. Classification and imaging of proximal humerus fractures. *Orthop Clin North Am.* 2008;39(4):393-403. Doi: 10.1016/j.ocl.2008.05.002
2. Mauro CS. Proximal humeral fractures. *Curr Rev Musculoskelet Med.* 2011;4(4):214-20.
3. Shrader MW, Sanchez-Sotelo J, Sperling JW, Rowland CM, Cofield RH. Understanding proximal humerus fractures: image analysis, classification, and treatment. *J Shoulder Elbow Surg.* 2005;14(5):497-505.
4. Foroohar A, Tosti R, Richmond JM, Gaughan JP, Ilyas AM. Classification and treatment of proximal humerus fractures: inter-observer reliability and agreement across imaging modalities and experience. *J Orthop Surg Res.* 2011;6:38.
5. Carofino BC, Leopold SS. Classifications in brief: the Neer classification for proximal humerus fractures. *Clin Orthop Relat Res.* 2013;471(1):39-43.
6. Neer CS 2nd. Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg Am.* 1970;52(6):1077-89.
7. Carrerra EF, Wajnsztein A, Lenza M, Archetti Netto N. Reproducibility of three classifications of proximal humeral fractures. *Einstein (São Paulo).* 2012;10(4):473-9.
8. Neer CS 2nd. Four-segment classification of proximal humeral fractures: purpose and reliable use. *J Shoulder Elbow Surg.* 2002;11(4):389-400.
9. Matsushigue T, Pagliaro Franco V, Pierami R, Jun Sugawara Tamaoki M, Archetti Netto N, Hide Matsumoto M. Do computed tomography and its 3D reconstruction increase the reproducibility of classifications of fractures of the proximal extremity of the humerus? *Rev Bras Ortop.* 2014;49(2):174-7.
10. Majed A, Macleod I, Bull AM, Zyto K, Resch H, Hertel R, et al. Proximal humeral fracture classification systems revisited. *J Shoulder Elbow Surg.* 2011;20(7):1125-32.
11. Rockwood CA, Green DP, Bucholz RW. Fractures in adults. In: Bucholz RW, Heckman JD, Court-Brown CM, Tornetta P, editors. *Proximal Humerus Fractures.* Vol. 1. Philadelphia: Lippincott Williams & Wilkins; 2010. p. 9-87.
12. Papakonstantinou MK, Hart MJ, Farrugia R, Gabbe BJ, Kamali Moaveni A, van Bavel D, et al. Interobserver agreement of Neer and AO classifications for proximal humeral fractures. *ANZ J Surg.* 2016;86(4):280-4.
13. Bernstein J, Adler LM, Blank JE, Dalsey RM, Williams GR, Iannotti JP. Evaluation of the Neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. *J Bone Joint Surg Am.* 1996;78(9):1371-5.
14. 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):1751-5.
15. Sidor ML, Zuckerman JD, Lyon T, Koval K, Cuomo F, Schoenberg N. The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. *J Bone Joint Surg Am.* 1993;75(12):1745-50.
16. Brorson S, Hróbjartsson A. Training improves agreement among doctors using the Neer system for proximal humeral fractures in a systematic review. *J Clin Epidemiol.* 2008;61(1):7-16.
17. Schwartz C, Cuny C. Fractures of the proximal humerus: a prospective review of 188 cases. *Eur J Orthop Surg Traumatol.* 2003;13(1):1-12.
18. Kristiansen B, Andersen ULS, Olsen CA, Varmarken JE. The Neer classification of fractures of the proximal humerus. An assessment of interobserver variation. *Skeletal Radiol.* 1988;17(6):420-2.
19. Sjöden GO, Movin T, Aspelin P, Guntner P, Shalabi A. 3D-radiographic analysis does not improve the Neer and AO classifications of proximal humeral fractures. *Acta Orthop Scand.* 1999;70(4):325-8.
20. Siegel S, Castellan Jr NJ. *Estatística não paramétrica para ciências do comportamento.* 2a ed. Porto Alegre: Artmed; 2006.