ORIGINAL ARTICLE

DOI: 10.15343/0104-7809.202650e19382025I
ISSN: 1980-3990

Intra/interobserver agreement of proximal humeral fracture classification systems

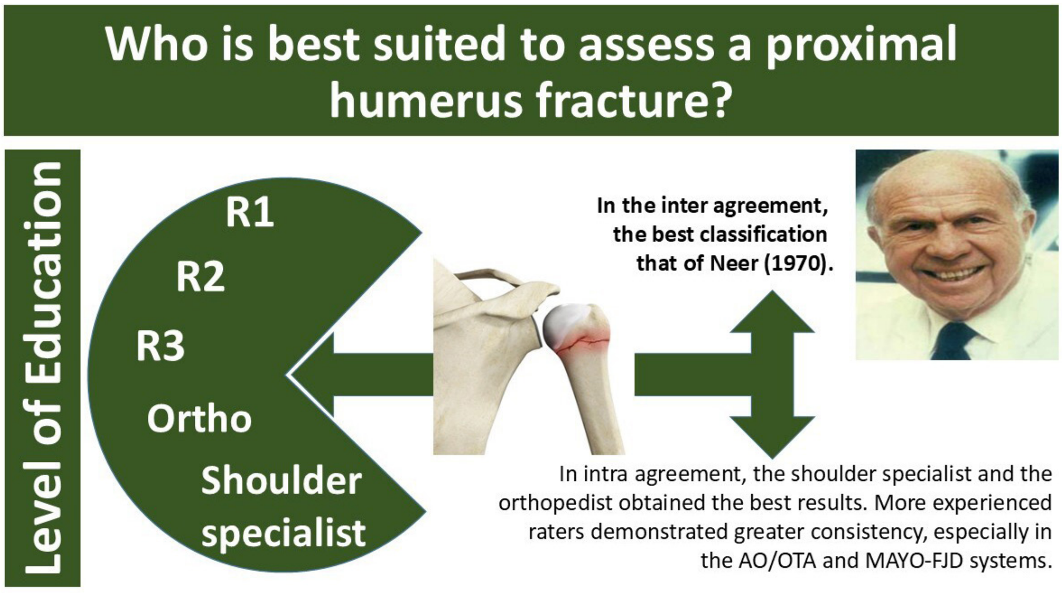
Douglas Barbosa Lima[1] Alexia Ribas Checchi[2] Igor Fiorese Vieira[2] José Luiz Lopes Vieira[3] Giulia Maria Jorge Dinies[1] Renata Mello Calandrini[1] João Vitor Vedam[1]

[1]Hospital Regional de Ponta Grossa. Ponta Grossa/PR, Brasil.
[2]Centro Universitário Campo Real. Guarapuava/PR, Brasil.
[3]Universidad Católica del Maule – UCM. Talca, Chile.
E-mail: jllopesvieira@gmail.com

Highlights

- Proximal humeral fracture is the fourth most common fracture in adults.
- The fracture classification method defines treatment and prognosis.
- Upper extremity specialists are consistent in the assessment of proximal humeral fractures.
- Residency programs are fundamental to the quality of proximal humeral fracture assessment.

Graphical Abstract



Abstract

This study assessed the inter- and intraobserver agreement of the NEER, LEGO/HERTEL, AO/OTA, and MAYO-FJD classification systems for proximal humeral fractures. Five evaluators — a first-year resident (R1), a second-year resident (R2), a third-year resident (R3), an orthopedic surgeon (Ortho), and a shoulder surgery specialist (Spec.) – assessed 50 radiographs of patients with proximal humeral fractures at two time points (T1 and T2). Cohen's and Fleiss's Kappa coefficients were used for statistical analysis. Regarding interobserver agreement, the NEER system at T1 and T2 (κ=0.444 and κ=0.496) demonstrated moderate agreement. The LEGO/HERTEL system at T1 and T2 (κ=0.376 and κ=0.387), the AO/OTA system at T1 and T2 (κ=0.269 and κ=0.223), and the MAYO-FJD system at T1 and T2 (κ=0.367 and κ=0.300) demonstrated low agreement. Moderate intraobserver agreement was observed for the NEER system. The LEGO/HERTEL system showed substantial agreement with κ=0.528 (R1) and near-perfect agreement with κ=0.867 (Spec.). For the AO/OTA system, Kappa values ranged from 0.347 (R1) to 0.793 (R3), oscillating between low and substantial. The MAYO-FJD system presented the lowest agreement (κ=0.154 [R1] to κ=0.665 [Spec.]). It is concluded that regarding interobserver agreement, the NEER system achieved a moderate result. Regarding intraobserver agreement, more experienced evaluators demonstrated greater consistency compared to residents.

Keywords: Fractures. Radiography. Humerus. Interobserver. Intraobserver.

Associate Editor: Edison Barbieri
Mundo Saúde. 2026,50:e19382025
O Mundo da Saúde, São Paulo, SP, Brasil.
https://revistamundodasaude.emnuvens.com.br

Received: 15 december 2025.
Accepted: 24 june 2026.
Published: 14 july 2026.

INTRODUCTION

Proximal humeral fracture is the fourth most common fracture in adults, ranking only behind distal radius, proximal femur, and osteoporotic vertebral fractures[1]. It represents approximately 80% of all humeral fractures and occurs predominantly in elderly women with osteoporosis, with an approximate ratio of 3:1 relative to men, generally resulting from ground-level falls; these patients also present an increased risk of subsequent fractures[2,3].

For adequate fracture management, it is essential that the fracture be classified through radiographs, as this defines the clinical approach, treatment, and patient prognosis. However, fracture classification systems currently used for the proximal humeral extremity demonstrate low agreement and reproducibility[4].

In 1970, Charles S. Neer II proposed a system for classifying proximal humeral fractures based on the number of displaced bone segments. The model divides the proximal humerus into four anatomical segments: the head, greater tuberosity, lesser tuberosity, and diaphysis, considering as a "fractured part" only those segments with displacement ≥1 cm or angulation ≥45°, except for the greater tuberosity, for which a displacement of 0.5 cm is considered. Thus, fractures may be described as two-, three-, or four-part, including variants such as fracture-dislocations or anatomical neck fractures. The NEER classification is widely used today, serving as a reference for treatment selection and prognostic assessment[5].

The LEGO/HERTEL system describes the morphology of proximal humeral fractures in up to 12 combinatory patterns, based on four main anatomical structures (head, greater tuberosity, lesser tuberosity, and diaphysis). This approach seeks to standardize fracture description, facilitating communication and reproducibility among evaluators. Hertel *et al.* (2004)[6] also proposed independent prognostic criteria to estimate the risk of humeral head ischemia, which include the assessment of the medial calcar, metaphyseal continuity, and head-diaphysis contact. Thus, while the LEGO system is a morphological classification, the so-called "Hertel criteria" correspond to clinical predictors of avascular necrosis and are not formally part of the classification.

On the other hand, the AO/OTA classification system refers to a universal fracture classification system created by the AO Foundation (Association for the Study of Osteosynthesis) in conjunction with the OTA (Orthopaedic Trauma Association), which divides fractures into three distinct groups, each subdivided into three subgroups, totaling nine fracture types.

Two-part fractures (A) may involve the tuberosities or the surgical neck. When involving the tuberosities, subtype A1.1 affects the greater tuberosity (GT), while subtype A1.2 affects the lesser tuberosity (LT). For surgical neck fractures, they may be simple (2.1), wedge (2.2), or shear/vertical (2.3). Three-part fractures (B) generally involve the surgical neck associated with one tuberosity. Subtype B1.1 affects the greater tuberosity (GT) and B1.2 affects the lesser tuberosity (LT). These subtypes may also be qualified regarding the presence of an intact (u) or comminuted (v) wedge, indicating the degree of fracture fragmentation.

Articular fractures (C) involve the anatomical neck and may occur in isolation or in association with the surgical neck. In subtype C1, the anatomical neck may undergo valgus impaction, being qualified according to the affected tuberosity: n → GT, o → LT, p → LT + GT, or may occur as an isolated anatomical neck fracture (1.3). In four-part fractures (C3), involving the anatomical and surgical neck, the metaphysis and the joint may be intact (3.1), the metaphysis intact with a comminuted joint (3.2), or both comminuted (3.3), representing more complex and unstable patterns[7].

For fractures of greater complexity, the MAYO-Fundación Jiménez Díaz (FJD) classification may be employed. This system uses three main anatomical components: the humeral head, humeral neck, and tuberosities. The classification encompasses fracture patterns including isolated greater and lesser tuberosity fractures, surgical neck fracture, posteromedial varus fracture, valgus impaction fracture, and fractures in which the humeral head is displaced – encompassing head dislocation, split head, and head impression[8].

Since proximal humeral fracture is the fourth most common fracture, it is necessary to standardize its classification within medical practice, aiming at an effective treatment plan for each specific type of injury. Currently in the literature, there are no relevant studies that use all of these classification systems; therefore, an inter- and intraobservational assessment of the classifications is necessary to identify the most suitable for managing proximal humeral fractures, particularly among professionals at different levels of orthopedic training.

DOI: 10.15343/0104-7809.202650e19382025I

METHODS

This is a retrospective, descriptive observational study with a cross-sectional approach, which through different proximal humeral fracture classification systems aims to assess interobserver agreement. Following the consent of the co-participating institution and approval by the Research Ethics Committee under No. 7,717,872, the study was conducted using radiographs already existing in the patient records of the Hospital Universitário Regional dos Campos Gerais (HU-UEPG) from July 2023 to January 2024. Inclusion criteria comprised radiographs of patients diagnosed with proximal humeral fracture (ICD S42.2), in two projections: anteroposterior and lateral view. Radiographs of proximal humeral fractures in only one projection, or any other type of fracture, were excluded.

Thus, 50 radiographic images of proximal humeral fractures were analyzed, covering the full spectrum of the NEER, LEGO/HERTEL, AO/OTA, and MAYO-FJD classifications, including at least one representative of each proximal humeral fracture type. For the assessment, the 50 cases were presented to a group of evaluators at different stages of orthopedic training: a first-year resident (R1), a second-year resident (R2), a third-year resident (R3), an orthopedic surgeon (Ortho) who is not a shoulder surgery specialist, and a shoulder surgery specialist (Spec.).

Radiographs were presented in a randomized order to the evaluators, with a time interval between the two assessments of 15 to 30 days. Evaluators had no access to their first-round responses, nor to the patient's clinical history or surgical outcome. The order in which radiographs were displayed was randomized in the second round. Evaluators received no prior training; that is, each evaluator based their assessment on their prior knowledge acquired through lectures, residency practice, and years of orthopedic practice.

The results obtained from the observers' assessments were tabulated using Microsoft Excel 365 (Microsoft Corp., Redmond, WA, USA) and submitted to Cohen's and Fleiss's Kappa tests for intra- and interobserver assessment, respectively, using the Statistical Package for the Social Sciences (SPSS, IBM Corp., Armonk, NY, USA), version 30.0, for statistical analyses. Interobserver agreement assessment tables were produced, evidencing the Kappa index measure for each evaluator class (R1, R2, R3, Ortho, and Spec.) for each of the two assessments (T1 and T2).

Intraobserver agreement assessment tables were produced comparing each assessment with the other, as well as the upper and lower confidence limits for a 95% CI. Kappa values with $p < 0.05$ were considered statistically significant. Interpretation of results followed the criteria proposed by Landis and Koch (1977)[9], wherein values below or equal to zero indicated poor agreement; slight, between 0 and 0.2; fair, between 0.2 and 0.4; moderate, between 0.4 and 0.6; substantial, between 0.6 and 0.8; and excellent or near-perfect, between 0.8 and 1.0.

RESULTS

Table 1 presents the interobserver agreement values obtained using Fleiss's Kappa coefficient, considering the different indicators assessed at two time points (T1 and T2).

Table 1 - Determination of interobserver agreement among the NEER, LEGO/HERTEL, AO/OTA, and MAYO-FJD classification systems for proximal humeral fractures (HU-UEPG – 07/2023 to 01/2024).

Classification System	% Agreement	Fleiss κ	95% CI	p-value	Agreement Level
NEER T1	44.40	0.444	0.394–0.493	<0.001	Moderate
NEER T2	49.60	0.496	0.445–0.546	<0.001	Moderate
LEGO/HERTEL T1	37.60	0.376	0.341–0.411	<0.001	Low
LEGO/HERTEL T2	38.70	0.387	0.351–0.423	<0.001	Low
AO/OTA T1	26.90	0.269	0.242–0.296	<0.001	Low
AO/OTA T2	22.30	0.223	0.196–0.251	<0.001	Low
MAYO-FJD T1	36.70	0.367	0.342–0.392	<0.001	Low
MAYO-FJD T2	30.00	0.300	0.276–0.325	<0.001	Low

Note: % = percentage; CI = confidence interval; T1 = first assessment; T2 = second assessment; p = probability; K = Kappa index.

The NEER T1 and NEER T2 indicators demonstrated moderate agreement, with Kappa values of 0.444 (95% CI: 0.394–0.493; $p < 0.001$) and 0.496 (95% CI: 0.445–0.546; $p < 0.001$), respectively. In contrast, the LEGO/HERTEL T1 ($\kappa = 0.376$; 95% CI: 0.341–0.411; $p < 0.001$), LEGO/HERTEL T2 ($\kappa = 0.387$; 95% CI: 0.351–0.423; $p < 0.001$), AO/OTA T1 ($\kappa = 0.269$; 95% CI: 0.242–0.296; $p < 0.001$), and AO/OTA T2 ($\kappa = 0.223$; 95% CI: 0.196–0.251; $p < 0.001$) indicators demonstrated low agreement, evidencing greater divergence among evaluators.

Similarly, the MAYO-FJD T1 and MAYO-FJD T2 indicators presented low agreement, with Kappa values of 0.367 (95% CI: 0.342–0.392; $p < 0.001$) and 0.300 (95% CI: 0.276–0.325; $p < 0.001$), respectively. Overall, the results reveal heterogeneity in agreement among evaluators, ranging from low to moderate depending on the indicator analyzed and the assessment time point.

Table 2 presents the Cohen's Kappa coefficients obtained in the intraobserver agreement analysis (comparison between T1 and T2) for each observer and indicator.

Table 2 - Level of intraobserver agreement (T1 vs. T2) for the NEER, LEGO/HERTEL, AO/OTA, and MAYO-FJD classification systems in proximal humeral fractures (HU-UEPG – 07/2023 to 01/2024).

Observer	NEER	LEGO/HERTEL	AO/OTA	MAYO-FJD
R1	0.523	0.528	0.347	0.154
R2	0.740	0.696	0.658	0.574
R3	0.876	0.866	0.793	0.349
Orthopedic surgeon (Ortho)	0.777	0.745	0.683	0.620
Specialist (Spec.)	0.766	0.867	0.763	0.665

Note: R1 = First-year resident; R2 = Second-year resident; R3 = Third-year resident.

Overall, the Kappa values indicate variable levels of intraobserver consistency across the different indicators, ranging from low to near-perfect. For the NEER indicator, agreement ranged from moderate to near-perfect among evaluators, with Kappa values between 0.523 (R1) and 0.876 (R3). These results suggest good stability of the classifications between the two assessment time points. Similarly, the LEGO/HERTEL indicator also presented agreement ranging from moderate to near-perfect, with Kappa values between 0.528 (R1) and 0.867 (Spec.).

These results suggest good stability of the classifications between the two assessment time points for both indicators. For the AO/OTA indicator, Kappa values ranged from 0.347 (R1) to 0.793 (R3), indicating agreement ranging from low to substantial depending on the evaluator. Finally, the MAYO-FJD indicator presented the lowest intraobserver stability indices, with Kappa values ranging from 0.154 (R1) to 0.665 (Spec.), reflecting low to moderate agreement between the two assessments.

Overall, it was observed that the more experienced evaluators (orthopedic surgeon and specialist) demonstrated greater intraobserver consistency compared to the resident evaluators, especially for the AO/OTA and MAYO-FJD indicators. These findings suggest that clinical experience may be associated with greater stability in assessments conducted at different time points.

DISCUSSION

The results obtained in this study demonstrated significant variations in inter- and intraobserver agreement across the different proximal humeral fracture classification systems, evidencing that the clinical experience of evaluators plays a determining role in the consistency of assessments. The NEER system presented the best interobserver agreement indices at both time points (T1 and T2), with moderate agreement – corroborating similar results[10] that reported fair to substantial reliability for NEER and HERTEL classifications, with the former being superior in terms of intraobserver stability, especially in three-dimensional reconstructions and even among observers with different levels of experience[11].

In contrast, the LEGO/HERTEL and MAYO-FJD systems presented low interobserver agreement, indicating greater divergence among evaluators. This limitation is frequently reported in the literature and attributed to the greater morphological complexity of these classifications, which involve multiple anatomical and prognostic parameters, making reproducibility difficult in two-dimensional radiograph

DOI: 10.15343/0104-7809.202650e19382025I

analyses[8].

Despite this, studies show that the MAYO-FJD system may perform better when combined with computed tomography (CT), since multiplanar visualization reduces interpretation ambiguities[8]. This finding suggests that the exclusive use of plain radiographs, although practical and widely available, may limit the diagnostic accuracy of more detailed systems that have demonstrated better agreement with the use of three-dimensional reconstructions[11].

Similarly, the AO/OTA system presented low interobserver agreement, suggesting considerable divergence among evaluators. This limitation is also observed in previous studies, which report lower interrater reliability for this system, especially in its earlier versions[12]. The AO/OTA classification update published in 2018 improved the clarity of criteria and simplified the coding process, maintaining adequate descriptive power and slightly improving observer agreement[13]. Similarly, it presents variable reproducibility according to the experience level of the evaluators, reinforcing the need for standardized training to optimize diagnostic consistency[14].

Regarding intraobserver agreement, the results of this study were consistent with those found by Díaz *et al.* (2022)[14], who observed that increased intraobserver agreement occurs with advancing professional experience, highlighting that experienced orthopedic surgeons achieve higher reliability indices compared to residents. This trend was also observed in the present study, in which the orthopedic surgeon and the shoulder surgery specialist obtained the highest intraobserver Kappa values, especially for the AO/OTA and MAYO-FJD systems. Such findings confirm that time in practice and specific training directly influence the ability to correctly interpret the morphological nuances of proximal humeral fractures.

Although the MAYO-FJD system has recently emerged as a more comprehensive and anatomical proposal for the classification of proximal humeral fractures, the results of this study showed that this classification system still presents low interobserver agreement. This finding may be related to the fact that it is a new system, proposed by Foruria *et al.* (2022)[8], while the other systems – NEER (1970), LEGO/HERTEL (2004), and AO/OTA (1990, revised in 2018) –have been consolidated for longer, presenting greater familiarity among professionals and broad diffusion in orthopedic practice.

Although the NEER system demonstrates overall better performance, its clinical value is still debated, as it presents limitations in complex fracture cases and in patients with significant anatomical alterations. In analyses based on plain radiographs, classifications such as NEER and AO/OTA maintain only moderate agreement among evaluators, suggesting that none of the current systems achieves ideal reliability for universal standardization[15].

In this regard, recent reviews argue that the incorporation of three-dimensional methods and artificial intelligence can substantially increase the reproducibility and clinical applicability of these classifications, reducing bias related to individual experience[16,17].

Adequate training during medical residency in orthopedics is essential for the development of diagnostic skills and the ability to correctly interpret radiographic images – factors that directly influence therapeutic choice and clinical outcomes[11,14]. A thorough radiographic assessment is decisive in defining the need for surgical or conservative treatment, enabling more precise and individualized decisions[15].

Furthermore, the correct interpretation of radiographs contributes to the reduction of hospital costs and to the functional and economic success of treatment, favoring the patient's full recovery[8]. Additionally, three-dimensional printing may be applied, especially in fracture planning and intraoperative guidance in sequelae management[17].

Although widely used in clinical practice, the classic classifications for proximal humeral fracture assessment present significant limitations. Studies have pointed to low interobserver agreement between the NEER and AO/OTA systems, which has generated uncertainty in the standardization of therapeutic approaches[5,18,19]. Due to these divergences, current decision-making emphasizes a tendency to value specific pathomorphological criteria – namely, medial calcar stability and the risk of vascular necrosis[6].

Despite the divergences among classification systems, the longevity and continued use of the NEER classification occur primarily due to three factors: a common orthopedic language, as it provides terminology understood by surgeons worldwide[20]; necrosis prevention, as it was one of the first systems to critically address the vascularization of the humeral head and the risk of avascular necrosis[21]; and treatment guidance, by applying the practical rule of "displacement" (fragments displaced >1 cm or angulated >45°), which provides an easily applicable guideline for indicating conservative or surgical management[5].

On the other hand, the MAYO-FJD classification has the following main potential advantages: a) prediction of conservative treatment failure, as it groups patients into clear prognostic categories, accurately identifying patterns that will evolve well clinically from those with a high risk of collapse[22]; b) lower failure rate, assisting in predictive decision-making regarding when non-surgical treatment is appropriate[22]; and c) high reliability based on patterns of substantial interobserver agreement (Kappa index between 0.69 and 0.81) when using computed tomography[8].

In conclusion, human interpretation of complex proximal humeral fractures is frequently heterogeneous, and a window of opportunity emerges with the use of artificial intelligence (AI): Deep Learning algorithms are being trained to standardize and improve this assessment. This new context will enable high precision, with research showing that convolutional neural networks (CNNs) achieve over 95% accuracy in identifying the presence or absence of fractures on plain radiographs[23].

Another advance will be the classification of complex patterns, as AI operates in the automated identification of 3- and 4-part fractures, often surpassing the accuracy of general practitioners and equaling the performance of shoulder surgery specialists[23] – generating greater consistency, since in the future AI may remove human subjectivity, providing standardization in image reading that rapidly assists the surgeon in formulating the surgical plan[24].

Furthermore, three-dimensional models have demonstrated greater interobserver agreement (specialists versus residents) in the selection of proximal humeral fracture treatment compared to radiography, tomography, and holography. The agreement in treatment selection among specialists who used tomography and holography as diagnostic methods was twice that observed among residents[16].

Specifically, using highly complex systems (MAYO-FJD and AO/OTA) without CT support and 3D reconstructions may induce classification errors. Without these tools, the classifications are impaired in their main function – to guide surgical planning and reliably predict patient prognosis[13].

Among the limitations of this study, the use of plain radiographs as the assessment method stands out. Although widely available and low-cost, they may present variations in image quality and projection positioning, interfering with interpretation and reducing agreement among evaluators. The absence of three-dimensional computed tomography also limited the analysis of complex fractures, which could have been better characterized by more detailed methods.

Another limitation concerns the limited number of evaluators and the inclusion of only one observer per group, which may have introduced bias arising from intrasubject variability and the level of familiarity with each classification system. Additionally, the time each evaluator dedicated to image analysis may have influenced the accuracy and consistency of the results.

As practical applications, the results of this study reinforce the need to improve radiographic teaching and training methods during medical education and orthopedic residency. Standardization of assessment criteria and the use of complementary tools – such as 3D models and reconstruction software – may improve the reliability of classifications and reduce variability among observers.

Furthermore, the implementation of continuous training programs in radiographic interpretation and fracture classification may assist in making more precise decisions between surgical and conservative treatment, reducing reoperation rates and optimizing hospital resources.

Finally, in clinical practice, the incorporation of digital technologies and artificial intelligence into the diagnostic routine may favor the standardization of clinical approaches, improving functional prognosis and greater efficiency in the management of proximal humeral fractures. Thus, different surgeons looking at the same radiograph may classify the fracture into distinct categories and propose completely different approaches for the same patient – a fact that impacts the ability to provide the patient with a clear prognosis regarding complications or the future need for revision surgeries, representing a challenge in the clinical environment.

CONCLUSION

The present study demonstrated that, among the evaluated systems, the NEER classification presented the best interobserver agreement, with moderate results, while the LEGO/HERTEL, AO/OTA, and MAYO-FJD systems exhibited low agreement. In the intraobserver analysis, it was observed that the more experienced evaluators presented greater consistency in the classifications, indicating that professional experience is a determining factor for diagnostic stability.

The findings highlight the importance of improving orthopedic teaching and training processes, especially during medical residency, while also suggesting that the use of complementary imaging tools may increase accuracy and reduce the subjectivity of radiographic classifications. Although the NEER system remains the most reproducible among those evaluated, the pursuit of more objective, integrated, and reproducible classification methods continues to be an essential challenge for modern orthopedic practice.

For future studies, more in-depth analyses are suggested, considering the evaluation of simple versus complex fractures, two-, three-, and four-part fractures, and the impact of professional experience on each proximal humeral fracture classification system.

CRediT author statement

Conceptualization: Checchi AR; Vieira, IF; Vedam, JV. Methodology: Vieira, IF; Lima, DB; Checchi AR; Dinies, GMJ. Validation: Vieira, JLL; Lima, DB. Statistical analysis: Vieira, JLL; Vieira, IF. Formal analysis: Vieira, JLL; Vieira, IF; Lima, DB; Calandrini, RM. Investigation: Vieira, IF; Checchi AR; Lima, DB; Calandrini, RM. Writing-original draft preparation: Checchi AR; Vieira, IF; Vieira, JLL; Dinies, GMJ. Writing-review and editing: Vieira, IF; Checchi AR; Vieira, JLL; Calandrini, RM; Vedam, JV. Visualization: Checchi AR; Vieira, JLL; Dinies, GMJ. Supervision: Vieira, IF; Lima, DB; Vedam, JV. Project administration: Vieira, IF; Checchi AR; Lima, DB.

All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

1. Court-Brown CM, Caesar B. Epidemiology of adult fractures: A review. Injury. 2006; 37(8): 691-697. DOI: 10.1016/j.injury.2006.04.130
2. Olsson C, Nordqvist A, Petersson CJ. Increased fragility in patients with fracture of the proximal humerus: a case control study. Bone. 2004; 34(6): 1072-1077. DOI: 1016/j.bone.2004. 1001.009
3. Taskesen A, Göçer A, Uzel K, Yaradılmış YU. Effect of Osteoporosis on Proximal Humerus Fractures. Geriatr Orthop Surg Rehabil. 2020; 11:2151459320985399. DOI: 10.1177/2151459320985399
4. Carrerra EF, Wajnsztejn A, Lenz M, Archetti Netto N. Reproducibility of three classifications of proximal humeral fractures. Einstein (São Paulo), 2012; 10(4): 473–479. DOI: 10.1590/S1679-45082012000400014
5. Neer CS 2nd. Displaced proximal humeral fractures. I. Classification and evaluation. J Bone Joint Surg Am. 1970; 52(6): 1077-1089. DOI: 10.2106/00004623-197052060-00001
6. Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. J Shoulder Elbow Surg. 2004; 13(4): 427-433. DOI: 10.1016/j.jse.2004.01.034
7. Muller ME, Koch P, Nazarian S, Schatzker J. The comprehensive classification of fractures of long bones. Berlin: Springer-Verlag, 1990.
8. Foruria AM, Martinez-Catalan N, Pardos B, Larson D, Barlow J, Sanchez-Sotelo J. Classification of proximal humerus fractures according to pattern recognition is associated with high intraobserver and interobserver agreement. JSES Int. 2022; 6(4): 563-568. DOI: 10.1016/j.jseint.2022.03.005
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1): 159-74. DOI: 10.2307/2529310
10. Iordens GI, et al. The reliability and reproducibility of the Hertel classification for comminuted proximal humeral fractures compared with the Neer classification. J Orthop Sci. 2016; 21(5): 596-602. DOI: 10.1016/j.jos.2016.05.011
11. Papakonstantinou MK, et al. Interobserver agreement of Neer and AO classifications for proximal humeral fractures. ANZ J Surg. 2016; 86(4): 280-284. DOI: 10.1111/ans.13451
12. Marmor MT, et al. Comparison of the Neer classification to the 2018 update of the Orthopedic Trauma Association/AO fracture classification for classifying proximal humerus fractures. OTA Int. 2022; 5(3): e200. DOI: 10.1097/OI9.0000000000000200
13. Marongiu G, Leinardi L, Congia S, Frigau L, Mola F, Capone A. Reliability and reproducibility of the new AO/OTA 2018 classification system for proximal humeral fractures: a comparison of three different classification systems. J Orthop Traumatol. 2020; 21(1): 4. DOI: 10.1186/s10195-020-0543-1
14. Diaz LEP, Dada Neto F, Lofrano L, Garcia JVDC, Santana MVF, Dobashi ET. Intra- and inter-observer agreement of proximal humeral fractures classifications in adults. Acta Ortop Bras. 2022; 30(6): e257229. DOI: 10.1590/1413-785220223006e257229
15. Martínez-Sola R, et al. 'Absolute' inter-observer classifications agreement for proximal humeral fractures with a single shoulder anteroposterior X-ray. J Orthop Surg (Hong Kong). 2021; 29(2): 23094990211010520. DOI: 10.1177/23094990211010520
16. Cocco LF, Aihara AY, Franciozi C, Dos Reis FB, Luzo MVM. Three-dimensional models increase the interobserver agreement for the treatment of proximal humerus fractures. Patient Saf Surg. 2020; 14:33. DOI: 10.1186/s13037-020-00258-2
17. de Giovanni R, Coppola M, Rossi V, Mariconda M, Cozzolino A. A Comprehensive Review of 3D Imaging and Printing in Proximal Humerus Fractures and Sequelae. Journal of Clinical Medicine. 2025; 14(21): 7711. DOI: 10.3390/jcm14217711
18. 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. DOI: 10.2106/00004623-199312000-00002
19. 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
20. Vieira AC, Vieira A, Évora PT, Porto PT. Neer Classification – how do we use it and why is it important? European Society Radiology Congress. 2018, Poster section number C-2074. DOI: 10.1594/ecr2018/C-2074
21. Matsumura N, Furuhata R, Seto T, Takada Y, Shirasawa H, Oki S, Kawano Y, Shiono S. Reproducibility of the modified Neer classification defining displacement with respect to the humeral head fragment for proximal humeral fractures. J Orthop Surg Res. 2020; 23;15(1): 438. DOI: 10.1186/s13018-020-01966-2
22. Tafoya-Arreguín GA, Castillo-González JR, Sepúlveda-Cadena CO, de Jesus Martinez-Ruiz J, Briseno-Ramírez J, Prognostic value of the Mayo-FJD classification for nonsurgical proximal humerus fractures: a morphological approach to predicting functional outcomes, JSES International.

2026. DOI: 10.1016/j.jseint.2026.101743
23. Chung SW, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop. 2018; 89(4): 468-473. DOI: 10.1080/17453674.2018.1453714
24. Sperling Jr. JW, et al. A deep learning algorithm to detect proximal humerus fractures on radiographs. JSES Reviews, Reports, and Techniques. 2025; 5:635e640, DOI: 10.1016/j.xrrt.2025.07.025

How to cite this article: Lima, D.B., Checchi, A.R., Vieira, I.F., Vieira, J.L.L., Dinies, G.M.J., Calandrini, R.M., Vedam, J.V. (2026). Intra/interobserver agreement of proximal humeral fracture classification systems. O Mundo Da Saúde, 50. https://doi.org/10.15343/0104-7809.202650e19382025I. Mundo Saúde. 2026,50:e19382025.