ORIGINAL ARTICLE
SHOULDER
DOI: http://dx.doi.org/10.1590/1413-785220223006e257229

INTRA- AND INTER-OBSERVER AGREEMENT OF PROXIMAL HUMERAL FRACTURES CLASSIFICATIONS IN ADULTS

CONCORDÂNCIA INTRA E INTEROBSERVADORES DAS CLASSIFICAÇÕES NAS FRATURAS DO ÚMERO PROXIMAL EM ADULTOS

Luis Eduardo Plumacher Diaz[1], Francisco Dada Neto[1], Lucas Lofrano[1], João Vitor da Cruz Garcia[1], Marcos Vinicius Felix Santana[1,2], Eiffel Tsuyoshi Dobashi[1,2]

1. Rede D'Or São Luiz, Hospital IFOR, São Bernardo do Campo, SP, Brazil.
2. Universidade Federal de São Paulo, Paulista School of Medicine, São Paulo, SP, Brazil.

ABSTRACT

Objective: Evaluating intra- and inter-observer agreement of the Neer, AO, and AO/OTA proximal humerus fractures classification systems in adults. Methods: In total, 100 X-rays of patients with proximal humerus fractures were selected according to the inclusion and exclusion criteria established in this study. They were evaluated by four evaluators with different levels of expertise. The evaluation was performed at two distinct moments, with an interval of 21 days between each analysis. Images were randomized for the second evaluation by a researcher who did not participate in the image selection process. A Fleiss Kappa test was performed to evaluate intra- and inter-observer agreement. Results: We observed a substantial agreement with k = 0.669, k = 0.715, and k = 0.780 for the Neer, AO, and AO/OTA classification systems, respectively. Conclusion: In the second evaluation, intra-observer agreement improved. In the first evaluation, we obtained values of k = 0.724, k = 0.490, and k = 0.599 for the evaluation of the Neer, AO, and AO/OTA classifications. In the second evaluation, the values k = 0.759, k = 0.772, and k = 0.858. Therefore, the evaluations went from moderate to substantial for the AO classification and from moderate to practically perfect for the AO/OTA classification. The level of inter-observer agreement was substantial (0.61–0.80), with k = 0.669, k = 0.715, and k = 0.780 for the Neer, AO, and AO/OTA classifications, respectively. ***Level of Evidence III, Cross-Sectional Observational Study.***

Keywords: Radiography. Shoulder Fractures. Fractures, Bone. Classification. Observer Variation.

RESUMO

Objetivo: Avaliar a concordância intra e interobservadores entre os sistemas de classificação Neer, AO e AO/OTA nas fraturas do úmero proximal de indivíduos adultos. Métodos: Após a aplicação dos critérios de inclusão e exclusão determinados para a realização deste trabalho, foram selecionadas 100 radiografias de pacientes com fratura do úmero proximal. Estas foram submetidas à avaliação de quatro examinadores com níveis diferentes de expertise. A avaliação foi realizada em dois momentos distintos, com intervalo de 21 dias entre cada análise. As imagens foram randomizadas para a segunda avaliação por um pesquisador que não participou da seleção de imagens. Foi aplicado o teste kappa de Fleiss para verificar a concordância intra e interobservador. Resultados: Na primeira avaliação obtivemos valores de k = 0,724, k = 0,490 e k = 0,599, enquanto na segunda avaliação, os valores k = 0,759, k = 0,772 e k = 0,858 para as avaliações de Neer, AO e AO/OTA, respectivamente. Isso indica que a concordância intraobservador melhorou na segunda avaliação. Conclusões: As avaliações passaram de moderada para substancial para a classificação AO e de moderada para praticamente perfeita para o sistema AO/OTA. O nível de concordância interobservadores foram considerados substanciais (0,61-0,80) com k = 0,669, k = 0,715 e k = 0,780 para as classificações de Neer, AO e AO/OTA, respectivamente. ***Nível de Evidência III, Estudo Transversal Observacional.***

Descritores*: Radiografia. Fraturas do Ombro. Fraturas Ósseas. Classificação. Variações Dependentes do Observador.*

Citation: Diaz LEP, Dada Neto F, Lofrano L, Garcia JVC, Santana MVF, Dobashi ET. Intra- and inter-observer agreement of proximal humeral fractures classifications in adults. Acta Ortop Bras. [online]. 2022;30(6): Page 1 of 5. Available from URL: http://www.scielo.br/aob.

INTRODUCTION

The treatment of any fracture depends on a detailed evaluation of the extension and characteristics of bone lesions, as well as of the damage affecting adjacent soft tissues. The success of a treatment depends on the anatomical restoration and biomechanics of the involved structures for the reconstruction of the joint.[1]

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

The study was conducted at Hospital IFOR, Rede D'Or São Luiz.
Correspondence: Luis Eduardo Plumacher Diaz. Rua Vergueiro, 266, apt 87P, São Paulo, SP, Brazil, 01504000. luisplumacherdiaz@gmail.com

Article received on 10/12/2021, approved on 12/10/2021.

CC BY NC

Proximal humerus fractures correspond to about 4% of all lesions affecting a full-grown skeleton and is the most common fracture of the upper limb.[1]

To characterize the types of a proximal humerus fracture, classification systems with multiple objectives, such as defining the severity, treatment, and prognosis of the damage, are systematically used. Classification systems become efficient when they are easily understood, with a simple language, as well as reliable and reproducible, providing a high intra- and inter-observer agreement rate.[2,3]

Neer is a fracture classification system developed in the early 1970s based on the number of parts of the cephalic portion of the humerus, dividing them into two, three, or four parts, which correspond to the involvement and displacement of the greater tubercle, lesser tubercle, head, and shaft.[4,5]

The AO/ASIF[3] group based its methodology on an alphanumeric system that grades the severity of the lesion in increasing order, considering its site, pattern, and involved kinetic energy. For surgeons to make full use of this system, identifying what Müller called the "essence" of the fracture is crucial. This attribute provides a specific identification of bone lesions, which depends on a very accurate description. Each bone or bone region is numbered and these structures are divided by their site and segments. Each bone segment is divided in types, groups, and subgroups, creating a hierarchical organization, in which the morphological complexity of fractures establishes the difficulties inherent to its treatment and prognosis.

In 2018, the AO/OTA Fracture and Dislocation Classification Compendium[6] was published. It was the second review of a first publication made in 1996, as a combination of the efforts of the AO Foundation and the Orthopedic Trauma Association (OTA). This compendium created a standardized and rational methodology to describe all fractures and dislocations, to establish a consistent system of clinical interaction and research. After 20 years of use, this 2018 review presented a series of suggestions aimed at the systematic improvement of the application of this system.

We found in the literature numerous studies that evaluate inter-observer agreement of radiographic classification systems for proximal humerus fractures, but their results were conflicting.[7-9] Most studies showed reasonable or moderate agreement. Moreover, the number of studies evaluating inter-observer agreement of the AO classification was considerably smaller. This lack of studies, especially on the 2018 AO/OTA, stimulated our group to perform this study.

Some authors used other more specialized imaging modalities, such as conventional computed tomography and the use of three-dimensional images to improve the interpretation of proximal humerus fractures.[10,11] However, their results did not improve agreement rates. The current literature suggests that failures of agreement may be related to failures of the classification systems, as well as to the experience of evaluators.[7-9]

Thus, this study aimed to evaluate intra- and inter-observer agreement of the Neer, AO, and AO/OTA proximal humerus fractures classification systems in adults. Moreover, we aimed to observe the differences between different intra-observer analyses.

METHODS

This study was approved by the Research Ethics Committee of the Brazil Platform under no. CAAE 51482521.9.0000.5625.

Patients were selected based on the following inclusion criteria:

1. Adults from 18 to 60 years of age;
2. Patients of both sexes;
3. Patients with two or more incidences of X-ray of the shoulder;
4. Not having previous rheumatic diseases, active or previous infectious diseases in the segment of study, progressive neurological diseases, bone fragility (e.g., osteogenesis imperfecta), pathological fractures, history of chronic use of alendronate, corticosteroids, immunosuppressants, or chemotherapy and radiotherapy agents, osteoblastic diseases, sickle cell anemia, bone dysplasia, active endocrine and metabolic diseases, and bone loss.

Patients were excluded based on the following exclusion criteria:

1. Patients who refused to sign the informed consent form specially prepared for this study;
2. Incomplete and low-quality X-ray.

With the aid of a medical statistician, sample calculation was performed[12] to establish the number of X-rays necessary for this study. A 5% significance level and 80% power were considered. Thus, the minimum number of X-rays would be 95 consecutive cases for the result of the analysis to be statistically significant.

The sample was selected from images stored in the Viewer and Webris programs from December 2018 to January 2020. For the use of data, a prior authorization was obtained from the hospital. All X-rays were selected by a researcher who did not participate in their classification process.

In total, 100 X-rays of patients who met the inclusion criteria were used. For the first evaluation, a file was created to distribute these X-rays. For the second evaluation, a second file was created, in which X-rays were randomly distributed, and delivered to each researcher.

X-rays were classified by four evaluators, with different levels of experience: a first-year resident physician in Orthopedics and Traumatology (evaluator 1), a third-year resident physician in Orthopedics and Traumatology (evaluator 2), an orthopedist with a title of specialist given by the Brazilian Society of Orthopedics and Traumatology (evaluator 3), and a subspecialist in shoulder and elbow surgery (evaluator 4).

To minimize the interpretation bias, evaluators received a prior explanation of the classification systems to be used. Each evaluator received a file with the images of each classification system. The four evaluators performed evaluations independently at two different moments, with an interval of 21 days between each analysis. To preserve confidentiality during analysis, evaluators could not talk to each other about the studied X-rays during the entire evaluation process. Data on patient name, age, sex, fracture time, and trauma mechanism were also not presented. Moreover, evaluators did not have access to the patients' clinical history.

Fracture classification system

In this study, three fracture classification systems were used: Neer, AO, and AO/OTA.

The Neer classification considers that proximal humerus fractures can, in a reproducible way, result in four anatomical segments, with or without additional fracture lines, which are arbitrarily defined as those in which a segment undergoes a translation of at least 1 cm or a minimum angulation of 45°.[4,5] The resulting four-segment classification provides a descriptive proximal humerus fracture classification system, which mainly aims to conceptualize the pathological anatomy of these fractures and the terminology for each category. For displaced fractures, the number of displaced segments and the main displaced segment are considered. Displaced fractures with less than 1 cm of displacement and 45° of angulation are considered non-displaced and commonly called "one-part fractures."[4,5]

The AO classification is an alphanumeric system based on the fracture site and its morphology. It was especially created for fractures of long bones and their respective segments and subsegments.[3] The fracture morphology is described by a specific code representing the fracture pattern, a severity code, and an additional code used in certain types of specific fractures. This classification

system considers the site and presence of impaction, angulation, translation, or comminution of fractures, as well as the presence or absence of dislocation. They are classified as belonging to bone segment 11 (1 for humerus, 1 for proximal segment) and subclassified in types, groups, and subgroups.[3]

Type A fractures are extra-articular unifocal fractures associated with a single fracture line; type B fractures are extra-articular bifocal fractures associated with two fracture lines; and type C fractures involve the humerus head or anatomical neck. Type A fractures are grouped in fractures of the greater tubercle (A1), surgical neck fractures with metaphyseal impaction (A2), and surgical neck fractures without metaphyseal impaction (A3). Type B fractures are grouped in surgical neck fractures with metaphyseal impaction and a fracture of the greater tubercle or lesser tubercle (B1), surgical neck fractures without impaction and a fracture of the greater tubercle or lesser tubercle (B2), and surgical neck fractures with a fracture of the greater tubercle or lesser tubercle and glenohumeral dislocation (B3). Type C fractures are grouped in anatomical neck fractures with mild displacement (C1), anatomical neck fractures with significant displacement (C2), and anatomical neck fractures with glenohumeral dislocation (C3). Each type of fracture is subgrouped according to displacement, valgus or varus angulation of the humerus head, comminution, and presence and direction of the dislocation of the glenohumeral joint.[3]

The 2018 AO/OTA[6] considers the same bone segment and groups of the AO classification, however, the subgroups are more detailed. Fractures of the tubercles belong to group 11A1 and its subgroups are A1.1 (fractures of the greater tubercle) and A1.2 (fractures of the lesser tubercle). Surgical neck fractures belong to group 11A2 and its subgroups are classified as: simple fracture (A2.1), wedge fragment (A2.2), multifragmentary fracture (A2.3), and vertical fracture (A2.4). Extra-articular bifocal surgical neck fractures belong to group 11B1 and its subgroups include B1.1 (greater tubercle) and B1.2 (lesser tubercle). Finally, articular or four-part anatomical neck fractures belong to group 11C1 and its subgroups include valgus-impacted fractures (C1.1) and isolated anatomical neck fractures (C1.2). Anatomical neck fractures involving the metaphysis belong to group C3 and are subclassified in: articular multifragmentary metaphyseal fractures (C3.1), intra-articular multifragmentary metaphyseal fractures (C3.2), and fractures with extension to the shaft (C3.3).[6]

Statistical analysis

The statistical analysis was performed by a professional specialized in health statistics using the IBM® SPSS® software, which offers a specific statistical analysis.

According to Altman,[13] the interpretation of results is a "chance-corrected proportional agreement." Thus, a kappa test using a coefficient of agreement with a value that varies from +1 (perfect agreement) to −1 (complete disagreement), passing by 0 (agreement equivalent to chance), is used. In this study, the Fleiss kappa coefficient was considered the most appropriate considering the multiple evaluators or evaluations and the many categories of the evaluated scales.[14] This method was used to evaluate intra- and inter-observer agreement for each scale.[15,16]

The intervals used as cut-off values were:

1. 0.0–0.2, representing a very low agreement.
2. 0.21–0.40, representing a poor agreement.
3. 0.41–0.60, representing a moderate agreement.
4. 0.61–0.80, representing a substantial agreement.
5. < 0.80, representing a practically perfect agreement.[12,14,17]

RESULTS

Table 1 (intra-observer evaluation) and Table 2 (inter-observer evaluation) present the results obtained.

Table 1. Results of the intra-observer Fleiss Kappa test.

Classification	Evaluator 1	Evaluator 2	Evaluator 3	Evaluator 4
Neer	0.790 (Substantial)	0.771 (Substantial)	0.844 (Practically perfect)	0.737 (Substantial)
AO	0.730 (Substantial)	0.769 (Substantial)	0.813 (Practically perfect)	0.694 (Substantial)
AO/OTA	0.341 (Poor correlation)	0.760 (Substantial)	0.811 (Practically perfect)	0.706 (Substantial)

Table 2. Results of the inter-observer Fleiss kappa test.

Classification	First evaluation	Second evaluation	First evaluation + second evaluation
Neer	0.724 (Substantial)	0.759 (Substantial)	0.669 (Substantial)
AO	0.490 (Moderate)	0.772 (Substantial)	0.715 (Substantial)
AO/OTA	0.599 (Moderate)	0.858 (Practically perfect)	0.780 (Substantial)

Regarding the intra-observer evaluation, for evaluator 1, the Fleiss kappa index value was k = 0.790 (substantial agreement) for the Neer classification, k = 0.730 (substantial agreement) for the AO classification, and k = 0.341 (poor correlation) for the AO/OTA classification. For evaluator 2, it was k = 0.771 (substantial agreement) for the Neer classification, k = 0.769 (substantial agreement) for the AO classification, and k = 0.760 (substantial agreement) for the AO/OTA classification. For evaluator 3, it was k = 0.844 (practically perfect) for the Neer classification, k = 0.813 (practically perfect) for the AO classification, and k = 0.811 (practically perfect) for the AO/OTA classification. For evaluator 4, it was k = 0.737 (substantial agreement) for the Neer classification, k = 0.694 (substantial agreement) for the AO classification, and k = 0.706 (substantial agreement) for the AO/OTA classification.

The inter-observer correlations were k = 0.724 (substantial agreement) for the Neer classification, k = 0.490 (moderate correlation), and k = 0.599 (moderate correlation). In the second evaluation, the Fleiss kappa indexes were k = 0.759 (substantial agreement), k = 0.772 (substantial agreement), and k = 0.858 (practically perfect) for the Neer, AO, and AO/OTA classifications, respectively.

DISCUSSION

This study evaluated intra- and inter-observer agreement of proximal humerus fractures classification systems, based on the analysis performed by four physicians with different orthopedic experiences. We carefully elaborated this study so that its results could present the understanding and application of each different fracture classification system. In the selection of X-rays, which was made by a researcher who did not participate in their classification process, we sought to resolve the possible biases, allowing an impartial analysis to be performed at two different moments. We performed the second evaluation after randomization of images after three weeks. Considering the results of intra-observer agreement, evaluators 1, 2, and 4 obtained a substantial Fleiss kappa value for the Neer, AO and AO/OTA classifications and evaluator 3 obtained a practically perfect value. For the AO/OTA classification, the less experienced evaluator presented poor correlation index in comparison with the other evaluators, possibly due to the lower ability to interpret images.

This fact can positively influence the agreement of proximal humerus fractures classifications. Although Sidor et al.[18] obtained kappa values of 0.83 when a shoulder surgeon used the Neer

classification, in comparison with residents (k = 0.48), other authors did not support the hypothesis that a greater experience would be equivalent to a better inter-observer agreement.

This finding corroborates the idea that the interpretation skills of residents in Orthopedics with less experience improve when they are guided by subspecialist surgeons. The current practice in the UK, in particular, recognizes it. Increasingly, upper limb trauma is treated by surgeons with maximum experience in this anatomical region even in district general practice.

Regarding the results of the global inter-observer analysis, we could observe a substantial correlation for all classification systems. However, when considering the first and second inter-observer evaluations separately, the results for each classification system improved. The familiarization associated with an improvement in the potential to interpret fractures can be considered a determining factor to justify this result.

However, some opinions does not agree that the Neer classification is sufficiently reliable. The variability in the quality of the X-rays used in previous studies with this evaluation method justifies this fact. This divergence may result from an incorrect positioning of the fractured limb during the X-ray examination, making it difficult to interpret the fracture lines, which can be multiple in comminuted lesions. Obese patients or with too much muscle mass cannot reach the projection of soft tissues and, often by intense pain, the ideal positioning.

The interpretation of the degree of displacement based on the displacement and angulation between fragments can allow this classification to be more consistently applied. It offers treatment options in a systematized way and has the potential to eliminate the poor result of the treatment of this fracture, which remains highly variable.

In 2011, Mahadeva et al.[19] presented kappa mean values of 0.617 to 0.730 in the evaluation of X-rays showing proximal humerus fractures, classifying them according to the Neer classification. Their results are very similar to ours, even considering the difference in experience among evaluators.

However, we found studies that suggest that the results of inter-observer agreement tests are slightly better for the Neer classification system, which would make the categorization of cases more useful in clinical practice.

To date, reducing the number of categories in each classification system have been showing minimal improvements. Moreover, the more advanced imaging modalities failed to significantly improve the inter-observer agreement.

A small number of researchers used the simplified AO classification. Majed et al.[20] simplified the AO classification to three categories and found an inter-observer kappa value of 0.30 in comparison with the value of 0.11 for the 27-category system. Siebenrock and Gerber[21] also found this result. In their study, agreement improved with the three-category system (k = 0.53) in comparison with the nine-category system (the AO classification) (k = 0.42). However, simplifying the complete classification system for nine- or three -category systems presented no substantial improvement.

Considering the variable agreement of all classification systems, unsurprisingly, the management of proximal humerus fractures can be quite varied and challenging.

The agreement of the Neer and AO classification systems improved, but the use of more advanced imaging modalities could not significantly improve inter-observer agreement.

Papakonstantinou et al.[8] observed a moderate agreement for the evaluation of 104 X-rays performed by three orthopedic surgeons with experience of two to 15 years. Siebernrock and Gerber[21] also found a moderate agreement (k = 0.40 for Neer and k = 0.42 for AO) in the inter-observer evaluation while the mean kappa for intra-observer reliability was 0.60 for the Neer classification and 0.58 for the AO classification.

Therefore, some studies, by incorporating computed tomography to adequately evaluate lesions and recognize the different fracture patterns, aimed to resolve the interpretation bias of images and, later, allow the application of the classifications, seeking to improve the levels of agreement. Moreover, the use of routine computed tomography would increase the cost of managing lesions beyond the exposure inherent to their use of patient exposure to an excessive amount of radiation.[9,11,22]

The development of other technologies and software capable of making a customized reproduction of 3D models for the evaluation of proximal humerus fractures improved the understanding and treatment regimen of some patients. The use of 3D models to better understand complex fractures in the pelvis, acetabulum, and tibial plateau was incorporated as an adjuvant diagnostic method to define the schedule of surgical treatments. These models are also useful in teaching and training in the medical field.[11]

They are considered a potential imaging method to improve diagnostic agreement by resident specialists or physicians. Those who use augmented reality presents a substantial diagnostic agreement, thus, it may be a potential option, as well as radiography and tomography. However, a study showed that the use of this resource in a given period of medical experience among different evaluators did not increase diagnostic agreement between the proposed methods. This study showed that the highest inter-observer agreement was among upper-limb surgeons with 3D reconstruction.[9,11,22]

We believe that an appropriate classification system should present a higher possible level of intra- and inter-observer agreement. It should also predict which treatment method and which type of fracture would present the best results, along with lower complication rates. Therefore, a search for a system that met these attributes still exists, as the studies found do not seem to have solved it. However, our results present a substantial agreement index, which shows that the use of the applied systems is satisfactory regardless of the level of experience of evaluators.

Therefore, we understand that an ideal classification system has not yet been found. Indexes show that these lesions have been systematically and similarly treated or, at least, discussed and receiving appropriate treatment.

A weakness of this study was its inability to provide a better classification, with reliability greater than those commonly used. We also did not evaluate the physicians' responses regarding agreement in subgroups when analyzing the AO classification, considering only its general aspects.

As a strength, this study included the AO/OTA classification in the agreement evaluation between the proximal humerus fracture classification systems available. Only one of the studies found presented this evaluation.

CONCLUSION

In the second evaluation, intra-observer agreement improved and we can attribute this increase to the greater familiarization to the classification systems used. In the first evaluation, we obtained the values k = 0.724, k = 0.490, and k = 0.599 for the Neer, AO, and AO/OTA classifications, respectively. In the second evaluation, kappa values were k = 0.759, k = 0.772, and k = 0.858. The evaluations went from moderate to substantial for the AO classification and from moderate to practically perfect for the AO/OTA classification. Therefore, the studied classifications presented substantial intra-observer levels of agreement, by the Fleiss kappa statistical method, with k = 0.669, k = 0.715, and k = 0.780 for the Neer, AO, and AO/OTA classifications, respectively.

Finally, this study significantly contributes to a better understanding of classification systems and their application by orthopedists. However, further studies capable of evaluating agreement between classifications in detail, as well as capable of elaborating a classification with greater intra- and inter-observer reliability, are necessary.

AUTHORS' CONTRIBUTIONS: Each author contributed individually and significantly to the development of this article. LEPD: literature review, data collection and analysis, writing of the article; FDN: data collection and analysis, writing of the article; LL: literature review and review of the article to be published; JVCG: literature review, data collection; MVFS: literature review, development of the research project; ETD: development of the research project, data analysis, review of the article.

REFERENCES

1. Tornett P 3rd, Ricci WM, Ostrum RF, McQueen MM, McKee MD, Court-Brown CM, editors. Rockwood and Green's fractures in adults. 9th ed. Philadelphia: Wolters Kluwer Health; 2020.
2. Utino AY, Alencar DR, Maringolo LF, Negrão JM, Blumetti FC, Dobashi ET. Intra and inter-observer agreement of the AO classification system for fractures of the long bones in the pediatric population. Rev Bras Ortop. 2015;50(5):501-8.
3. Müller ME, Koch P, Nazarian S, Schatzker J. The comprehensive classification of fractures in long bones. Berlin: Springer-Verlag; 1990.
4. Neer CS 2nd. Displaced proximal humeral fractures. I. Classification and evaluation. J Bone Joint Surg Am. 1970;52(6):1077-89.
5. Neer CS 2nd. Displaced proximal humeral fractures. II. Treatment of three-part and four-part displacement. J Bone Joint Surg Am. 1970;52(6):1090-103.
6. Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF. Fracture and Dislocation Classification Compendium—2018. J Orthop Trauma. 2018;32(Suppl 1):S1-170.
7. Cocco LF, Yazzigi JA Jr, Kawakami EFKI, Alvachian HJF, Reis FB, Luzo MVM. Inter-observer reliability of alternative diagnostic methods for proximal humerus fractures: a comparison between attending surgeons and orthopedic residents in training. Patient Saf Surg. 2019;13:12.
8. 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.
9. Foroohar A, Tosti R, Richmond JM, Gaughan JP, Ilyas AM. Classification and treatment of proximal humeral fractures: inter-observer reliability and agreement across imaging modalities and experience. J Orthop Surg Res. 2011;6:38.
10. Sjödén GOJ, Movin T, Aspelin P, Güntner 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.
11. Matsushigue T, Franco VP, Pierami R, Tamaoki MJS, Archetti Netto N, Matsumoto MH. 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.
12. Cantor AB. Sample-size calculations for Cohen's Kappa. Psychol Methods. 1996;1(2):150-3.
13. Altman DG. Practical statistic for medical research. 3rd ed. London: Chapman and Hall; 1995.
14. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360-3.
15. Rosner BA. Fundamentals of biostatistics. 4th ed. Belmont: Duxbury Press; 1995.
16. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: Wiley; 1981.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
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.
19. Mahadeva D, Dias RG, Deshpande SV, Datta A, Dhillon SS, Simons AW. The reliability and reproducibility of the Neer classification system – digital radiography (PACS) improves agreement. Injury. 2011;42(4):339-42.
20. Majed A, Macleod I, Bull AMJ, Zyto K, Resch H, Hertel R, et al. Proximal humeral fracture classification systems revisited. J Shoulder Elbow Surg. 2011;20(7):1125-32.
21. 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.
22. Sjödén GOJ, Movin T, Güntner P, Aspelin P, Ahrengart L, Ersmark H, Sperber A. Poor reproducibility of classification of proximal humeral fractures: additional CT of minor value. Acta Orthop Scand. 1997;68(3):239-42.