

INTER- AND INTRA-OBSERVER AGREEMENT IN THE MILCH AND WEISS SYSTEMS

CONCORDÂNCIA INTER E INTRAOBSERVADOR DOS SISTEMAS MILCH E WEISS

SERGIO CHARLES LOZOYA,¹ JOAQUÍN DARÍO TREVIÑO BÁEZ,² JESÚS MARÍA RANGEL FLORES,¹ JESÚS MIGUEL BRIZUELA VENTURA,¹ OMAR ARAIZA TOPETE,¹ ANDREA ALCÁZAR JUÁREZ³

1. Specialized Orthopedics and Traumatology Unit, Hospital de Traumatología y Ortopedia No. 21 del Instituto Mexicano del Seguro Social, Monterrey, NL, Mexico.

2. High-Speciality, Medical Unit No. 23 del Instituto Mexicano del Seguro Social, Monterrey, NL, Mexico.

3. Health Sciences Division, Universidad de Monterrey, Monterrey, NL, Mexico.

ABSTRACT

Objective: The aim of this study is to estimate the inter- and intra-observer agreement of the Weiss and Milch classification systems in radiological studies of fractures of the lateral humeral condyle in pediatric patients. **Methods:** An agreement study was performed with non-probability sampling of consecutive cases with a sample size of a hundred radiological studies, which were evaluated by three experienced orthopedic surgeons and three resident physicians; following a thorough comparison of both inter- and intra-observer agreements over a six-week period based on the Fleiss' kappa, which was used to determine the inter- and intra-observer agreement rates of both classifications. **Results:** The overall reliability of the entire group of evaluators for the Milch classification in the inter-observer evaluation was $\kappa = 0.13$, 95% CI, 0.08–0.18, and the intra-observer evaluation was $\kappa = 0.08$, 95% CI, 0.06–0.11. For the Weiss classification, the overall evaluation had an inter-observer agreement of $\kappa = 0.53$, 95% CI, 0.50–0.57, and an intra-observer agreement of $\kappa = 0.22$, 95% CI, 0.20–0.24. **Conclusion:** In the present study, the Weiss classification system demonstrated greater agreement than the Milch classification; however, the latter may require complementary studies, such as an arthrography to enhance classification accuracy. **Level of Evidence II, Diagnostic Studies – Investigating Diagnostic Examination.**

Keywords: Condyle. Children. Fractures, Salter-Harris. Elbow joint. Observer variation.

RESUMO

Objetivo: Este estudo visa estimar a concordância intra e interobservador da classificação de Weiss e Milch, em estudos radiológicos de pacientes pediátricos com fratura da parte lateral do côndilo do úmero. **Métodos:** Estudo de concordância realizado com amostragem não probabilística de casos consecutivos com amostra de cem estudos radiológicos que foram avaliados por três observadores especialistas e por três médicos residentes; depois de uma comparação completa da concordância inter e intraobservador durante seis semanas, com base no kappa de Fleiss, que foi usado para determinar a concordância inter e intraobservador das duas classificações. **Resultados:** A confiabilidade global de todo o grupo de avaliadores da classificação de Milch na avaliação interobservador foi $\kappa = 0,13$, IC 95% 0,08-0,18 e intraobservador foi $\kappa = 0,08$, IC 95% 0,06-0,11. A avaliação geral da classificação de Weiss obteve concordância interobservador de $\kappa = 0,53$, IC 95% 0,50-0,57 e intraobservador de $\kappa = 0,22$, IC 95% 0,20-0,24. **Conclusão:** No presente estudo, o sistema de classificação de Weiss mostrou concordância maior do que a classificação Milch; contudo, esta última pode exigir estudos complementares, como artrografia, para ampliar a acurácia da classificação. **Nível de Evidência II, Estudos Diagnósticos – Investigação de um Exame para Diagnóstico.**

Descritores: Côndilo. Criança. Fraturas Salter-Harris. Articulação do cotovelo. Variações dependentes do observador.

Citation: Charles-Lozoya S, Treviño-Baéz JD, Rangel-Flores JM, Brizuela-Ventura JM, Topete-Araiza O, Juárez-Alcázar A. Inter- and intra-observer agreement in the milch and weiss systems. *Acta Ortop Bras.* [online]. 2018;26(4):218-21. Available from URL: <http://www.scielo.br/aob>.

INTRODUCTION

The ideal system for classifying fractures is one that enhances communication between doctors, standardizes research terminology, and guides the decision-making process on the best treatment for each patient.^{1,2} An increase in the incidence of pediatric fractures of the lateral humeral condyle has recently been reported. Some

authors report an increase of 9.3%,^{3,4} while others report a range between 12 and 13.5%.^{5,6} Needless to say, the accurate classification of this type of fracture displacement and joint stability is extremely important to assess the best treatment for these patients, which could be surgical or conservative. In addition, open reduction and internal fixation could be an unnecessary treatment that could

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

Work conducted at the Mexican Social Security Institute, Hospital de Traumatología y Ortopedia No. 21, Department of Pediatrics Orthopedics, Monterrey, NL, Mexico.

Correspondence: Sergio Charles Lozoya. Ave. José Ma. Pino Suárez and Juan Ignacio Ramón, Downtown area, Monterrey, NL, México. 64000. scharleslozoya@yahoo.com.mx

Article received in 02/09/2018, approved in 05/29/2018.



potentially increase the risk of avascular necrosis due to extensive dissection. Nevertheless, conservative treatment, or a percutaneous fixation on an unstable or displaced fracture, could result in consolidation failure and residual deformities.⁷ Despite being one of the most widely used classification systems in pediatric orthopedic hospital services, the original Milch classification has not yet demonstrated its usefulness in defining treatment and prognosis,⁸ and has been reported to have an inter-observer agreement with a Cohen kappa $\kappa = 0.12$.⁹ The recently-introduced Weiss classification system, which is also based on joint displacement and congruence, may be able to predict fracture complications, and could indicate the best treatment to follow.¹⁰ The use of an adequate classification could reduce indications for surgery, especially in fractures with minimal displacement, provided the displacement and follow-up can be ensured,¹¹ hence the importance of using a proper classification system.

There have been no repeatability and reproducibility tests among different observers with different levels of experience and training in the use of these classification systems. Therefore, the main objective of this study was to determine the repeatability and reproducibility of the Weiss and Milch classification systems for pediatric fractures of the lateral humeral condyle in pediatric patients by evaluating the intra- and inter-observer agreements in radiological studies with multiple observers and outcomes.

MATERIAL AND METHODS

In 2017, an agreement study was performed at the Pediatric Orthopedics Service of the High-specialty Medical Unit of the Trauma and Orthopedics Hospital (*Hospital de Traumatología y Ortopedia No. 21*) of the Mexican Social Security Institute in Monterrey, Mexico. Prior approval from the Research Ethics Committee (R-2017-1903-11) was gained before conducting the study; all the parents signed an informed consent form, and radiological studies from the daily admission sheets of patients between 2015 and 2016 were taken from our hospital's imaging database, for use in the study. Patients aged between 3 and 14 years of age with anteroposterior and lateral radiological studies of the affected elbow were included. Patients with both fracture and elbow luxation, ipsilateral fracture in another area of the elbow, or with repeat fractures, were excluded. In order to measure agreement, the studies were blindly sorted and individually evaluated by three experienced pediatric orthopedists, and by three fourth-year orthopedic surgical residents. The evaluators were summoned on different days and were given the definitions and diagrams of the Milch⁸ and Weiss¹⁰ classifications. Figure 1 describes the Milch classification for lateral humeral condyle fractures, Type I having a fracture line that is lateral to the trochlear groove, and Type II having a fracture line into the trochlear groove. Figure 2 describes the Weiss classification system for lateral humeral condyle fractures; Type I fractures have a displacement of less than 2 mm, while Type II fractures have a displacement greater than or equal to 2mm with an intact articular cartilage, and Type III fractures have a displacement greater than or equal to 2 mm with an affected articular surface. The radiographs were shown to the evaluators. Each evaluator was given a test-like form and asked to classify the radiological studies, without receiving any information about the treatment that was previously prescribed to each patient. The amount of time considered adequate for measurement of intra-observer agreement was six weeks, during which time the images from the same cohort of patients were randomly shown to the evaluators. To estimate the sample size, previously reported data of Cohen's kappa from the Pennington et al.⁹ study was taken into account. An estimation of the confidence interval in studies seeking agreement among multiple observers was used, as reported by Rotondi et al.¹² and Donner et al.¹³ In the present study, which had three

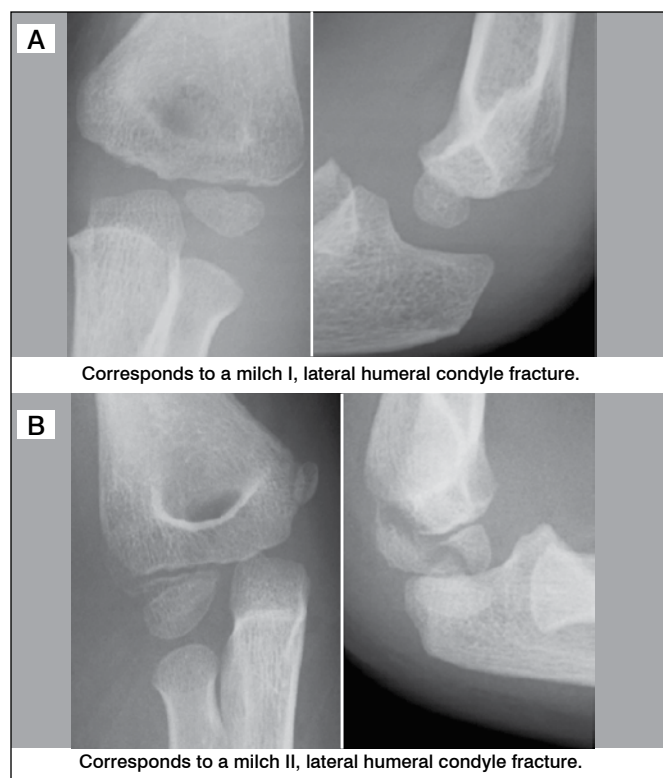


Figure 1. Milch classification for lateral humeral condyle fractures.

observers in each group and an $\alpha = 0.05$, with a lower limit of 0.70 and an upper limit of 0.90, the results for the sample size was 65 patients, but the decision to evaluate 100 radiographs was made in order to increase the variability. The data were both categorically and mutually exclusive for each category. For the analysis, the general kappa index was calculated as described by Fleiss.¹⁴ The agreement levels for the kappa coefficient were in accordance with those proposed by Landis and Koch,¹⁵ who classified them in categories ranging from: < 0.00 , poor agreement; to 0.00–0.20 slight agreement; 0.21–0.40 fair agreement; 0.41–0.60 moderate agreement; 0.61–0.80 substantial agreement, and 0.81–1.00 almost perfect agreement.¹⁶ All the statistical analyses were performed with SPSS statistical software version 22 (SPSS, Chicago, IL, USA).

RESULTS

One hundred radiological tests were analyzed by the evaluators, from which the intra- and inter-observer agreements were determined. Among the three experienced orthopedic surgeons, the overall inter-observer reproducibility for the Milch classification system resulted in moderate agreement ($\kappa = 0.41$; 95% CI, 0.30–0.53). On the other hand, the same group of observers obtained substantial agreement ($\kappa = 0.61$; 95% CI, 0.51–0.69) for the Weiss classification in overall inter-observer reproducibility. For the evaluations performed by the three orthopedic surgical residents, the Milch classification gained very poor, null and negative agreement ($\kappa = -0.04$; 95% CI, -0.07–0.15) in the overall inter-observer reproducibility. In the same group of residents, the results of the overall inter-observer reproducibility for the Weiss classification indicated moderate agreement ($\kappa = 0.47$; 95% CI, 0.39–0.55). The details of each assessment, as well as the types of classification and fracture patterns, are shown in Tables 1 and 2. The estimation of the intra-observer agreement performed by the group of experienced observers, specialists in pediatric orthopedics, using the Milch

classification, showed agreement of $\kappa = 0.20$; 95% CI, 0.15–0.25. In the same group of observers, the results obtained for the Weiss classification were $\kappa = 0.22$; 95% CI, 0.19–0.28, which translated into fair agreement in both evaluations. The intra-observer agreement in the orthopedic residents group showed poor agreement ($\kappa = -0.01$; 95% CI, -0.06–0.04) when using the Milch classification. Finally, the evaluation of intra-observer agreement among the same group of residents using the Weiss classification showed

slight agreement ($\kappa = 0.17$; 95% CI, 0.13–0.20). The details of the intra-observer evaluation are shown in Tables 3 and 4; Table 5 shows the agreement of the entire group of evaluators. Overall, the Weiss classification obtained moderate inter-observer agreement ($\kappa = 0.53$; 95% CI, 0.50–0.57) and fair intra-observer agreement ($\kappa = 0.22$; 95% CI, 0.20–0.24). Using the Milch classification, the inter-observer agreement was $\kappa = 0.13$; 95% CI, 0.08–0.18 and the intra-observer agreement was $\kappa = 0.08$; 95% CI, 0.06–0.11, which translates into slight agreement for both results.

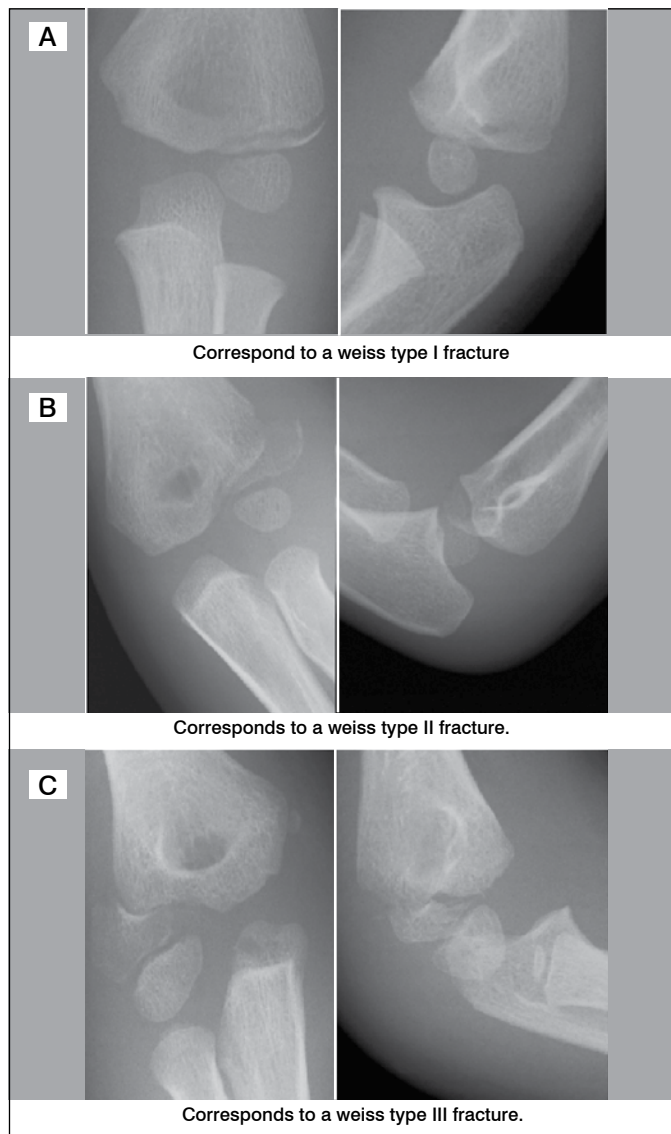


Figure 2. Weiss classification for lateral humeral condyle fractures.

Table 1. Inter-observer agreement for the Milch and Weiss classification (Group of expert evaluators).

Classification	Kappa	95% CI
Milch		
I	0.41	0.30 – 0.53
II	0.41	0.30 – 0.53
Overall	0.41	0.30 – 0.53
Weiss		
I	0.72	0.61 – 0.83
II	0.37	0.26 – 0.49
III	0.68	0.57 – 0.80
Overall	0.61	0.51 – 0.69

Table 2. Inter-observer agreement for the Milch and Weiss classification (Medical residents group).

Classification	Kappa	95% CI
Milch		
I	0.04	- 0.08 – 0.15
II	0.04	- 0.08 – 0.15
Overall	0.04	- 0.08 – 0.15
Weiss		
I	0.47	0.38 – 0.60
II	0.27	0.15 – 0.38
III	0.62	0.51 – 0.73
Overall	0.47	0.39 – 0.55

Table 3. Intra-observer agreement for the Milch and Weiss classification (Group of expert evaluators).

Classification	Kappa	95% CI
Milch		
I	0.20	0.15 – 0.25
II	0.20	0.15 – 0.25
Overall	0.20	0.15 – 0.25
Weiss		
I	0.27	0.21 – 0.32
II	0.12	0.07 – 0.17
III	0.26	0.21 – 0.32
Overall	0.22	0.19 – 0.28

Table 4. Intra-observer agreement for the Milch and Weiss classification (Medical residents group).

Classification	Kappa	95% CI
Milch		
I	- 0.01	- 0.06 – 0.04
II	- 0.01	- 0.06 – 0.04
Overall	- 0.01	- 0.06 – 0.04
Weiss		
I	0.17	0.12 – 0.22
II	0.05	0.001 – 0.10
III	0.25	0.20 – 0.30
Overall	0.17	0.13 – 0.20

Table 5. Inter-observer and Intra-observer agreement for the Milch and Weiss classification (Entire group).

Classification	Intra-observer agreement			Inter-observer agreement	
	Kappa	95% CI	Kappa	95% CI	
Milch					
I	0.08	0.06 – 0.11	0.13	0.08 – 0.18	
II	0.08	0.06 – 0.11	0.13	0.08 – 0.18	
Overall	0.08	0.06 – 0.11	0.13	0.08 – 0.18	
Weiss					
I	0.25	0.23 – 0.27	0.59	0.54 – 0.64	
II	0.10	0.07 – 0.12	0.30	0.25 – 0.35	
III	0.30	0.27 – 0.32	0.67	0.62 – 0.72	
Overall	0.22	0.20 – 0.24	0.53	0.50 – 0.57	

DISCUSSION

Multiple attempts have been made to design a classification system for lateral humeral condyle in pediatric patients that is able to answer questions regarding the mechanism of injury, anatomical difference in the displacement, and the best treatment for each fracture.¹⁷ To assess such questions, the concordance of both classification systems was compared. In the present study, the Milch classification, when used by medical residents, obtained intra-observer agreement of $\kappa = -0.01$ and inter-observer agreement of $\kappa = 0.04$, indicating poor and fair agreements, respectively. This corroborates with the previous study performed by Pennington et al.⁹ which reported fair inter-observer agreement ($\kappa = 0.12$); however, that same study obtained moderate intra-observer agreement ($\kappa = 0.45$) when comparing the original and modified Milch classifications. The evaluations performed by the expert physicians with the Milch classification showed a higher degree of inter-observer agreement, achieving moderate agreement ($\kappa = 0.41$) along with slight intra-observer agreement ($\kappa = 0.20$), which could possibly reflect repeatability with variable consistency and stability. Moreover, the Weiss classification in the current study demonstrated higher degrees of agreement than those demonstrated with the Milch classification in the group of expert physicians, and the inter-observer estimation was substantial ($\kappa = 0.61$). On the contrary, the intra-observer agreement fell to fair levels ($\kappa = 0.22$), which may be attributed to external reliability in the interpretations of the classification, since the evaluations were performed under similar conditions.

On the other hand, attempts to enhance evaluation of these fractures, as performed by Song et al.^{7,18} using the Jakob classification¹⁷ conclude that in order to properly classify the degree of displacement and prescribe the most adequate treatment, an internal oblique radiograph in type I, and an arthrography in types II and III of said classification, must be performed. A report evaluating fractures that showed minimal displacement in a radiograph, therefore being treated conservatively, recommended that the routine use of magnetic resonance imaging in follow-up is consistent with a stable fracture,

thus preventing the need for further imaging or clinical studies.¹⁹ Another recent report concluded that apart from assessing joint stability, fractures with a displacement greater than 2 mm should be assessed for surgery and that the majority of them should be treated with an open reduction and internal fixation.²⁰

One of the main limitations of our study is that in the classification developed by Weiss et al.¹⁰ for patients whose fracture caused uncertainty regarding joint congruence and had a displacement greater than 2 mm, arthrography was used to determine the best treatment: closed reduction with nails, or open reduction with internal fixation. Arthrography was not used in the present study, which may have affected the correct classification of the joint congruence of the fracture. This limitation is possibly reflected in the difficulty of classifying fractures as a type II, which obtained fair reliability ($\kappa = 0.27$). However, it was reported in these evaluations that in fractures classified as type III with displacement greater than 4 mm, arthrography was not performed to corroborate the joint congruence. In view of this, arthrography that have been used in fractures with displacement greater than 4 mm that could have an intact articular surface and may be treated with closed reduction and nails. Another limitation of this study is that no internal oblique radiographs were taken; these have proven to be a good method to determine the greatest displacement of the fracture,¹⁸ and this knowledge is useful for classifying and treating fractures.

CONCLUSION

In hospitals where resources are limited, there is a need for an easy-to-interpret classification system that does not require the costly studies. Based on our results, conservative treatment for lateral humeral condyle pediatric fractures in these environments should only be indicated when follow-up is guaranteed. In the present study, the Weiss classification showed greater agreement than the Milch classification; nonetheless, complementary studies, such as arthrography, may be needed to classify fractures adequately.

AUTHORS' CONTRIBUTIONS: Each author made significant individual contributions to this manuscript. SCHL (0000-0003-3866-0517)*, JDDB (000-0003-1465-4593)*, and JMRF (0000-0002-6206-9893)* were the main contributors to the writing of the manuscript. JMBV (0000-0002-4271-4429)*, OAT (0000-0001-7760-3879)* and AAJ (0000-0003-1505-8531)*, were responsible for collecting the imaging tests for evaluation and the clinical data. SCHL and JDDB, evaluated the data for the statistical analysis and revised the manuscript. JMBV and OAT, contributed to evaluating and classifying the X-ray images, and SCHL conducted the bibliographic research and contributed to the intellectual concept of the study. All authors participated in drafting the article or revising it critically for important intellectual content and final approval of the version to be submitted. *ORCID (Open Researcher and Contributor ID).

REFERENCES

1. Muller ME, Nazarian S, Koch P, Schatzker J. The comprehensive classification of fractures of the long bones. Berlin: Springer; 1990.
2. Urrutia J, Zamora T, Besa P, Zamora M, Schweitzer D, Klaber I. Inter and intra-observer agreement evaluation of the AO and the Tronzo classification systems of fractures of the trochanteric area. *Injury*. 2015;46(6):1054-8.
3. Mardani-Kivi M, Zohrevandi B, Saheb-Ekhtiari K, Hashemi-Motlagh K. How much are emergency medicine specialists' decisions reliable in the diagnosis and treatment of pediatric fractures? *Arch Bone Jt Surg*. 2016;4(1):60-4.
4. Behdad A, Behdad S, Hosseinpour M. Pediatric elbow fractures in a major trauma center in Iran. *Arch Trauma Res*. 2013;1(4):172-5.
5. Landin LA, Danielsson LG. Elbow fractures in children. An epidemiological analysis of 589 cases. *Acta Orthop Scand*. 1986;57(4):309-12.
6. Emery KH, Zingula SN, Anton CG, Salisbury SR, Tamai J. Pediatric elbow fractures: a new angle on an old topic. *Pediatr Radiol*. 2016;46(1):61-6.
7. Song KS, Waters PM. Lateral condylar humerus fractures: which ones should we fix? *J Pediatr Orthop*. 2012;32 (Suppl 1):S5-9.
8. Milch H. Fractures and fracture dislocations of the humeral condyles. *J Trauma*. 1964;4:592-607.
9. Pennington RG, Corner JA, Brownlow HC. Milch's classification of paediatric lateral condylar mass fractures: analysis of inter- and intraobserver reliability and comparison with operative findings. *Injury*. 2009;40(3):249-52.
10. Weiss JM, Graves S, Yang S, Mendelsohn E, Kay RM, Skaggs DL. A new classification system predictive of complications in surgically treated pediatric humeral lateral condyle fractures. *J Pediatr Orthop*. 2009;29(6):602-5.
11. Marcheix PS, Vacquerie V, Longis B, Peyrou P, Fourcade L, Moulies D. Distal humerus lateral condyle fracture in children: when is the conservative treatment a valid option? *Orthop Traumatol Surg Res*. 2011;97(3):304-7.
12. Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *J Clin Epidemiol*. 2012;65(7):778-84.
13. Donner A, Eliasziw M. Sample size requirements for reliability studies. *Stat Med*. 1987;6(4):441-8.
14. Fleiss JL. The measurement of interrater agreement. *Statistical methods for rates and proportions*. New York, NY: John Wiley & Sons; 1981.p.212-36
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.
16. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360-3.
17. Jakob R, Fowles JV, Rang M, Kassab MT. Observations concerning fractures of the lateral humeral condyle in children. *J Bone Joint Surg Br*. 1975;57(4):430-6.
18. Song KS, Kang CH, Min BW, Bae KC, Cho CH. Internal oblique radiographs for diagnosis of nondisplaced or minimally displaced lateral condylar fractures of the humerus in children. *J Bone Joint Surg Am*. 2007;89(1):58-63.
19. Hailotte G, Bachy M, Delpont M, Kabbaj R, Ducou le Pointe H, Vialle R. The Use of Magnetic Resonance Imaging in Management of Minimally Displaced or Nondisplaced Lateral Humeral Condyle Fractures in Children. *Pediatr Emerg Care*. 2017;33(1):21-5.
20. Bakarman KA, Alsiddiky AMM, Alzain KO, Alkawahski HM, Bin Nasser AS, Alsaleh KA, et al. Humeral lateral condyle fractures in children: redefining the criteria for displacement. *J Pediatr Orthop Part B*. 2016;25(5):429-33.