Original Article

Intra and interobserver concordance of the AO classification system for fractures of the long bones in the pediatric population[☆]

Artur Yudi Utino*, Douglas Rene de Alencar, Leonardo Fernadez Maringolo, Julia Machado Negrão, Francesco Camara Blumetti, Eiffel Tsuyoshi Dobashi

Escola Paulista de Medicina, Universidade Federal de São Paulo (Unifesp) , São Paulo, SP, Brazil

ARTICLE INFO

ABSTRACT

Objective: The AO classification for fractures of the long bones in the pediatric population was developed and validated in 2006. However, the complexity of this system has limited its use in clinical practice and few studies in the literature have evaluated its reproducibility and applicability. The present study had the objective of determining the intra and interobserver agreement using the pediatric AO system, among physicians with different levels of experience.

Methods: After making the sample calculation, 108 consecutive radiographs on long-bone fractures in patients aged 0–16 years, coming from the digital files of the quaternary-level hospital, were selected. The radiographs were classified by five examiners with different levels of experience after prior explanations about the system. A chart containing images from the classification was made available for consultation. The evaluations were made at two different times by each observer. The Fleiss kappa index was used to ascertain the intra and interobserver agreement.

Results: Intraobserver agreement that was at least substantial was obtained for all the items of the classification and it reached excellent levels for all observers in relation to five of the seven items considered. The interobserver evaluation presented excellent levels of agreement in two items, substantial in two items, moderate to substantial in one item and poor to moderate in one item. No influence from the observer's experience was observed with regard to obtaining higher or lower levels of agreement, either in the intraobserver or in the interobserver evaluation.

Conclusions: In this study, the intra and interobserver agreement was considered to be good or excellent for the pediatric AO classification system, for the parameters of bone, segment, paired bone, subsegment, standard and deviation. However, the intra and interobserver

agreement was statistically unsatisfactory for the parameter of severity/side of avulsion. The levels of agreement obtained did not depend on the observer's level of experience within pediatric orthopedics.

Concordância intra e interobservadores do sistema de classificação AO para fraturas dos ossos longos na população pediátrica

R E S U M O

Palavras-chave:
Fraturas ósseas/classificação
Criança
Ortopedia
Método

Objetivo: A classificação AO para fraturas dos ossos longos na população pediátrica foi desenvolvida e validada em 2006. Entretanto, a complexidade desse sistema tem limitado o seu uso na prática clínica. Poucos estudos na literatura avaliam sua reprodutibilidade e aplicabilidade. Este trabalho teve como objetivo determinar a concordância intra e interobservadores com o uso do sistema de classificação AO pediátrica entre médicos de diferentes níveis de experiência.

Métodos: Após a feitura do cálculo amostral, foram selecionadas 108 radiografias consecutivas de fraturas de ossos longos de pacientes de 0–16 anos, provenientes do arquivo digital de um hospital de nível quaternário. As radiografias foram classificadas por cinco examinadores com diferentes níveis de experiência após uma explicação prévia sobre o sistema. Foi mostrada uma planilha que continha as imagens da classificação para consulta. As avaliações foram feitas em dois momentos distintos por cada observador. O índice Kappa de Fleiss foi usado para verificar a concordância intra e interobservadores.

Resultados: Foram obtidas concordâncias intraobservadores no mínimo substanciais em todos os itens da classificação, alcançaram níveis excelentes por todos os observadores em cinco dos sete itens considerados. A avaliação interobservadores apresentou níveis de concordância excelentes em dois itens, substancial em dois itens, moderada a substancial em um item e pobre a moderada em um dos itens. Não se observou influência da experiência do observador na obtenção de maiores ou menores níveis de concordância, intra ou interobservadores.

Conclusões: Neste estudo, a concordância intra e interobservadores foi considerada boa ou excelente para o sistema de classificação AO pediátrico para os parâmetros: osso, segmento, osso pareado, subsegmento, padrão e desvio. No entanto, a concordância intra e interobservadores foi estatisticamente insatisfatória no parâmetro gravidade/lado da avulsão. Os níveis de concordância obtidos independem da experiência do observador em ortopedia pediátrica.

Introduction

The main reason for hospitalization within pediatric orthopedics is fractures of the long bones.[1] Classification of the fractures is essential for determining the epidemiology, facilitating communication between orthopedists and defining treatment algorithms.[2] Several classification systems have been developed based on the location and morphology of the injuries, in order to categorize each type of injury of long bones in children.[1]

The AO classification for long-bone fractures in adults is not used for the pediatric population because it does not take into consideration the bone elasticity, presence of the growth plate and anatomical characteristics of the epiphysis.[1] The same trauma mechanism may produce different fracture patterns in children, such as plastic deformities, greenstick fractures and complex fractures.[1] Another important characteristic is the greater fragility of the growth plate, which is less resistant than the surrounding bone, thus meaning that this structure is more easily injured.

Any orthopedic classification system needs to be clinically relevant, reproducible and valid. To meet these objectives, the system needs to go through three investigative stages, as proposed by Audigé et al.[2] In the case of pediatric fractures, the first stage should involve experienced pediatric orthopedists, in order to define a common language for describing the fracture patterns and classification process. The second stage relates to developing international multicenter agreement studies that involve surgeons with different levels of experience. The third stage relates to implementation of a prospective clinical study.[3]

The pediatric AO classification takes into consideration the AO system for long-bone fractures in adults and the most relevant pediatric fractures. The location of the fracture and its morphology are taken into consideration. The bone is subdivided into three segments: proximal (epiphysis + metaphysis), diaphyseal and distal (epiphysis + metaphysis). Regarding morphology, the disease code for the child and the fracture severity and displacement, which depend on the type of fracture, are considered.[1]

The authors of the pediatric AO classification have already reached the third stage of the validation process, i.e. application of the proposed system within the context of a prospective clinical study.[2] However, the degree of complexity of this method and the difficulty in incorporating it into clinical practice lead us to believe that studies evaluating its reproducibility and accuracy are still needed, especially if less experienced orthopedists are taken into consideration.

Thus, we conceived this study with the aim of estimating the intra and interobserver agreement of the AO classification system for long bones in children, among examiners with different levels of experience.

Materials and methods

This research project was submitted to the research ethics committee of the Brazil Platform for assessment and approval (approval number: 29073114.3.0000.5505).

Sample calculation

Firstly, we determined the number of radiographs that would be needed to obtain kappa values greater than 0.70, through tests with a significance level of 5% and power of 80%. The calculation showed that we would need to evaluate at least 95 radiographs. The formula used for this calculation was as follows[4]:

$$N = \{(z(alpha) * root(Q_0) + z(beta) * root(Q_1))/(K_1 - K_0)\}^2$$

In which $z(alpha)$ and $z(beta)$ are obtained from the normal distribution; Q_0 and Q_1 are obtained from the table of the reference article for the sample size; and K_1 and K_0 are the kappa values obtained from the hypotheses of the test. For this analysis, we obtain:

$$N = \{(1.64 * root(0.817) + 0.84 * root(0.301))/(0.9 - 07)\}^2 = 94.92$$

Sample selection

These examinations were obtained consecutively between January 2013 and March 2014 in the imaging diagnostics department of a quaternary-level university hospital, with prior authorization. All the radiographs produced during this period that were identified in the digital files as images of segments of the appendicular skeleton were obtained for evaluation. These segments included the pelvis, thigh, knee, lower leg, ankle, shoulder, upper arm, elbow, forearm and wrist. Examinations performed on children aged 0–16 years who presented fracture of the long bones were included. The radiographs were selected so as to include examinations with two views and good radiographic quality. This was done by two orthopedists who did not participate in the classification process. Thus, 119 radiographs on fractures of the long bones were collected, in anteroposterior and lateral views. Among these, six were excluded due to poor quality and five because the growth plate had already closed. The study therefore included 108 radiographs.

Process of classifying the radiographs

The radiographs were classified by five examiners with different levels of experience. One was at expert level (>10 years of experience as a pediatric orthopedist – examiner 5), one was at advanced level (>5 years of experience as a pediatric orthopedist – examiner 4), one was at medium level (>1 year of experience as a pediatric orthopedist – examiner 3) and two were at basic level (general orthopedists – examiners 1 and 2).

With the aim of minimizing bias due to difficulties in interpretation and inexperience with the classification system, the observers were given prior explanations regarding the classification systems used. Furthermore, during the classification process, a brochure containing the entire AO classification for pediatric long-bone fractures was available for each participant.

The radiographs were organized in chronological order in a closed digital file. The classifications were made by five observers, at two different times, with an interval of 15 days between one evaluation and the other. Each of the five researchers evaluated and classified the radiographs independently. The observers were given all the time that they needed to evaluate the radiographs.

The participants were instructed not to discuss the classification systems until the end of the classification stage. Furthermore, they did not have access to the patients' histories or to any clinical data.

Statistical analysis

The statistical analysis on the results obtained was performed by a specialist professional in the field of medical statistics. The Fleiss kappa test was used to assess the intra and interobserver agreement for each scale.[5,6] It is considered that using the Fleiss kappa coefficient is the most appropriate method for situations in which multiple examinations or evaluations are made and when the scale evaluated presents several categories.[7]

The tests were interpreted in accordance with Altman[8] as "proportional agreement with correction of random occurrences". The kappa agreement coefficient has values ranging from +1 (perfect agreement), through 0 (agreement equal to chance) to –1 (complete discordance). There are no definitions regarding the agreement levels that are accepted, but some studies have suggested that results in the range of 0–0.2 show very low agreement, 0.21–0.40 poor agreement, 0.41–0.60 moderate agreement and 0.61–0.80 substantial agreement. Values greater than 0.80 are considered to be practically perfect agreement.[4,7–9]

Fracture classification system

The overall structure of the classification is based on the location of the fracture and its morphology. The fracture locations covered are the different long bones and their respective segments and subsegments. The morphology of the fracture is described by a specific code that represents the fracture pattern, with a code for the severity and an additional code that is used for certain types of fractures (displaced supracondylar fractures of the humerus, displaced fractures of the head and neck of the radius and fractures of the femoral neck).[10]

The numbering system for the long bones (1–4) and for the segments (proximal = 1, diaphyseal = 2 and distal = 3) is similar to that of the AO system described by Müller for fractures of the long bones in adults.[11] It differs in relation to the coding for malleolar fractures, such as fractures of the distal tibia or fibula. Moreover, the definitions of the three bone segments differ from those of adults. The letters R, U, T and F refer to the radius, ulna, tibia and fibula and are added to the code for the segment, in relation to paired bones, when only one bone is fractured or when both bones are fractures but with different patterns.[10]

With regard to the subsegments, segments 1 and 3 are subdivided into two subsegments: the epiphysis (E) and the metaphysis (M). Segment two is the same as the diaphyseal subsegment (D).[10]

The metaphysis is defined as a square in which the sides have the same length as the widest part of the growth place. In relation to paired bones such as the radius/ulna and tibia/fibula, both bones should be included in the square. The proximal femur is an exception: its metaphysis is not defined as a square but is located between the growth plate and the subtrochanteric line.[10]

If the center of the fracture line is located inside the above-mentioned square, it is a metaphyseal fracture. If the epiphysis and the respective growth plate (physis) are included, it is an epiphyseal fracture. Intra and extra-articular ligament avulsions are epiphyseal and metaphyseal injuries, respectively.[10]

A certain number of fracture patterns that are important in children are described by the so-called "child code". These fracture patterns are specific for the subsegments in which they are located and thus are grouped as E, M or D. This code also takes into consideration some internationally accepted classification systems for pediatric fractures (such as the classification of Salter–Harris).[3,10,12]

The severity code distinguishes between two grades: simple (.1) and multifragmented (.2). To describe the side of the avulsion, when necessary, the letter M would indicate medial ligament avulsion and the letter L, lateral.

Supracondylar fractures of the humerus, which are classified as 13-M/3, are described using an additional code that takes into consideration the degree of displacement (I–IV), which is very similar to the classification of Gartland.[13]

When the paired bones (radius/ulna or tibia/fibula) both present the same fracture pattern, they should be documented by only one classification code. In this case, the severity code will be that of the bone that is more severely fractured. On the other hand, when only one bone is fractured, a lower-case letter defines this bone (r, u, t or f) and should be added to the code for the segment. For example, 22u describes a diaphyseal fracture of the ulna in isolation. Furthermore, when the two bones are fractured with different fracture patterns, each fracture should be classified separately and a lower-case letter should be included in the classification. For example, a complete spiral fracture of the radius and plastic deformity of the ulna are classified as 22r-D/5.1 and 22u-D/1.1.[10]

Fractures of the head and neck of the radius are described by an additional code (I–III) that takes into account the angle and grade of displacement. Fractures of the femoral neck are proximal metaphyseal fractures (M), with an intertrochanteric line that limits the metaphysis. These metaphyseal fractures can be divided into three types, which are represented by an additional code (I–III) that takes into account the position of the fracture in the proximal metaphysis: transcervical, basicervical and transtrochanteric.[10]

Results

Intraobserver agreement

The data relating to the statistical evaluation on intraobserver agreement and the respective results according to the Fleiss kappa index are shown in Tables 1–7. Each item that forms part of the classification was analyzed independently and is presented in a specific table.

In general terms, substantial correlation of agreement was found in relation to practically all the items addressed in the classification. Excellent agreement levels were obtained by all

Table 1 – Statistical analysis on intraobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the bone. CI, confidence interval.

AO classification	Fleiss kappa index	95% CI	
		Lower	Upper
Bone			
Examiner 1	1	1	1
Examiner 2	1	1	1
Examiner 3	1	1	1
Examiner 4	0.99	0.9704	1
Examiner 5	1	1	1

Table 2 – Statistical analysis on intraobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the segment. CI, confidence interval.

AO classification	Fleiss kappa index	95% CI	
		Lower	Upper
Segment			
Examiner 1	0.9141	0.8461	0.9822
Examiner 2	1	1	1
Examiner 3	0.9864	0.9597	1
Examiner 4	0.9864	0.96	1
Examiner 5	1	1	1

Table 3 – Statistical analysis on intraobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the paired bone. CI, confidence interval.

AO classification	Fleiss kappa index	95% CI	
		Lower	Upper
Paired bone			
Examiner 1	0.95	0.8935	1
Examiner 2	0.9811	0.9441	1
Examiner 3	0.9199	0.8429	0.9918
Examiner 4	1	1	1
Examiner 5	1	1	1

Table 4 – Statistical analysis on intraobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the subsegment. CI = confidence interval.

AO classification	Fleiss kappa index	95% CI	
		Lower	Upper
Subsegment			
Examiner 1	0.8467	0.7495	0.9439
Examiner 2	1	1	1
Examiner 3	0.9890	0.9673	1
Examiner 4	0.9483	0.8953	1
Examiner 5	0.8685	0.7890	0.9480

Table 5 – Statistical analysis on intraobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the pattern. CI, confidence interval.

AO classification	Fleiss kappa index	95% CI	
		Lower	Upper
Pattern			
Examiner 1	0.8035	0.7110	0.8959
Examiner 2	0.9142	0.8496	0.9788
Examiner 3	0.9612	0.9279	0.9945
Examiner 4	0.9113	0.8444	0.9781
Examiner 5	0.8597	0.7828	0.9367

Table 6 – Statistical analysis on intraobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the severity and side of the avulsion. CI = confidence interval.

AO classification	Fleiss kappa index	95% CI	
		Lower	Upper
Severity and side of the avulsion			
Examiner 1	0.8178	0.6196	1
Examiner 2	0.6524	0.2857	1
Examiner 3	0.7391	0.4650	1
Examiner 4	0.8347	0.6776	0.9917
Examiner 5	0.7554	0.5899	0.9209

Table 7 – Statistical analysis on intraobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the displacement. CI, confidence interval.

AO classification	Fleiss kappa index	95% CI	
		Lower	Upper
Displacement			
Examiner 1	0.9068	0.7632	1
Examiner 2	1	1	1
Examiner 3	0.9524	0.9025	1
Examiner 4	0.8779	0.7560	0.9998
Examiner 5	0.9361	0.8716	1

the observers in relation to the items bone, segment, paired bone, subsegment, pattern and displacement. On the other hand, the severity and side of avulsion presented substantial agreement for three observers and excellent for the other two. Lastly, it was seen that greater observer experience did not necessarily imply a higher level of agreement.

Interobserver agreement

Tables 8–14 show the results from the Fleiss kappa index relating to the interobserver analysis on the first and second assessments by the examiners involved in this study.

The interobserver agreement index was considered to be excellent for the items of bone and segment and substantial for the items of paired bone and subsegment. The item pattern showed moderate agreement only for one of the observers in comparison with the others, excellent for two other examiners and substantial agreement in the correlation between the remaining observers. Lastly, the item of severity and side of the injury was the one that presented greatest disparity of results. It reached an excellent agreement index only in the comparative analysis between two of the observers, while the others ranged from poor to moderate, at most. Once again, the results do not allow any correlation between the agreement levels obtained and the observers' experience.

Discussion

The pediatric AO classification is a relatively new method for grouping and standardizing the descriptions of different types of long-bone fractures in children. In the orthopedic literature, only a very limited number of studies have addressed this topic. This stimulated our group to conduct the present study, with the aim of assessing the applicability and reproducibility of this system within our setting.

An ideal classification system should conform to very well defined criteria, such as being easy to apply, being highly reproducible, having high accuracy, being capable of adequately guiding the treatment and being capable of indicating the prognosis for the injuries.[2,14–17] In addition, an ideal classification should enable comparisons between the results obtained from different series, and should allow better documentation of epidemiological data.[2]

The AO group put forward a systematic method that covered all long-bone injuries in children and used Müller's

Table 8 – Statistical analysis on interobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the bone. CI, confidence interval.

Bone	Examiner 2	Examiner 3	Examiner 4	Examiner 5
Examiner 1	1(95%CI : 1–1)	1 (95% CI: 1–1)	1 (95% CI: 1–1)	0.99 (95% CI: 0.97–1)
Examiner 2		1 (95% CI: 1–1)	1 (95% CI: 1–1)	0.99 (95% CI: 0.97–1)
Examiner 3			1 (95% CI: 1–1)	0.99 (95% CI: 0.97–1)
Examiner 4				0.99 (95% CI: 0.97–1)

Table 9 – Statistical analysis on interobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the segment. CI, confidence interval.

Segment	Examiner 2	Examiner 3	Examiner 4	Examiner 5
Examiner 1	0.8886 (95% CI: 0.81–0.96)	0.8886 (95% CI: 0.81–0.96)	0.86 (95% CI: 0.77–0.94)	0.83 (95% CI: 0.74–92)
Examiner 2		0.9729 (95% CI: 0.93–1)	0.9727 (95% CI: 0.93–1)	0.9457 (95% CI: 0.89–0.99)
Examiner 3			0.9454 (95% CI: 0.89–0.99)	0.9186 (95% CI: 0.85–0.98)
Examiner 4				0.9453 (95% CI: 0.89–99)

Table 10 – Statistical analysis on interobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the paired bone. CI, confidence interval.

Paired bone	Examiner 2	Examiner 3	Examiner 4	Examiner 5
Examiner 1	0.7988 (95% CI: 0.69–0.91)	0.6593 (95% CI: 0.52–0.79)	0.95 (95% CI: 0.89–1)	0.95 (95% CI: 0.89–1)
Examiner 2		0.6439 (95% CI: 0.51–0.78)	0.8497 (95% CI: 0.75–0.94)	0.8497 (95% CI: 0.75–0.94)
Examiner 3			0.6510 (95% CI: 0.52–0.78)	0.6510 (95% CI: 0.52–0.78)
Examiner 4				1 (95% CI: 1–1)

Table 11 – Statistical analysis on interobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the subsegment. CI, confidence interval.

Subsegment	Examiner 2	Examiner 3	Examiner 4	Examiner 5
Examiner 1	0.8378 (95% CI: 0.74–0.94)	0.8114 (95% CI: 0.71–0.91)	0.7977 (95% CI: 0.68–0.91)	0.6445 (95% CI: 0.51–0.78)
Examiner 2		0.8718 (95% CI: 0.79–0.95)	0.9585 (95% CI: 0.90–1)	0.7442 (95% CI: 0.63–0.86)
Examiner 3			0.8318 (95% CI: 0.74–0.93)	0.7414 (95% CI: 0.62–0.86)
Examiner 4				0.7464 (95% CI: 0.64–0.86)

Table 12 – Statistical analysis on interobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the pattern. CI, confidence interval.

Pattern	Examiner 2	Examiner 3	Examiner 4	Examiner 5
Examiner 1	0.7567 (95% CI: 0.64–0.87)	0.7118 (95% CI: 0.60–0.82)	0.7531 (95% CI: 0.64–0.87)	0.4327 (95% CI: 0.34–0.53)
Examiner 2		0.7117 (95% CI: 0.59–0.84)	0.8971 (95% CI: 0.82–0.97)	0.4534 (95% CI: 0.36–0.55)
Examiner 3			0.7486 (95% CI: 0.63–0.86)	0.4451 (95% CI: 0.35–0.54)
Examiner 4				0.4924 (95% CI: 0.4–0.59)

Table 13 – Statistical analysis on interobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the severity and side of the avulsion. CI, confidence interval.

Severity and side of the avulsion	Examiner 2	Examiner 3	Examiner 4	Examiner 5
Examiner 1	0.1547 (95% CI: –7 to 0.38)	0.4992 (95% CI: 0.18–0.82)	0.8347 (95% CI: 0.68–0.99)	0.3286 (95% CI: 0.09–0.57)
Examiner 2		–0.03 (95% CI: –0.05 to 0)	0.27 (95% CI: 0–0.54)	0.1818 (95% CI: 0–0.38)
Examiner 3			0.4296 (95% CI: 0.08–0.77)	0.0912 (95% CI: –0.07 to 0.25)
Examiner 4				0.3223 (95% CI: 0.09–0.55)

Table 14 – Statistical analysis on interobserver agreement according to the Fleiss kappa index, described for each examiner and for each of the parameters analyzed in the pediatric AO classification: in this table, the displacement. CI, confidence interval.

Displacement	Examiner 2	Examiner 3	Examiner 4	Examiner 5
Examiner 1	0.8160 (95% CI: 0.67–0.96)	0.7840 (95% CI: 0.65–0.92)	0.8160 (95% CI: 0.67–0.96)	0.7584 (95% CI: 0.61–0.91)
Examiner 2		0.7850 (95% CI: 0.63–0.94)	1 (95% CI: 1–1)	0.6611 (95% CI: 0.46–0.86)
Examiner 3			0.8056 (95% CI: 0.66–0.95)	0.6084 (95% CI: 0.41–0.80)
Examiner 4				0.6611 (95% CI: 0.46–0.86)

classification for adults as its basis. This method is based on an alphanumeric system and aims to categorize the main descriptive elements of these fractures, such as their location and type.[11] This classification was validated in a study published by Slongo et al.,[3] and started to be used in studies conducted by the authors who conceptualized it.

Until then, each body segment of the immature skeleton had been studied in isolation. The classification of each body segment of the immature skeleton was studied separately and the classifications of the different types of fracture were determined by authors with particular interest in each of the regions studied. We observed that for this reason, there was a large number of classifications for childhood and adolescence, guided by different criteria. For example, we can cite the systems of Poland,[18] Bergenfeldt,[19] Aitken,[20] Salter and Harris,[12] and Peterson[21] for growth plate injuries. We are aware that this multiplicity of classification methods is found for fractures of a variety of segments of the immature skeleton. However, Slongo et al.[2] emphasized that almost none of these systems have been subjected to proper validation for subsequent clinical application.

Independent of the classification method, it is ideally expected that there should be a high level of agreement among the professionals who use these methods. We observed in our study that for the variables of severity and pattern in the AO classification system for children, the level of agreement achieved was lower among some of the examiners. For the variable of pattern, there are nine subtypes for the length of the epiphysis and seven for the length of the diaphysis. Therefore, we take the view that the large number of options for each of these variables allows each examiner to have more choices that can be made, and that this is independent of the expertise and/or experience of those involved. The inference that we can make is that, despite the logic of the classification systems available, as advocated by their respective authors, they can be considered to be very complex, regardless of the detailing of each category. Therefore, this did not allow there to be an adequate level of confidence between the observers, when applied.

A smaller number of options may also generate a more reliable classification system, but this may not resolve the problem of the classifications, in a general manner. For example, in the study by Sidor et al.,[22] reduction of the number of types of fracture in order to apply the modified Neer classification for the proximal humerus was not found to provide any increase in agreement.

We believe that, in a general manner, our study presents several important points. Firstly, we brought together a large number of cases (108) that presented great variability of injuries. We observed that other studies presented series ranging in size from 10 to 275 cases.[10,14] In studies in the literature that involved the type of analysis used in our study, there was an average participation of five evaluators for every 50 cases.[14] Secondly, our observers had a variety of levels of experience, which also made it possible to ascertain whether the degree of learning might interfere with the application of the different classification systems. In our study, greater experience among the examiners did not increase the agreement among the items evaluated, which denotes that it may be possible to make general use of the classification system for the entire community of orthopedic surgeons, independent of their experience of managing pediatric fractures.[2]

We support the idea that simplified classification systems would be expected to present higher levels of intra and interobserver agreement than would the systems evaluated in this study. They would also be expected to more efficiently predict what the best treatment method would be and what type would give rise to the lowest late complication rates. Thus, a system that encompasses the predicates of an ideal classification needs to be planned for long-bone fractures of the immature skeleton.

In this manner, in our opinion, an ideal classification system has not yet been achieved. The complexity of the analysis on fractures that involve the locomotor apparatus during childhood and adolescence is directly related to several factors: age; differences in growth between different bone segments; growth patterns; bone remodeling rates; mechanical action on the bone; state of the adjacent structures; difference in growth rates between the proximal and distal growth plates; growth of the epiphysis; status of the circulation; energy of the trauma involved, etc. The need for comprehension of the influence of all these variables that change with growth of the locomotor apparatus makes creation of a single acceptable classification system a very complex task.

Conclusions

In this study, the intra and interobserver agreement for the pediatric AO classification system was considered to be good or excellent for the parameters of bone, segment, paired bone, subsegment, pattern and displacement. However, the intra and interobserver agreement relating to the parameters of severity and side of the avulsion was statistically unsatisfactory.

Conflicts of interest

The authors declare no conflicts of interest.

REFERENCES

1. Meling T, Harboe K, Enoksen HC. Reliable classification of children's fractures according to the comprehensive classification of long bone fractures by Müller. Acta Orthop. 2013;84(2):207–12.

2. Audigé L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. J Orthop Trauma. 2005;19(6):401–6.

3. Slongo T, Audigé L, Schlickewei W, Clavert JM, Hunter J. Development and validation of the AO pediatric comprehensive classification of long bone fractures by the Pediatric Expert Group of the AO Foundation in collaboration with AO Clinical Investigation and Documentation and the International Association for Pediatric Traumatology. J Pediatr Orthop. 2006;26(1):43–9.

4. Cantor AB. Sample-size calculations for Cohen's kappa. Psychol Methods. 1996;1(2):150–3.

5. Rosner BA. Fundamentals of biostatistics. 4th ed. Belmont: Duxbury Press; 1995. p. 426.

6. Fleiss JL. Statistical methods for rates and proportion. 2nd ed. New York: Wiley; 1981.

7. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360–3.

8. Altman DG. Practical statistic for medical research. 3rd ed. London: Chapman and Hall; 1995. p. 403–9.

9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.

10. Slongo T, Audigé L, Clavert JM, Lutz N, Frick S, Hunter J. The AO comprehensive classification of pediatric long-bone fractures: a web-based multicenter agreement study. J Pediatr Orthop. 2007;27(2):171–80.

11. Müller M, Narzarian S, Koch P, Schatzker J. The comprehensive classification for fractures of long bones. Berlin; Heidelberg; New York: Springer-Verlag; 1990.

12. Salter RB, Harris WR. Injuries involving the epiphyseal plate. J Bone Joint Surg Am. 1963;45(3):587–622.

13. Gartland JJ. Management of supracondylar fractures of the humerus in children. Surg Gynecol Obstet. 1959;109(2):145–54.

14. Audigé L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. Acta Orthop Scand. 2004;75(2):184–94.

15. Garbuz DS, Masri BA, Esdaile J, Duncan CP. Classification systems in orthopaedics. J Am Acad Orthop Surg. 2002;10(4):290–7.

16. Burstein AH. Fracture classification systems: do they work and are they useful? J Bone Joint Surg Am. 1993;75(12):1743–4.

17. Martin JS, Marsh JL. Current classification of fractures. Rationale and utility. Radiol Clin North Am. 1997;35(3):491–506.

18. Poland J. Traumatic separation of the epiphysis. London: Smith, Elder & Co; 1898. p. 144–62.

19. Bergenfeldt E. Beitrage zur Kenntnis der traumatischen Epiphysenlosungen an den langen Rohrenknochen der Extremitaten: eine klinisch-rontgenologische Studie. Acta Chir Scand. 1933;73 Suppl. 28.

20. Aitken AP. The end result of the fractured distal tibial epiphysis. J Bone Joint Surg. 1936;18(3):685–91.

21. Peterson HA. Physeal fractures. Part 3: Classification. J Pediatr Orthop. 1994;14(4):439–48.

22. 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.