



Artigo original

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



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

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

INFORMAÇÕES SOBRE O ARTIGO

Histórico do artigo:

Recebido em 8 de agosto de 2014

Aceito em 21 de agosto de 2014

On-line em 3 de abril de 2015

Palavras-chave:

Fraturas ósseas/classificação

Criança

Ortopedia

Método

R E S U M O

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

[☆] Trabalho desenvolvido na Disciplina de Ortopedia Pediátrica, Departamento de Ortopedia e Traumatologia, Universidade Federal de São Paulo (Unifesp), São Paulo, SP, Brasil.

* Autor para correspondência.

E-mail: artur.utino@gmail.com (A. Y. Utino).

<http://dx.doi.org/10.1016/j.rbo.2014.08.010>

0102-3616/© 2014 Sociedade Brasileira de Ortopedia e Traumatologia. Publicado por Elsevier Editora Ltda. Todos os direitos reservados.

interobservadores foi estatisticamente insatisfatória no parâmetro gravidade/lado da avulsão. Os níveis de concordância obtidos independem da experiência do observador em ortopedia pediátrica.

© 2014 Sociedade Brasileira de Ortopedia e Traumatologia. Publicado por Elsevier Editora Ltda. Todos os direitos reservados.

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

A B S T R A C T

Keywords:

Bone fractures/classification
Child
Orthopedics
Method

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 concordance 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 concordance.

Results: Intraobserver concordance 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 concordance 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 concordance, either in the intraobserver or in the interobserver evaluation.

Conclusions: In this study, the intra and interobserver concordance 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 concordance was statistically unsatisfactory for the parameter of severity/side of avulsion. The levels of concordance obtained did not depend on the observer's level of experience within pediatric orthopedics.

© 2014 Sociedade Brasileira de Ortopedia e Traumatologia. Published by Elsevier Editora Ltda. All rights reserved.

Introdução

A principal razão que motiva uma internação em ortopedia pediátrica é a fratura dos ossos longos.¹ A classificação das fraturas é essencial para determinação da epidemiologia, comunicação entre os ortopedistas e definição de algoritmos de tratamento.² Múltiplos sistemas de classificação foram desenvolvidos com base na localização e morfologia das lesões para categorizar cada tipo de lesão dos ossos longos na criança.¹

A classificação AO de fraturas dos ossos longos em adultos não é usada na população pediátrica, pois não leva em consideração a elasticidade óssea, a placa fisária e as características anatômicas da epífise.¹ O mesmo mecanismo de trauma pode produzir padrões diferentes

de fraturas em crianças, como as deformidades plásticas, a fratura em galho verde e as fraturas complexas.¹ Outra característica importante é a maior fragilidade da placa de crescimento, que é menos resistente do que o osso ao redor, o que faz com que essa estrutura seja lesada mais facilmente.

Qualquer sistema de classificação ortopédica deve ser clinicamente relevante, confiável, reprodutível e válido. Para atingir esses objetivos tal sistema deve passar por três fases de pesquisa, conforme proposto por Audige et al.² No caso das fraturas pediátricas, a primeira fase deve envolver ortopedistas pediátricos experientes para definir uma linguagem comum para descrever os padrões de fratura e o processo de classificação. A segunda fase relaciona-se com desenvolver estudos de concordância multicêntricos internacionais que envolvam cirurgiões com diferentes níveis de experiência.

A terceira fase está relacionada com a execução de um estudo clínico prospectivo.³

A classificação AO pediátrica leva em consideração o sistema AO para fraturas dos ossos longos em adultos e classificações de fraturas pediátricas mais relevantes. São levadas em consideração a localização da fratura e sua morfologia. O osso é subdividido em três segmentos: proximal (epífise + metáfise), diafisário e distal (epífise + metáfise). Quanto à morfologia, consideramos o código da criança, a gravidade e o deslocamento, dependendo do tipo de fratura.¹

Os autores da classificação AO pediátrica já alcançaram a terceira fase do processo de validação, ou seja, o sistema proposto deve ser aplicado no contexto de um estudo clínico prospectivo.² Entretanto, o grau de complexidade desse método e a dificuldade de incorporá-lo na prática clínica nos levam a acreditar que ainda são necessários trabalhos que avaliem sua reprodutibilidade e acurácia, principalmente se considerarmos os ortopedistas menos experientes.

Dessa forma, idealizamos este trabalho com o objetivo de estimar a concordância intra e interobservadores do sistema de classificação AO para as fraturas dos ossos longos em crianças, em examinadores com diferentes níveis de experiência.

Material e métodos

Este projeto de pesquisa foi submetido à avaliação e aprovação pelo comitê de ética em pesquisa da Plataforma Brasil (nº CAAE: 29073114.3.0000.5505).

Cálculo amostral

Inicialmente, determinamos o número de radiografias necessárias para obter valores de Kappa maiores do que 0,70, por meio de teste com nível de significância de 5% e poder de 80%. O cálculo obtido foi de que necessitaríamos de no mínimo 95 radiografias avaliadas. A fórmula usada para esse cálculo é dada por:⁴

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

Na qual $z(\alpha)$ e $z(\beta)$ são obtidos da distribuição normal; Q_0 e Q_1 são obtidos da tabela do artigo de referência do tamanho da amostra; K_1 e K_0 são valores de Kappa obtidos das hipóteses do teste.

Para essa análise temos:

$$N = \{(1,64 * \text{raiz}(0,817) + 0,84 * \text{raiz}(0,301)) / (0,9 - 0,7)\}^2 \\ = 94,92$$

Seleção da amostra

Esses exames foram obtidos de maneira consecutiva de janeiro de 2013 a março de 2014 no departamento de diagnóstico por imagem de um hospital universitário de nível quaternário, com autorização prévia. Todas as radiografias feitas nesse período foram obtidas para avaliação quando identificadas no arquivo digital como de segmentos do

esqueleto apendicular, incluindo: bacia, coxa, joelho, perna, tornozelo, ombro, braço, cotovelo, antebraço e punho. Foram incluídos exames feitos em crianças entre 0-16 anos e que apresentassem fraturas dos ossos longos. As radiografias foram selecionadas por dois ortopedistas que não participaram do processo de classificação, de forma a incluir exames com duas incidências e boa qualidade radiográfica. Foram coletadas 119 radiografias nas incidências anteroposterior e perfil de fraturas de ossos longos. Desses exames, seis foram excluídos por baixa qualidade e cinco em virtude de a fise estar fechada. Restaram 108 radiografias.

Processo de classificação das radiografias

As radiografias foram classificadas por cinco examinadores com diferentes níveis de experiência. Um é nível expert (> 10 anos de experiência como ortopedista pediátrico - examinador 5), um é avançado (> 5 anos de experiência como ortopedista pediátrico - examinador 4), um é nível médio (> 1 ano de experiência como ortopedista pediátrico - examinador 3) e dois são nível básico (ortopedista geral - examinadores 1 e 2).

Com a finalidade de minimizar o viés devido às dificuldades de interpretação e inexperiência com a classificação, os observadores obtiveram uma explanação prévia dos sistemas de classificação usados. Além disso, durante o processo de classificação o participante tinha à sua disposição a brochura que continha toda a classificação AO pediátrica de fraturas de ossos longos.

As radiografias foram organizadas em ordem cronológica em um arquivo digital fechado. As classificações foram feitas por cinco observadores e em dois momentos diferentes, com um intervalo de 15 dias entre uma avaliação e a outra. Cada um dos cinco pesquisadores avaliou independentemente as radiografias e as classificou. Os observadores dispuseram de todo o tempo necessário para avaliar as radiografias.

Os participantes foram instruídos a não discutir os sistemas de classificação até o fim da fase de classificação. Além disso, não obtiveram acesso à história dos pacientes ou a qualquer dado clínico.

Análise estatística

A análise estatística dos resultados obtidos foi feita por um profissional especializado da área de estatística médica. Foi usado o teste Kappa de Fleiss para avaliar a concordância intra e interobservadores para cada escala.^{5,6} O uso do coeficiente Kappa de Fleiss é considerado o mais apropriado quando nos defrontamos com a situação em que múltiplos examinadores ou avaliações são feitas e quando a escala avaliada apresenta muitas categorias.⁷

O teste foi interpretado de acordo com Altman⁸ como "concordância proporcional com correção do acaso". Kappa é o coeficiente de concordância cujo valor varia de +1 (concordância perfeita), passa por 0 (concordância igual ao acaso) e vai até -1 (discordância completa). Não há definições quanto aos níveis de concordância aceitos, mas alguns estudos sugerem que resultados entre 0-0,2 apresentam uma concordância muito pequena, 0,21-0,40 uma concordância pobre, 0,41-0,60 uma concordância moderada e 0,61-0,80 uma concordância

substancial. Um valor acima de 0,80 é considerado concordância praticamente perfeita.^{4,7-9}

Sistema de classificação das fraturas

A estrutura global da classificação é baseada na localização da fratura e na sua morfologia. A localização da fratura compreende os diferentes ossos longos e seus respectivos segmentos e subsegmentos. A morfologia da fratura é descrita por um código específico que representa o padrão de fratura, um código de gravidade e um código adicional que é usado em certos tipos de fraturas (supracondilar umeral desviada, da cabeça e colo do rádio desviado e do colo do fêmur).¹⁰

A numeração dos ossos longos (1-4) e dos segmentos (proximal = 1, diafisário = 2 e distal = 3) é similar à da classificação AO de Müller para fraturas dos ossos longos no adulto,¹¹ difere na codificação das fraturas maleolares como fraturas da tibia/fíbula distal. Além disso, a definição dos três segmentos ósseos é diferente da do adulto. As letras R, U, T e F se referem a rádio, ulna, tibia e fíbula e são adicionadas ao código do segmento, em ossos pareados, quando somente um osso está fraturado ou quando ambos os ossos estão fraturados com padrões diferentes.¹⁰

Com relação aos subsegmentos, os segmentos 1 e 3 são subdivididos em dois subsegmentos, a epífise (E) e a metáfise (M). O segmento dois é o mesmo que o subsegmento diafisário (D).¹⁰

A metáfise é determinada por um quadrado no qual os lados têm o mesmo comprimento que a parte mais larga da placa de crescimento. Nos ossos pareados, como o rádio/ulna e a tibia/fíbula, ambos os ossos devem ser incluídos no quadrado. O fêmur proximal é uma exceção, sua metáfise não é definida pelo quadrado, mas localizada entre a placa de crescimento e a linha subtrocantérica.¹⁰

Se o centro da linha da fratura está localizado dentro do quadrado mencionado, é uma fratura metafisária. Se a epífise e a respectiva placa de crescimento (fise) estão incluídas, é uma fratura epifisária. Avulsões ligamentares intra e extra-articulares são lesões epifisárias e metafisárias, respectivamente.¹⁰

Há um certo número de padrões de fraturas importantes em crianças que estão descritos pelo chamado "código de criança". Esses padrões de fraturas são específicos dos subsegmentos nos quais estão localizados e, dessa forma, agrupados de acordo com E, M ou D. Esse código ainda leva em consideração algumas classificações de fraturas pediátricas internacionalmente aceitas (por exemplo, a classificação de Salter-Harris).^{3,10,12}

O código de severidade distingue entre dois graus: simples (.1) e multifragmentado (.2). Para descrever o lado da avulsão, quando necessário, a letra M indicaria avulsão ligamentar medial e a letra L, lateral.

Fraturas supracondilares umerais, que são classificadas como 13-M/3, são descritas por um código adicional que leva em consideração o grau de desvio (I até IV), bastante similar à classificação de Gartland.¹³

Quando nos ossos pareados (rádio/ulna ou tibia/fíbula) ambos apresentam o mesmo padrão de fratura, as duas devem ser documentadas por somente um código de classificação. Nesse caso, o código de severidade será o do osso mais

Tabela 1 – Análise estatística da concordância intraobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o osso) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Classificação AO	Índice de Kappa de Fleiss	IC (95%)	
		Inferior	Superior
Osso			
Examinador 1	1	1	1
Examinador 2	1	1	1
Examinador 3	1	1	1
Examinador 4	0,99	0,9704	1
Examinador 5	1	1	1

gravemente fraturado. Em contrapartida, quando apenas um osso estiver fraturado, uma letra minúscula determina esse osso (r, u, t ou f) e deve ser adicionada ao código do segmento. Por exemplo, 22u descreve uma fratura diafisária isolada da ulna. Ademais, quando ambos os ossos estiverem fraturados com padrões de fraturas diferentes, cada fratura deve ser classificada separadamente e uma letra minúscula deve ser incluída na classificação. Por exemplo, uma fratura completa em espiral do rádio e uma deformidade plástica da ulna são classificadas como 22r-D/5.1 e 22u-D/1.1.¹⁰

Fraturas da cabeça e do colo do rádio são descritas por um código adicional (I-III) que leva em conta a angulação e o grau de desvio. Fraturas do colo femoral são fraturas metafisárias proximais (M), com a linha intertrocantérica que limita a metáfise. Essas fraturas metafisárias podem ser divididas em três tipos, que são representados por um código adicional (I-III) que leva em conta a posição da fratura na metáfise proximal: transcervical, basocervical e transtrocantérica.¹⁰

Resultados

Concordância intraobservador

Os dados relacionados à avaliação estatística da concordância intraobservadores e o respectivo resultado de acordo com o índice de Kappa de Fleiss estão demonstrados nas tabelas 1-7.

Tabela 2 – Análise estatística da concordância intraobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o segmento) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Classificação AO	Índice de Kappa de Fleiss	IC (95%)	
		Inferior	Superior
Segmento			
Examinador 1	0,9141	0,8461	0,9822
Examinador 2	1	1	1
Examinador 3	0,9864	0,9597	1
Examinador 4	0,9864	0,96	1
Examinador 5	1	1	1

Tabela 3 – Análise estatística da concordância intraobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o osso pareado) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Classificação AO	Índice de Kappa de Fleiss	IC (95%)	
		Inferior	Superior
<i>Osso pareado</i>			
Examinador 1	0,95	0,8935	1
Examinador 2	0,9811	0,9441	1
Examinador 3	0,9199	0,8429	0,9918
Examinador 4	1	1	1
Examinador 5	1	1	1

Tabela 4 – Análise estatística da concordância intraobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o subsegmento) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Classificação AO	Índice de Kappa de Fleiss	IC (95%)	
		Inferior	Superior
<i>Subsegmento</i>			
Examinador 1	0,8467	0,7495	0,9439
Examinador 2	1	1	1
Examinador 3	0,9890	0,9673	1
Examinador 4	0,9483	0,8953	1
Examinador 5	0,8685	0,7890	0,9480

Cada item constante da classificação foi analisado de forma independente e apresentado em uma tabela específica.

Em termos gerais, uma correlação substancial da concordância foi encontrada em praticamente todos os itens abordados na classificação. Níveis excelentes de concordância foram obtidos por todos os observadores nos itens *osso*, *segmento* e *osso pareado* e *subsegmento*, *padrão* e *desvio*. A *gravidade/lado da avulsão*, por sua vez, apresentou concordância substancial para três observadores e excelente para os dois demais. Por fim, verifica-se que a maior experiência do

Tabela 5 – Análise estatística da concordância intraobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o padrão) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Classificação AO	Índice de Kappa de Fleiss	IC (95%)	
		Inferior	Superior
<i>Padrão</i>			
Examinador 1	0,8035	0,7110	0,8959
Examinador 2	0,9142	0,8496	0,9788
Examinador 3	0,9612	0,9279	0,9945
Examinador 4	0,9113	0,8444	0,9781
Examinador 5	0,8597	0,7828	0,9367

Tabela 6 – Análise estatística da concordância intraobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria a gravidade/lado da avulsão) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Classificação AO	Índice de Kappa de Fleiss	IC (95%)	
		Inferior	Superior
<i>Gravidade/lado da avulsão</i>			
Examinador 1	0,8178	0,6196	1
Examinador 2	0,6524	0,2857	1
Examinador 3	0,7391	0,4650	1
Examinador 4	0,8347	0,6776	0,9917
Examinador 5	0,7554	0,5899	0,9209

Tabela 7 – Análise estatística da concordância intraobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o desvio) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Classificação AO	Índice de Kappa de Fleiss	IC (95%)	
		Inferior	Superior
<i>Desvio</i>			
Examinador 1	0,9068	0,7632	1
Examinador 2	1	1	1
Examinador 3	0,9524	0,9025	1
Examinador 4	0,8779	0,7560	0,9998
Examinador 5	0,9361	0,8716	1

observador não implicou necessariamente maiores níveis de concordância.

Concordância interobservador

As tabelas 8–14 demonstram os resultados do índice de Kappa de Fleiss da análise interobservadores da primeira e segunda avaliações pelos examinadores envolvidos neste estudo.

O índice de concordância interobservador foi considerado excelente para os itens *osso* e *segmento* e substancial nos itens *osso pareado* e *subsegmento*. O item *padrão* obteve concordância moderada apenas para um dos observadores em comparação com os demais, excelente entre dois outros examinadores e substancial na correlação dos demais observadores entre si. Por fim, o item *gravidade/lado da lesão* foi o que apresentou os resultados mais díspares, alcançou um índice de concordância excelente apenas na análise comparativa entre dois dos observadores e as demais variaram entre pobre e no máximo moderada. Novamente, os resultados não permitem inferir qualquer correlação ente os níveis de concordância obtidos e a experiência dos observadores.

Discussão

A classificação AO pediátrica é um método relativamente novo de agrupar e padronizar a descrição dos diversos tipos de fratura dos ossos longos na criança. Na literatura ortopédica, observamos um número muito limitado de trabalhos

Tabela 8 – Análise estatística da concordância interobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o osso) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Osso	Examinador 2	Examinador 3	Examinador 4	Examinador 5
Examinador 1	1 (IC 95%: 1-1)	1 (IC 95%: 1-1)	1 (IC 95%: 1-1)	0,99 (IC 95%: 0,97-1)
Examinador 2		1 (IC 95%: 1-1)	1 (IC 95%: 1-1)	0,99 (IC 95%: 0,97-1)
Examinador 3			1 (IC 95%: 1-1)	0,99 (IC 95%: 0,97-1)
Examinador 4				0,99 (IC 95%: 0,97-1)

Tabela 9 – Análise estatística da concordância interobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o segmento) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Segmento	Examinador 2	Examinador 3	Examinador 4	Examinador 5
Examinador 1	0,8886 (IC 95%: 0,81-0,96)	0,8886 (IC 95%: 0,81-0,96)	0,86 (IC 95%: 0,77-0,94)	0,83 (IC 95%: 0,74-0,92)
Examinador 2		0,9729 (IC 95%: 0,93-1)	0,9727 (IC 95%: 0,93-1)	0,9457 (IC 95%: 0,89-0,99)
Examinador 3			0,9454 (IC 95%: 0,89-0,99)	0,9186 (IC 95%: 0,85-0,98)
Examinador 4				0,9453 (IC 95%: 0,89-0,99)

Tabela 10 – Análise estatística da concordância interobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o osso pareado) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Osso pareado	Examinador 2	Examinador 3	Examinador 4	Examinador 5
Examinador 1	0,7988 (IC 95%: 0,69-0,91)	0,6593 (IC 95%: 0,52-0,79)	0,95 (IC 95%: 0,89-1)	0,95 (IC 95%: 0,89-1)
Examinador 2		0,6439 (IC 95%: 0,51-0,78)	0,8497 (IC 95%: 0,75-0,94)	0,8497 (IC 95%: 0,75-0,94)
Examinador 3			0,6510 (IC 95%: 0,52-0,78)	0,6510 (IC 95%: 0,52-0,78)
Examinador 4				1 (IC 95%: 1-1)

Tabela 11 – Análise estatística da concordância interobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria subsegmento) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Subsegmento	Examinador 2	Examinador 3	Examinador 4	Examinador 5
Examinador 1	0,8378 (IC 95%: 0,74-0,94)	0,8114 (IC 95%: 0,71-0,91)	0,7977 (IC 95%: 0,68-0,91)	0,6445 (IC 95%: 0,51-0,78)
Examinador 2		0,8718 (IC 95%: 0,79-0,95)	0,9585 (IC 95%: 0,90-1)	0,7442 (IC 95%: 0,63-0,86)
Examinador 3			0,8318 (IC 95%: 0,74-0,93)	0,7414 (IC 95%: 0,62-0,86)
Examinador 4				0,7464 (IC 95%: 0,64-0,86)

Tabela 12 – Análise estatística da concordância interobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o padrão) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Padrão	Examinador 2	Examinador 3	Examinador 4	Examinador 5
Examinador 1	0,7567 (IC 95%: 0,64-0,87)	0,7118 (IC 95%: 0,60-0,82)	0,7531 (IC 95%: 0,64-0,87)	0,4327 (IC 95%: 0,34-0,53)
Examinador 2		0,7117 (IC 95%: 0,59-0,84)	0,8971 (IC 95%: 0,82-0,97)	0,4534 (IC 95%: 0,36-0,55)
Examinador 3			0,7486 (IC 95%: 0,63-0,86)	0,4451 (IC 95%: 0,35-0,54)
Examinador 4				0,4924 (IC 95%: 0,4-0,59)

Tabela 13 – Análise estatística da concordância interobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria a gravidade/lado da avulsão) analisados na classificação AO pediátrica. (IC) Intervalo de confiança

Gravidade/Lado da avulsão	Examinador 2	Examinador 3	Examinador 4	Examinador 5
Examinador 1	0,1547 (IC 95%: -0,7-0,38)	0,4992 (IC 95%: 0,18-0,82)	0,8347 (IC 95%: 0,68-0,99)	0,3286 (IC 95%: 0,09-0,57)
Examinador 2		-0,03 (IC 95%: -0,05-0)	0,27 (IC 95%: 0-0,54)	0,1818 (IC 95%: 0-0,38)
Examinador 3			0,4296 (IC 95%: 0,08-0,77)	0,0912 (IC 95%: -0,07-0,25)
Examinador 4				0,3223 (IC 95%: 0,09-0,55)

Tabela 14 – Análise estatística da concordância interobservadores de acordo com o índice de Kappa de Fleiss, descrito para cada examinador e para cada um dos parâmetros (nesse caso seria o desvio) analisados na classificação AO pediátrica. (IC Intervalo de confiança)

Desvio	Examinador 2	Examinador 3	Examinador 4	Examinador 5
Examinador 1	0,8160 (IC 95%: 0,67-0,96)	0,7840 (IC 95%: 0,65-0,92)	0,8160 (IC 95%: 0,67-0,96)	0,7584 (IC 95%: 0,61-0,91)
Examinador 2		0,7850 (IC 95%: 0,63-0,94)	1 (IC 95%: 1-1)	0,6611 (IC 95%: 0,46-0,86)
Examinador 3			0,8056 (IC 95%: 0,66-0,95)	0,6084 (IC 95%: 0,41-0,80)
Examinador 4				0,6611 (IC 95%: 0,46-0,86)

que abordam esse tema. Isso estimulou este grupo a fazer o presente estudo, com o intuito de avaliar a aplicabilidade e reprodutibilidade desse sistema em nosso meio.

Um sistema ideal de classificação deve obedecer a critérios muito bem definidos, como: ser de fácil aplicação; ser altamente reprodutível; ter elevada acurácia; orientar o tratamento adequado; e indicar o prognóstico das lesões.^{2,14-17} Além disso, uma classificação ideal deve permitir que comparações entre resultados obtidos em diferentes séries sejam feitas, bem como possibilitar uma melhor documentação de dados epidemiológicos.²

O grupo AO propôs uma sistemática que abrange todas as lesões dos ossos longos em crianças e usou como base a classificação de Müller para adultos. Esse método é baseado em um sistema alfanumérico, visa a categorizar os principais elementos descritivos das fraturas, como a localização e a personalidade.¹¹ Essa classificação foi validada em um estudo publicado em 2006, por Slongo et al.,³ quando passou a ser usada em trabalhos feitos pelos seus idealizadores.

Até então, cada segmento corporal do esqueleto imaturo era estudado isoladamente, a classificação dos diferentes tipos de fratura era determinada por autores com particular interesse em cada uma das regiões estudadas. Por esse motivo, observamos que há, para a infância e adolescência, um grande número de classificações que se norteiam por diferentes critérios. Como exemplo, podemos citar os sistemas de Poland (1898),¹⁸ Bergenfeldt (1933),¹⁹ Aitken (1936)²⁰ Salter e Harris (1963)¹² e Peterson (1994)²¹ para as lesões fisárias. Sabemos que essa multiplicidade de métodos de classificação é encontrada para fraturas de diversos segmentos do esqueleto imaturo. Entretanto, Slongo et al.² ressaltaram que a quase totalidade desses sistemas não foi submetida a uma validação adequada para sua posterior aplicação clínica.

Independentemente dos métodos de classificação, espera-se idealmente um alto nível de concordância entre os profissionais que os usam. Observamos em nosso estudo que as variáveis *gravidade* e *padrão* do sistema de classificação AO infantil obtiveram um índice de concordância mais baixo entre alguns dos examinadores. Para a variável *padrão*, existem nove subtipos para o comprometimento epifisário e sete para o comprometimento diafisário. Logo, consideramos que o grande número de opções para cada uma dessas variáveis proporcionou que cada examinador pudesse ter mais opções de escolha, o que ocorreu independentemente da perícia e/ou experiência dos envolvidos. Podemos inferir que, apesar da lógica dos sistemas de classificação disponíveis defendida por seus respectivos autores, essa pode ser considerada bastante complexa, apesar do detalhamento de cada categoria. Portanto, isso não permitiu que houvesse um índice de confiança adequado entre os observadores quando foi aplicada.

Um reduzido número de opções também pode gerar um sistema de classificação mais confiável, mas esse pode não resolver a problemática das classificações de modo geral. Por exemplo, no trabalho de Sidor et al.²² a redução do número de tipos de fratura para aplicar a classificação modificada de Neer para o úmero proximal não proporcionou aumento da concordância.

Acreditamos que, de modo geral, nossa pesquisa apresenta vários aspectos relevantes. Primeiramente, compilamos um grande número de casos (108) que apresentaram uma grande variabilidade de lesões. Em outros estudos observamos a apresentação de séries desde 10 a 275 casos.^{10,14} A literatura expõe que, nos estudos que envolvem o tipo de análise usada em nosso trabalho, há participação em média de cinco avaliadores para 50 casos.¹⁴ Em segundo lugar, os observadores têm variados níveis de experiência, o que permitiu verificar também se o grau de aprendizado interfere na aplicação dos diferentes sistemas de classificação. No nosso trabalho, a maior experiência do examinador não elevou a concordância nos itens avaliados, o que denota o possível uso generalizado da classificação para toda a comunidade de cirurgiões ortopédicos, independentemente da experiência desses no manejo de fraturas pediátricas.²

Apoiamos a ideia de que uma classificação simplificada deve apresentar um maior nível de concordância intra e interobservador do que os sistemas avaliados neste trabalho. Deve também predizer qual o melhor método de tratamento e qual tipo determinaria as menores taxas de complicação tardia de modo mais eficaz. Portanto, uma sistemática que contemple os predicados de uma classificação ideal deveria ser planejada para as fraturas dos ossos longos no esqueleto imaturo.

Logo, em nossa opinião, um sistema de classificação ideal ainda não foi alcançado. A complexidade na análise de fraturas que envolvem o aparelho locomotor na infância e adolescência está diretamente relacionada com vários aspectos: idade; diferenças de crescimento de cada segmento ósseo; padrão de crescimento; taxa de remodelação óssea; ação mecânica sobre o osso; estado das estruturas adjacentes; diferença das taxas de crescimento da fise proximal e distal; crescimento da epífise; status circulatório; energia do trauma envolvido etc. A compreensão da influência de todas essas variáveis que mudam com o crescimento do aparelho locomotor torna a criação de um sistema de classificação único e aceitável uma tarefa bastante complexa.

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, na concordância intra e interobservadores foi estatisticamente insatisfatório o parâmetro gravidade/lado da avulsão.

Conflitos de interesse

Os autores declaram não haver conflitos de interesses.

REFERÊNCIAS

- Meling Terje, 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.
- Audigé L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma*. 2005;19(6):401-6.
- 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 col laboration with AO Clinical Investigation and Documentation and the International Association for Pediatric Traumatology. *J Pediatr Orthop*. 2006;26(1):43-9.
- Cantor AB. Sample-size calculations for Cohen's Kappa. *Psychol Methods*. 1996;1(2):150-3.
- Rosner BA. *Fundamentals of biostatistics*. 4 ed. Belmont: Duxbury Press; 1995. p. 426.
- Fleiss JL. *Statistical methods for rates and proportion*. 2 ed. New York: Wiley; 1981.
- Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360-3.
- Altman DG. *Practical statistic for medical research*. 3 ed. London: Chapman and Hall; 1995. p. 403-9.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.
- 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.
- Müller M, Narzarian S, Koch P, Schatzker J. *The comprehensive classification for fractures of long bones*. Berlin, Heidelberg, New York: Springer-Verlag; 1990.
- Salter RB, Harris WR. Injuries involving the epiphyseal plate. *J Bone Joint Surg Am*. 1963;45(3):587-622.
- Gartland JJ. Management of supracondylar fractures of the humerus in children. *Surg Gynecol Obstet*. 1959;109(2):145-54.
- 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.
- Garbuz DS, Masri BA, Esdaile J, Duncan CP. Classification systems in orthopaedics. *J Am Acad Orthop Surg*. 2002;10(4):290-7.
- Burstein AH. Fracture classification systems: do they work and are they useful? *J Bone Joint Surg Am*. 1993;75(12):1743-4.
- Martin JS, Marsh JL. Current classification of fractures. Rationale and utility. *Radiol Clin North Am*. 1997;35(3):491-506.
- Poland J. *Traumatic separation of the epiphysis*. London: Smith, Elder & Co; 1898. p. 144-62.
- Bergensfeldt E. Beitrage zur Kenntnis der traumatischen Epiphysenlosungen an den langen Rohrenknochen der Extremitaten: eine klinisch-rontgenologische Studie. *Acta Chir Scand*. 1933;73 Suppl 28.
- Aitken AP. The end result of the fractured distal tibial epiphysis. *J Bone Joint Surg*. 1936;18(3):685-91.
- Peterson HA. Physeal fractures. Part 3: classification. *J Pediatr Orthop*. 1994;14(4):439-48.
- 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.