



Artigo Original

Avaliação de concordância interobservador da classificação de Albertoni para dedo em martelo[☆]



Vinícius Alexandre de Souza Almeida^{a,*}, Carlos Henrique Fernandes^a,
João Baptista Gomes dos Santos^a, Francisco Alberto Schwarz-Fernandes^b,
Flavio Faloppa^a e Walter Manna Albertoni^a

^a Departamento de Ortopedia e Traumatologia, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, SP, Brasil

^b College of Medicine Orthopedic, University of South Florida, Tampa, Estados Unidos

INFORMAÇÕES SOBRE O ARTIGO

Histórico do artigo:

Recebido em 22 de setembro de 2016

Aceito em 10 de novembro de 2016

On-line em 8 de fevereiro de 2017

Palavras-chave:

Traumatismos dos tendões

Traumatismos dos dedos

Reprodutibilidade dos testes

Classificação

Ruptura

Deformidades adquiridas da mão

R E S U M O

Objetivo: Avaliar a reprodutibilidade da classificação de Albertoni para dedo em martelo.

Métodos: Foi feita uma avaliação por meio de questionário no qual foram avaliadas 43 radiografias em perfil da articulação interfalângica distal de dedos da mão, com lesão tipo dedo em martelo. Todas as lesões foram caracterizadas pela classificação de Albertoni, por 19 entrevistados (12 cirurgiões de mão e sete residentes). Foi então avaliada a concordância com o coeficiente Kappa generalizado, separadas por grupos – (A) avulsão tendínea; (B) fratura avulsão; (C) fratura do lábio dorsal e (D) lesão fisária – e por subgrupos (cada grupo dividido em 1 e 2).

Resultados: A concordância foi excelente para o grupo A ($k=0,95$ [0,93-0,97]) e manteve-se boa quando separados em A1 e A2. No grupo B, a concordância foi moderada ($k=0,42$ [0,39-0,44]), e foi ruim quando separada em B1 e B2. No grupo C, a concordância foi boa ($k=0,72$ [0,70-0,74]), mas quando separada em C1 e C2 se tornou moderada. No grupo D foi sempre ruim ($k=0,16$ [0,14-0,19]). A concordância geral foi moderada ($k=0,57$ [0,56-0,58]).

Conclusão: Pela avaliação da concordância geral, a classificação de Albertoni é considerada reprodutível pelo método usado na pesquisa.

© 2016 Sociedade Brasileira de Ortopedia e Traumatologia. Publicado por Elsevier Editora Ltda. Este é um artigo Open Access sob uma licença CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Evaluation of interobserver agreement in Albertoni's classification for mallet finger

A B S T R A C T

Objective: To measure the reliability of Albertoni's classification for mallet finger.

Methods: Agreement study. Forty-three radiographs of patients with mallet finger were assessed by 19 responders (12 hand surgeons and seven residents). Injuries were classified by

Keywords:

Tendon injuries

Finger injuries

[☆] Trabalho desenvolvido na Disciplina de Cirurgia da Mão e Membro Superior, Departamento de Ortopedia e Traumatologia, Universidade Federal de São Paulo, São Paulo, SP, Brasil.

* Autor para correspondência.

E-mail: vinicius.asa@hotmail.com (V.A. Almeida).

<https://doi.org/10.1016/j.rbo.2016.11.006>

0102-3616/© 2016 Sociedade Brasileira de Ortopedia e Traumatologia. Publicado por Elsevier Editora Ltda. Este é um artigo Open Access sob uma licença CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Reproducibility of results
Classification
Rupture
Acquired hand deformities

Albertoni's classification. For agreement comparison, lesions were grouped as: (A) tendon avulsion; (B) avulsion fracture; (C) fracture of the dorsal lip; and (D) physis injury—and subgroups (each group divided into two subgroups). Agreement was assessed by Fleiss's modification for kappa statistics.

Results: Agreement was excellent for Group A ($k=0.95$ (0.93 - 0.97)) and remained good when separated into A1 and A2. Group B was moderate ($k=0.42$ (0.39 - 0.44)) and poor when separated into B1 and B2. In the Group C, agreement was good ($k=0.72$ (0.70 - 0.74)), but when separated into C1 and C2, it became moderate. Group D was always poor ($k=0.16$ (0.14 - 0.19)). The general agreement was moderate, with $k=0.57$ (0.56 - 0.58).

Conclusion: Albertoni's classification evaluated for interobserver agreement is considered a reproducible classification by the method used in the research.

© 2016 Sociedade Brasileira de Ortopedia e Traumatologia. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introdução

Lesões do mecanismo extensor dos dedos estão dentre as lesões mais prevalentes na prática clínica do ortopedista. O tendão extensor terminal, formado pela união das duas bandas laterais, insere-se na região dorsal da base da falange distal. A lesão desse tendão, ou fraturas intrarticulares da base da falange distal, levam a uma deformidade em flexão da articulação interfalângica distal (IFD), conhecida como “dedo em martelo”.¹ Essa lesão acomete principalmente a população jovem, comum em práticas esportivas, e pode levar a um déficit funcional significativo se tratada inadequadamente.

Várias classificações clínicas foram descritas com a finalidade de categorizar essa afecção. Pratt *et al.*,² em 1957, classificaram o dedo em martelo baseado na etiologia: laceração, esmagamento e trauma indireto. Wehbe e Schneider, em 1984, descreveram um sistema que categorizava essas lesões em três tipos.³ Doyle *et al.*⁴ descreveram outro sistema muito usado na literatura. No Brasil, a classificação clínico-radiológica de Albertoni,⁵ descrita em 1986, é amplamente usada.

Uma classificação de boa qualidade deve ter, primeiramente, uma linguagem simples e fornecer diretrizes confiáveis que direcionem tratamento, prognóstico e possibilidade de complicações. Além disso, deve ser viável, confiável e reprodutível, essa última característica é dada pela concordância interobservador.^{1,6} Dizer que uma classificação é reprodutível significa dizer que vários indivíduos são aptos a reproduzir o mesmo resultado a qualquer momento, em qualquer lugar.¹ Assim, torna-se possível a comparação em diferentes centros de diferentes pacientes e os respectivos desfechos para cada tipo de tratamento.

Estudos de reprodutibilidade são clássicos na literatura para medir a qualidade de sistemas de classificação, principalmente em ortopedia. Esses estudos usam normalmente poucos observadores pela dificuldade de manter uma avaliação confiável. Qualquer sistema de classificação piora sua concordância à medida que aumentamos a quantidade de observadores e também quanto mais categorias temos que diferenciar. A baixa experiência dos observadores na afecção

avaliada e estudos multicêntricos também tem a tendência de diminuir a concordância.⁷

Não encontramos na literatura estudo prévio sobre a reprodutibilidade da classificação de Albertoni, nem mesmo qualquer estudo de reprodutibilidade de qualquer classificação sobre dedo em martelo.

Temos como hipótese que essa classificação tem boa concordância interobservador. O objetivo deste estudo é avaliar a concordância interobservador na classificação de Albertoni para dedo em martelo, e com isso, quantificar a reprodutibilidade dela no manejo dessa afecção.

Material

Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da instituição em que foi feita (sob o CAAE 49960815.8.0000.5505).

Foi feita uma pesquisa por meio de questionário na qual constavam 43 fotografias de radiografias em perfil da IFD de dedos da mão com lesão tipo dedo em martelo. Todas as radiografias foram consideradas de boa qualidade pelos pesquisadores.

A classificação de Albertoni⁸ constava no início do questionário. Ela divide a lesão de acordo com achados de uma radiografia em perfil da articulação IFD, e categoriza em quatro tipos: (A), lesão tendinosa pura, sem fratura; (B), lesão com avulsão óssea; (C), lesão associada à fratura da região dorsal da base da falange distal, compreende um terço ou mais da superfície articular; (D), descolamento epifisário em crianças. Cada tipo é subdividido em 1 e 2. Nos tipo A e B, o subtipo 1 é dado por uma deformidade em flexão menor do que 30 graus e o subtipo 2 por uma deformidade em flexão maior ou igual a 30 graus. A deformidade maior do que 30 graus infere que ocorreu lesão dos ligamentos retinaculares e de estruturas capsulares nos tipo A2 e B2. O tipo C diferencia-se em C1, com articulação congruente (estável), e C2, com articulação subluxada ou luxada (instável). Já o tipo D diferencia-se em D1, por descolamento epifisário (lesão de Salter e Harris tipo 1) e D2, como fratura-descolamento (Salter e Harris Tipo 3).^{8,9}

Tabela 1 – Nível de concordância segundo Landis e Koch

Valor de k	Força de concordância
< 0,00	Não há concordância
0,01-0,20	Muito ruim
0,21-0,40	Ruim
0,41-0,60	Moderada
0,61-0,80	Boa
0,81-1,00	Excelente

Abaixo de cada fotografia de radiografia eram apresentadas as opções A1, A2, B1, B2, C1, C2, D, para que o observador escolhesse apenas uma delas. Devido à raridade do tipo D não subdividimos em D1 e D2. Para determinação dos subgrupos 1 ou 2 nos tipo A e B foi fornecido ao avaliador um goniômetro e uma caneta para as medições necessárias. Consideramos que todos os observadores estavam aptos à medição das angulações, visto que todos têm título de especialista em ortopedia e traumatologia.

O questionário foi aplicado a 19 observadores, todos da mesma instituição, divididos em 12 cirurgiões de mão e sete residentes em cirurgia da mão. Cada observador respondeu ao questionário em separado, não havendo discussão entre eles.

Métodos estatísticos

A concordância interobservador foi feita por meio do coeficiente de concordância de Fleiss – k, uma generalização para mais de dois avaliadores, baseada na medida de concordância de Scott.^{10,11} O erro padrão e consequentemente os intervalos de confiança foram calculados de acordo com o algoritmo de Fleiss.¹²

Foram comparadas as concordâncias das sete opções possíveis da classificação. Foram também comparadas as concordâncias com agrupamento da classificação em quatro categorias maiores: A1 e A2 em A; B1 e B2 em B; C1 e C2 em C; e D. Dessa forma podemos avaliar a dificuldade de diferenciar os grupos (A,B,C,D) ou os subgrupos (1 e 2).

Em todas as avaliações, além dos resultados para o conjunto dos 19 profissionais, também foram comparados os resultados separadamente para médicos residentes e cirurgiões de mão. Essa comparação foi feita para avaliar a existência de diferenças de concordância entre diferentes níveis de experiência profissional. Para todos os testes inferenciais o valor de erro alfa foi fixado em 0,05.

O valor de k varia de -1 a 1, em que 1 significa total concordância, -1 significa total discordância e zero significa que os avaliadores classificaram aleatoriamente os itens. Usamos a escala de classificação de concordância segundo Landis e Koch¹³ (tabela 1).

Resultados

De acordo com as respostas dos observadores foi confeccionada a tabela 2.

Em relação à distribuição dos tipos no grupo de observadores, o tipo mais recorrente foi C1 nos grupos de cirurgiões e residentes e o menos recorrente foi D para os cirurgiões e B2 para os residentes (tabela 3).

Tabela 2 – Número de respostas para cada tipo de Albertoni, para cada radiografia

Radiografias	Opções de acordo com a classificação de Albertoni						
	A1	A2	B1	B2	C1	C2	D
1	19	0	0	0	0	0	0
2	0	19	0	0	0	0	0
3	0	0	1	0	12	6	0
4	12	1	0	0	0	0	6
5	0	19	0	0	0	0	0
6	5	14	0	0	0	0	0
7	18	1	0	0	0	0	0
8	0	19	0	0	0	0	0
9	6	13	0	0	0	0	0
10	0	19	0	0	0	0	0
11	0	0	0	0	13	6	0
12	0	0	7	0	8	4	0
13	0	0	14	1	4	0	0
14	0	0	12	1	6	0	0
15	0	0	0	0	18	1	0
16	0	19	0	0	0	0	0
17	19	0	0	0	0	0	0
18	0	0	3	1	7	8	0
19	0	0	4	10	1	4	0
20	0	0	2	0	9	8	0
21	0	0	0	1	11	7	0
22	0	0	1	1	1	16	0
23	0	0	1	1	9	8	0
24	0	0	0	0	8	11	0
25	0	0	16	0	3	0	0
26	0	0	0	0	18	1	0
27	0	0	0	0	11	8	0
28	0	0	1	0	18	0	0
29	0	0	0	0	18	1	0
30	0	0	2	0	17	0	0
31	0	0	0	0	18	1	0
32	0	0	3	0	2	14	0
33	3	16	0	0	0	0	0
34	0	0	7	0	10	2	0
35	0	0	9	0	7	0	3
36	0	0	8	0	8	0	3
37	0	0	0	0	17	2	0
38	12	0	7	0	0	0	0
39	0	0	1	1	1	16	0
40	19	0	0	0	0	0	0
41	0	0	2	0	14	3	0
42	14	5	0	0	0	0	0
43	0	0	1	0	1	14	3

Com o agrupamento da classificação de Albertoni em categorias maiores pelos tipos A, B, C e D, o tipo mais prevalente foi o C, seguido pelo A, B e D, tanto no grupo de cirurgiões de mão quanto no de residentes (tabela 4).

Na tabela 5 podem ser observados os resultados para a concordância na classificação de Albertoni segundo o coeficiente de concordância de Fleiss.

Dentre os cirurgiões de mão as classificações do tipo A1 (k=0,75 (0,71-0,78) e A2 (k=0,84 (0,80-0,87) apresentaram maior concordância do que os demais grupos, foram considerados com concordâncias boa e excelente, respectivamente. No tipo B1 (k=0,34 (0,30-0,37) e B2 (k=0,19 (0,15-0,23), a concordância foi ruim e muito ruim, respectivamente. A do tipo C1 (k=0,51 (0,48-0,55) e C2 (k=0,44 (0,41-0,48) tiveram a concordância moderada, enquanto no tipo D (k=0,10 (0,06-0,14) a

Tabela 3 – Distribuição dos tipos da classificação de Albertoni atribuídos pelos cirurgiões de mão e residentes às 43 imagens

Observador	Classificação de Albertoni													
	A1		A2		B1		B2		C1		C2		D	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Cirurgiões														
1	6	14,0	8	18,6	1	2,3	0	0,0	12	27,9	16	37,2	0	0,0
2	7	16,3	7	16,3	1	2,3	0	0,0	19	44,2	7	16,3	2	4,7
3	8	18,6	6	14,0	5	11,6	0	0,0	15	34,9	9	20,9	0	0,0
4	5	11,6	9	20,9	6	14,0	1	2,3	14	32,6	8	18,6	0	0,0
5	9	20,9	6	14,0	1	2,3	0	0,0	11	25,6	16	37,2	0	0,0
6	8	18,6	6	14,0	10	23,3	2	4,7	14	32,6	2	4,7	1	2,3
7	8	18,6	6	14,0	8	18,6	0	0,0	16	37,2	5	11,6	0	0,0
8	6	14,0	8	18,6	7	16,3	1	2,3	12	27,9	8	18,6	1	2,3
9	7	16,3	7	16,3	2	4,7	0	0,0	22	51,2	5	11,6	0	0,0
10	5	11,6	10	23,3	8	18,6	5	11,6	11	25,6	4	9,3	0	0,0
11	5	11,6	9	20,9	7	16,3	1	2,3	12	27,9	8	18,6	1	2,3
12	9	20,9	6	14,0	3	7,0	3	7,0	18	41,9	4	9,3	0	0,0
Total	83	16,1	88	17,1	59	11,4	13	2,5	176	34,1	92	17,8	5	1,0
Residentes														
13	7	16,3	7	16,3	12	27,9	0	0,0	11	25,6	5	11,6	1	2,3
14	7	16,3	8	18,6	2	4,7	0	0,0	13	30,2	11	25,6	2	4,7
15	6	14,0	9	20,9	4	9,3	1	2,3	11	25,6	12	27,9	0	0,0
16	6	14,0	9	20,9	7	16,3	1	2,3	15	34,9	4	9,3	1	2,3
17	5	11,6	8	18,6	5	11,6	1	2,3	12	27,9	10	23,3	2	4,7
18	6	14,0	8	18,6	7	16,3	0	0,0	17	39,5	3	7,0	2	4,7
19	7	16,3	8	18,6	6	14,0	1	2,3	15	34,9	4	9,3	2	4,7
Total	44	14,6	57	18,9	43	14,4	4	1,3	94	31,2	49	16,3	10	3,3
Total geral	127	15,5	145	17,7	102	12,6	17	2,1	270	33,0	141	17,3	15	1,8

Tabela 4 – Distribuição dos tipos da classificação de Albertoni, agrupada em categorias maiores, atribuídos pelos cirurgiões de mão e residentes às 43 radiografias

	Classificação de Albertoni agrupada							
	A		B		C		D	
	n	%	n	%	n	%	n	%
Cirurgiões								
	14	32,6	1	2,3	28	65,1	–	–
	14	32,6	1	2,3	26	60,5	2	4,7
	14	32,6	5	11,6	24	55,8	–	–
	14	32,6	7	16,3	22	51,2	–	–
	15	34,9	1	2,3	27	62,8	–	–
	14	32,6	12	27,9	16	37,2	1	2,3
	14	32,6	8	18,6	21	48,8	–	–
	14	32,6	8	18,6	20	46,5	1	2,3
	14	32,6	2	4,7	27	62,8	–	–
	15	34,9	13	30,2	15	34,9	–	–
	14	32,6	8	18,6	20	46,5	1	2,3
	15	34,9	6	14,0	22	51,2	–	–
Total	171	33,1	72	14,0	268	51,9%	5	1,0
Residentes								
	14	32,6	12	27,9	16	37,2	1	2,3
	15	34,9	2	4,7	24	55,8	2	4,7
	15	34,9	5	11,6	23	53,5	–	–
	15	34,9	8	18,6	19	44,2	1	2,3
	13	30,2	6	14,0	22	51,2	2	4,7
	14	32,6	7	16,3	20	46,5	2	4,7
	15	34,9	7	16,3	19	44,2	2	4,7
Total	101	33,6	47	15,6	143	47,5	10	3,3
Total geral	272	33,3	119	14,6	411	50,3	15	1,8

Tabela 5 – Coeficiente generalizado de concordância para a avaliação das imagens com p-valor < 0,0001

Classificação	k (intervalo de confiança 95%)	Classificação	k (intervalo de confiança 95%)
Cirurgião			
A	0,95 (0,91-0,99)	A1	0,75 (0,71-0,78)
B	0,34 (0,31-0,38)	A2	0,84 (0,80-0,87)
C	0,71 (0,67-0,75)	B1	0,34 (0,30-0,37)
D	0,10 (0,06-0,14)	B2	0,19 (0,15-0,23)
Geral	0,72 (0,69-0,74)	C1	0,51 (0,48-0,55)
		C2	0,44 (0,41-0,48)
		D	0,10 (0,06-0,14)
			0,56 (0,54-0,58)
Residente			
A	0,96 (0,89-1,00)	A1	0,82 (0,76-0,89)
B	0,55 (0,48-0,61)	A2	0,90 (0,83-0,96)
C	0,77 (0,70-0,83)	B1	0,46 (0,39-0,52)
D	0,24 (0,18-0,31)	B2	0,49 (0,43-0,56)
Geral	0,76 (0,72-0,81)	C1	0,49 (0,43-0,56)
		C2	0,37 (0,30-0,43)
		D	0,24 (0,18-0,31)
			0,59 (0,55-0,62)
Cirurgião + Residente			
A	0,95 (0,93-0,97)	A1	0,77 (0,74-0,79)
B	0,42 (0,39-0,44)	A2	0,86 (0,84-0,88)
C	0,72 (0,70-0,74)	B1	0,38 (0,36-0,40)
D	0,16 (0,14-0,19)	B2	0,28 (0,26-0,30)
Geral	0,73 (0,71-0,74)	C1	0,52 (0,50-0,54)
		C2	0,42 (0,39-0,44)
		D	0,16 (0,14-0,19)
			0,57 (0,56-0,58)

concordância foi muito ruim. A concordância geral foi $k=0,56$ (0,54-0,58), considerada moderada.

Dentre os residentes, os tipos A1 ($k=0,82$ (0,76-0,89)) e A2 ($k=0,90$ (0,83-0,96)) também apresentam melhor concordância que os demais, foram consideradas excelentes. Os grupos B1 ($k=0,46$ (0,39-0,52)), B2 ($k=0,49$ (0,43-0,56)) e C1 ($k=0,49$ (0,43-0,56)) tiveram concordâncias moderadas, enquanto C2 ($k=0,37$ (0,30-0,43)) e D ($k=0,24$ (0,18-0,31)) tiveram a concordância ruim. Porém, seguindo o intervalo de confiança, podemos considerar que os grupos B1 e B2 e C1 e C2 têm a mesma concordância. A concordância geral foi $k=0,59$ (0,55-0,62), considerada moderada.

Quando analisados em conjunto os cirurgiões de mão e residentes, os resultados foram semelhantes. Em A1 ($k=0,77$ (0,74-0,79)) e A2 ($k=0,86$ (0,84-0,88)) a concordância foi considerada boa e excelente, respectivamente. Já em B1 ($k=0,38$ (0,36-0,40)) e B2 ($k=0,28$ (0,26-0,30)) foram consideradas ruins. Os tipos C1 ($k=0,52$ (0,50-0,54)) e C2 ($k=0,42$ (0,39-0,44)) tiveram concordância moderada e o tipo D $k=0,16$ (0,14-0,19) teve concordância muito ruim. A concordância geral foi $k=0,57$ (0,56-0,58), considerada moderada.

Quando agrupados pela classificação em grandes grupos (A, B, C, D), houve uma melhoria em todos os grupos, a mais expressiva foi no grupo C. Em todas as análises, se considerarmos o intervalo de confiança de 95%, com a união dos subtipos 1 e 2 a concordância melhora.

Dentre os cirurgiões, a concordância em A ($k=0,95$ (0,91-0,99)) foi excelente, em B foi ruim ($k=0,34$ (0,31-0,38)), em C

($k=0,71$ (0,67-0,75)) foi boa e em D manteve-se muito ruim. A concordância geral foi boa, com $k=0,72$ (0,69-0,74).

Dentre os residentes, a concordância em A ($k=0,96$ (0,89-1,00)) foi excelente, em B ($k=0,55$ (0,48-0,61)) foi moderada, em C ($k=0,77$ (0,70-0,83)) foi boa e em D manteve-se ruim. A concordância geral foi boa, com $k=0,76$ (0,72-0,81).

Quando agrupados cirurgiões de mão e residentes, a concordância em A ($k=0,95$ (0,93-0,97)) foi excelente, em B ($k=0,42$ (0,39-0,44)) foi moderada, em C ($k=0,72$ (0,70-0,74)) foi boa e em D manteve-se ruim. A concordância geral nessa segunda avaliação foi boa, com $k=0,73$ (0,71-0,74). Os dados estão diagramados nos gráficos a seguir, mostram os respectivos intervalos de confiança (figs. 1 e 2).

Discussão

O dedo em martelo é uma afecção muito prevalente, acomete principalmente uma população em idade economicamente ativa. Uma classificação eficiente e reprodutível guia melhor um tratamento mais eficiente.

Dentre as classificações previamente descritas para esse tipo de deformidade, a de Pratt *et al.*² apenas divide a etiologia e não guia tratamento ou prognóstico. Da mesma forma, a classificação de Wehbe e Schneider,³ que usa a extensão do comprometimento da superfície articular. Ela separa cada um dos três tipos em três subtipos (A, menos de 1/3 da superfície articular acometida; B, entre 1/3 e 2/3 da superfície articular; e

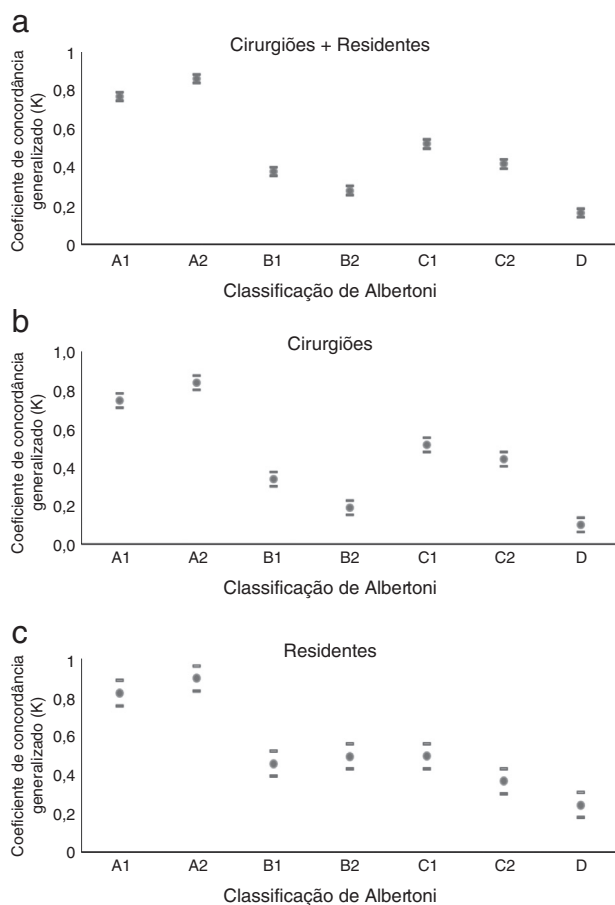


Figura 1 – Coeficiente de concordância generalizado para cada subgrupo. Os pontos representam o valor do coeficiente e os traços delimitam o intervalo de confiança 95%. (a) Cirurgiões e residentes, (b) Cirurgiões de mão, (c) Médicos residentes em cirurgia da mão.

C mais de 2/3 da superfície acometida). Porém, isso não dita o tratamento ou o prognóstico. Na classificação de Doyle, existe uma divisão por parâmetros clínicos não contemplados nos outros sistemas, mas não há categorização dos padrões radiográficos para todos os tipos.⁴ Essa última é a mais usada na literatura mundialmente.¹

Comparada com outras classificações, a de Albertoni define a programação terapêutica, que muda de acordo com a lesão em questão. Cada tipo tem um tratamento específico. Lesões do tipo A1 e B1 são tratadas classicamente com imobilização. As lesões C1 e D são tratadas com redução de forma incruenta e com imobilização com tala metálica. As lesões do tipo A2, B2 e C2 normalmente requerem tratamento cirúrgico.^{5,9} Essa classificação é muito usada entre os ortopedistas brasileiros, daí a importância da avaliação de sua reprodutibilidade.

Na literatura não foi encontrado estudo sobre a reprodutibilidade da classificação de Albertoni, nem mesmo qualquer estudo de reprodutibilidade sobre classificações para dedo em martelo. Isso demonstra a originalidade do nosso trabalho.

A classificação de Albertoni é baseada em radiografias. Na literatura vários estudos de concordância interobservador avaliam classificações radiográficas. Audigé et al.¹⁴ avaliaram

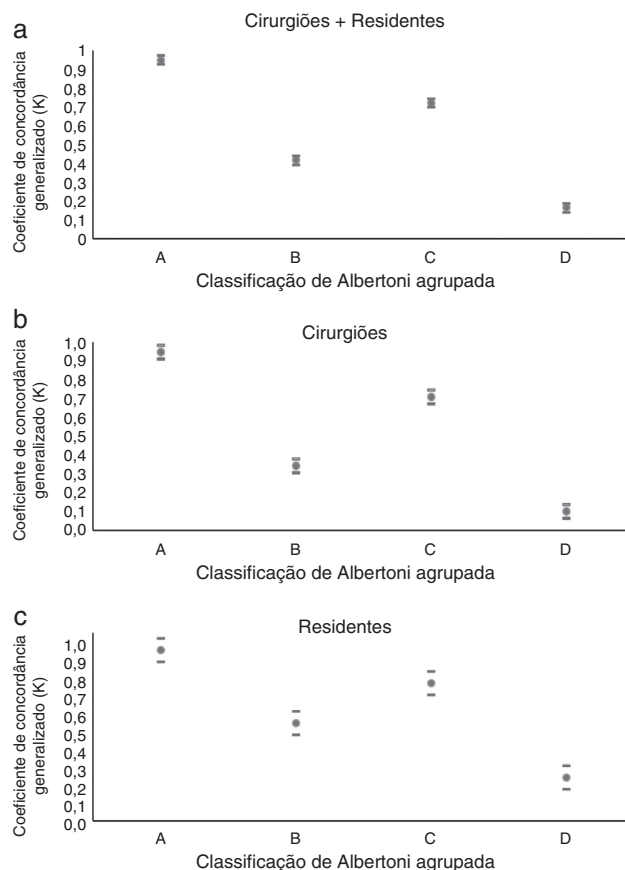


Figura 2 – Coeficiente de concordância generalizado para dados agrupados. Os pontos representam o valor do coeficiente e os traços delimitam o intervalo de confiança 95%. (a) Cirurgiões e residentes, (b) Cirurgiões de mão, (c) Médicos residentes em cirurgia da mão.

44 estudos de reprodutibilidade sobre classificações em ortopedia com o uso de critérios de imagem e encontraram pouca uniformidade na metodologia, o que dificulta a comparação da reprodutibilidade. Belloti et al.⁶ avaliaram a reprodutibilidade de classificações de fraturas da extremidade distal do rádio, enquanto Utino et al.¹⁵ estudaram a concordância para a classificação AO para ossos longos na população pediátrica.

O número de radiografias a serem avaliadas é um fator importante para avaliação da concordância. Tanto um número pequeno demais quanto um número grande demais de avaliações tendem a piorar a concordância.¹⁶ Na revisão sistemática de Audigé sobre estudos de reprodutibilidade em ortopedia, foi observada uma grande variação na quantidade de radiografias por estudo, de 14 até 200 avaliações.¹⁴ Berger et al.,¹⁷ numa revisão sistemática sobre a reprodutibilidade da classificação de Eaton para rizartrose, avaliaram quatro trabalhos. Nesses trabalhos a variação da quantidade de radiografias foi de 40 a 43. Baseados nesses trabalhos, consideramos que as 43 radiografias usadas em nosso trabalho foram suficientes para avaliação da concordância interobservador na classificação de Albertoni.

A quantidade de observadores é outro fator que interfere nos coeficientes de concordância. Quanto maior a quantidade

desse, menor a probabilidade de concordarem. Na literatura, também não encontramos uniformidade no número de observadores. Thomsen et al.¹⁸ usaram apenas quatro observadores, enquanto Randsborg et al.⁷ trabalharam com 12 observadores e Audigé et al.¹⁴ analisaram trabalhos com a quantidade de observadores de 2 a 36 (mediana = 5). Em nosso trabalho, usamos 19 observadores e temos dois subgrupos, de residentes e cirurgiões de mão. A quantidade de radiografias associada à quantidade de observadores neste estudo está, portanto, dentro dos padrões encontrados na literatura.^{6,7,14,15,17-24}

A concordância piora quanto maior o número de categorias em uma classificação.^{6,7} A classificação de Albertoni, por ter sete possíveis opções, teria a tendência a ter uma pior concordância, comparada com outras classificações. Isso é comprovado quando agrupamos os tipos A1 e A2 em A; B1 e B2 em B; C1 e C2 em C e há um aumento da concordância em todas as análises, uma vez que diminuimos para quatro categorias.

Não há consenso na literatura acerca de quanto deve ser considerado um valor de k para considerar uma classificação como reprodutível ou não.^{14,16} Esses valores são definidos arbitrariamente pelos autores.¹⁶ Fleiss e Joseph¹⁰ consideraram como moderada a boa concordância resultados de k entre 0,40 e 0,75. Svanholm et al.²⁵ consideraram como bom apenas valores de k maiores do que 0,75. Brage et al.²² consideraram como reprodutíveis valores de k acima de 0,50. Landis e Koch,¹³ padrão usado em nosso trabalho (tabela 1) e o mais usado nos dias de hoje,¹⁴ consideraram moderada no intervalo de 0,4 a 0,6 e como boa concordância acima de 0,6.

Os tipos A1 e A2 de Albertoni avaliados como um único grupo tiveram o coeficiente de concordância $k=0,95$ (0,93-0,97), o que indica uma excelente concordância. Separados em A1 ($k=0,77$ (0,74-0,79)) e A2 ($k=0,84$ (0,80-0,87)), o coeficiente de concordância diminuiu pouco, mas se manteve bom, com queda de 16% no coeficiente k . Ou seja, com a união das categorias e supressão do parâmetro que as diferencia, a concordância altera pouco. Isso mostra que a mensuração da angulação de < 30 graus (A1) ou > 30 graus (A2) pode ser considerado um parâmetro reprodutível.

Os tipos B1 e B2 de Albertoni avaliados como um único grupo tiveram o coeficiente de concordância moderado de $k=0,42$ (0,39-0,44). Separados em B1 ($k=0,38$ (0,36-0,40)) e B2 ($k=0,28$ (0,26-0,30)), o coeficiente de concordância diminuiu pouco com queda de 21% no coeficiente k e se tornou ruim. Analogamente ao encontrado no tipo A, não há dificuldade na diferenciação entre B1 e B2, assim como entre A1 e A2, o que corrobora o fato de o parâmetro da angulação ser reprodutível. Quanto à concordância ruim, acreditamos que o motivo principal desse resultado seja a baixa prevalência do tipo B no questionário aplicado (14,6%) de acordo com a tabela 4.

Os tipos C1 e C2 de Albertoni avaliados como um único grupo tiveram o coeficiente de concordância bom, com $k=0,77$ (0,70-0,83). Separados em C1 ($k=0,52$ (0,50-0,54)) e C2 ($k=0,42$ (0,39-0,44)), o coeficiente de concordância diminuiu muito, com queda de 41% do coeficiente k , que se tornou moderado para C1 e ruim para C2. Isso nos faz interpretar que existe uma dificuldade dos observadores de definir a congruência articular da IFD. Ou seja, em relação à congruência articular, a concordância cai consideravelmente. Acreditamos que uma das formas de melhorar a concordância no tipo C seria definir melhor o

critério de avaliação da congruência articular. Isso pode guiar modificações futuras na classificação.

Outro fator relevante é que, das 43 radiografias, em 22 houve dúvida na escolha entre B e C (tabela 2). Já no tipo A (A1 ou A2), quando escolhido, sempre era concordante, com exceção das radiografias 4 e 38 (tabela 2). Isso evidenciou a menor concordância nos tipos B e C em relação ao tipo A. Nossa interpretação é que o parâmetro para separar avulsão óssea (tipo B) de fratura da região dorsal da base da falange distal (tipo C) não é bem entendido pelos avaliadores.

A classificação do tipo D não foi separada em D1 e D2 devido à pequena incidência desse tipo de lesão. Houve uma concordância muito ruim, com $k=0,16$ (0,14-0,19). Houve uma baixa prevalência desse tipo (1,8%), foi classificado em apenas quatro casos, o que também justifica a baixa concordância.²³

A experiência dos avaliadores na afecção avaliada tem a tendência a alterar a concordância.⁷ Mattos et al.²⁴ concluíram que a baixa experiência dos observadores diminuiu a concordância. Porém, em nosso estudo, entre os residentes, a concordância geral foi de $k=0,76$ (0,72-0,81), enquanto entre os cirurgiões de mão foi de $k=0,72$ (0,69-0,74). Apesar de o grupo de residentes mostrar valores de k maiores do que o grupo de cirurgiões de mão, considerando o intervalo de confiança de 95% (figs. 1 e 2), não podemos inferir que são mais concordantes do que os cirurgiões. Isso mostra uma vantagem na classificação de Albertoni, que não altera concordância com a experiência dos observadores. Acreditamos que há uma tendência a maior concordância entre os residentes, pelo fato de terem níveis de conhecimento e experiências parecidos, e por estarem em treinamento no mesmo centro, o que leva à uniformidade. Entretanto, isso não foi comprovado em nosso trabalho.

Audigé cita que na avaliação de 44 trabalhos sobre reprodutibilidade de classificações em ortopedia, dos 86 coeficientes de concordância calculados apenas quatro foram excelentes ($k > 0,80$), 17 foram boas (entre 0,60-0,80) e 32 foram moderadas (entre 0,40-0,60) e 33 foram razoáveis ou fracas ($< 0,40$).¹⁴ A concordância geral da classificação de Albertoni foi moderada, com $k=0,57$ (0,56-0,58), baseada na classificação de Landis e Koch.¹³ Apesar de moderada, quando comparada com a literatura, e se levarmos em conta todos os fatores discutidos, consideramos a classificação de Albertoni como reprodutível.

Há de se considerar ainda um viés de seleção, dado que as radiografias não são randomizadas, e foram escolhidas pelos pesquisadores pela qualidade do exame. Outra característica relevante é que não podemos garantir que as medições foram feitas por métodos corretos, apesar de orientados corretamente e fornecido material correto para a medição. Este foi um estudo de apenas um centro, o que tende a uniformizar respostas e melhorar a concordância.

Conclusão

A classificação de Albertoni tem concordância interobservador boa ou excelente para os tipos A1 e A2, moderada para tipos C1 e C2 e ruim para tipos B1, B2 e D. Consideramos a classificação de Albertoni como reprodutível, segue os métodos estatísticos empregados, e comparada com a pesquisa na literatura. Acreditamos que uma melhor definição dos

critérios de congruência articular aperfeiçoaria substancialmente a concordância.

Conflito de interesse

Os autores declaram não haver conflitos de interesse.

REFERÊNCIAS

1. Alla SR, Deal ND, Dempsey JJ. Current concepts: mallet finger. *Hand*. 2014;9(2):138-44.
2. Pratt DR, Bunnell S, Howard LD. Mallet finger: classification and methods of treatment. *Am J Surg*. 1957;93(4):573-9.
3. Wehbé MA, Schneider LH. Mallet fractures. *J Bone Joint Surg Am*. 1984;66(5):658-69.
4. Doyle JR, Green DP, Hotchkiss RN, Pederson WC. Extensor tendons: acute injuries. In: Green DP, Hotchkiss RN, Pederson WC, editors. *Green's operative hand surgery*. 4th ed. New York: Churchill Livingstone; 1999. p. 195-8.
5. Albertoni WM. Estudo crítico de tratamento do dedo em martelo. Análise de 200 caso [tese]. São Paulo: Universidade Federal de São Paulo, Escola Paulista de Medicina; 1986.
6. Belloti JC, Tamaoki MJ, Franciozi CE, Santos JB, Balbachevsky D, Chap EC, et al. Are distal radius fracture classifications reproducible? Intra and interobserver agreement. *Sao Paulo Med J*. 2008;126(3):180-5.
7. Randsborg PH, Sivertsen EA. Classification of distal radius fractures in children: good inter- and intraobserver reliability, which improves with clinical experience. *BMC Musculoskelet Disord*. 2012;13:6.
8. Albertoni WM. Mallet finger: classification. *Rev Hosp Sao Paulo Esc Paul Med*. 1989;1(3):133-6.
9. Albertoni WM. The Brooks-Graner procedure for correction of mallet finger. *The Hand*. 1988;3:97-100.
10. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76(5):378-82.
11. Scott WA. Reliability of content analysis: the case of nominal scale coding. *Public Opin Q*. 1955;19(3):321-5.
12. Fleiss JL, Nee JC, Landis JR. Large sample variance of kappa in the case of different sets of raters. *Psycho Bull*. 1979;86(5):974-7.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.
14. Audigé L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. *Acta Orthop*. 2004;75(2):184-94.
15. Utino AY, de Alencar DR, Maringolo LF, Negrão JM, Blumetti FC, Dobashi ET. Concordância intra e interobservadores do sistema de classificação AO para fraturas dos ossos longos na população pediátrica. *Rev Bras Ortop*. 2015;50(5):501-8.
16. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257-68.
17. Berger AJ, Momeni A, Ladd AL. Intra and interobserver reliability of the Eaton classification for trapeziometacarpal arthritis: a systematic review. *Clin Orthop Relat Res*. 2014;472(4):1155-9.
18. Thomsen NO, Overgaard S, Olsen LH, Hansen H, Nielsen ST. Observer variation in the radiographic classification of ankle fractures. *Bone Joint J*. 1991;73(4):676-8.
19. Van Embden D, Rhemrev SJ, Genelin F, Meylaerts SAG, Roukema GR. The reliability of a simplified Garden classification for intracapsular hip fractures. *Orthop Traumatol Surg Res*. 2012;98(4):405-8.
20. Valderrama-Molina CO, Estrada-Castrillón M, Hincapie JA, Lugo-Agudelo LH. Intra and interobserver agreement on the Oestern and Tscherné classification of soft tissue injury in periarticular lower-limb closed fractures. *Colomb Med (Cali)*. 2014;45(4):173-8.
21. Kim JK, Kim DJ. The risk factors associated with subluxation of the distal interphalangeal joint in mallet fracture. *J Hand Surg Eur Vol*. 2015;40(1):63-7.
22. Brage ME, Rockett M, Vraney R, Anderson R, Toledano A. Ankle fracture classification: a comparison of reliability of three X-ray views versus two. *Foot Ankle Int*. 1998;19(8):555-62.
23. Liggieri AC, Tamanaha MJ, Abechain JJK, Ikeda TM, Dobashi ET. Concordância intra e interobservadores das diferentes classificações usadas na doença de Legg-Calvé-Perthes. *Rev Bras Ortop*. 2015;50(6):680-5.
24. Mattos CA, Jesus AAK, dos Santos Floter M, Nunes LFB, de Baptista Sanches B, Zabeu JLA. Reprodutibilidade das classificações de Tronzo e AO para fraturas transtrocanterianas. *Rev Bras Ortop*. 2015;50(5):495-500.
25. Svanholm H, Starklint H, Gundersen HJ, Fabricius J, Barlebo H, Olsen S. Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic. *APMIS*. 1989;97(8):689-98.