Avaliação da reprodutibilidade e concordância da classificação IDEAL para fraturas de extremidade distal de rádio

Assessment of Reproducibility and Agreement of the IDEAL Classification for Distal Radius Fractures

João Victor da Rocha Lima[1] Lucas Araújo Silva[1] Antonio Guilherme Chagas Silva Feitosa[1] Rafael Lima Medeiros[1] Luis Fernando Martins Carvalho[1] Bruno Wilson da Silva Moura[1]

[1] Unidade de Traumato-Ortopedia, Hospital Universitário da Universidade Federal do Piauí (HUUFPI), Teresina, PI, Brasil

Rev Bras Ortop 2024;59(6):e901–e906.

Endereço para correspondência João Victor da Rocha Lima, M.D., Orthopedic Surgeon HUUHPI – Unidade de Traumato-OrtopediaCampus Universitário Ministro Petrônio Portela, s/n - Ininga, Teresina - PI, 64049-550, Brasil (e-mail: joaovictorce@gmail.com).

Resumo

Objetivo Analisar a reprodutibilidade e concordância intra- e interobservadores da classificação IDEAL para fraturas de extremidade distal do rádio.

Métodos Estudo qualitativo, analítico, realizado pela avaliação de 50 pares de radiografias em duas incidências de pacientes com fratura da extremidade distal do rádio. As radiografias foram avaliadas por dez observadores de diferentes níveis de treinamento em ortopedia em três avaliações distintas. Os resultados foram submetidos aos testes Kappa de Cohen e Fleiss para obtenção dos níveis de concordância intra- e interobservadores. Foram utilizados os programas Excel e SPSS, versão 26.0, para cálculos estatísticos.

Resultados Os valores do índice Kappa de Cohen para avaliação intraobservadores indicaram pobre a pouca concordância (-0,177–0,259), com significância estatística em apenas uma instância. Os valores do índice Kappa de Fleiss revelaram pouca concordância entre o grupo de residentes (0,277–0,383) com significância estatística; concordância pobre a pouca entre o grupo de ortopedistas gerais (0,114–0,225) com significância estatística na maioria das instâncias; e concordância moderada entre o grupo de cirurgiões de mão (0,449–0,533) com significância estatística.

Conclusão A classificação IDEAL obteve níveis de concordância interobservadores variando entre pobre e moderado, com influência do nível de treinamento dos profissionais. Os outros níveis de concordância intraobservadores variaram de pobre a pouca concordância, mas sem significância estatística.

Palavras-chave

- classificação
- fraturas do rádio
- reprodutibilidade dos testes

Trabalho desenvolvido na Unidade de Traumato-Ortopedia, Hospital Universitário da Universidade Federal do Piauí (HUUFPI), Teresina, PI, Brasil.

recebido
22 de fevereiro de 2024
aceito
05 de setembro de 2024

DOI https://doi.org/10.1055/s-0044-1792121.
ISSN 0102-3616.

© 2024. The Author(s).
This is an open access article published by Thieme under the terms of the Creative Commons Attribution 4.0 International License, permitting copying and reproduction so long as the original work is given appropriate credit (https://creativecommons.org/licenses/by/4.0/).

Thieme Revinter Publicações Ltda., Rua do Matoso 170, Rio de Janeiro, RJ, CEP 20270-135, Brazil

Abstract

Objective To analyze the reproducibility and intra- and interobserver agreement of the IDEAL classification for distal radius fractures.

Methods This qualitative, analytical study evaluated 50 pairs of radiographs in two views of patients with distal radius fractures. There were ten observers with different levels of orthopedic training who assessed the radiographs in three different evaluations. The results underwent the Cohen and Fleiss Kappa tests to determine intra- and interobserver agreement levels. Statistical calculations used Excel and SPSS, version 26.0.

Results The Cohen Kappa index values for intraobserver evaluation indicated poor to little agreement (-0.177–0.259), with statistical significance in only one instance. The Fleiss Kappa index values revealed little agreement among the resident group (0.277–0.383) with statistical significance, poor to little agreement among the general orthopedists (0.114–0.225) with statistical significance in most instances, and moderate agreement among hand surgeons (0.449–0.533) with statistical significance.

Conclusion The IDEAL classification had interobserver agreement levels ranging from poor to moderate, influenced by the physicians' training level. Other intraobserver agreement levels ranged from poor to little agreement but without statistical significance.

Keywords

- classification
- radius fractures
- reproducibility of results

Introdução

As fraturas da extremidade distal do rádio são extremamente prevalentes, representando 16% das fraturas do corpo e 74% do antebraço, com distribuição bimodal, acometendo adolescentes/adultos jovens (traumas de alta energia) e idosos (traumas de baixa energia). O mecanismo de trauma mais comum é a queda ao solo com punho em hiperextensão.[1–4]

Apesar da prevalência elevada, nunca houve bastante consenso na literatura em relação a uma classificação de escolha para fraturas de extremidade distal do rádio. Os primeiros conceitos se iniciaram antes do advento da radiografia, com a descrição dada por Colles para a fratura com desvio dorsal em 1814. Em 1951, Gartland e Werley propuseram a primeira classificação para fraturas da extremidade distal de rádio, seguidos por Frykman em 1967, pelo grupo AO proposta por Muller em 1986, a de Fernandez em 1991, a Universal proposta por Cooney em 1993, chegando à classificação IDEAL, proposta pela Divisão de Cirurgia da Mão da Universidade Federal de São Paulo (UNIFESP), em 2013.[1,3,5]

A classificação IDEAL se baseia em cinco parâmetros (dois epidemiológicos e três radiográficos), sendo eles: idade (maior ou menor que 60 anos), energia do trauma que ocasionou a fratura, desvio dos fragmentos (presença ou não), incongruência articular (incongruência ou afastamento >2 milímetros) e lesões associadas (presença ou ausência). Cada um dos parâmetros pontua zero ou um e o somatório deles dá a classificação da fratura, sendo dos tipos I (0–1 pontos), II (2–3) ou III (4–5). Cada tipo na descrição sugere um tratamento e o prognóstico da lesão.[1]

Estudos anteriores evidenciam baixos a moderados níveis de concordância intra e interobservadores para as classificações mais antigas disponíveis na literatura, como as classificações de Frykman, Fernandez e AO. As classificações Universal e IDEAL apresentaram resultados melhores em comparação às demais.[1,2] Verificou-se que aquelas com mais subtipos e divisões apresentaram menor concordância interobservadores, podendo suscitar problemas em relação à concordância intraobservador pelo maior tempo necessário para se habituar ao uso do instrumento.[1,5–12]

Em razão da existência de várias classificações para as fraturas de extremidade distal do rádio, se torna importante definir a melhor, através de estudos como este, que avaliam sua reprodutibilidade e confiabilidade. Este estudo tem por objetivo analisar a reprodutibilidade e concordância intra- e interobservadores da classificação IDEAL para fraturas de extremidade distal do rádio, assim como verificar a influência do nível de treinamento dos observadores.

Materiais e Métodos

Trata-se de um estudo qualitativo, analítico, retrospectivo e de documentação direta, realizado a partir da avaliação de radiografias de pacientes acometidos por fratura de extremidade distal de rádio por observadores de diferentes níveis de experiência em traumatologia. A pesquisa foi realizada nas dependências de um Hospital Universitário, onde foram obtidas as radiografias para realização do estudo e conduzidas as entrevistas para obtenção dos dados, nos meses de novembro e dezembro de 2022.

De acordo com o método de Giraudeau e Mary[13] para determinação da amostra de acordo com o nível de concordância esperado, a quantidade de avaliadores e o intervalo de confiança (IC), diversas quantidades mínimas para a amostra são esperadas. Conforme demonstrado na ►**Tabela 1**, para

Tabela 1 Estimativa do tamanho da amostra com uso do coeficiente de correlação intraclasse baseado em Giraudeau e Mary

		Número de participantes requeridos para um IC 95% em três níveis de confiança		
Número de avaliadores	ICC esperado	± 0.05	± 0.10	± 0.15
2	0,9	56	14	4
	0,8	200	50	13
	0,7	400	100	25
	0,6	630	158	40
	0,5	865	217	55
4	0,9	36	9	3
	0,8	119	30	8
	0,7	222	56	14
	0,6	322	81	21
	0,5	401	101	26
6	0,9	31	8	2
	0,8	103	26	7
	0,7	187	47	12
	0,6	263	66	17
	0,5	314	79	20
10+	0,9	29	8	2
	0,8	92	23	6
	0,7	164	41	11
	0,6	224	56	14
	0,5	259	65	17

Abreviações: IC, intervalo de confiança; ICC, coeficiente de correlação intraclasse.
Nota: Adaptado de Karanicolas et al.[14]

um Kappa esperado de 0.70 e um intervalo de confiança de 90%, um mínimo de 41 amostras são necessárias.[14] Foram obtidos 50 pares de radiografias (incidências anteroposterior e lateral) que evidenciem fraturas de extremidade distal de rádio dentre os prontuários eletrônicos dos pacientes atendidos neste Hospital Universitário entre os anos de 2019 e 2022.

Os critérios de inclusão foram: pacientes cujos prontuários contenham o CID de fratura de extremidade distal do rádio (S52.5); pacientes que tenham sido atendidos no Hospital Universitário. Os critérios de exclusão foram: pacientes que tenham sido submetidos a algum tipo de tratamento, cirúrgico ou não, para a fratura de extremidade distal do rádio antes da obtenção da radiografia; pacientes que não apresentem em seus prontuários a disponibilidade do exame de imagem da fratura de extremidade distal do rádio.

Foram convidados a participar da pesquisa como observadores três ortopedistas especialistas em cirurgia da mão, quatro ortopedistas gerais pertencentes ao serviço de ortopedia do Hospital Universitário e três residentes de ortopedia, um de cada ano, do Hospital Universitário. Avaliaram as radiografias para classificá-las de acordo com a classificação IDEAL. A avaliação se deu em três momentos diferentes, separados por 15,3 ± 4,34 dias.

Os resultados obtidos com as avaliações dos observadores foram tabulados utilizando o Microsoft Excel 2019 (Microsoft Corp., Redmond, WA, EUA) e submetidos aos testes Kappa de Cohen e Fleiss para avaliação intra- e interobservadores, respectivamente, através do software Statistical Package Social Sciences (SPSS, IBM Corp., Armonk, NY, USA), versão 26.0, para análises estatísticas.[15] Obtiveram-se as tabelas de avaliação de concordância interobservadores evidenciando a medida do índice de Kappa para cada classe de avaliadores (residentes, ortopedistas gerais e cirurgiões de mão) para cada uma das três avaliações. As tabelas de avaliação de concordância intraobservador, comparando cada uma das avaliações com as outras duas, bem como a presença dos valores limites superior e inferior para um IC de 90%.

Foram considerados significativos os valores de Kappa com $p < 0,1$. A interpretação dos resultados seguindo o entendimento proposto por Landis e Koch, no qual valores menores que e até zero indicavam concordância pobre, pouca entre 0 e 0,2, razoável entre 0,2 e 0,4, moderada entre 0,4 e 0,6, substancial entre 0,6 e 0,8, e excelente ou quase perfeita entre 0,8 e 1.[16]

Esta pesquisa foi autorizada pelo Comitê de Ética em Pesquisa sob o número CAAE: 63490322.8.0000.8050, parecer de n° 5.726.415.

Resultados

Os valores dos índices Kappa de Cohen obtidos para a concordância intraobservadores (►**Tabela 2**) foram de 0,259 (pouca concordância) em uma instância (CM1 T1 x T2), com significância estatística ($p = 0.021$), e valores de 0,140 ou menos (concordância pobre) em todas as outras, sem significância estatística em nenhuma delas ($p > 0,1$).

Na ►**Tabela 3**, observa-se índices de Kappa de Fleiss para concordância interobservadores do grupo de Residentes variando entre 0,277 e 0,383 nas três avaliações, com significância estatística. Os ICs não contém o valor-parâmetro e $p \leq 0,008$.

A ►**Tabela 4** traz índices de Kappa de Fleiss do grupo de Ortopedistas Gerais com valores entre 0,114 e 0,225, com significância estatística. Os ICs não contém o valor-parâmetro e $p \leq 0,008$.

A ►**Tabela 5** traz índices de Kappa de Fleiss do grupo de Cirurgiões de Mão apresentando valores entre 0,449 e 0,553, todos com significância estatística. Os resultados observados para cada tipo da classificação apresentam valores maiores para o tipo III em todas as avaliações em comparação com os tipos I e II, mas todas as avaliações apresentaram excelentes níveis de significância. Os ICs não contém o valor-parâmetro e $p < 0,001$.

Discussão

As fraturas da extremidade distal do rádio, sendo um diagnóstico tão prevalente, necessitam de uma total compreensão de toda a complexidade dos padrões de fratura que podem surgir e ampliação de seu escopo para outros fatores que implicam sobre o prognóstico da patologia.[10,12] A clas-

Tabela 2 Índices de Kappa de Cohen e valores-p dos mesmos nos testes de concordância intraobservadores

	T1 x T2		T2 x T3		T1 x T3	
	κ	*p*	κ	*p*	κ	*p*
R1	0,091	0,384	0,028	0,786	0,049	0,660
R2	0,140	0,174	-0,078	0,419	-0,026	0,805
R3	-0,151	0,139	-0,043	0,673	0,015	0,885
OG1	-0,022	0,838	-0,053	0,624	-0,017	0,872
OG2	0,009	0,940	-0,177	0,159	0,006	0,963
OG3	-0,121	0,255	0,054	0,646	-0,069	0,570
OG4	0,108	0,352	-0,032	0,779	-0,029	0,797
CM1	0,259	0,021	-0,078	0,463	-0,009	0,933
CM2	0,138	0,178	0,028	0,791	0,042	0,683
CM3	0,028	0,791	-0,053	0,646	0,006	0,956

Abreviações: CM, cirurgião de mão; OG, ortopedista geral; R, residente de ortopedia e traumatologia.

Tabela 3 Índices de Kappa de Fleiss para concordância interobservadores do grupo de Residentes em Ortopedia e Traumatologia de cada avaliação e para cada tipo descrito na classificação IDEAL

	Índice κ	*p*	Limite inferior IC 90%	Limite superior IC 90%
κ de Fleiss T1	0,305	< 0,001	0,206	0,404
Tipo 1	0,349	< 0,001	0,214	0,483
Tipo 2	0,215	0,008	0,081	0,350
Tipo 3	0,386	< 0,001	0,252	0,521
κ de Fleiss T2	0,383	< 0,001	0,282	0,483
Tipo 1	0,435	< 0,001	0,301	0,569
Tipo 2	0,302	< 0,001	0,167	0,436
Tipo 3	0,452	< 0,001	0,317	0,586
κ de Fleiss T3	0,277	< 0,001	0,175	0,378
Tipo 1	0,308	< 0,001	0,173	0,442
Tipo 2	0,250	0,002	0,116	0,384
Tipo 3	0,292	< 0,001	0,157	0,426

Abreviação: IC, intervalo de confiança.

Tabela 4 Índices de Kappa de Fleiss interobservadores do grupo de Ortopedistas de cada avaliação e para cada tipo descrito na classificação IDEAL

	Índice κ	*p*	Limite inferior IC 90%	Limite superior IC 90%
κ de Fleiss T1	0,186	< 0,001	0,115	0,258
Tipo 1	0,472	< 0,001	0,377	0,567
Tipo 2	0,112	0,053	0,017	0,207
Tipo 3	0,065	0,261	-0,030	0,160
κ de Fleiss T2	0,114	0,008	0,043	0,184
Tipo 1	0,330	< 0,001	0,235	0,425
Tipo 2	0,011	0,849	-0,084	0,106
Tipo 3	0,090	0,119	-0,005	0,185
κ de Fleiss T3	0,225	< 0,001	0,154	0,295
Tipo 1	0,359	< 0,001	0,335	0,454
Tipo 2	0,148	< 0,001	0,361	0,243
Tipo 3	0,223	< 0,001	0,485	0,318

Abreviação: IC, intervalo de confiança.

Tabela 5 Índices de Kappa de Fleiss interobservadores do grupo de Cirurgiões da Mão de cada avaliação e para cada tipo descrito na classificação IDEAL

	Índice κ	*p*	Limite inferior IC 90%	Limite superior IC 90%
κ de Fleiss T1	0,533	< 0,001	0,430	0,637
Tipo 1	0,469	< 0,001	0,335	0,604
Tipo 2	0,495	< 0,001	0,361	0,629
Tipo 3	0,620	< 0,001	0,485	0,754
κ de Fleiss T2	0,449	< 0,001	0,347	0,550
Tipo 1	0,365	< 0,001	0,231	0,500
Tipo 2	0,430	< 0,001	0,296	0,564
Tipo 3	0,525	< 0,001	0,391	0,659
κ de Fleiss T3	0,531	< 0,001	0,430	0,631
Tipo 1	0,627	< 0,001	0,493	0,761
Tipo 2	0,470	< 0,001	0,336	0,604
Tipo 3	0,542	< 0,001	0,407	0,676

Abreviação: IC, intervalo de confiança.

sificação IDEAL cumpre este aspecto, pois envolve idade e energia do trauma em seus parâmetros.

Entre as limitações deste estudo, pode-se citar o baixo número de avaliadores para cada categoria, além da ausência da categoria residentes em cirurgia da mão.

Observou-se primariamente uma tendência à resultados com pouca ou nenhuma concordância, ou até mesmo discordância na avaliação intraobservadores, com a maioria das instâncias não apresentando significância estatística considerável, achado que difere do que foi encontrado na literatura, com a maioria dos estudos obtendo resultados com concordâncias de moderada a alta.[1,3–5]

O grau de concordância interobservadores medido pelo índice Kappa de Fleiss mostrou um resultado com mais solidez estatística do que o de Cohen na análise de concordância intraobservadores. Pode-se verificar uma dificuldade dos observadores em concordar no tipo intermediário da classificação, sendo mais fácil nos extremos. Dada a baixa ou até ausente concordância entre os ortopedistas em relação aos residentes, é possível inferir que a diferença no nível de treinamento aparentemente não implicou em um maior nível de concordância para o grupo com mais experiência em ortopedia, observando-se até mesmo uma maior concordância entre os residentes. Nos estudos de Andersen et al.[17] e Belloti et al.[18] não houve influência do nível de experiência do observador sobre os resultados obtidos, o que corrobora com os achados deste estudo, visto que os observadores menos experientes obtiveram resultados com maiores níveis de concordância e melhores níveis de significância estatística quando comparados aos mais experientes.

O grupo dos cirurgiões da mão, por outro lado, obteve os melhores níveis de concordância interobservadores entre os três grupos avaliados, com níveis moderados em todas as três instâncias da pesquisa. Com os resultados obtidos pelo grupo dos cirurgiões de mão, infere-se que o treinamento adicional específico possibilitou ao grupo obter resultados mais concordantes entre si quando comparados com os demais. Neste cenário, diferindo do que foi proposto por Illarramendi et al.,[8] Andersen et al.,[17] Jayakumar et al.[3] e Belloti et al.,[18] a experiência adicional do grupo de cirurgiões de mão foi o fator preponderante para a obtenção dos melhores níveis de concordância interobservadores deste estudo.

O objetivo das classificações, de forma geral, é fornecer uma ferramenta que consiga enquadrar a fratura em um tipo ou classe de forma acurada e guiar o tratamento dessa fratura enquanto define prognóstico para a mesma, e também possibilite a comunicação efetiva entre profissionais das mais diferentes origens.[4] Apesar de não se verificar níveis elevados ($> 0,8$) de concordância intra- e interobservadores para as classificações neste ou nos demais estudos presentes na literatura,[1,3–8] verifica-se que há a possibilidade de cumprir com o objetivo supracitado.

Conclusão

O presente estudo encontrou níveis de concordância interobservadores variando entre pobre e moderado, além de demonstrar que o nível de treinamento só interferiu nos resultados no grupo dos cirurgiões de mão, sem diferença significativa entre residentes e ortopedistas. Em conclusão, a classificação se mostrou, até certo ponto, irreprodutível e não concordante.

Não obstante, é essencial que sejam realizados novos estudos deste tipo, com esta ou outras classificações, com o objetivo de fornecer cada vez mais solidez às evidências científicas e permitir a escolha da melhor classificação.

Suporte Financeiro
Os autores declaram que não receberam suporte financeiro de agências nos setores público, privado ou sem fins lucrativos para realizar este estudo.

Conflito de Interesses
Os autores declaram não haver conflito de interesses.

Agradecimentos
Os autores agradecem a todos os participantes da pesquisa por doarem de seu tempo e disponibilizarem de sua expertise para contribuir com este estudo, a todos os pacientes que possibilitaram a realização das avaliações com seus exames e à professora mestra Thaís Cristina Araújo Moreira, cujo auxílio científico foi imensurável.

Referências

1 Belloti JC, dos Santos JB, de Moraes VY, Wink FV, Tamaoki MJ, Faloppa F. The IDEAL classification system: a new method for classifying fractures of the distal extremity of the radius - description and reproducibility. Sao Paulo Med J 2013;131(04): 252–256
2 Azad A, Kang HP, Alluri RK, Vakhshori V, Kay HF, Ghiassi A. Epidemiological and Treatment Trends of Distal Radius Fractures across Multiple Age Groups. J Wrist Surg 2019;8(04):305–311
3 Mauck BM, Swigler CW. Evidence-Based Review of Distal Radius Fractures. Orthop Clin North Am 2018;49(02):211–222
4 Nellans KW, Kowalski E, Chung KC. The epidemiology of distal radius fractures. Hand Clin 2012;28(02):113–125
5 Kleinlugtenbelt YV, Groen SR, Ham SJ, et al. Classification systems for distal radius fractures. Acta Orthop 2017;88(06):681–687
6 Jayakumar P, Teunis T, Giménez BB, Verstreken F, Di Mascio L, Jupiter JB. AO Distal Radius Fracture Classification: Global Perspective on Observer Agreement. J Wrist Surg 2017;6(01):46–53
7 Siripakarn Y, Niempoog S, Boontanapibul K. The comparative study of reliability and reproducibility of distal radius' fracture classification among: AO frykman and Fernandez classification systems. J Med Assoc Thai 2013;96(01):52–57
8 Illarramendi A, González Della Valle A, Segal E, De Carli P, Maignon G, Gallucci G. Evaluation of simplified Frykman and AO classifications of fractures of the distal radius. Assessment of interobserver and intraobserver agreement. Int Orthop 1998;22(02):111–115
9 Shehovych A, Salar O, Meyer C, Ford DJ. Adult distal radius fractures classification systems: essential clinical knowledge or abstract memory testing? Ann R Coll Surg Engl 2016;98(08): 525–531
10 Naqvi SG, Reynolds T, Kitsis C. Interobserver reliability and intraobserver reproducibility of the Fernandez classification for distal radius fractures. J Hand Surg Eur Vol 2009;34(04):483–485
11 Moloney M, Kåredal J, Persson T, Farnebo S, Adolfsson L. Poor reliability and reproducibility of 3 different radiographical classification systems for distal ulna fractures. Acta Orthop 2022; 93:438–443
12 Yinjie Y, Gen W, Hongbo W, et al. A retrospective evaluation of reliability and reproducibility of Arbeitsgemeinschaftfür Osteosynthesefragen classification and Fernandez classification for distal radius fracture. Medicine (Baltimore) 2020;99(02):e18508
13 Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. Stat Med 2001;20(21):3205–3214
14 Karanicolas PJ, Bhandari M, Kreder H, et al; Collaboration for Outcome Assessment in Surgical Trials (COAST) Musculoskeletal Group. Evaluating agreement: conducting a reliability study. J Bone Joint Surg Am 2009;91(Suppl 3):99–106
15 Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of Kappa and weighted Kappa. Psychol Bull 1969;72(05):323–327
16 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(02):184–194
17 Andersen DJ, Blair WF, Steyers CM Jr, Adams BD, el-Khouri GY, Brandser EA. Classification of distal radius fractures: an analysis of interobserver reliability and intraobserver reproducibility. J Hand Surg Am 1996;21(04):574–582
18 Belloti JC, Tamaoki MJ, Franciozi CE, et al. Are distal radius fracture classifications reproducible? Intra and interobserver agreement. Sao Paulo Med J 2008;126(03):180–185