


# Concordância intra e interobservador com relação ao sistema de classificação de Walch para artrose da articulação do ombro\*

## *Intra- and Interobserver Agreement Regarding the Walch Classification System for Shoulder Joint Arthritis*

Lauro José Rocchetti Pajolli<sup>1</sup> Marcelo Casciato Carlini<sup>1</sup>  Isabella Ferrari<sup>1</sup> Fábio Teruo Matsunaga<sup>1</sup>  
Nicola Archetti Netto<sup>1</sup> Marcel Jun Sugawara Tamaoki<sup>1</sup>

<sup>1</sup>Departamento de Ortopedia e Traumatologia, Escola Paulista de Medicina, Universidade Federal de São Paulo (Unifesp), São Paulo, SP, Brasil

Endereço para correspondência Lauro José Rocchetti Pajolli, MD, Departamento de Ortopedia e Traumatologia, Escola Paulista de Medicina, Universidade Federal de São Paulo (Unifesp), São Paulo, SP, Brasil (e-mail: laurojrpajolli@gmail.com).

Rev Bras Ortop 2019;54:644–648.

### Resumo

#### Palavras-chave

- ▶ articulação do ombro
- ▶ osteoartrite/classificação
- ▶ reprodutibilidade dos testes
- ▶ tomografia computadorizada por raios x

### Abstract

**Objetivo** Avaliar a concordância inter e intraobservador com relação ao sistema de classificação de Walch para artrose do ombro.

**Materiais e Métodos** Foram selecionadas tomografias computadorizadas da articulação do ombro de pacientes adultos entre 2012 e 2016, que foram classificadas por médicos com diferentes níveis de experiência em ortopedia. As imagens foram examinadas em três momentos distintos, e a análise foi avaliada pelo índice Kappa de Fleiss para verificar a concordância intra e interobservador.

**Resultados** O índice Kappa na concordância intraobservador variou entre 0,305 e 0,545. A concordância interobservador se mostrou muito baixa no fim das três avaliações ( $\kappa = 0,132$ ).

**Conclusão** A concordância intraobservador com relação à classificação de Walch modificada mostrou-se variável, entre moderada e baixa. A concordância interobservador foi baixa.

**Objective** To evaluate the inter- and intraobserver agreement regarding the Walch classification system for shoulder arthritis.

**Materials and Methods** Computed tomography scans of the shoulder joint of adult patients were selected between 2012 and 2016, and they were classified by physicians with different levels of expertise in orthopedics. The images were examined at three different times, and the analyses were evaluated by the Fleiss Kappa index to verify the intra- and interobserver agreement.

\* Trabalho desenvolvido no Hospital da Pontifícia Universidade Católica de Campinas, Campinas, SP, Brasil. Publicado Originalmente por Elsevier Editora Ltda.

**Keywords**

- ▶ shoulder joint
- ▶ osteoarthritis/ classification
- ▶ reproducibility of results

**Results** The Kappa index for the intraobserver agreement ranged from 0.305 to 0.545. The interobserver agreement was very low at the end of the three evaluations ( $\kappa = 0.132$ ).

**Conclusion** The intraobserver agreement regarding the the modified Walch classification varied from moderate to poor. The inter-observer agreement was low.

## x-ray computed tomography Introdução

A osteoartrose (OA) é definida como degeneração articular tanto de origem primária quanto secundária. Tal limitação gera dificuldade para realizar atividades diárias, e pode se tornar incapacitante.

Na população idosa, a artrose do ombro pode acometer até 20% dos indivíduos.<sup>1</sup> A forma primária se apresenta de maneira insidiosa, sem afecções prévias no ombro, geralmente com acometimento de outras articulações. Em contrapartida, na secundária há relato de histórico prévio.<sup>1</sup>

Seu tratamento inicial se baseia em medidas clínicas e medicamentosas. A indicação de tratamento cirúrgico é frequente para pacientes com prejuízo das atividades de vida diária e que não responderam ao tratamento não cirúrgico.

O número de artroplastias e hemiartroplastias do ombro tem crescido nos últimos anos. Estudos prévios mostram um aumento de 10,6% no número de artroplastias totais, e de 6,7% de hemiartroplastias do ombro, entre 1993 e 2007.<sup>2</sup>

Os exames de imagem são métodos de auxílio no diagnóstico e estadiamento da doença, e também contribuem para a indicação do tratamento. Rotineiramente são usadas radiografias em três incidências: anteroposterior, perfil escapular, e perfil axilar.<sup>1</sup> A tomografia computadorizada tem como principal objetivo evidenciar a anteversão da glenoide e proporcionar visão detalhada do acometimento articular.<sup>3</sup>

As classificações têm o objetivo principal de permitir a comunicação entre os profissionais envolvidos no estudo de determinada doença, e de padronizar diagnósticos e tratamentos nas pesquisas clínicas. Assim, uma boa classificação deve ser reprodutível e ter a capacidade de prever o prognóstico de determinada afecção.<sup>4</sup>

Um método usado para a avaliação da reprodutibilidade de uma classificação é a análise da concordância intra e interobservador. Concordância intraobservador é a correspondência entre observações feitas pelo mesmo observador em diferentes intervalos de observação. O acordo entre observadores define a concordância interobservador.

Existem diversas classificações para a OA do ombro, e a mais usada mundialmente é a de Walch et al,<sup>3</sup> proposta em 1999 e modificada em 2016,<sup>4</sup> que estadia e avalia a progressão da OA do ombro baseada em cortes tomográficos da articulação dos pacientes. Tal estudo leva em consideração a morfologia da glenoide, seu ângulo de retroversão, e sua relação com a cabeça umeral. A partir desses dados, é proposto o melhor tipo de artroplastia a ser empregado no tratamento dessa afecção.

Contudo, são escassas as informações quanto à reprodutibilidade e concordância, principalmente referentes à modificação feita em 2016.

O objetivo deste estudo é avaliar a concordância intra e interobservador da classificação de Walch modificada para artrose da articulação do ombro.

## Materiais e Métodos

Este é um estudo analítico transversal retrospectivo de concordância para classificações. O projeto de pesquisa foi aprovado pelo Comitê de Ética em Pesquisa da Plataforma Brasil (C.A.A.E nº 66863817.3.0000.5505).

### Cálculo amostral

Inicialmente, foi determinado o número de 62 tomografias 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%.

### Seleção da amostra

As imagens selecionadas foram obtidas entre 2012 e 2016, no setor de Cirurgia de Ombro e Cotovelo. Foram incluídos exames feitos em adultos com idade  $\geq 18$  anos que apresentassem OA do ombro. Os exames foram selecionados por dois ortopedistas que não participaram do processo de classificação da doença, de forma a incluir exames de boa qualidade.

Foram excluídas imagens de pacientes com fraturas da região proximal do úmero, fraturas da glenoide, fraturas do corpo da escápula, e luxações da articulação do ombro, assim como todos os pacientes que apresentavam qualquer material de síntese nas imagens.

Inicialmente, 62 tomografias do ombro foram analisadas. Após serem submetidas aos critérios de exclusão, dez exames foram excluídos do estudo. Assim, 52 tomografias foram usadas para a avaliação e classificação da OA do ombro.

### Processo de classificação das imagens

Os exames foram classificados por cinco examinadores com diferentes níveis de experiência.

Dois examinadores de nível *expert* (ENE1 e ENE2, com mais de seis anos de experiência como ortopedista especialista em ombro e cotovelo), um examinador de nível avançado (ENA, com um ano de experiência como ortopedista especialista em ombro e cotovelo), um examinador de nível básico (ENB, médico residente em ortopedia), e um graduando de medicina (GM).

Com a finalidade de minimizar o viés devido às dificuldades de interpretação e inexperiência com a classificação, os observadores tiveram um treinamento prévio sobre o sistema de classificação usado. Além disso, durante o processo

de classificação, o participante teve à sua disposição a brochura com toda a classificação de Walch.

As imagens foram organizadas em um arquivo digital fechado. As classificações foram feitas pelos observadores em três momentos, com um intervalo de três semanas entre uma avaliação e outra. Na primeira avaliação (T1), as imagens foram visualizadas em ordem numérica. Na segunda avaliação (T2), três semanas depois, houve randomização da sequência das imagens, assim como na terceira (T3). A sequência das imagens em cada avaliação foi randomizada por uma pessoa alheia às avaliações, não ligada diretamente ao estudo, e essa sequência foi revelada somente na análise final estatística.

Cada um dos avaliadores classificou independentemente as imagens. Não houve limite de tempo para a avaliação.

Os participantes foram instruídos a não discutir os sistemas até o fim da fase de classificação. Além disso, não houve acesso ao histórico dos pacientes ou a qualquer dado clínico deles.

### Análise estatística

A análise estatística dos resultados obtidos foi feita por um profissional especializado da área de estatística médica. Usou-se o teste Kappa de Fleiss para avaliar a concordância intra e interobservador para cada escala. O uso do coeficiente Kappa de Fleiss é considerado o mais apropriado para análise de situações em que múltiplos examinadores estão envolvidos ou várias avaliações são feitas, e quando a escala avaliada apresenta muitas categorias.<sup>5</sup>

O teste foi interpretado de acordo com Altman<sup>6</sup> 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 e 0,2 apresentam uma concordância muito pequena; de 0,21 a 0,40, uma concordância pequena; de 0,41 a 0,60, uma concordância moderada; e de 0,61 a 0,80, uma concordância substancial. Para um valor acima de 0,80, considera-se que a concordância foi praticamente perfeita.<sup>6,7</sup>

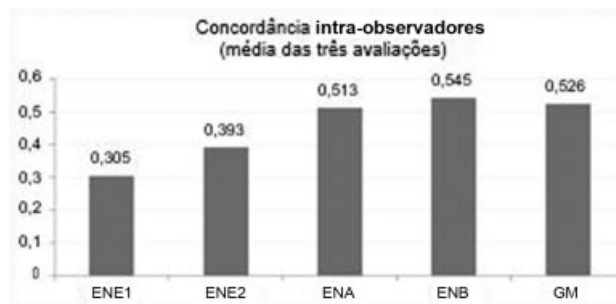
### Sistema de classificação de artrose do ombro

Segundo a classificação de Walch, a OA da articulação do ombro é dividida em quatro tipos, com subdivisões:

(A) artrose com a cabeça umeral centralizada (sem deslocamento); (A1) erosão pequena; (A2) erosão grande; (B) artrose com subluxação posterior da cabeça umeral; (B1) diminuição do espaço articular, com presença de osteófitos e esclerose subcondral; (B2) retroversão da glenoide e acometimento do lábio posterior (glenoide bicôncava); (B3) retroversão > 14°, com ou sem subluxação; (C) retroversão da glenoide > 25° independente da erosão; e (D) anteversão da glenoide e/ou subluxação anterior da cabeça umeral.<sup>4</sup>

### Resultados

Não foi estabelecida uma resposta correta, mas observou-se a concordância intra e interobservador (maior concordância e maior discordância).



**Fig. 1** Média da concordância intraobservador feita ao término das três avaliações. Abreviaturas: ENE1 e ENE2, examinadores de nível expert; ENA, examinador de nível avançado; ENB, examinador de nível básico; GM, graduando de medicina.

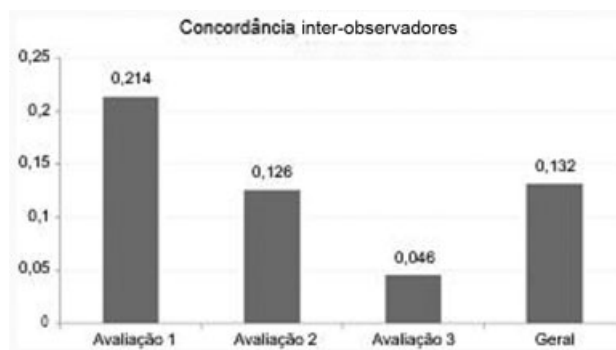
A ► **Figura 1** mostra o índice Kappa para a concordância intraobservador entre as três avaliações distintas, com o uso de sete níveis (A1, A2, B1, B2, B3, C e D). O melhor resultado revelou uma concordância moderada ( $\kappa = 0,545$ ).

A ► **Figura 2** mostra o índice Kappa para a concordância interobservador em cada avaliação separadamente, e a concordância geral no fim das três avaliações, com o uso dos mesmos sete níveis. A melhor concordância foi obtida na primeira avaliação, mas foi uma concordância pequena ( $\kappa = 0,214$ ). Após as três avaliações, a concordância interobservador foi muito pequena ( $\kappa = 0,132$ ).

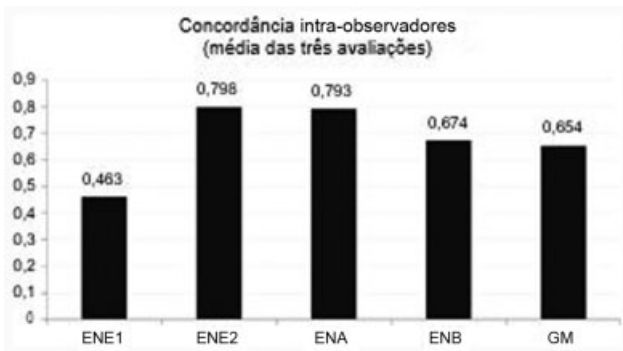
Foram feitos cálculos de concordância com o uso somente dos quatro níveis básicos da classificação de Walch (A, B, C e D). Imagens classificadas como A1 e A2 foram agrupadas como A; imagens classificadas como B1, B2 e B3 foram agrupadas como B.

A ► **Figura 3** mostra o índice Kappa para a concordância intraobservador quando usados somente os quatro níveis básicos. Nesse cenário, o melhor resultado obteve concordância substancial, quase perfeita ( $\kappa = 0,798$ ).

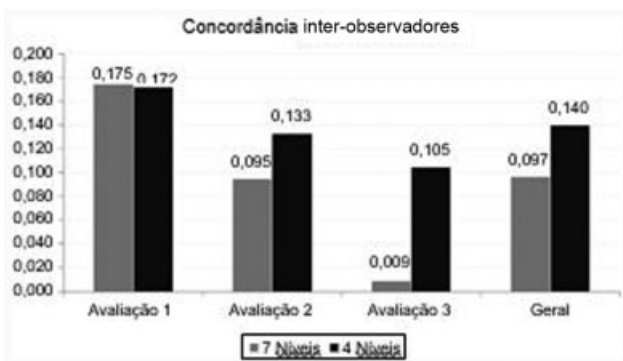
A ► **Figura 4** apresenta a comparação do índice Kappa interobservador quando usados os sete níveis (A1, A2, B1, B2, B3, C e D), e após o agrupamento para os quatro níveis básicos. Nota-se que, mesmo após a simplificação do sistema de classificação, a melhor concordância interobservador se manteve muito pequena ( $\kappa = 0,172$ ).



**Fig. 2** Concordância interobservador nas três avaliações distintas e avaliação geral da concordância.



**Fig. 3** Média da concordância intraobservador, com o uso de quatro níveis. Abreviaturas: ENE1 e ENE2, examinadores de nível *expert*; ENA, examinador de nível avançado; ENB, examinador de nível básico; GM, graduando de medicina.



**Fig. 4** Comparação da concordância interobservador nas três avaliações distintas e avaliação geral da concordância, com o uso de sete e quatro níveis.

## Discussão

A classificação de Walch foi escolhida por se tratar de um método bastante usado por ortopedistas para determinar o acometimento da articulação do ombro em pacientes com artrose primária. A concordância intra e interobservador tem grande importância para a avaliação de qualquer classificação ortopédica.

O índice Kappa na concordância intraobservador variou de 0,305 (ENE1) a 0,545 (ENB), o que revela que houve pouca ou moderada concordância para um mesmo avaliador. A grande variação entre os resultados decorre, provavelmente, da complexidade da classificação. A experiência profissional não teve o efeito esperado na concordância intraobservador, visto que o maior índice foi obtido pelo ENB, e o menor, pelo ENE1.

A concordância interobservador se mostrou muito baixa no fim das três avaliações ( $\kappa = 0,132$ ). Houve diminuição do índice entre os três momentos de avaliação. Isso mostrou que o tempo e a familiarização com a classificação não tiveram efeito relevante no fim das avaliações, além do fato de que houve treinamento prévio à primeira avaliação, o que pode ter influenciado os resultados.

Quando comparado a estudos que avaliaram a concordância de outras classificações, o presente estudo demons-

trou índices Kappa inferiores aos demais, e evidenciou menor concordância intra e interobservador da classificação de Walch em relação às demais. Matsunaga et al<sup>8</sup> demonstraram uma concordância satisfatória para a classificação de Mason para fraturas da porção proximal do rádio, tanto intra ( $\kappa = 0,582$ ) quanto interobservador ( $\kappa = 0,429-0,560$ ).

Ao usar quatro níveis básicos de avaliação, notou-se melhor concordância intraobservador, e obteve-se concordância substancial para a maioria dos examinadores. Esse resultado evidencia a dificuldade de avaliar as subdivisões dos tipos da classificação de Walch, e o fato de que simplificar a classificação faz com que melhores concordâncias sejam alcançadas.

Belotti et al<sup>9</sup> demonstraram em seu estudo que a concordância intra e interobservador para classificações da porção distal do rádio foi maior em classificações com menor número de variáveis. Tal fato corrobora o presente estudo, no qual houve aumento da concordância quando usado menor número de variáveis.

Os resultados deste estudo revelam uma diferença importante em comparação com os relatos do estudo de Bercik et al,<sup>4</sup> que demonstraram uma concordância muito boa interobservador e praticamente perfeita intraobservador. Essa diferença possivelmente se explica pelo uso de softwares especializados para a determinação do ângulo de versão da glenoide, e pelo uso de imagens com reconstrução tridimensional (3D) das tomografias nos estudos supracitados, que não foram usadas nas avaliações do presente estudo.

O uso de imagens com reconstrução 3D parece melhorar o entendimento da morfologia da glenoide. A OA pode causar degeneração óssea nos planos sagital, coronal e também no axial, e, assim, apresentar-se como um defeito 3D de difícil visualização em imagens bidimensionais.

Scalise et al<sup>10</sup> e Budge et al<sup>11</sup> fizeram estudos que usaram a tomografia com reconstrução 3D. Ambos demonstraram que houve melhor entendimento morfológico da glenoide e, assim, melhor concordância interobservador quando as imagens 3D foram oferecidas.

É importante mencionar que o presente estudo se limitou a avaliar as opiniões dos observadores; não foi estabelecida uma resposta correta para cada exame de imagem avaliado. Portanto, não foi avaliada a acurácia de cada observador. Para tanto, seriam necessárias a avaliação das respostas de cada observador e a comparação com um método que fosse padrão-ouro (alta especificidade e sensibilidade) para o diagnóstico.

## Conclusão

A concordância intraobservador da classificação de Walch modificada mostrou-se variável, entre moderada e pouca. Já a concordância interobservador mostrou-se baixa.

### Conflitos de Interesse

Os autores declaram não haver conflitos de interesse.

## Referências

- 1 Cofield RH, Briggs BT. Glenohumeral arthrodesis. Operative and long-term functional results. *J Bone Joint Surg Am* 1979;61(05): 668-677

- 2 Day JS, Lau E, Ong KL, Williams GR, Ramsey ML, Kurtz SM. Prevalence and projections of total shoulder and elbow arthroplasty in the United States to 2015. *J Shoulder Elbow Surg* 2010;19(08):1115–1120
- 3 Walch G, Badet R, Boulahia A, Khoury A. Morphologic study of the glenoid in primary glenohumeral osteoarthritis. *J Arthroplasty* 1999;14(06):756–760
- 4 Bercik MJ, Kruse K II, Yalozis M, Gauci MO, Chaoui J, Walch G. A modification to the Walch classification of the glenoid in primary glenohumeral osteoarthritis using three-dimensional imaging. *J Shoulder Elbow Surg* 2016;25(10):1601–1606
- 5 Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37(05):360–363
- 6 Altman DG. *Practical statistics for medical research*. 3rd ed. London: Chapman and Hall; 1995
- 7 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(08):689–698
- 8 Matsunaga FT, Tamaoki MJ, Cordeiro EF, et al. Are classifications of proximal radius fractures reproducible? *BMC Musculoskelet Disord* 2009;10:120
- 9 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
- 10 Scalise JJ, Codsi MJ, Bryan J, Brems JJ, Iannotti JP. The influence of three-dimensional computed tomography images of the shoulder in preoperative planning for total shoulder arthroplasty. *J Bone Joint Surg Am* 2008;90(11):2438–2445
- 11 Budge MD, Lewis GS, Schaefer E, Coquia S, Flemming DJ, Armstrong AD. Comparison of standard two-dimensional and three-dimensional corrected glenoid version measurements. *J Shoulder Elbow Surg* 2011;20(04):577–583