Original Article

Reproducibility assessment of different descriptions of the Kellgren and Lawrence classification for osteoarthritis of the knee[☆]

Felipe Borges Gonçalves, Felipe Almeida Rocha, Rodrigo Pires e Albuquerque[*]*,
Alan de Paula Mozella, Bernardo Crespo, Hugo Cobra*

Instituto Nacional de Traumatologia e Ortopedia, Centro de Cirurgia do Joelho, Rio de Janeiro, RJ, Brazil

A R T I C L E   I N F O

A B S T R A C T

Objective: To assess the inter- and intraobserver reproducibility of the original version and different descriptions of the Kellgren and Lawrence classification used in epidemiological studies for osteoarthritis of the knee.
Methods: The study included 72 patients with osteoarthritis of the knee. Three medical members of the Brazilian Society of Knee Surgery were invited to evaluate the images. An intra- and interobserver analysis was conducted, with an interval of one month. The intraobserver agreement was analyzed using the weighted Cohen's Kappa coefficient. The interobserver agreement was analyzed using the Krippendorff alpha coefficient (α).
Results: The intraobserver assessment indicated conflicting results. In the interobserver analysis, the level of agreement was superficial.
Conclusions: The classification of Kellgren and Lawrence and its variants generated a low reproducibility between observers. The intraobserver analysis showed a lack of uniformity in the use of this classification and its variants, even among experienced observers.

Avaliação da reprodutibilidade das diferentes descrições da classificação de Kellgren e Lawrence para osteoartrite do joelho

R E S U M O

Objetivo: Avaliar a reprodutibilidade inter e intraobservador da versão original e das diferentes descrições da classificação de Kellgren e Lawrence usadas em estudos epidemiológicos para osteoartrite do joelho.
Métodos: Foram estudados 72 pacientes com diagnóstico de osteoartrite do joelho. Três médicos membros da Sociedade Brasileira de Cirurgia do Joelho foram convidados para avaliar

as imagens. Análises intra e interobservador foram feitas com intervalo de um mês. A concordância intraobservador foi analisada por meio do coeficiente Kappa de Cohen ponderado. Na análise interobservador foi usado o coeficiente alpha de Krippendorff (α).

Resultados: A avaliação intraobservador apresentou resultados discordantes. Na análise interobservador, o grau de concordância foi superficial.

Conclusões: A classificação de Kellgren e Lawrence e suas variantes geraram uma baixa reprodutibilidade entre os observadores. A análise intraobservador apresentou resultados discordantes, demonstrou que há falta de uniformidade no uso dessa classificação e de suas variantes mesmo entre observadores experientes.

Introduction

Osteoarthritis is one of the most common diseases worldwide, with no distinction or ethnic preference.[1] The knee, being a load-bearing joint, is a frequent involved site.[1]

The radiological evaluation is paramount in patients with osteoarthritis of the knee. In the radiographic study, one is able to grade the severity of joint involvement, measure the axis, assess ligament instability or bone loss, and also indicate the type of treatment, as well as the necessary implant when surgery is needed.

The Kellgren and Lawrence grading is the most widely used classification for knee osteoarthritis when X-rays are assessed[2]; however, five versions of this classification have been described in epidemiological studies.[2]

In order to be reproducible, a classification should be simple, easy to remember, and helpful in guiding treatment and defining the prognosis of these injuries.[3] A characteristic that must be present in any classification is reproducibility.[3]

This study aimed to assess the interobserver and intraobserver reproducibility of the original version and of the different variants of the Kellgren and Lawrence classification used in epidemiological studies for osteoarthritis of the knee.

Material and methods

The study was presented in detail to and approved by the Ethics Committee under CAAE No. 31378714.6.0000.5273. All participants signed a informed consent prior to enrollment. They were also offered a financial incentive to participate.

In this hospital's outpatient clinic, 200 patients with osteoarthritis of the knee were selected. There was no age limitation. Exclusion criteria were: patients who underwent prior surgical procedures in the knee to be assessed or who underwent joint replacement on the contralateral knee, as well as patients with other rheumatologic diseases. After applying the exclusion criteria, 72 patients and their radiographic studies were selected to comprise the sample.

The study consisted of three observers, members of the Brazilian Society of Knee Surgery and part of the hospital staff, who conducted the radiographic analysis.

Knee radiographs in anteroposterior (AP) with bipedal load, lateral, axial patellar at 30°, and Rosenberg views were obtained from all patients, following a standard protocol. The AP view was made with the knee in extension and bipedal support. The tube-film distance was 1 m, and the radius was centered at the lower pole of the patella. The lateral view was achieved with the knee in 20° of flexion with patient standing; the tube-film distance was 1 m. Rosenberg view was made in posteroanterior (PA), under load and 45° of flexion. Feet were positioned parallel and aligned forward. The patella touched the film. X-rays were centered at the level of the inferior pole of the patella, with a craniocaudal inclination of 10° and a tube-film distance of 1 m.

A Shimatzo X-ray device, rated at 50 kV and 40 mA, was used. The exams were overseen by the main investigator regarding image quality and were repeated if considered of poor technical quality; patient positioning, knee, and X-ray device angulation were also observed. Angles were measured with a goniometer.

Scanned images were delivered on a CD-ROM to the observers. In order to minimize bias due to the difficulty of interpretation or possible forgetfulness, the classification and its variants are described in Table 1.

Radiographic analyses were performed blindly on two occasions, with a one-month interval, and the interpretations of the three observers were scanned for subsequent statistical analysis.

Data were analyzed with statistical analysis software R version 3.1.0, and SPSS (Statistical Package for the Social Sciences) version 22.0. The intraobserver agreement, which compared both assessments from the same observer for each of the five classifications, was analyzed by the weighted Cohen's Kappa coefficient.

The weighted Cohen's Kappa coefficient ranges from −1 to 1; values less than or equal to 0 represent no agreement and 1 represents total agreement. In this study, the classification adopted was the one proposed by Byrt,[4] as described in Table 2. The coefficients were calculated using the "psy" package of R.

In the interobserver analysis, another measure of agreement was used, Krippendorf's alpha coefficient (α). The rating of the agreement, given the value of α, was the same as that presented in Table 2. The coefficients were calculated using the Kalpha macro in SPSS.

Table 1 – Classification and its variants.

	Original	Variant 1	Variant 2	Variant 3	Variant 4
Grade 0 (0)	Normal	Normal	Normal	Normal	Normal
Grade I	Doubtful JSN and minute osteophytes on the border	Minute osteophytes	Minimal osteophytes dubious significance	Only minute osteophytes	Minute marginal osteophytes
Grade II	Possible JSN and definite osteophytes	Definite osteophytes	Definite osteophytes without JSN	Definite osteophytes and minute JSN	Definite osteophytes and minute JSN
Grade III	Moderate JSN, multiple osteophytes, a certain degree of subchondral sclerosis and possible deformity of the bone contour	Osteophytes and JSN	Moderate JSN (with osteophytes)	Osteophytes in moderate quantity and/or definite JSN	Multiple osteophytes of moderate size, definite JSN, and possible deformity in the bone contour (bone friction)
Grade IV	Notable JSN, severe subchondral sclerosis, definite deformity of the bone contour, and presence of large osteophytes	Large osteophytes, definite JSN, and deformity	Substantial JSN with subchondral sclerosis	Large osteophytes, severe JSN and/or bone sclerosis	Large osteophytes, considerable JSN, severe sclerosis, definite bone contour deformity (bone friction)

JSN, joint space narrowing.

Table 2 – Kappa coefficient values (K) and agreement classification.

K-value	Concordance rating
−1 to 0.00	None
0.0 to 0.20	Poor
0.21 to 0.40	Superficial
0.41 to 0.60	Reasonable
0.61 to 0.80	Good
0.81 to 0.92	Very good
0.93 to 1.0	Excellent

Results

Table 3 shows the values of the weighted Kappa coefficient (K) and its confidence interval (CI) at 95% confidence, which assesses the intraobserver agreement of each observer for each of the ratings. The values indicate that Observer 1 presented a "superficial" agreement between first and second observation for the original classification and for all its variants, with Kappa values equal to 0.34 or 0.35. Observer 2 presented a "very good" agreement between first and second observation for the original classification and for all its variants, with Kappa values between 0.85 and 0.92. Finally, Observer 3 showed an "excellent" agreement between first and second observation for the original classification and for all variants, with Kappa values equal to 0.97 for variants 1 and 4

and perfect agreement ($K = 1$) between the two evaluations in the original classification and in variants 2 and 3.

Table 4 shows the values of Krippendorff's alpha coefficient, which was used to assess interobserver agreement, in the first and second evaluation, for each of the ratings. Values show that, both in the first and second evaluation, for all ratings, the agreement between observers was "superficial". It is noteworthy that the agreement was lower in the first evaluation.

Discussion

Classifying diseases is a common practice. A good rating system is designed to be simple, reproducible, and able to group different stages of a lesion into homogeneous subgroups, allowing for comparisons, treatment algorithms, and prognosis.[3] What usually happens is that once a classification for a particular injury is established, based on a relevant and representative sample, a case that does not fit the described or classified types appears.[5] Weber, in his study of malleolar fractures, reserved a subgroup for "unclassifiable" injuries, i.e., those that, due to their peculiarity, could not be fitted into classes or groups.[5] Over time, some ratings have been replaced by more complete ones.

In the literature, there is still no consensus on which classification should be used for the study of osteoarthritis of the knee. Weidow et al.[6] reported that knee radiographic

Table 3 – Weighted Kappa coefficients of the intraobserver agreement between the first and the second evaluation, for each classification.

Observer	Classification				
	Original	Variant 1	Variant 2	Variant 3	Variant 4
1	0.35 (0.15; 0.55)	0.34 (0.14; 0.54)	0.34 (0.14; 0.54)	0.34 (0.14; 0.54)	0.34 (0.14; 0.54)
2	0.92 (0.84; 0.99)	0.85 (0.75; 0.95)	0.90 (0.82; 0.98)	0.90 (0.82; 0.98)	0.90 (0.82; 0.98)
3	1.0 (1.0; 1.0)	0.97 (0.92; 1.0)	1.0 (1.0; 1.0)	1.0 (1.0; 1.0)	0.97 (0.92; 1.0)

Table 4 – Krippendorff's alpha coefficient of the interobserver agreement in the first and second evaluation for each classification.		
Classification	First evaluation	Second evaluation
Original	0.25	0.33
Variant 1	0.23	0.28
Variant 2	0.21	0.32
Variant 3	0.22	0.34
Variant 4	0.26	0.35

classifications must be reviewed and improved through the examination technique or method employed. Sun et al.,[7] in a review study of 16 classifications for osteoarthritis of the knee, concluded that there was no unanimous choice among the various medical specialties. The Kellgren and Lawrence classification[8] values the presence or absence of osteophytes. In contrast, the Ahlbäck[9] classification assesses reduction of the joint space; some studies consider it to be the best method for analyzing progression of osteoarthritis.[10,11] Studies as such as that by Danielsson and Hernborg[12] demonstrated that osteophytes did not change over 16 years of evolution. In turn, Kijowski et al.[13] concluded that osteoarthritis of the knee should be diagnosed by marginal osteophytes. In fact, it is the progression of the disease that must be assessed by joint space narrowing, subchondral sclerosis, and subchondral cysts.[12] Felson et al.[14] observed that osteophytes are associated with poor alignment of the ipsilateral lower limb. Poor alignment is a powerful risk factor for the progression of osteoarthritis.[14] The present study used the Kellgren and Lawrence[8] classification, as it is routinely used by orthopedic surgeons and rheumatologists.

Albuquerque et al.[3] observed that the Kellgren and Lawrence[8] classification had lower level of agreement on a intra- and interobserver analysis of three different classifications: those by Dejour et al.,[15] Ahlbäck *apud* Keyes et al.,[10] and Kellgren and Lawrence.[8] The present research confirms the poor results of the Kellgren and Lawrence classification.[8] Rodrigues et al.[16] analyzed the interobserver reproducibility of the original Kellgren and Lawrence classification[8] and did not observe a statistically significant difference between knee specialists and general orthopedists. Furthermore, they observed a regular Kappa coefficient index. The present study performed an intra- and interobserver analysis and attempted to achieve a more accurate assessment when compared with studies such as those by Rodrigues et al.[16]

The literature features some studies comparing the Kellgren and Lawrence classification[8] and its different variants.[2,17] However, none of these studies used a radiographic analysis described by Rosenberg et al.[18] nor included patients with advanced stages of osteoarthritis of the knee. For this reason, the present research included these two variables, thus representing an unprecedented and extremely important study. Some studies indicate that the Rosenberg view provides a better evidence of joint wear.[18,19] Furthermore, authors believe that, for a classification to be assessed accurately, it must feature the pathology studied in its various grades.

Villardi et al.[20] and Galli et al.[21] observed a low degree of interobserver agreement in the use of Ahlbäck classification modified *apud* Keyes et al.[10] The present study, although

using a different classification system, also observed a weak agreement among observers.

The observers of the present research are experienced specialists in knee surgery; in order to reproduce a more accurate assessment, a response time was not stipulated.[21,22] Vilalta et al.[23] found that experienced observers generated individual variability and caused differences in results and confusion in the literature, a belief that the authors of the present study proved and defend. Brandt et al.[24] and Kijowski et al.,[25] when assessing patients with osteoarthritis, compared the AP view in loaded knee extension with arthroscopic findings. They emphasize that, in patients with osteoarthritis, the assessment of joint space and osteophytes are not suitable parameters for analyzing the disease. They suggest that further research should be conducted in order to find a complementary test with better accuracy. The present authors believe that knee arthroscopy is an excellent therapeutic method, but it is an invasive procedure and, therefore, should not be used as a diagnostic method. In the future, magnetic resonance with load may perhaps become a superior imaging exam in comparison with radiography.

Osteoarthritis of the knee is a common and fascinating disease. The radiographic analysis and the classification used are crucial points of controversy on this subject. The present study suggests that the original Kellgren and Lawrence classification[8] and its variants generated disagreement among observers. Thus, it is important to research and develop a radiographic classification of the knee to obtain a consensus or, perhaps, to improve agreement.

Conclusions

The Kellgren and Lawrence classification and its variants generated low reproducibility among observers.

In the intraobserver analysis, discordant results were observed. This demonstrates the lack of uniformity in the use of this classification and its variants, even among experienced observers.

Conflicts of interest

The authors declare no conflicts of interest.

REFERENCES

1. Cobra H, Barretto JM, Albuquerque RP. Artroplastia unicompartimental do joelho. In: Palma IM, editor. Artroplastia total do joelho. Rio de Janeiro: Guanabara; 2008. p. 105–12.
2. Schiphof D, de Klerk BM, Kerkhof HJ, Hofman A, Koes BW, Boers M, et al. Impact of different descriptions of the Kellgren and Lawrence classification criteria on the diagnosis of knee osteoarthritis. Ann Rheum Dis. 2011;70(8):1422–7.
3. Albuquerque RP, Giordano V, Sturm L, Azevedo Júnior V, Leão A, Amaral N. Análise da reprodutibilidade de três classificações para a osteoartrose do joelho. Rev Bras Ortop. 2008;43(8):329–35.
4. Byrt T. How good is the agreement? Epidemiology. 1996;7(5):561.

5. Franco JS, Musafir M. Classificação das fraturas mais freqüente. Rio de Janeiro: Schering-Plough; 1991.

6. Weidow J, Cederlund CG, Ranstam J, Kärrholm J. Ahlbäck grading of osteoarthritis of the knee. Poor reproducibility and validity based on visual inspection of the joint. Acta Orthop. 2006;77(2):262–6.

7. Sun Y, Guenther KP, Brenner H. Reliability of radiographic grading of osteoarthritis of the hip and knee. Scand J Rheumatol. 1997;26(3):155–65.

8. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957;16(4):494–502.

9. Ahlbäck S. Osteoarthrosis of the knee. A radiographic investigation. Acta Radiol Diagn. 1968; Suppl. 277:7–72.

10. Keyes GW, Carr AJ, Miller RK, Goodfellow JW. The radiographic classification of medial gonarthrosis – correlation with operation methods in 200 knees. Acta Orthop Scand. 1992;63(5):497–501.

11. Petersson IF, Boegard T, Saxne T, Silman AJ, Svensson B. Radiographic osteoarthritis of the knee classified by the Ahlback and Kellgren & Lawrence systems for the tibiofemoral joint in people aged 35–54 years with chronic knee pain. Ann Rheum Dis. 1997;56(8):493–6.

12. Danielsson L, Hernborg J. Clinical and roentgenologic study of knee joints with osteophytes. Clin Orthop Relat Res. 1970;(69):302–12.

13. Kijowski R, Blankenbaker DG, Stanton PT, Fine JP, De Smet AA. Radiographic findings of osteoarthritis versus arthroscopic findings of articular cartilage degeneration in the tibialfemoral joint. Radiology. 2006;239(3):818–24.

14. Felson DT, Gale DR, Elon Gale M, Niu J, Hunter DJ, Goggins J, et al. Osteophytes and progression of knee osteoarthritis. Rheumatology (Oxford). 2005;44(1):100–4.

15. Dejour H, Carret JP, Walch G. Les Gonarthroses. In: 7 émes Journées Lyonnaises de Chirurgie de Genou. 1991.

16. Rodrigues AA, Karam FC, Scorsatto C, Martins C, Pires LAS. Análise da reprodutibilidade da classificação de Kellgren e Lawrence para aosteoartrose do joelho. Rev AMRIGS. 2012;56(2):107–10.

17. Schiphof D, Boers M, Bierma-Zeinstra SM. Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. Ann Rheum Dis. 2008;67(7):1034–6.

18. Rosenberg TD, Paulos LE, Parker RD, Coward DB, Scott SM. The forty-five-degree posteroanterior flexion weight-bearing radiograph of the knee. J Bone Joint Surg Am. 1988;70(10):1479–83.

19. Mason RB, Horne JG. The posteroanterior 45 flexion weight-bearing radiograph of the knee. J Arthroplasty. 1995;10(6):790–2.

20. Villardi AE, Mandarino M, Veiga LT. Avaliação da reprodutibilidade da classificação de Ahlbäck modificada para osteoartrose do joelho. Rev Bras Ortop. 2006;41(5):157–61.

21. Galli M, De Santis V, Tafuro L. Reliability of the Ahlbäck classification of knee osteoarthritis. Osteoarthritis Cartilage. 2003;11(8):580–4.

22. Khan FA, Koff MF, Noiseux NO, Bernhardt KA, O'Byrne MM, Larson DR, et al. Effect of local alignment on compartmental patterns of knee osteoarthritis. J Bone Joint Surg Am. 2008;90(9):1961–9.

23. Vilalta C, Nunez M, Segur JM, Domingo A, Carbonell JA, Maculé F. Knee osteoarthritis: interpretation variability of radiological signs. Clin Rheumatol. 2004;23(6):501–4.

24. Brandt KD, Fife RS, Braustein EM, Katz B. Radiographic grading of the severity of knee osteoarthritis: relation of the Kellgren and Lawrence grade to a grade based on joint space narrowing and correlation with arthroscopic evidence of articular cartilage degeneration. Arthritis Rheum. 1991;34(11):1381–6.

25. Kijowski R, Blankenbaker D, Stanton P, Fine J, De Smet A. Arthroscopic validation of radiographic grading scales of osteoarthritis of the tibiofemoral joint. Am J Roentgenol. 2006;187(3):794–9.