Published online: 2019-08-20

Reliability of the Arthroscopic Classifications of Hip Chondral Lesions*

Confiabilidade das classificações artroscópicas das lesões condrais do quadril

Antônio Augusto Guimarães Barros[1] Carlos César Vassalo[1] Lincoln Paiva Costa[1] Juan Gómez-Hoyos[2] Vinícius de Oliveira Paganini[1] Marco Antônio Percope de Andrade[3]

[1] Hospital Madre Teresa, Belo Horizonte, MG, Brazil
[2] Hip Preservation Center, Baylor Scott and White Research Institute, Dallas, TX, USA
[3] Locomotor System Department, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

Address for correspondence Antônio Augusto Guimarães Barros, Master, Hospital Madre Teresa, Belo Horizonte, MG, Brazil (e-mail: antonioagbarros@gmail.com).

Rev Bras Ortop 2019;54:440–446.

Abstract

Objective To evaluate the inter- and intraobserver reliability of the Outerbridge, Beck, and Haddad classifications for acetabular joint cartilage lesions through the arthroscopic procedure.

Methods A total of 60 hip arthroscopy videos were evaluated twice by 4 surgeons at 2 different times to assess the inter- and intraobserver reproducibility of the classifications, and the data was analyzed by means of the weighted Cohen Kappa index.

Results The mean weighted Kappa values in the interobserver assessment of the Outerbridge, Beck, and Haddad classifications were, respectively, 0.72, 0.78, and 0.68. The three classifications were considered as presenting good interobserver agreement. Regarding the intraobserver assessment of the Outerbridge, Beck, and Haddad classifications, the weighted Kappa values were, respectively, 0.9, 0.9, and 0.93. The three classifications were considered as presenting excellent intraobserver agreement.

Conclusion In the present series, the Outerbridge, Beck, and Haddad classifications presented good interobserver reproducibility and excellent intraobserver reproducibility when evaluating acetabular chondral lesions by the arthroscopic approach.

Keywords
- arthroscopy
- hip
- cartilage diseases/classification
- reproducibility of results

Resumo

Objetivo Avaliar a confiabilidade inter- e intraobservador das classificações de Outerbridge, Beck e Haddad para lesões da cartilagem articular acetabular com o uso da via artroscópica.

Métodos Foram avaliados 60 vídeos de artroscopias do quadril por 4 cirurgiões em 2 momentos para avaliar a reprodutibilidade inter- e intraobservador das classificações. Os dados foram analisados a partir do cálculo do índice Kappa de Cohen ponderado.

Antônio Augusto Guimarães Barros's ORCID is https://orcid.org/0000-0003-3701-1937.

received
November 12, 2017
accepted
January 9, 2018

DOI https://doi.org/10.1016/j.rbo.2018.01.002.
ISSN 0102-3616.

Palavras-chave
- artroscopia
- quadril
- doenças das cartilagens/ classificação
- reprodutibilidade dos testes

Resultados Os valores do Kappa ponderado médio na avaliação interobservador das classificações de Outerbridge, Beck e Haddad foram, respectivamente, 0,72, 0,78 e 0,68. As três classificações foram consideradas como de boa concordância interobservador. Com relação à avaliação intraobservador das classificações de Outerbridge, Beck e Haddad, os valores Kappa foram, respectivamente, 0,9, 0,9 e 0,93. As três classificações foram consideradas excelentes na comparação intraobservador.

Conclusão Na presente série, as classificações de Outerbridge, Beck e Haddad apresentaram boa reprodutibilidade interobservador e excelente reprodutibilidade intraobservador ao avaliar lesões condrais acetabulares por via artroscópica.

Introduction

Hip arthroscopy has emerged as a minimally invasive method for the diagnosis and treatment of several intra- and extra-articular hip conditions.[1] This approach provides excellent visualization of the joint surface and allows the surgeon to identify early stages of cartilage damage.[2] Hip joint cartilage lesions are caused mainly by trauma, inflammatory arthritis, acetabular dysplasia, or femoroacetabular impingement (FAI).[3]

After treating young adults with idiopathic hip arthrosis, Ganz et al[4] realized that certain joint morphological characteristics resulted from the abnormal contact between the proximal femur and the acetabular ridge, the so-called FAI. Next, it was observed that this abnormal contact leads to damage patterns of the acetabular labrum and adjacent cartilage. Anterosuperior labral lesions associated with posteroinferior cartilaginous damage due to backlash in the femoral head and in the acetabular surface are associated with a Pincer-type impingement. The separation of the anterior superior labrum from the acetabular ridge and the shear or delamination of the adjacent acetabular joint cartilage are more indicative of the CAM-type impingement.[5] Patients with FAI usually present both deformities. Recent studies point to a strong association between hip osteoarthritis (OA) and FAI.[6] As such, currently there is a great interest in the treatment of hip labral and chondral lesions to assess whether these interventions can delay or even prevent the development of hip OA in these patients.

The rapid increase in FAI diagnosis and hip arthroscopy as a therapeutic access route resulted in an exponential growth of publications regarding this subject.[7] However, there is great variability in the literature on indications, outcomes, assessment methods, joint capsule and soft tissues management, as well as on rehabilitation. This variability is partially associated with the comparison of patients with labral and chondral lesions at different severity stages. The reduction of this variability requires the comparison of patients with lesions in similar stages.

Classification systems help communication, treatment planning, and prognostic assessment. An ideal classification system should be easily applicable, reliable, reproducible, include the entire lesional spectrum, be mutually exclusive, logical, and clinically useful.[8]

The present study aims to evaluate the inter- and intraobserver reliability of classifications of lesions of the hip joint cartilage through the arthroscopic route, and to define which are the most reproducible classification systems based on the weighted Cohen Kappa index calculation. In the present study, we attempt to find the arthroscopic classifications that present greater reproducibility so that communication, treatment outcomes and prognostic evaluations are reliable. The hypothesis is that the three classification systems studied have moderate inter- and intraobserver reliability.

Materials and Methods

From a database of 278 hip arthroscopy videos, whose recordings were previously authorized by the patients, 70 videos were initially selected. These patients were invited to participate in the present study and signed an informed consent form. The inclusion criterion consisted of all patients submitted to hip arthroscopy for FAI. The exclusion criteria included presence of radiological arthrosis (Tonnis > 2),[9] history of previous hip surgery (either in an open or in an arthroscopic procedure), videos with poor image quality, inadequate visualization, inadequate lesion palpation, and refusal to participate in the study. The application of these criteria resulted in a final sample of 60 videos

The procedures were performed by 2 surgeons experienced in hip arthroscopy (each surgeon performed > 200 procedures) and followed the pattern already routinely used by both professionals. Arthroscopies were performed with the patient in the supine position on a traction table and used anterolateral and medialanterior portals. The intra-articular assessment of the acetabulum was made with a camera in the anterolateral portal and with a palpation instrument in the medialanterior access for inspection of the articular surface, of the labrum and of the chondrolabral junction.

One of the authors, an arthroscopy-trained hip surgeon who did not participate in the surgeries, reviewed each recording and edited them. After the edition, each recording lasted ~ 1 minute and included the inspection of the articular surface, of the labrum, and of the chondrolabral junction. This process resulted in 60 digital recordings of hip arthroscopic procedures.

Four hip surgeons experienced in hip arthroscopy (all with at least 2 years of experience and participation in > 150 hip

arthroscopy cases) contributed in the study. During a meeting, the four surgeons discussed the classification systems described for joint cartilage injuries: Outerbridge,[10] Beck et al,[11] and Konan et al[12] (►**Table 1**). After 15 days, in a new meeting, the authors practiced the use of the classification systems in videos not included in the present study. After this training, each surgeon classified separately the lesions from the videos according to the proposed systems. If there was > 1 finding, the reviewers were instructed to report only the most severe. The initial values from each reviewer were used at the interobserver analysis. In a second moment, 1 month later, the reviewers were asked to classify again the lesions from the same videos, but sorted differently. These values were matched by case and reviewer and were compared with the first results for an intraobserver analysis.

The statistical evaluation of the reliability of the classification systems included the percentage of agreement and the weighted Kappa values as proposed by Cohen.[13] The Kappa values were classified as described by Landis et al;[14] as such, values from 0.81 to 1.0 indicated excellent agreement; 0.61 to 0.80, good agreement; 0.41 to 0.60, moderate agreement; 0.21 to 0.40, reasonable agreement; and 0 to 0.20, weak agreement. Statistical analyzes were performed in the software programs R, version 3.3.2, and Minitab version 18 (Minitab, LLC, State College, PA, USA).

The research was previously approved by the Research Ethics Committee and complied with the Helsinki Declaration from the World Medical Association.

Results

The demographic characteristics of the participants of the study are summarized in ►**Table 2**. The descriptive analysis of the classification of the lesions by each observer using the Outerbridge, Beck and Haddad systems at the first assessment is presented in ►**Table 3**. These results are illustrated in ►**Figs. 1–3**. These data show that the interobserver agreement using the Outerbridge classification system was moderate in one case (I and IV), good in four cases (I and II; I and III; II and IV; III and IV), and excellent in one case (II and III). The Kappa values ranged from 0.58 to 0.82, with an average value of 0.72; as such, the mean agreement between observers was considered good.

Table 2 Demographic features

Gender, n (%)	
Male	32 (53%)
Female	28 (47%)
Age, mean (standard deviation)	33 years old (±7)
Deformity, n (%)	
Combined (CAM and Pincer)	31 (52%)
CAM	27 (45%)
Pincer	2 (3%)

Table 1 Outerbridge, Beck and Haddad classification systems

Types	Description
Outerbridge	
0	No gross cartilaginous alterations
1	Softening or edema
2	Fragmentation/lesion < 1.3 cm
3	Fragmentation/lesion > 1.3 cm
4	Bone exposure
Beck	
0	No gross cartilaginous alterations
1	Malacia; superficial rugosity, fibrillation
2	Detachment; Loss of subchondral bone attachment, solid cartilage at gross examination, carpet phenomenon
3	Cleavage; Loss of subchondral bone attachment, effaced borders, cartilage thinning, flap
4	Total thickness defect
Haddad	
0	No gross cartilaginous alterations
1	Wave signal; Loss of subchondral bone attachment
2	Cleavage; Obvious separation at the chondrolabral junction, but arthroscopic probing shows the joint cartilage adhered to the underlying bone with no delamination evidence
3	Delamination; Cartilage grossly detached from acetabular bone
4	Bone exposure

Table 3 Assessment of the three classification systems for cartilage lesions at the first moment of evaluation

Classification system	Observer							
	I		II		III		IV	
	n	%	*n*	%	*n*	%	*n*	%
Outerbridge								
0	8	13.33	8	13.33	5	8.33	19	31.67
1	25	41.67	28	46.67	30	50.00	16	26.67
2	11	18.33	13	21.67	8	13.33	15	25.00
3	8	13.33	7	11.67	11	18.33	2	3.33
4	8	13.33	4	6.67	6	10.00	8	13.33
Beck et al.								
0	8	13.33	6	10.00	7	11.67	14	23.33
1	22	36.67	13	21.67	12	20.00	8	13.33
2	11	18.33	19	31.67	21	35.00	15	25.00
3	10	16.67	16	26.67	14	23.33	14	23.33
4	9	15.00	6	10.00	6	10.00	9	15.00
Haddad et al.								
0	10	16.67	10	16.67	13	21.67	15	25.00
1	22	36.67	15	25.00	18	30.00	5	8.33
2	10	16.67	13	21.67	9	15.00	17	28.33
3	10	16.67	16	26.67	13	21.67	15	25.00
4	8	13.33	6	10.00	7	11.67	8	13.33

n, number of observations; %, percentage.

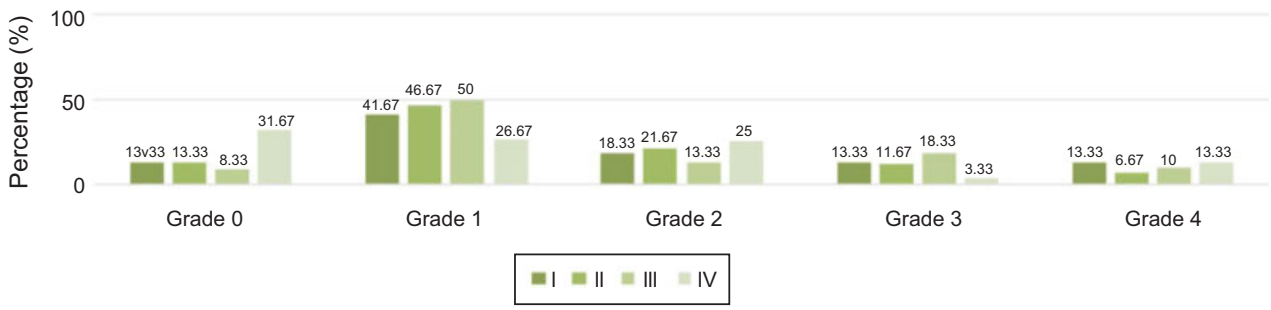

Fig. 1 Frequency of cartilage lesions, according to the four observers, using the Outerbridge classification system.

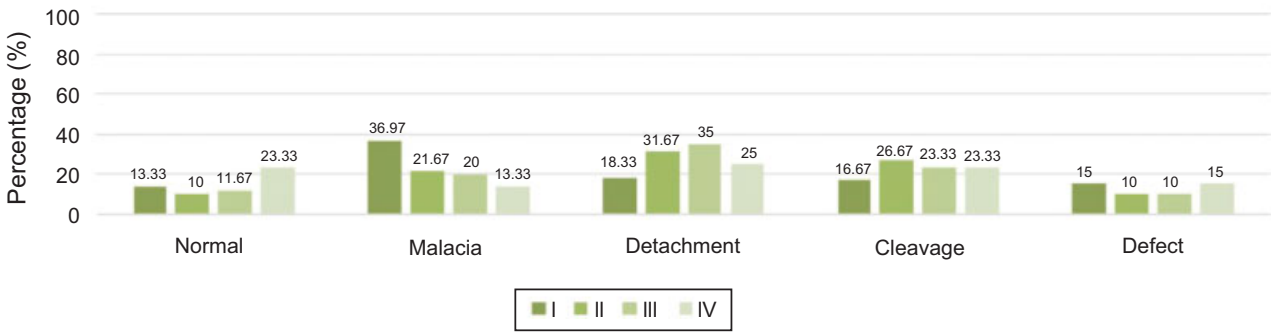

Fig. 2 Frequency of cartilage lesions, according to the four observers, using the Beck classification system.

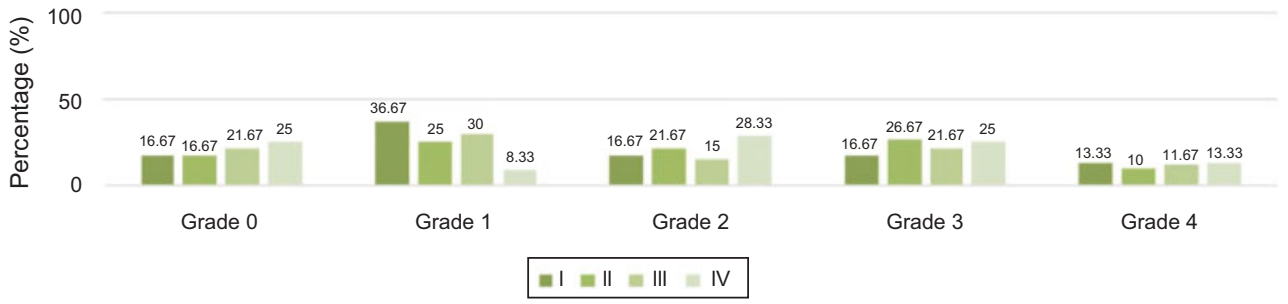

Fig. 3 Frequency of cartilage lesions, according to the four observers, using the Haddad classification system.

Regarding the Beck classification system, agreement was good in four cases (I and II; I and III; I and IV; II and IV), and excellent in two cases (II and III; III and IV). The Kappa values ranged from 0.74 to 0.84, with an average value of 0.78; as such, the mean agreement between observers was considered good.

The interobserver agreement of the Haddad classification system was moderate in one case (III and IV), and good in five cases (I and II; I and III; I and IV; II and III; II and IV). The Kappa values ranged from 0.60 to 0.78, with an average value of 0.68; as such, the mean agreement between observers was considered good. The interobserver assessment is summarized in ►**Table 4**.

In the intraobserver evaluation, the Kappa values for the Outerbridge system ranged from 0.83 to 0.94, with an average value of 0.90, therefore, the mean agreement between observers was considered excellent. For the Beck classification system, the agreement between the four observers was considered excellent, with Kappa values ranging from 0.87 and 0.92, with an average value of 0.90. Similarly, the Haddad classification system was considered excellent, with Kappa values ranging from 0.89 and 0,97, with an average value of 0.93. The intraobserver assessment is summarized in ►**Table 5**.

Discussion

The present study showed that the classification systems for arthroscopic assessment of chondral acetabular lesions is reproducible both in the inter- and intraobserver level. Using

Table 4 Interobserver assessment with weighted Kappa values

Classification system	Kappa	Kappa classification	95%CI	Agreement rate	95%CI
Outerbridge	**0.72**	**Good**			
Observer I and II	0.73	Good	0.53–0.94	0.62	0.48–0.74
Observer I and III	0.72	Good	0.52–0.92	0.58	0.45–0.71
Observer I and IV	0.58	Moderate	0.37–0.80	0.43	0.31–0.57
Observer II and III	0.82	Excellent	0.72–0.89	0.63	0.50–0.75
Observer II and IV	0.72	Good	0.60–0.87	0.52	0.39–0.65
Observer III and IV	0.73	Good	0.58–0.83	0.43	0.31–0.57
Beck	**0.78**	**Good**			
Observer I and II	0.77	Good	0.64–0.90	0.57	0.43–0.69
Observer I and III	0.74	Good	0.59–0.90	0.53	0.40–0.66
Observer I and IV	0.74	Good	0.62–0.86	0.50	0.37–0.63
Observer II and III	0.84	Excellent	0.77–0.91	0.68	0.55–0.79
Observer II and IV	0.78	Good	0.66–0.91	0.62	0.48–0.74
Observer III and IV	0.82	Excellent	0.81–0.84	0.63	0.50–0.75
Haddad	**0.68**	**Good**			
Observer I and II	0.68	Good	0.48–0.88	0.50	0.37–0.63
Observer I and III	0.78	Good	0.66–0.90	0.60	0.47–0.72
Observer I and IV	0.61	Good	0.39–0.82	0.42	0.29–0.55
Observer II and III	0.75	Good	0.54–0.96	0.67	0.53–0.78
Observer II and IV	0.66	Good	0.45–0.86	0.50	0.37–0.63
Observer III and IV	0.60	Moderate	0.37–0.82	0.43	0.31–0.57

Abbreviation: CI,,confidence interval.

Table 5 Intraobserver assessment with weighted Kappa value

Classification system	Kappa	Kappa classification	95%CI	Agreement rate	95%CI
Outerbridge	**0.90**	**Excellent**			
Observer I	0.83	Excellent	0.70–0.96	0.72	0.58–0.82
Observer II	0.90	Excellent	0.84–0.94	0.73	0.60–0.84
Observer III	0.94	Excellent	0.89–0.96	0.83	0.71–0.91
Observer IV	0.91	Excellent	0.82–0.96	0.78	0.65–0.88
Beck	**0.90**	**Excellent**			
Observer I	0.90	Excellent	0.83–0.95	0.77	0.64–0.86
Observer II	0.92	Excellent	0.79–1.00	0.87	0.75–0.94
Observer III	0.87	Excellent	0.70–1.00	0.80	0.67–0.89
Observer IV	0.91	Excellent	0.80–1.00	0.87	0.75–0.94
Haddad	**0.93**	**Excellent**			
Observer I	0.93	Excellent	0.87–0.97	0.80	0.67–0.89
Observer II	0.89	Excellent	0.75–1.00	0.85	0.73–0.93
Observer III	0.94	Excellent	0.88–0.98	0.85	0.73–0.93
Observer IV	0.97	Excellent	0.94–0.99	0.88	0.77–0.95

Abbreviation: CI,,confidence interval.

weighted Kappa coefficients, the interobserver evaluation showed good agreement for the three classifications studied, whereas the intraobserver agreement was considered excellent for all systems evaluated.

Studies evaluating the reproducibility of hip chondral lesion scores have variable results. Amenabar et al,[3] when evaluating the reproducibility of the Outerbridge, Beck and Haddad classification systems in the diagnosis of arthroscopic chondral lesions, observed interobserver Kappa values of 0.28, 0.33 and 0.47, respectively. The agreement of the Outerbridge system was considered poor, while the agreement of the Beck and Haddad classifications was deemed moderate. The intraobserver Kappa values were 0.62, 0.63 and 0.68 for the Outerbridge, Beck and Haddad systems, respectively, all showing moderate agreement. In the present study, the weighted Kappa coefficient values were higher, which can be partially explained by the two previous meetings for discussion and training. In addition, Amenabar et al[3] performed a second round of assessment 4 months after the first one. In the present study, the time between assessment rounds was only 1 month; if the surgeons had classified the same lesions again after a longer time, the intraobserver reproducibility might have been possibly smaller.

In a retrospective study involving 40 videos and 4 observers, Nepple et al[15] evaluated the reproducibility of the Beck classification for labral and chondral acetabular lesions. Regarding chondral lesions, mean Kappa values of 0.65 (0.49 to 0.74) and 0.80 (0.68 to 0.86) were found for interobserver and intraobserver reproducibility, respectively. As in the present study, the interobserver agreement for the Beck classification was considered good, but the intraobserver reproducibility was also good, not excellent as in the present study. In the present study, the Beck classification had the highest interobserver reproducibility, with an average Kappa value of 0.78. Studying the Haddad classification, Konan et al[12] observed a high reproducibility with an intraclass correlation coefficient of 0.88 for the interobserver evaluation, and of 0.81 for the intraobserver assessment. In the present study, the Haddad classification presented the highest intraobserver reproducibility, with an average Kappa value of 0.93.

Lasmar et al[16] evaluated the reproducibility of the Outerbridge classification for arthroscopically diagnosed knee chondral lesions and found an interobserver Kappa value of 0.43, and an intraobserver Kappa value of 0.31, indicating moderate agreement. However, among the six observers, two residents were included, and the experience in knee surgery from another two observers was not clear. The presence of a larger interobserver Kappa value compared with the intraobserver value was unexpected. In the present study, all of the observers are hip surgeons experienced in arthroscopic procedures, which theoretically increases the likelihood of agreement among them. In another study that also evaluated the reproducibility of the Outerbridge classification for knee injuries, Cameron et al[17] found an interobserver Kappa value of 0.72, and an intraobserver Kappa value of 0.91 among physicians with $>$ 5 years of experience. These values are very close to those observed in the present study, which also included as observers only surgeons experienced in hip arthroscopy.

The present study shows that the Outerbridge, Beck and Haddad classification systems are reproducible enough to be applied in the clinical practice. However, in addition to reproducibility, a classification should adequately describe the lesions and provide therapeutic and prognostic guidance. The Outerbridge classification was not developed to evaluate acetabular chondral lesions and does not consider the hallmark IFA lesions, such as the wave signal (**fig. 4**) or the cartilage detachment with flap formation (**fig. 5**). Therefore, in our view, the Beck and Haddad classification

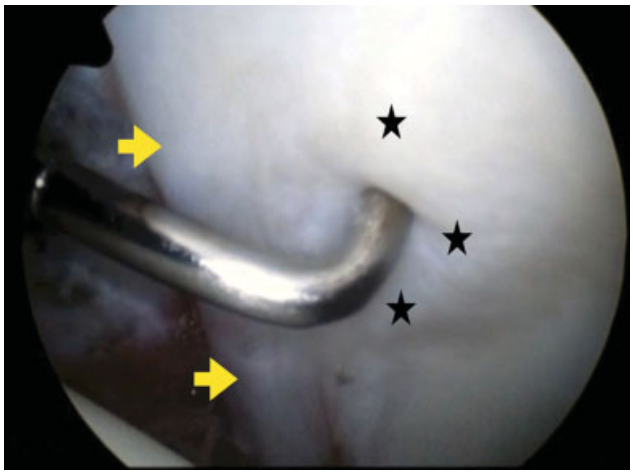

Fig. 4 Arthroscopic image of the hip indicating an acetabular cartilage lesion through a "wave sign" (stars) and acetabular labrum lesions (yellow arrow).

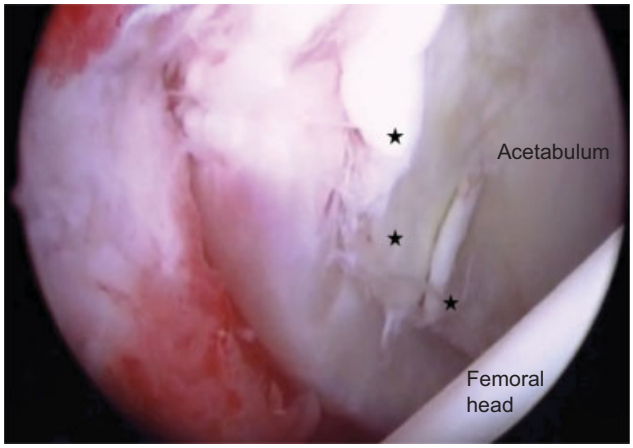

Fig. 5 Arthroscopic image of the hip indicating an acetabular cartilage lesion with flap formation (stars).

systems provide more accurate and consistent information regarding the types of acetabular lesions normally found; both can be routinely used depending on the familiarity of each surgeon.

The present study has some limitations. Being a retrospective study, it is subject to selection bias during the choice of the videos. Secondly, the observers included are part of the same team and attended regular meetings to discuss the topic, which may have increased the agreement rate among them. Finally, the videos were edited to last for a relatively short time compared with the total duration of the surgical procedure, so the observers had limited information for the classification of the lesions. If the observers had classified the lesions based on the complete surgical videos, the greater number of information could decrease the obtained reproducibility.

Conclusion

In our series, the Outerbridge, Beck and Haddad classification systems presented good interobserver reproducibility and excellent intraobserver reproducibility in the arthroscopic evaluation of acetabular chondral lesions.

Conflicts of Interests
The authors have no conflicts of interests to declare.

References

1 de Amorim Cabrita HA, de Castro Trindade CA, de Campos Gurgel HM, Leal RD, de Souza Marques RdaF. Hip arthroscopy. Rev Bras Ortop 2014;50(03):245–253
2 Sampson TG. Arthroscopic treatment for chondral lesions of the hip. Clin Sports Med 2011;30(02):331–348
3 Amenabar T, Piriz J, Mella C, Hetaimish BM, O'Donnell J. Reliability of 3 different arthroscopic classifications for chondral damage of the acetabulum. Arthroscopy 2015;31(08):1492–1496
4 Ganz R, Parvizi J, Beck M, Leunig M, Nötzli H, Siebenrock KA. Femoroacetabular impingement: a cause for osteoarthritis of the hip. Clin Orthop Relat Res 2003;(417):112–120
5 Bedi A, Kelly BT. Femoroacetabular impingement. J Bone Joint Surg Am 2013;95(01):82–92
6 Kowalczuk M, Yeung M, Simunovic N, Ayeni OR. Does femoroacetabular impingement contribute to the development of hip osteoarthritis? A systematic review. Sports Med Arthrosc Rev 2015;23(04):174–179
7 Yeung M, Khan M, Schreiber VM, et al. Global discrepancies in the diagnosis, surgical management, and investigation of femoroacetabular impingement. Arthroscopy 2014;30(12):1625–1633
8 Oliveira FA, Basile R, Pereira BC, Cunha RL. Avaliação da reprodutibilidade da classificação de Tronzo para fraturas intertrocantéricas do fêmur. Rev Bras Ortop 2014;49(06):581–585
9 Tönnis D. Congenital dysplasia and dislocation of the hip in children and adults. New York: Springer-Verlag; 1987
10 Outerbridge RE. The etiology of chondromalacia patellae. J Bone Joint Surg Br 1961;43-B:752–757
11 Beck M, Kalhor M, Leunig M, Ganz R. Hip morphology influences the pattern of damage to the acetabular cartilage: femoroacetabular impingement as a cause of early osteoarthritis of the hip. J Bone Joint Surg Br 2005;87(07):1012–1018
12 Konan S, Rayan F, Meermans G, Witt J, Haddad FS. Validation of the classification system for acetabular chondral lesions identified at arthroscopy in patients with femoroacetabular impingement. J Bone Joint Surg Br 2011;93(03):332–336
13 Cohen J. A coeficiente of agrément for nominal scales. Educ Psychol Meas 1960;20(01):37–46
14 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(01):159–174
15 Nepple JJ, Larson CM, Smith MV, et al. The reliability of arthroscopic classification of acetabular rim labrochondral disease. Am J Sports Med 2012;40(10):2224–2229
16 Lasmar NP, Lasmar RC, Vieira RB, de Oliveira JR, Scarpa AC. Assessment of the reproducibility of the Outerbridge and FSA classifications for chondral lesions of the knee. Rev Bras Ortop 2015;46(03):266–269
17 Cameron ML, Briggs KK, Steadman JR. Reproducibility and reliability of the outerbridge classification for grading chondral lesions of the knee arthroscopically. Am J Sports Med 2003;31(01):83–86