

In Brief

Classifications in Brief: Outerbridge Classification of Chondral Lesions

Casey Slattery BS, Christopher Y. Kweon MD

Received: 13 November 2017 / Accepted: 14 February 2018 / Published online: 15 March 2018
Copyright © 2018 by the Association of Bone and Joint Surgeons

History

Although cartilage lesions had been directly examined and described as far back as the early 20th century, the etiology of chondromalacia of the patella was not well understood when Outerbridge published his first paper on the subject in 1961 [15]. In this initial study, he evaluated the cartilage of the patella during 196 medial meniscectomies to better understand how chondromalacia progressed and which areas of the patella were primarily affected. He found that chondromalacia was most common on the medial facet as a result of constant friction with a rim on the upper border of the medial femoral condyle. He also noted the incidence of chondromalacia of the patella to be approximately 50% in patients who underwent open medial meniscectomy, even in the absence of symptoms. To better understand the etiology of chondromalacia of the patella, Outerbridge developed his classification system describing varying severity of cartilage lesions by direct visualization, which

he continued to use in his subsequent papers [15-17]. Since the introduction of Outerbridge's classification system originally designed for chondromalacia of the patella, it has been adapted to include the entire knee in 1989 and other joints since then [2, 8, 13].

In addition to Outerbridge's scheme, there are several other classification schemes describing chondral lesions. These include the modified Collins [6] and French Society of Arthroscopy (FSA) systems [13] designed for the knee as well as Beck's [3] and Konan's [9] designed for the hip. Aside from the studies referenced in this review, there is very little reported on the Collins or FSA classification systems. Collins' system was published before Outerbridge's original paper but, along with the FSA system, has failed to gain widespread popularity. The Beck scheme is based on findings during surgical dislocation of the hip and Konan's classification is fairly new with only two studies assessing its reliability [1]. Despite other proposed systems, the Outerbridge system continues to be the most widely used, which warrants investigation into its reliability.

Each author certifies that neither he, nor any member of his immediate family, has funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request. Each author certifies that his institution approved the reporting of this investigation and that all investigations were conducted in conformity with ethical principles of research.

Department of Orthopaedics & Sports Medicine, University of Washington, Seattle, WA, USA

C. Y. Kweon (✉), University of Washington Department of Orthopaedics & Sports Medicine 3800 Montlake Boulevard NE Campus Box 354060 Seattle, WA 98195-4060, USA email: ckweon@uw.edu

Purpose

In 1961, when the Outerbridge system was originally developed, it was used as a purely descriptive system to better understand the etiology of chondromalacia of the patella. Since then, it has been used to describe cartilage lesions in the knee, hip, and shoulder [2, 7, 8, 19]. The system is largely used to facilitate communication between surgeons. Although it has not been demonstrated to guide treatment, several studies have used the Outerbridge scheme to group patients for clinical research and for prognostic purposes [2, 7, 8, 19].

Accurately defining defect severity is also important for surgical planning and patient education.

Description

Based on direct visualization of the joint, either arthroscopic or open, the Outerbridge classification system was developed to be a simple, easy-to-use, and reproducible grading system of articular cartilage lesions. The system assigns a grade of 0 through IV to the chondral area of interest (Fig. 1). Grade 0 signifies normal cartilage. Grade I chondral lesions are characterized by softening and swelling, which often require tactile feedback with a probe or other instrument to assess. A Grade II lesion describes a partial-thickness defect with fissures that do not exceed 0.5 inches in diameter or reach subchondral bone. Grade III is fissuring of the cartilage with a diameter > 0.5 inches with an area reaching subchondral bone. The most severe is Grade IV, which includes erosion of the articular cartilage that exposes subchondral bone [15, 16].

Validation

Studies that have evaluated the reliability of Outerbridge's classification system either use arthroscopic video or another

imaging modality for comparison. The studies that have looked at the reproducibility of the scheme using arthroscopy videos have shown interobserver reliability ranging from a κ coefficient of 0.28 to 0.52 and intraobserver reproducibility ranging from a κ coefficient of 0.29 to 0.8 (Table 1) [1, 4, 5, 10, 11]. In these studies, Brismar et al. [4], Cameron et al. [5], Marx et al. [11], and Amenabar et al. [1] all used fully trained orthopaedic surgeons for reviewers, whereas Lasmar et al. [10] had two third-year residents along with four orthopaedic surgeons review their videos, demonstrating a clear intraobserver reliability discrepancy between the levels of training ($\kappa = -0.06$ versus 0.50). Cameron et al. [5] also found a discrepancy in reliability based on level of experience with the two surgeons in practice for > 5 years having an interobserver agreement of $\kappa = 0.72$ and those surgeons with less experience averaging $\kappa = 0.50$. This study also found a 68% concordance between the participating observers' arthroscopic evaluation and direct measurement with calipers (depth and width of lesions) at arthrotomy made by those same observers [5].

Brismar et al.'s study [4] compared the modified Collins and FSA classification systems as well as Outerbridge and found no difference among the three, concluding that none

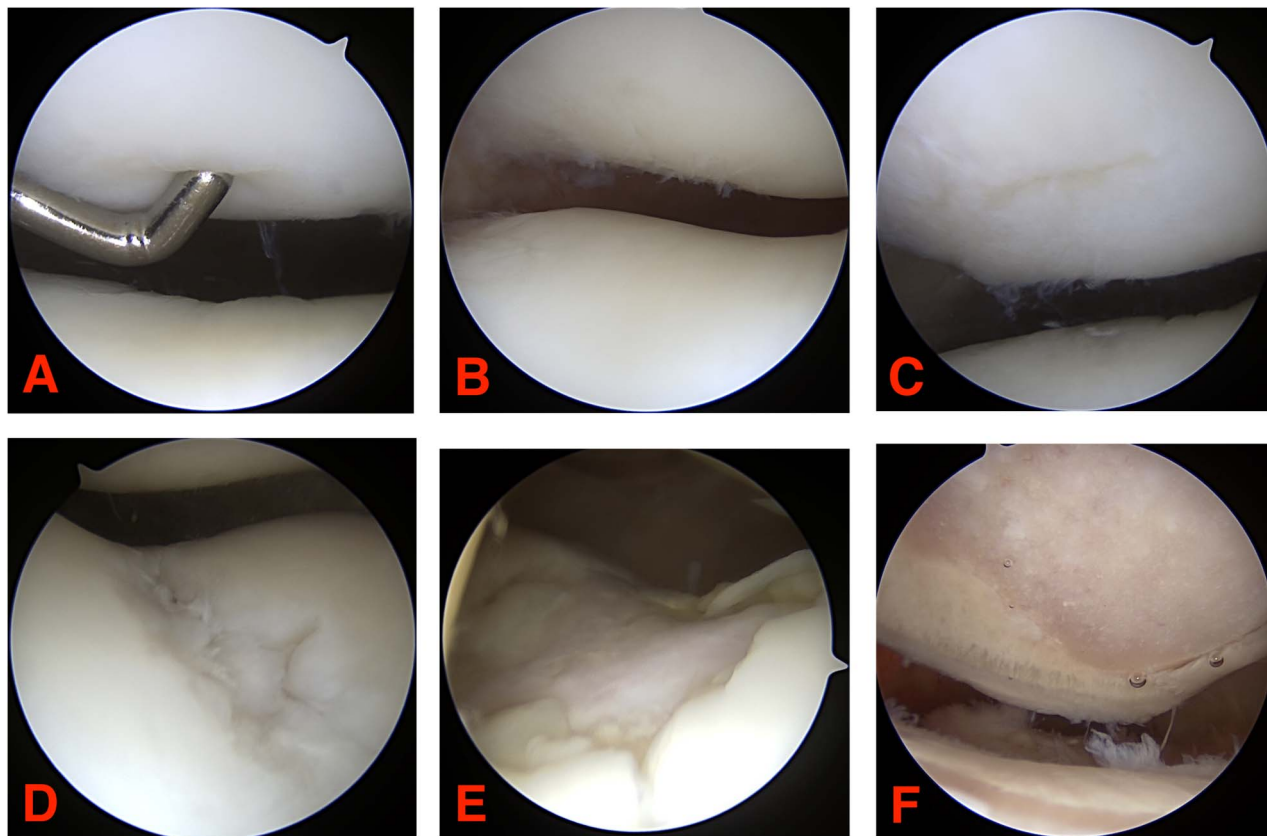


Fig. 1 A-F These intraoperative arthroscopic images demonstrate examples of Outerbridge classification grades: (A) Grade I; (B-C) Grade II; (D) Grade III; (E-F) Grade IV.

Table 1. Studies evaluating the reliability of the Outerbridge classification system of chondral lesions

Study	Year published	Number of reviewers	Number of patients	Method of comparison	Interobserver reliability*	Intraobserver reproducibility*
Omoumi et al. [14]	2017	5	5	CT arthrograms	0.35, fair	0.59–0.92
Amenabar et al. [1]	2015	4	40	Arthroscopic video	0.28, fair	0.62, substantial
Lasmar et al. [10]	2011	6	30	Arthroscopic video	0.43, moderate	0.29, fair
Marx et al. [11]	2005	6	22	Arthroscopic video	0.47, moderate	NA
Cameron et al. [5]	2003	9	6	Arthroscopic video	0.52, moderate	0.8, substantial
Brismar et al. [4]	2002	4	19	Arthroscopic video	0.47, moderate	0.54, moderate
Potter et al. [18]	1998	2	88	MR images	0.93, almost perfect	NA

*Reliability measured using κ coefficient; NA = not applicable.

of these classifications was sufficiently reliable for use in clinical research. Lasmar et al.'s study [10] also compared Outerbridge and FSA schemes with no difference between either interobserver or intraobserver reliability. The study by Amenabar et al. [1] evaluated chondral lesions of the hip using Outerbridge and two other classification systems designed for the hip (Beck [3] and Konan [9]). They found no difference between the systems regarding intraobserver reliability, but Konan's system was noted to have superior interobserver reliability in the hip. Lower reliability with the Outerbridge system compared with other schemes was believed to be a result of the specific chondral damage pattern usually caused by femoroacetabular impingement and the anatomy of the chondrolabral junction [1].

Studies that used imaging as a method of comparison (Table 1) found an interobserver reliability ranging from fair ($\kappa = 0.35$, CT arthrograms) to almost perfect ($\kappa = 0.93$, MR images) [14, 18]. Among these studies, Omoumi et al. [14], who used radiologists to evaluate CT arthrograms without a direct visual comparison, was the only study to test intraobserver reliability ($\kappa = 0.59$ – 0.92). This study found that more experienced radiologists in general had higher κ values for intraobserver reliability. The highest interobserver reliability for Outerbridge's scheme comes from Potter et al.'s [18] study that compared MR images of the knee with an arthroscopic evaluation. The two radiologists and three orthopaedic surgeons found an almost perfect (0.93) κ statistic.

The Outerbridge system has also proven to have some prognostic value. Sofu et al. [19] has shown Grade III and IV knee lesions to have worse visual analog scores and Lysholm scores after arthroscopic partial meniscectomy. Bateman et al. [2] demonstrated worse functional outcomes after arthroscopic shoulder posterior labral tear repairs in patients with Grade III lesions or higher. Kemp et al. [8] also found that patients who had Outerbridge Grade III and IV lesions found during hip arthroscopy for femoroacetabular impingement had worse pain and function at 18 months postsurgery compared with lower grade chondral lesions.

Limitations

Although widely used both in clinical and research settings over the past several decades, the Outerbridge classification system has several limitations. The most common criticism of this classification is its inconsistent and poor reproducibility among orthopaedic surgeons. The overall interobserver reliability ranged only from weak ($\kappa = 0.28$) [1] to moderate ($\kappa = 0.52$) [5], whereas intraobserver agreement was slightly better ranging from weak ($\kappa = 0.29$) [10] to substantial ($\kappa = 0.8$) [5]. However, some studies have mentioned that the amount of experience among reviewers affects the reliability of the system with more experienced surgeons having better reliability [5, 10]. Arthroscopy may also make it somewhat difficult to adequately differentiate the size of the lesion between Grades 2 and 3 as well as visualizing the softness and swelling needed to assign a Grade 1 [1]. Such variations in reliability suggest that the criteria for the Outerbridge system needs modification and/or advanced imaging (MRI) implemented into the scheme. The current crude macroscopic method used in Outerbridge grades may work to communicate cartilage lesion severity between surgeons, but the literature does not support its reliability for research purposes.

In the studies evaluating the reliability of the Outerbridge classification system through arthroscopic videos, there was a common limitation of small sample sizes, which ranged from six patients to 40 [1]. Additionally, there has been a relatively small number of studies validating the reliability of the Outerbridge classification system. In studies using direct visualization to assess this system, only five studies measured interobserver agreement and only four measured intraobserver agreement. Each study that evaluated the Outerbridge classification as the reference grading system used video recordings of knee arthroscopy, thus preventing grading surgeons from using tactile feedback as a cartilage assessment tool. This tactile feedback is especially critical because roughness and softening of the cartilage are important for appropriate grading [11]. Any future studies on the present Outerbridge

system's reliability should incorporate tactile feedback into the methodology, which may limit the study to only assessing interobserver reliability during arthroscopic surgery or the use of cadaver knees.

The Outerbridge classification system also does not provide a clear correlation with disease prognosis or a guide to treatment. There are only a few studies that have shown some prognostic value to the Outerbridge system [2, 8, 12, 19] and no studies were found in this review that discuss treatment guidance. Because these are two key features that a classification system should incorporate, their absence remains a major limitation for this system.

Conclusions

The inter- and intraobserver agreement for the Outerbridge classification system for chondral lesions ranges from fair to substantial. This inconsistent reliability remains a substantial limitation of this system. Although the Outerbridge scheme remains the most widespread classification system for grading cartilage lesions, it fails to guide treatment decisions and there is little evidence that it provides much prognostic information. To further evaluate Outerbridge's system, future research should include validation studies with larger sample sizes, methodology that allows for tactile feedback, and evaluation in a variety of joints for more accurate assessment of articular cartilage morphology. Outerbridge and similar macroscopic classification schemes that evaluate chondral lesions fail to provide the confidence needed for use in research settings. This system is > 50 years old and does not incorporate the advances in imaging technology over that timeframe. The best reliability found in this review compared arthroscopic and MR images. The authors recommend that the Outerbridge system and any future macroscopic grading system of chondral lesions need to incorporate advanced imaging (MRI) to achieve the reliability needed for a successful classification system.

References

1. 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;31:1492–1496.
2. Bateman DK, Black EM, Lazarus MD, Abboud JA. Outcomes following arthroscopic repair of posterior labral tears in patients older than 35 years. *Orthopedics*. 2017;40:e305–e311.
3. 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;87:1012–1018.
4. Brismar BH, Wredmark T, Movin T, Leandersson J, Svensson O. Observer reliability in the arthroscopic classification of osteoarthritis of the knee. *J Bone Joint Surg Br*. 2002;84:42–47.
5. 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;31:83–86.
6. Collins D. *The Pathology of Articular and Spinal Diseases*. London, UK: Edward Arnold and Co; 1949.
7. Curl WW, Krome J, Gordon ES, Rushing J, Smith BP, Poehling GG. Cartilage injuries: a review of 31,516 knee arthroscopies. *Arthroscopy*. 1997;13:456–460.
8. Kemp JL, Makdissi M, Schache AG, Pritchard MG, Pollard TCB, Crossley KM. Hip chondropathy at arthroscopy: prevalence and relationship to labral pathology, femoroacetabular impingement and patient-reported outcomes. *Br J Sports Med*. 2014;48:1102–1107.
9. 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;93:332–336.
10. Lasmar NP, Lasmar RCP, 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*. 2011;46:266–269.
11. Marx RG, Connor J, Lyman S, Amendola A, Andrish JT, Kaeding C, McCarty EC, Parker RD, Wright RW, Spindler KP. Multirater agreement of arthroscopic grading of knee articular cartilage. *Am J Sports Med*. 2005;33:1654–1657.
12. Moon H-K, Koh Y-G, Kim YC, Park Y-S, Jo S-B, Kwon S-K. Prognostic factors of arthroscopic pull-out repair for a posterior root tear of the medial meniscus. *Am J Sports Med*. 2012;40:1138–1143.
13. Noyes FR, Stabler CL. A system for grading articular cartilage lesions at arthroscopy. *Am J Sports Med*. 1989;17:505–513.
14. Omoumi P, Michoux N, Larbi A, Lacoste L, Lecouvet FE, Perlepe V, Vande Berg BC. Multirater agreement for grading the femoral and tibial cartilage surface lesions at CT arthrography and analysis of causes of disagreement. *Eur J Radiol*. 2017;88:95–101.
15. Outerbridge RE. The etiology of chondromalacia patellae. *J Bone Joint Surg Br*. 1961;43:752–757.
16. Outerbridge RE. Further studies on the etiology of chondromalacia patellae. *J Bone Joint Surg Br*. 1964;46:179–190.
17. Outerbridge RE, Dunlop JA. The problem of chondromalacia patellae. *Clin Orthop Relat Res*. 1975;110:177–196.
18. Potter HG, Linklater JM, Allen AA, Hannafin JA, Haas SB. Magnetic resonance imaging of articular cartilage in the knee. An evaluation with use of fast-spin-echo imaging. *J Bone Joint Surg Am*. 1998;80:1276–1284.
19. Sofu H, Oner A, Camurcu Y, Gursu S, Ucpunar H, Sahin V. Predictors of the clinical outcome after arthroscopic partial meniscectomy for acute trauma-related symptomatic medial meniscal tear in patients more than 60 years of age. *Arthroscopy*. 2016;32:1125–1132.