Reliability and Validity of the Arthroscopic International Cartilage Repair Society Classification System: Correlation With Histological Assessment of Depth

Tim Dwyer, M.B.B.S., Ph.D., F.R.A.C.S., F.R.C.S.C., C. Ryan Martin, M.D., F.R.A.C.S., Rita Kendra, M.D., F.R.C.P.C., Corey Sermer, M.Sc., Jaskarndip Chahal, M.D., M.Sc., F.R.C.S.C., M.B.A., Darrell Ogilvie-Harris, M.D., M.Sc., F.R.C.S.C., Daniel Whelan, M.D., M.Sc., F.R.C.S.C., Lucas Murnaghan, M.D., M.Ed., F.R.C.S.C., Aaron Nauth, M.D., M.Sc., F.R.C.S.C., and John Theodoropoulos, M.D., M.Sc., F.R.C.S.C.

Purpose: To determine the interobserver reliability of the International Cartilage Repair Society (ICRS) grading system of chondral lesions in cadavers, to determine the intraobserver reliability of the ICRS grading system comparing arthroscopy and video assessment, and to compare the arthroscopic ICRS grading system with histological grading of lesion depth. **Methods:** Eighteen lesions in 5 cadaveric knee specimens were arthroscopically graded by 7 fellowship-trained arthroscopic surgeons using the ICRS classification system. The arthroscopic video of each lesion was sent to the surgeons 6 weeks later for repeat grading and determination of intraobserver reliability. Lesions were biopsied, and the depth of the cartilage lesion was assessed. Reliability was calculated using intraclass correlations. **Results:** The interobserver reliability was 0.67 (95% confidence interval, 0.5-0.89) for the arthroscopic grading, and the intraobserver reliability with the video grading was 0.8 (95% confidence interval, 0.67-0.9). A high correlation was seen between the arthroscopic grading of depth and the histological grading of depth (0.91); on average, surgeons graded lesions using arthroscopy a mean of 0.37 (range, 0-0.86) deeper than the histological grade. **Conclusions:** The arthroscopic ICRS classification system has good interobserver and intraobserver reliability. A high correlation with histological assessment of depth provides evidence of validity for this classification system. **Clinical Relevance:** As cartilage lesions are treated on the basis of the arthroscopic ICRS classification, it is important to ascertain the reliability and validity of this method.

Accuracy in determining the size, location, and depth of chondral lesions in the knee is critical, as each of these factors is important in determining both the type and efficacy of any chosen treatment.[1,2] The most cartilage lesions are graded at arthroscopy, with several classification systems available; of these, the Outerbridge Classification[3] and the International Cartilage Repair Society (ICRS) grading system[4] are the most commonly used.[4,5]

The ICRS classification system is used to characterize cartilage injury on the basis of lesion area and depth.[4] Two previous studies have evaluated the reliability of the ICRS classification system during arthroscopy, with contrasting results. In 2010, Spahn et al.[6] reported the interobserver reliability of 4 experienced arthroscopy surgeons independently grading lesions in the medial compartment of the knee during live arthroscopy—in this study the interobserver agreement was poor (0.17). In 2011, Niemeyer et al.[1] compared arthroscopic ICRS grading of cartilage lesions with grading at arthrotomy, identifying an 80.9% consensus—in this study all lesions were ICRS grade 3 or 4.

From the Department of Surgery, Division of Orthopedics, University of Toronto Orthopaedic Sports Medicine (T.D., C.R.M., J.C., D.O-H., D.W., L.M., A.N., J.T.); Women's College Hospital (T.D., J.C., D.O-H., D.W., L.M., J.T.); and Mt Sinai Hospital (T.D., R.K., C.S., J.T.), Toronto, Ontario, Canada.

The authors report that they have no conflicts of interest in the authorship and publication of this article.

Received June 30, 2016; accepted December 5, 2016.

Address correspondence to Tim Dwyer, M.B.B.S., Ph.D., F.R.A.C.S., F.R.C.S.C., Women's College Hospital, 76 Grenville St., Toronto, Ontario M5S 1B2, Canada. E-mail: tim.dwyer@wchospital.ca

0749-8063/16614/$36.00

http://dx.doi.org/10.1016/j.arthro.2016.12.012

Table 1. Arthroscopic Grading of the Articular Lesions Based on International Cartilage Repair Society Score[4]

Grade	Description	Grade Subgroup
0	Normal intact cartilage	
1	Chondral softening and blistering, superficial lesions, fissures and cracks, soft indentation	Grade 1A—superficial lesions or softening Grade 1B—superficial fissures and lacerations
2	Fraying, lesions, and fissures extending down to <50% of cartilage depth	
3	Partial loss of cartilage thickness, cartilage lesions extending down >50% of cartilage depth as well as down to the calcified layer	Grade 3A—defect more than 50% but not down to the calcified layer Grade 3B—down to the calcified layer Grade 3C—down to but not through the subchondral bone plate Grade 3D—defect more than 50% with blisters
4	Full-thickness cartilage loss with exposure of the subchondral bone	Grade 4A—defect included the superficial subchondral bone plate Grade 4B—defect down to deep subchondral bone

There are a variety of histological grading systems developed for use in grading osteoarthritis[7-9] and cartilage repair.[10] These systems are used to determine the severity of chondral lesions, or to determine the quality of repair tissue in patients, especially in regard to animal and human trials of cartilage repair procedures.[10] Although many of the parameters used in these measures include descriptions of cell morphology and matrix staining, the ICRS 2 histology scoring system[10] includes grading of the depth of the cartilage lesion.

The purpose of this study was 3-fold: (i) to determine the interobserver reliability of the ICRS grading system of chondral lesions in cadavers, (ii) to determine the intraobserver reliability of the ICRS grading system comparing arthroscopy and video assessment, and (iii) to compare the arthroscopic ICRS grading system with histological grading of lesion depth. We hypothesized that the ICRS grading system would have good inter- and intraobserver reliability, as well as a good correlation with histological grading of lesion depth.

Methods

Ten whole, fresh-frozen, cadaveric legs were obtained from the anatomy department of the University of Toronto. Inclusion criteria were any age and sex of cadavers that had localized cartilage lesions. Exclusion criteria were any cadaveric knees with widespread cartilage changes, or those with no cartilage defects. A single surgeon (T.D.) performed arthroscopy of 10 cadaveric knees, and selected cartilage lesions that were estimated to be evenly distributed between all 4 ICRS grades—a minimum of 4 lesions in each grade were selected. No lesions were created by surgeons for the purpose of this study. The same surgeon also recorded an arthroscopic video for each lesion, using a probe to provide information regarding the depth of each lesion.

After this, each cartilage lesion was graded arthroscopically by 7 fellowship-trained arthroscopic surgeons (T.D., J.C., D.O-H., D.W., L.M., A.N., J.T.) using standard anterolateral and anteromedial portals. Grading was performed using the ICRS grading system; this system was reviewed in detail with each surgeon, with a description of the ICRS grading system provided at each arthroscopic station (Table 1).[4,11] Surgeons were also asked to consider the subgroups of each grade during the arthroscopic grading. The arthroscopic video of each lesion was sent to each surgeon 6 weeks later for repeat grading—all surgeons were blinded from their original grading of each lesion.

After arthroscopic grading, each of the cartilage lesions identified was biopsied at the deepest aspect of each chondral lesion using arthroscopic mosaicplasty instruments (Smith & Nephew, Andover, MA) by a single surgeon. Lesions in the tibial plateau were accessed using hyperflexion of the knee, using a spinal needle to identify the best orientation of an accessary arthroscopic portal. Full-thickness osteochondral cylinders including subchondral bone with a diameter of 4.5 mm were taken, and fixed in 10% formalin for 24 hours before being decalcified and paraffin embedded. After this, the embedded samples were cut into 5 μm, and placed onto saline coated glass slides, with standard H&E and Toluidine blue protocols used for sample staining. The depth of each cartilage lesion, in correlation with the arthroscopic ICRS classification system, was performed by a pathologist experienced at grading cartilage lesions (R.K.).

Statistical Analysis

The sample size was calculated using an estimated interobserver reliability of 0.6, taken from a study using the Outerbridge classification.[12] Using an alpha of 0.5 and a power of 80%, the sample size was calculated to be 18. A minimum of 4 cartilage lesions in each of the 4 ICRS grades were selected for study. Interobserver reliability for surgeons' arthroscopic measurement of cartilage lesions was calculated using the interclass correlation coefficients (ICC), whereas the intraobserver reliability was calculated by comparing each surgeon's arthroscopic measurement with his or her grading of each lesion via a video review. The correlation between the arthroscopic grading of each lesion and the histological grade of lesion depth was also calculated using ICC. ICC was calculated using a 2-way random effects

Table 2. Characteristics of Each Cartilage Lesion, With Mean Arthroscopic and Video ICRS Grading and Corresponding Histological Grading of Depth

Lesion	Location	ICRS Arthroscopic Grade Mean (SD)	ICRS Video Grade Mean (SD)	Histological Grading of Depth
6952	LTP	2 (0.58)	2.71 (0.49)	2
6952	Trochlea	1 (0)	1.57 (0.53)	1
6909	MFC	2.86 (0.38)	3.14 (0.69)	2
6069	MTP	2.86 (0.38)	2.86 (0.38)	3
6909	LTP	3 (0.58)	1.86 (0.69)	3
6967R	LTP	3.43 (0.53)	3.29 (0.49)	–
6967R	LFC	1.14 (0.69)	0.86 (0.38)	1
6967R	MFC	3.43 (0.53)	3.57 (0.53)	3
6967R	MTP	1.86 (0.38)	2.29 (0.38)	–
6967R	Trochlea	3.29 (0.49)	3.43 (0.53)	3
6982	MFC	2.86 (0.38)	3.14 (0.38)	2
6982	Trochlea	1.29 (0.49)	0.86 (0.38)	1
6982	LFC	1 (0)	1.14 (0.69)	1
6967L	MFC	2.71 (0.76)	3.14 (0.69)	2
6967L	MTP	2.71 (0.76)	2.57 (0.53)	–
6967L	LTP	1.86 (0.69)	2.00 (0.58)	1
6967L	LFC	1.57 (1.13)	1.29 (0.49)	1
6967L	Trochlea	1.71 (0.49)	2.14 (0.69)	1

ICRS, International Cartilage Repair Society; LFC, lateral femoral condyle; LTP, lateral tibial plateau; MFC, medial femoral condyle; MTP, medial tibial plateau; SD, standard deviation.

model (ICC 2:k). Cohen's kappa was used to assess the degree of classification of the subgrades.

Results

Eighteen lesions in 5 cadaveric knee specimens were selected. The locations of the 18 lesions identified at the time of arthroscopy were as follows: trochlea n = 4, medial tibial plateau n = 3, lateral tibial plateau n = 4, lateral femoral condyle n = 3, and medial femoral condyle n = 4. No focal patella chondral lesions were identified in the 5 cadaveric knees. The mean grading of each lesion by all 3 methods is seen in Table 2. The interobserver reliability was 0.67 (95% confidence interval, 0.5-0.89) for the arthroscopic grading, and the intraobserver reliability with the video grading was 0.8 (95% confidence interval, 0.67-0.9). With regard to subgrade classification, Cohen's kappa showed poor agreement (0.27).

Histologic analysis was not available for 3 of the lesions because the biopsy did not go down to the level of the bone. For the remaining 15 lesions, a high correlation was seen between the arthroscopic grading of depth and the histological grading of depth (0.91); on average, surgeons graded lesions using arthroscopy a mean of 0.37 (range, 0-0.86) deeper than the histological grade. Examples of arthroscopic images in comparison with their histological appearance are seen in Figures 1 to 3.

Discussion

The major finding of this study was evidence of good inter- and intraobserver reliability for the use of the ICRS cartilage grading system, when both arthroscopic and video assessment of cadaveric lesions were performed. There was also evidence of a high correlation with histological assessment of depth, providing evidence of validity for the ICRS classification system.

The ability to reliably grade cartilage lesions at arthroscopy is important, as this method of assessment is used to guide treatment options.[13] The arthroscopic grading of articular cartilage in the knee is known to be affected by multiple factors, including localization of lesions,[14] depth,[15] size,[13] as well as rater experience.[12,16] Other issues that can affect the grading of lesions include the natural variation of the cartilage thickness throughout the knee, as there can be up to a 3-fold difference between weight-bearing and non–weight-bearing regions.[17,18]

At this time it is thought that the arthroscopic evaluation of articular cartilage lesions is the gold standard.[19,20] However, some authors have questioned the use of arthroscopy,[5] due to evidence of inaccuracy of grading both defect size[1] and depth.[6] Potential reasons for this inaccuracy include the magnification of lesions at arthroscopy, as well as inherent difficulty evaluating the depth of lesions relative to subchondral bone, an important component of the ICRS grading system.[1]

The results of this study identified that there was good interobserver reliability when using arthroscopy to grade chondral lesions in cadavers using the ICRS classification. Two previous studies have evaluated the ICRS classification, with significant discrepancies in their findings. In 2010, Spahn et al.[6] asked 4 experienced arthroscopy surgeons to rate lesions in the medial compartment using the ICRS grading system—in only 10% (6/60) did all 4 surgeons grade the cartilage lesions

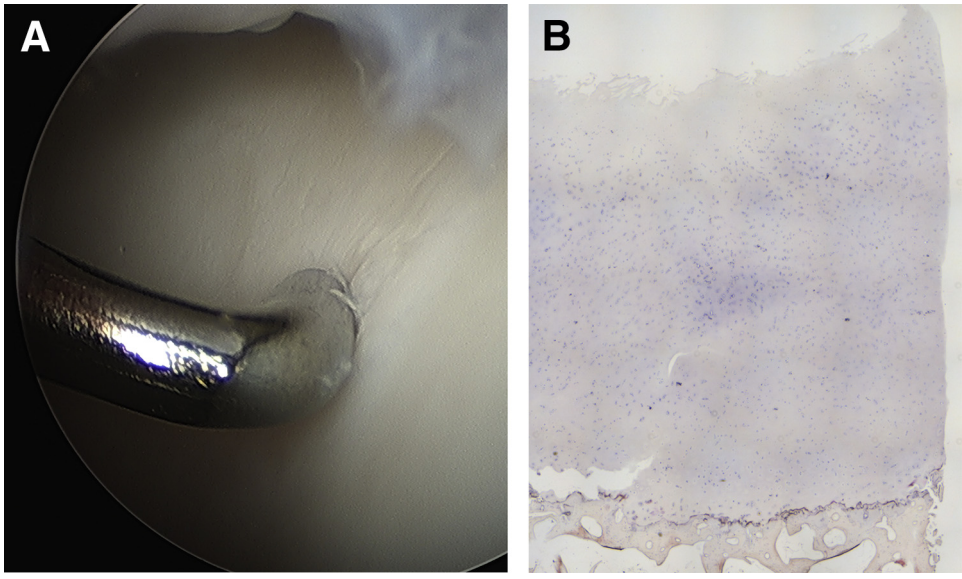

Fig 1. Screenshot from the video of a trochlea lesion (A) in a left knee, graded by surgeons as mean International Cartilage Repair Society (ICRS) grade 1.29 using arthroscopy, and deemed to be ICRS grade 1 using histology (B).

the same. This study was limited by a low number of high-grade lesions available for grading. In 2011, Niemeyer et al.[1] compared arthroscopic ICRS grading of cartilage lesions with grading at arthrotomy, identifying an 80.9% consensus—this high level of agreement was likely affected by the fact that all lesions were being treated with a cartilage implantation procedure, and were thus more likely to be only high-grade (3 and 4) lesions. In our study, a deliberative attempt was made to include lesions that covered the spectrum of the ICRS classification system; however, the reason for the difference between our findings and the low reliability identified by Spahn et al.[6] is unknown. The results of these studies may suggest that the classification of higher grade chondral lesions is more reliable; we did not perform differential reliability measures for the high- and low-grade lesions separately because of the limited numbers in our study, and thus cannot comment on this specifically.

There are also mixed results in the literature regarding the reliability of the Outerbridge classification system. Marx et al.[14] described substantial interobserver reliability (with the exclusion of tibial plateau lesions) when using videotaped assessment of chondral lesions. However, both Cameron et al.[12] and Brismar et al.[15] reported only fair to good interobserver reliability of the Outerbridge system; one possible explanation for this discrepancy is that the study by Marx et al.[14] assessed only grade 2 and 3 cartilage lesions, which may have increased the reliability of their study. The studies by both Cameron et al.[12] and Brismar et al.[15] also reported on intraobserver reliability of the Outerbridge classification; whereas the study by Cameron et al. identified excellent agreement (kappa 0.80), it was only fair to good in the study by Brismar et al. (kappa, 0.42-0.62). The reason for this discrepancy is unknown, although Cameron et al. did identify that the reliability increased with the experience of the surgeon. It is important to note that the major difference between the Outerbridge and ICRS classifications is that Outerbridge grade 2 and 3 lesions are distinguished by size, rather than by depth; as the decision to treat

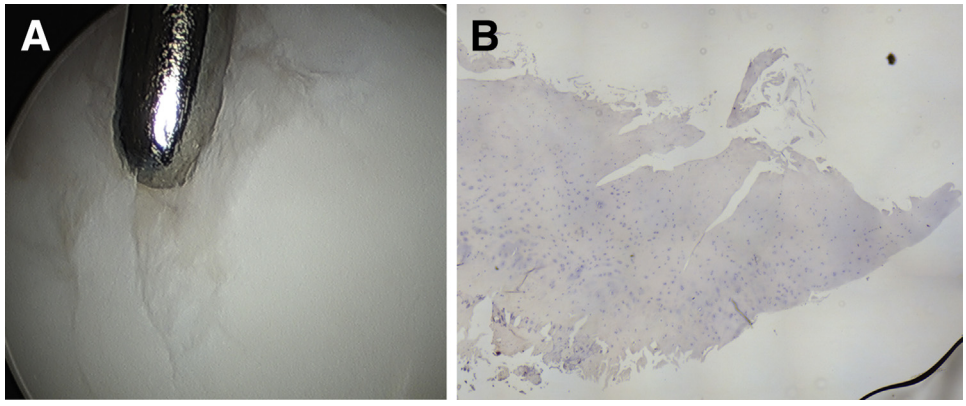

Fig 2. Screenshot from the video of a medial femoral condyle lesion (A) in a left knee, graded as mean International Cartilage Repair Society (ICRS) grade of 2.71 by surgeons using arthroscopy, and deemed to be ICRS grade 2 using histology (B).

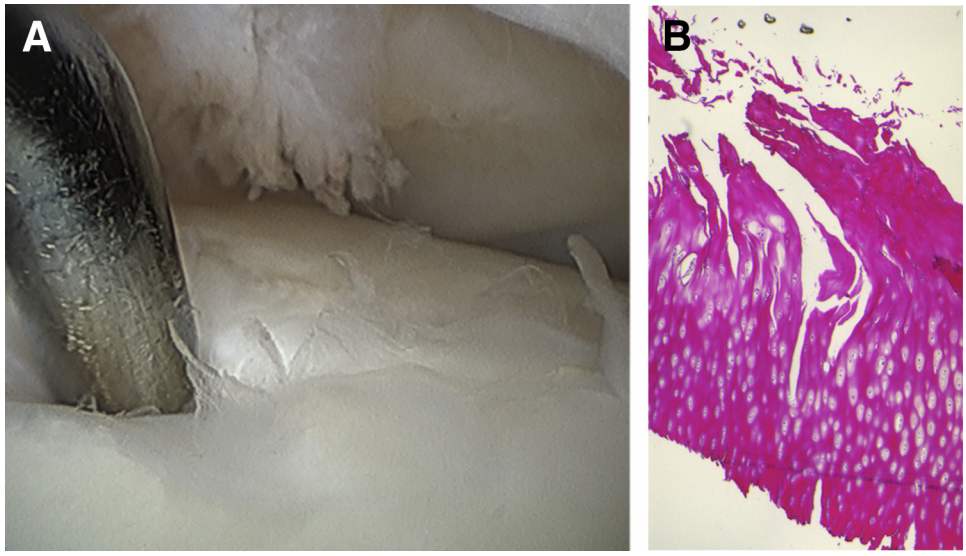

Fig 3. Screenshot from the video of a lateral tibial plateau image (A) in a left knee, graded as mean International Cartilage Repair Society (ICRS) grade of 3.43 by surgeons using arthroscopy, and deemed to be ICRS grade 3 using histology (B).

cartilage lesions surgically hinges primarily on their depth,[2,11,21] it may be that the ICRS grading system is more appropriate.

Our study identified that the agreement between surgeons with regard to the subgrades was poor. This suggests that although surgeons are able to reliably classify lesions using the main 4 categories of the ICRS classification, they have limited ability to use the subgrades with any degree of consistency. Surgeons should be aware of this finding, and place limited reliance on the use of these subgrades. Fortunately, it is unlikely that classifying a lesion as 3B rather than 3C would change the management of these types of chondral lesions.

An important finding of this study was that there was evidence of a high correlation between the arthroscopic grading of lesion depth and the histological grading of lesion depth. In this study, the pathologist was asked to grade the depth of each lesion in correlation with the ICRS arthroscopic classification. Although surgeons tended to grade the lesions slightly higher at arthroscopy in comparison with histology, a high correlation was seen between these 2 methods of assessing lesion depth. The tendency to overcall the depth of lesions using arthroscopy is relatively easy to explain; although the depth of lesions can be estimated using a probe, in the absence of lesions that extend to subchondral bone, surgeons are forced to estimate the depth of cartilage that remains. Despite this obvious technical challenge in the arthroscopic grading of lesions, the high correlation with histological assessment of depth encountered in this study is encouraging for surgeons, as the most treatment algorithms are based on the perceived cartilage depth of lesions. Although the reliability of histological grading of lesion depth is unknown, there is evidence for the high reliability of most histological grading systems used by experts.[22] For this reason, we believe that the results of our study allow the continued use of arthroscopic grading of cartilage lesions as a gold standard.

Limitations

In this study, lesions were graded from 1 to 4 using the ICRS grading system, which enabled a mean of the ICRS grade for all surgeons to be calculated for both the arthroscopic and video grading, and compared with the histological measurement of depth. It would also have been possible for surgeons to estimate the depth of the cartilage lesions as per the ICRS system, that is, 25%, 75%, and calculate a mean using these measurements. The accuracy of such a measurement technique is unknown. It is also important to note that although many histological scoring systems have been described and been shown to be reliable, when quantifying the degree of articular degeneration, as well as to grade articular cartilage repair, the most of these scoring systems are made up of multiple components, rather than just depth of the lesion.[7,22,23]

Other limitations of this study include the use of lesions in cadaveric specimens, as opposed to lesions in live patients; for this reason some of the lesions were focal, whereas some of the cartilage lesions were more widespread in keeping with osteoarthritic changes. It may be that osteoarthritic changes represent a continuum of changes, as opposed to more consistent changes in a focal cartilage lesion.[15] Furthermore, most of these lesions will have been degenerative, as opposed to the traumatic chondral lesions encountered in young patients with sports injuries. Although these results should be replicated in live patients, it is not appropriate to biopsy low-grade lesions in patients with asymptomatic and low-grade lesion. The lesions studied were in multiple locations through the knee, which may introduce some variability; because no patella chondral lesions were included, we cannot comment on the reliability of

grading lesions in this area. It is important to note that this study focused on the reliability of grading the depth of lesions—surgeons were not asked to size the lesions. Another limitation was the use of video recording to provide a measure of intraobserver reliability; although arthroscopic grading in real time allows the added benefit of tactile feedback provided by the use of a probe, in this study video grading was shown to be reliable, consistent with previous studies.[14] Reliability calculations for each grade of chondral lesion were not performed due to the low number of lesions in each group.

Finally, an important limitation was that only 15 of 18 lesions were sampled adequately. This was because the samples were taken using arthroscopic techniques, with some lesions hard to sample—the posterior aspect of the tibial plateau being an example. Lesions were not sampled open, as it was felt that arthroscopic biopsy may prove to be a useful adjunct to the management of cartilage lesions, should the ICRS grading system be found to be of limited reliability. The addition of a further 3 samples may have changed the findings of our study. It is also possible that the small volume of lesion sampled may not represent the entire lesion; however, the surgeon responsible deliberately sampled the deepest aspect of each lesion, and was responsible for providing the videotaped probing of each lesion.

Conclusions

The arthroscopic ICRS classification system has good interobserver and intraobserver reliability. A high correlation with histological assessment of depth provides evidence of validity for this classification system.

References

1. Niemeyer P, Pestka JM, Erggelet C, Steinwachs M, Salzmann GM, Sudkamp NP. Comparison of arthroscopic and open assessment of size and grade of cartilage defects of the knee. *Arthroscopy* 2011;27:46-51.
2. Cole BJ, Pascual-Garrido C, Grumet RC. Surgical management of articular cartilage defects in the knee. *J Bone Joint Surg Am* 2009;91:1778-1790.
3. Outerbridge RE. The etiology of chondromalacia patellae. *J Bone Joint Surg Br* 1961;43:752-757.
4. Brittberg M, Winalski CS. Evaluation of cartilage injuries and repair. *J Bone Joint Surg Am* 2003;85:58-69 (suppl 2).
5. Spahn G, Klinger HM, Hofmann GO. How valid is the arthroscopic diagnosis of cartilage lesions? Results of an opinion survey among highly experienced arthroscopic surgeons. *Arch Orthop Trauma Surg* 2009;129:1117-1121.
6. Spahn G, Klinger HM, Baums M, et al. Near-infrared spectroscopy for arthroscopic evaluation of cartilage lesions: Results of a blinded, prospective, interobserver study. *Am J Sports Med* 2010;38:2516-2521.
7. Mankin HJ, Dorfman H, Lippiello L, Zarins A. Biochemical and metabolic abnormalities in articular cartilage from osteo-arthritic human hips. II. Correlation of morphology with biochemical and metabolic data. *J Bone Joint Surg Am* 1971;53:523-537.
8. Ostergaard K, Petersen J, Andersen CB, Bendtzen K, Salter DM. Histologic/histochemical grading system for osteoarthritic articular cartilage: Reproducibility and validity. *Arthritis Rheum* 1997;40:1766-1771.
9. Pritzker KP, Gay S, Jimenez SA, et al. Osteoarthritis cartilage histopathology: Grading and staging. *Osteoarthritis Cartilage* 2006;14:13-29.
10. Mainil-Varlet P, Van Damme B, Nesic D, Knutsen G, Kandel R, Roberts S. A new histology scoring system for the assessment of the quality of human cartilage repair: ICRS II. *Am J Sports Med* 2010;38:880-890.
11. Brittberg M, Peterson L, Sjogren-Jansson E, Tallheden T, Lindahl A. Articular cartilage engineering with autologous chondrocyte transplantation. A review of recent developments. *J Bone Joint Surg Am* 2003;85:109-115 (suppl 3).
12. 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.
13. Hunt N, Sanchez-Ballester J, Pandit R, Thomas R, Strachan R. Chondral lesions of the knee: A new localization method and correlation with associated pathology. *Arthroscopy* 2001;17:481-490.
14. Marx RG, Connor J, Lyman S, et al. Multirater agreement of arthroscopic grading of knee articular cartilage. *Am J Sports Med* 2005;33:1654-1657.
15. 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.
16. Javed A, Siddique M, Vaghela M, Hui AC. Interobserver variations in intra-articular evaluation during arthroscopy of the knee. *J Bone Joint Surg Br* 2002;84:48-49.
17. Hall FM, Wyshak G. Thickness of articular cartilage in the normal knee. *J Bone Joint Surg Am* 1980;62:408-413.
18. Muller-Gerbl M, Schulte E, Putz R. The thickness of the calcified layer of articular cartilage: A function of the load supported? *J Anat* 1987;154:103-111.
19. 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.
20. Figueroa D, Calvo R, Vaisman A, Carrasco MA, Moraga C, Delgado I. Knee chondral lesions: Incidence and correlation between arthroscopic and magnetic resonance findings. *Arthroscopy* 2007;23:312-315.
21. Messner K, Gillquist J. Cartilage repair. A critical review. *Acta Orthop Scand* 1996;67:523-529.
22. Bonasia DE, Marmotti A, Massa AD, et al. Intra- and inter-observer reliability of ten major histological scoring systems used for the evaluation of in vivo cartilage repair. *Knee Surg Sports Traumatol Arthrosc* 2015;23:2484-2493.
23. Otte P. [Degeneration of joint cartilage. Clinical and radiological aspects]. *Munch Med Wochenschr* 1968;110:2677-2683 [in German].