The Reliability of Arthroscopic Classification of Acetabular Rim Labrochondral Disease
Jeffrey J. Nepple, Christopher M. Larson, Matthew V. Smith, Young-Jo Kim, Ira Zaltz, Rafael J. Sierra and John C. Clohisy

DOI: 10.1177/0363546512457157

The online version of this article can be found at:
http://ajs.sagepub.com/content/40/10/2224

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
American Orthopaedic Society for Sports Medicine

Additional services and information for The American Journal of Sports Medicine can be found at:

Email Alerts: http://ajs.sagepub.com/cgi/alerts
Subscriptions: http://ajs.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

Version of Record - Sep 28, 2012
OnlineFirst Version of Record - Aug 27, 2012
What is This?
The Reliability of Arthroscopic Classification of Acetabular Rim Labrochondral Disease

Jeffrey J. Nepple,* MD, Christopher M. Larson,† MD, Matthew V. Smith,* MD, Young-Jo Kim,‡ MD, Ira Zaltz,§ MD, Rafael J. Sierra,** MD, and John C. Clohisy,* MD

Investigation performed at the Department of Orthopaedic Surgery, Washington University School of Medicine, St Louis, Missouri, in association with the ANCHOR (Academic Network for Conservational Hip Outcomes Research) Study Group

Background: The results of surgical treatment for femoroacetabular impingement have been increasingly reported, and more advanced intra-articular disease has been identified as an important predictive factor of outcome. Yet, the reliability of arthroscopic hip disease classification has not been well defined.

Purpose/Hypothesis: To determine the intraobserver and interobserver reliability of the Beck classification of labral and articular cartilage disease (anterior-superior acetabular rim) encountered in hip arthroscopy. Secondly, we identified the sources of poor reliability that may be improved with future disease classification schemes. Our hypothesis was that the Beck classification of labral and chondral lesions would demonstrate substantial reliability, while the differentiation of early forms of disease would be a common source of disagreement.

Study Design: Cohort study; Level of evidence, 3.

Methods: Four experienced hip arthroscopists reviewed standardized arthroscopic videos of 40 cases. Arthroscopic findings at the anterior-superior acetabular rim were classified using the Beck classification of labral and articular cartilage disease. Repeat classification of videos was performed at least 2 weeks later. The reliability of arthroscopic classification was defined using the average weighted Cohen \( k \) values and agreement rates.

Results: Arthroscopic classification of labral disease using the Beck classification demonstrated moderate to substantial interobserver reliability (average \( k = 0.62 \); range, 0.48–0.78) and an overall agreement rate of 81.7%. Intraobserver reliability showed a similar level of reliability (average \( k = 0.65 \); agreement rate, 80.6%). The differentiation between labral degeneration and labral detachment was a common source of disagreement. Similarly, the Beck classification of articular cartilage disease had moderate to substantial interobserver reliability (average \( k = 0.65 \); range, 0.49–0.78) and overall agreement rate of 57.5%. Intraobserver reliability showed a slightly better level of reliability (average \( k = 0.80 \); agreement rate, 77.5%). The differentiation between articular cartilage malacia and debonding was a common source of disagreement.

Conclusion: The arthroscopic classification of acetabular rim disease with the Beck classification has substantial interobserver reliability. This level of reliability is similar to previously reported arthroscopic disease classifications in the knee and shoulder and seems appropriate for future outcome reporting. Future classifications that eliminate common sources of disagreement may further improve the reliability.

Keywords: hip arthroscopy; femoroacetabular impingement; reliability

Arthroscopic treatment of intra-articular hip disease has become increasingly common with the improved understanding of femoroacetabular impingement (FAI). In FAI, the labrochondral junction of the anterior-superior acetabular rim is the most common site of injury, leading to adjacent articular cartilage and labral lesions. Classification of cartilage and labral disease has also evolved. Articular cartilage injury has been classically described by Outerbridge, while our understanding of labral pathological abnormalities has evolved from initial studies by Lage et al., McCarthy et al., and Seldes et al. Abnormalities of the labrum and articular cartilage are now commonly classified using the Beck classification systems (Table 1). Labral condition is classified as (1) normal, (2) degeneration, (3) full-thickness tear, (4) detachment, or (5) ossification. Articular cartilage disease is classified as (1) normal, (2) malacia, (3) debonding, (4) cleavage, or (5) defect. The Beck classification schemes were initially described for open surgical dislocation but have been applied to arthroscopic interventions as well.
Short-term outcomes of the treatment of FAI are being frequently reported with clinical improvement noted in most patients. Byrd and Jones found improvements to be generally maintained at 10-year follow-up. Yet, a subgroup of patients has suboptimal outcomes. Advanced intra-articular disease has been proposed as a predictive factor for poorer clinical results. Therefore, establishing the reliability of arthroscopic classification of labral and articular cartilage conditions is important for identifying disease characteristics that have prognostic value. An effective classification scheme should not only adequately describe the extent of disease present but also help guide the treatment of disease and potentially predict outcome. Additionally, reliable disease classification systems will facilitate the comparison of results from different centers, surgeons, and surgical techniques. Investigation of the reliability of current standards of disease classification will also assist in determining which disease characteristics can and cannot be reliably classified and may allow for the development of more reliable classification systems. While the reliability of arthroscopic classification has been reported in the knee and shoulder, to our knowledge, it has not been previously established for the hip.

The purpose of this study was to determine the intraobserver and interobserver reliability of the Beck and modified Beck classification schemes of labral and articular cartilage disease (acetabular rim) encountered in hip arthroscopy. Secondly, we attempted to identify sources of poor reliability that may be improved with future disease classification schemes. Our hypothesis was that the Beck classification of labral and chondral conditions would demonstrate moderate reliability, while the differentiation of early forms of disease would be a common source of disagreement.

MATERIALS AND METHODS

A reliability study of arthroscopic findings of labrochondral disease at the anterior-superior acetabular rim was performed. All patients eligible for inclusion in the study were undergoing arthroscopic treatment for presumed intra-articular hip lesions, most commonly FAI, and did not have radiographic findings of advanced osteoarthritis (Tonnis ≥2). The senior author generally recorded video of a systematic diagnostic arthroscopic examination during all procedures.

A hip arthroscopic procedure was performed with the patient in the supine position, utilizing the midanterior and anterolateral portals. The standardized videos were obtained through an anterolateral viewing portal, while the midanterior portal was utilized for arthroscopic probing. All videos included a standardized arthroscopic examination focusing on the anterior-superior labrum and articular cartilage, which included probing and visualization of the acetabular labrum, labrochondral junction, and articular cartilage. Systematic probing of the articular cartilage, labrochondral junction, articular side of the labrum, and capsular side of the labrum was performed.

A large database of videos from approximately 100 consecutive arthroscopic surgeries was reviewed by 2 authors not involved in the actual readings. Cases were excluded for poor quality videos, inadequate visualization, or inadequate probing of pathological lesions. Videos were chosen to span the spectrum of intra-articular findings and to

### TABLE 1
Beck Classification of Labral and Articular Cartilage Lesions

<table>
<thead>
<tr>
<th>Lesion</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labrum</td>
<td>Macroscopically sound labrum</td>
</tr>
<tr>
<td>Normal</td>
<td>Macroscopically sound cartilage</td>
</tr>
<tr>
<td>Degeneration</td>
<td>Thinning or localized hypertrophy, fraying, discoloration</td>
</tr>
<tr>
<td>Full-thickness tear</td>
<td>Complete avulsion from the acetabular rim</td>
</tr>
<tr>
<td>Detachment</td>
<td>Separation between the acetabular and labral cartilage, preserved attachment to bone</td>
</tr>
<tr>
<td>Ossification</td>
<td>Osseous metaplasia, localized or circumferential</td>
</tr>
<tr>
<td>Articular cartilage</td>
<td>Macroscopically sound cartilage</td>
</tr>
<tr>
<td>Normal</td>
<td>Macroscopically sound cartilage</td>
</tr>
<tr>
<td>Malacia</td>
<td>Roughening of surface, fibrillation</td>
</tr>
<tr>
<td>Debonding</td>
<td>Loss of fixation to the subchondral bone, macroscopically sound cartilage, carpet phenomenon</td>
</tr>
<tr>
<td>Cleavage</td>
<td>Loss of fixation to the subchondral bone: frayed edges, thinning of cartilage, flap</td>
</tr>
<tr>
<td>Defect</td>
<td>Full-thickness defect</td>
</tr>
</tbody>
</table>

References 1, 4, 7, 9, 12, 17, 19, 21, 23.

Address correspondence to John C. Clohisy, MD, Department of Orthopaedic Surgery, Washington University School of Medicine, One Barnes-Hospital Plaza, Suite 11300 West Pavilion, Campus Box 8233, St Louis, MO 63110 (e-mail: clohisyj@wudosis.wustl.edu).

1Department of Orthopaedic Surgery, Washington University School of Medicine, St Louis, Missouri.

2Minnesota Orthopedic Sports Medicine Institute, Twin Cities Orthopedics, Edina, Minnesota.

3Department of Orthopaedic Surgery, Children’s Hospital Boston, Boston, Massachusetts.

4William Beaumont Hospital, Royal Oak, Michigan.

5Department of Orthopaedic Surgery, Mayo Clinic, Rochester, Minnesota.

One or more of the authors has declared the following potential conflict of interest or source of funding: John C. Clohisy received funding support for this study from a National Football League Charities grant and the Curing Hip Disease Fund. Dr Clohisy has a potential conflict of interest in Biomet and Zimmer Inc. Christopher M. Larson is a consultant for Smith & Nephew and A2 Surgical and is a stockholder in A2 Surgical.
ensure a high quality of visualization. Specific attention was paid to ensure several cases with a normal labrum, ossified labrum, normal cartilage, and articular cartilage defect were included. Forty patients were included in the study. Videos were generally 30 to 60 seconds in length.

Four surgeons familiar with hip arthroscopy (>100 hip arthroscopy cases) participated in the study. Readers participated in a telephone conference discussion of the relevant classification schemes as well as an in-person group session discussing the classification schemes and practice video cases. An initial reading was followed by a second reading of reordered cases more than 2 weeks later.

Arthroscopic findings were classified into the 5 types of labral and articular cartilage states by the Beck criteria2 (Table 1). If more than 1 finding was noted, the readers were instructed to report the most advanced finding. Because classification focused on only a segment of the labrum, the definition of labral ossification was modified to include only complete, nonfocal ossification. Additional post hoc analyses were performed to investigate specific sources of disagreement.

The initial read by each reviewer was utilized for interobserver analysis, while both reads were utilized for intraobserver analysis. Interobserver analysis included comparison of each of the 40 readings with each of the 3 other readers (6 reader combinations for a total of 240 comparisons). Similarly, intraobserver analysis included comparison of each of the 40 readings for 4 readers (160 comparisons).

Statistical measures of reliability including percentage agreement and weighted Cohen $\kappa$ values were calculated. Cohen $\kappa$ values were interpreted as described by Landis and Koch.15 The $\kappa$ values were classified as described by Landis and Koch15 with values of .81 to 1.0 indicating excellent agreement, .61 to .80 substantial agreement, .41 to .60 moderate agreement, .21 to .40 fair agreement, and 0 to .20 slight agreement. Absolute agreement on any given case was defined as the agreement of all 4 readers.

### RESULTS

Tables 2 and 3 show the distribution (by reader) of labral and articular cartilage findings as classified by the Beck classification. The most common labral classification was detachment (57.5%-82.5% by reader), while full-thickness tears were the least common (2.5%-7.5% by reader). The Beck articular cartilage classification was more variable, with all 5 diagnoses being present in greater than 10%.

### Interobserver Reliability

**Labral Classification.** The average weighted Cohen $\kappa$ value for the Beck labral classification was .62 (Table 4). Individual weighted $\kappa$ values between any 2 readers ranged from .48 to .78. Absolute agreement of all 4 readers occurred in 67.5% (27/40) of cases. Categorization of labral status as normal2 (Beck 1) or abnormal (Beck 2-5) resulted in an average weighted $\kappa$ value of .69 (range, .36-.84). The overall agreement rate for normal/abnormal classification was 94.6% (227/240). Labral ossification was noted by all 4 readers in 3 cases, while 1 additional case was classified as labral ossification by 1 reader. Full-thickness labral tears in the Beck classification were uncommonly recorded (2.5%-7.5% of cases), with a single case having agreement of all 4 readers. Only 2 other cases were classified as full-thickness tears (both a single reader).

The percentage of cases classified as labral degeneration was highly variable between 0% and 17.5% of cases.

### TABLE 2 Distribution of the Beck Labrum Classification by Reader$^a$

<table>
<thead>
<tr>
<th>Normal</th>
<th>Degeneration</th>
<th>Full-Thickness Tear</th>
<th>Detachment</th>
<th>Ossification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader 1</td>
<td>3 (7.5)</td>
<td>2 (5.0)</td>
<td>1 (2.5)</td>
<td>31 (77.5)</td>
</tr>
<tr>
<td>Reader 2</td>
<td>2 (5.0)</td>
<td>5 (12.5)</td>
<td>1 (2.5)</td>
<td>28 (70.0)</td>
</tr>
<tr>
<td>Reader 3</td>
<td>4 (10.0)</td>
<td>7 (17.5)</td>
<td>3 (7.5)</td>
<td>23 (57.5)</td>
</tr>
<tr>
<td>Reader 4</td>
<td>3 (7.5)</td>
<td>0</td>
<td>1 (2.5)</td>
<td>33 (82.5)</td>
</tr>
<tr>
<td>Average</td>
<td>7.5%</td>
<td>8.8%</td>
<td>3.8%</td>
<td>71.9%</td>
</tr>
</tbody>
</table>

$^a$Values are expressed as n (%) unless otherwise indicated.

### TABLE 3 Distribution of the Beck Cartilage Classification by Reader$^a$

<table>
<thead>
<tr>
<th>Normal</th>
<th>Malacia</th>
<th>Debonding</th>
<th>Cleavage</th>
<th>Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader 1</td>
<td>12 (30.0)</td>
<td>12 (30.0)</td>
<td>7 (17.5)</td>
<td>5 (12.5)</td>
</tr>
<tr>
<td>Reader 2</td>
<td>6 (15.0)</td>
<td>10 (25.0)</td>
<td>8 (20.0)</td>
<td>9 (22.5)</td>
</tr>
<tr>
<td>Reader 3</td>
<td>12 (30.0)</td>
<td>2 (5.0)</td>
<td>14 (35.0)</td>
<td>9 (22.5)</td>
</tr>
<tr>
<td>Reader 4</td>
<td>6 (15.0)</td>
<td>7 (17.5)</td>
<td>12 (30.0)</td>
<td>11 (27.5)</td>
</tr>
<tr>
<td>Average</td>
<td>22.5%</td>
<td>19.4%</td>
<td>25.6%</td>
<td>21.3%</td>
</tr>
</tbody>
</table>

$^a$Values are expressed as n (%) unless otherwise indicated.
Nine cases were classified as labral degeneration by at least 1 reader, with none having absolute agreement. In comparison, 27 of the remaining 31 cases (87.1%) had agreement on labral classification by all 4 readers. Discrepancy in 1 case resulted from disagreement of classification between degeneration and normal, while in 8 cases, disagreement between degeneration and labral detachment occurred. The average weight \( \kappa \) value for these 9 cases was .18 (range, –.17 to .44) compared with .80 (range, .68-.90) for the remaining 31 cases.

Subgroup analysis of the 9 cases labeled labral degeneration by the Beck classification found an average weighted \( \kappa \) of .42. Similarly, the modified Beck classification had an average weighted \( \kappa \) of .42 for these 9 cases. Additional post hoc grouping of labral status as (A) normal, (B) labral tear or severe degeneration, or (C) complete ossification resulted in an average weighted \( \kappa \) value of .62 for these cases. The overall weighted \( \kappa \) for this grouping among all cases was .80 (range, .69-.92). The overall agreement rate for this grouping was 92.5% (222/240).

**Articular Cartilage Classification.** The average weighted Cohen \( \kappa \) value for the Beck articular cartilage classification was .65 (Table 4). Individual weighted \( \kappa \) values between any 2 readers ranged from .49 to .74. Absolute agreement of all 4 readers occurred in 32.5% (13/40) of cases. Simplified grouping of articular cartilage findings as normal or abnormal (Beck 2-5) resulted in an average weighted \( \kappa \) value of .60 for these cases. The overall weighted \( \kappa \) for this grouping among all cases was .80 (range, .69-.92). The overall agreement rate for this grouping was 85.8% (206/240). The percentage of cases classified as malacia was highly variable: between 5% and 30% of cases. Among cases classified by the senior author as normal, malacia, or debonding (n = 19), the presence of debonding (with or without associated malacia) had only slight interobserver reliability (average \( \kappa = .30 \); range, .11-.60) among other readers. The percentage of cases classified as articular cartilage cleavage varied from 12.5% to 27.5% compared with 7.5% to 17.5% for cartilage defects. Among cases classified by the senior author as articular cartilage cleavage or defects (n = 15), interobserver reliability of the other readers was moderate (average \( \kappa = .414 \); range, .237-.561). A simplified post hoc grouping of articular cartilage findings as (A) normal, (B) malacia or debonding, or (C) cleavage or defect resulted in an average weighted \( \kappa \) value of .68 (range, .53-.76). The overall agreement rate for this subclassification was 74.1% (178/240).

**Interobserver Reliability**

**Labral Classification.** The Beck classification of labral status had substantial combined interobserver reliability \( (\kappa = .65) \) (Table 4). Individual reader reliability ranged from .31 to .83. The overall agreement rate was 75.8%. The reader with an intraobserver reliability of .31 had 17 cases of disagreement, with 76.5% (13/17) regarding differentiation of labral degeneration or labral detachment. The intraobserver reliabilities of the other 3 readers were .68, .78, and .83.

**Articular Cartilage Classification.** The Beck classification of articular cartilage status had substantial to excellent reliability (average \( \kappa = .80 \); range, .68-.86) (Table 4), with an overall agreement rate of 77.5%.

**DISCUSSION**

To our knowledge, there are no published studies reporting the reliability of arthroscopic classification of labral and articular cartilage hip disease in a clinically relevant setting. In the present study, we observed substantial interobserver reliability of the Beck classification of labral and articular cartilage disease in hip arthroscopy.

While the Beck classification of labral disease did show substantial reliability \( (\kappa = .62) \), we found the differentiation of labral degeneration from labral detachment to be a major source of disagreement (Figure 1). A labral lesion in FAI begins at the labrochondral junction.** Early disease at the labrochondral junction may result in partial detachment and mild associated changes in the adjacent labrum and cartilage. These findings are difficult to fit into the Beck labral classification. The Beck classification has been criticized for not being a progressive classification as well as frequently having difficulties in classifying disease when several types of injury coexist. Labral tears in the hip are generally detachments at the labrochondral junction, with a preserved labral attachment to bone.2,3,10,11,14,18,22 While full-thickness labral tears do occur, they are actually very uncommon. Often, by the time full-thickness labral tears occur, they are accompanied by significant labral degeneration. The treatment of a full-thickness labral tear does not differ significantly from that of a labral detachment; rather, the quality of the labral tissue determines its repairability.

The classification of articular cartilage disease using the Beck classification also had moderate reliability \( (\kappa = .65) \). Acetabular rim cartilage disease in FAI is linked to the status of the labrochondral junction.** In early disease, repetitive contact at the labrochondral junction may result in adjacent cartilage debonding. Further damage may cause malacia or partial-thickness cartilage flap formation, until eventually complete disruption of the labrochondral junction results in full articular cartilage disease. Readers demonstrated poor reliability in differentiating articular cartilage malacia and debonding (Figure 2). This is not surprising because early disease at the labrochondral junction

**References 2, 3, 10, 11, 13, 14, 18, 22.
can result in varying degrees of both malacia and debonding. Differentiating between which of these 2 factors is more prominent is likely to be very subjective. Readers also demonstrated poor reliability in differentiating cleavage lesions from full-thickness defects. This is likely because when a cleavage lesion is present, there is often significant cartilage fraying as well as adjacent labral degeneration and hypermobility, which can make judging a small amount of articular cartilage loss very difficult. Large articular cartilage defects are relatively uncommon, as the articular cartilage sleeve in a cleavage lesion often remains in continuity until a late stage. Based on these 2 findings, a post hoc articular cartilage classification system including (1) normal, (2) early articular disease: malacia/debonding, or (3) advanced articular disease: cleavage/defect was investigated. The agreement rate improved from 57.5% to 73.2% for interobserver comparisons and from 75.8% to 81.8% for intraobserver comparisons. This modified classification resulted in a similar level of substantial reliability ($\kappa = .68$), as it accounts for the number of categories.

Our findings with hip arthroscopy can be compared with those of similar studies investigating the classification of knee and shoulder disease. Multiple studies have previously reported the reliability of arthroscopic disease classification in the knee (meniscus/articular cartilage) and shoulder (rotator cuff/superior labrum anterior-posterior [SLAP]). These studies have generally concluded that this level of interobserver reliability is clinically acceptable (with $\kappa$ values generally .40-.80), and these classification systems have become accepted in the peer-reviewed literature.

There are several potential limitations to the current study. First, the classification of disease utilizing recorded intraoperative videos fails to provide the viewer with tactile feedback and the ability to further probe areas of question. While actual arthroscopic examinations would be feasible in cadaveric specimens, it is not in patients with FAI. The classification of disease based on arthroscopic videos may be a source of bias. While readers in the current study were aware of the purpose of the study, they were not provided with other clinical information. The arthroscopic visualization and probing could have been potentially influenced by the pathological abnormalities present. However, the standardization of videos in our study attempted to provide a consistent means to classify intra-articular findings. Second, the current study focused on findings at the anterior-superior acetabular rim only. The findings in our study may not be applicable to other regions of the central compartment where visualization is less optimal or to findings in the peripheral compartment. Additionally, the current study utilized surgeons familiar with hip arthroscopy. The reliability demonstrated might not extrapolate to surgeons with less experience.

The current study demonstrates that arthroscopic classification of acetabular rim disease using the Beck classification has substantial interobserver reliability. We have also identified areas in the labrochondral disease classification that pose a challenge for interobserver reliability; namely, the distinction between labral degeneration and tear and the distinction between acetabular malacia and debonding remain sources of disagreement. In summary, we propose that the Beck classification is appropriate for documenting hip disease characteristics present at the anterior-superior acetabular rim during hip arthroscopy. Future grading systems should focus on improved reliability for these labrochondral disease characteristics.

REFERENCES


