


A Comprehensive Histological Assessment of Osteoarthritis Lesions in Mice

Cartilage
2(4) 354–363
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1947603511402665
http://cart.sagepub.com


Margaret A. McNulty¹, Richard F. Loeser², Cynthia Davey³,
Michael F. Callahan⁴, Cristin M. Ferguson⁴, and Cathy S. Carlson¹

Abstract

Objective: Accurate histological assessment of osteoarthritis (OA) is critical in studies evaluating the effects of interventions on disease severity. The purpose of the present study was to develop a histological grading scheme that comprehensively and quantitatively assesses changes in multiple tissues that are associated with OA of the stifle joint in mice. **Design:** Two representative midcoronal sections from 158 stifle joints, including naturally occurring and surgically induced OA, were stained with H&E and Safranin-O stains. All slides were evaluated to characterize the changes present. A grading scheme that includes both measurements and semiquantitative scores was developed, and principal components analysis (PCA) was applied to the resulting data from the medial tibial plateaus. A subset of 30 tibial plateaus representing a wide range of severity was then evaluated by 4 observers. Reliability of the results was evaluated using intraclass correlation coefficients (ICCs) and area under the receiver operating characteristic (ROC) curve. **Results:** Five factors were retained by PCA, accounting for 74% of the total variance. Interobserver and intraobserver reproducibilities for evaluations of articular cartilage and subchondral bone were acceptable. The articular cartilage integrity and chondrocyte viability factor scores were able to distinguish severe OA from normal, minimal, mild, and moderate disease. **Conclusion:** This newly developed grading scheme and resulting factors characterize a range of joint changes in mouse stifle joints that are associated with OA. Overall, the newly developed scheme is reliable and reproducible, characterizes changes in multiple tissues, and provides comprehensive information regarding a specific site in the stifle joint.

Keywords

osteoarthritis, histology, murine, morphometry

Introduction

Accurate morphological assessment of osteoarthritis (OA) severity in both human and animal tissues is critical for evaluating studies designed to slow the progression of the disease. Because of the many factors (e.g., cost, environmental factors, diet, and activity) complicating human studies, animal models of OA are used to ensure a more uniform population and to allow histological assessments of joint tissues. For many years, the Mankin Histological-Histochemical Grading System (HHGS)¹ has been the most widely used grading scheme for assessing OA severity in animal models. The Mankin HHGS includes 4 parameters (structure, cells, Safranin-O staining, and tidemark integrity) that identify changes within the articular cartilage. Several studies have attempted to validate the Mankin HHGS in nonrodent species, and these have identified low interobserver and intraobserver reproducibilities, indicating that the scheme is neither reliable nor reproducible^{2–4} and lacks specificity (i.e., distinguishes between normal and severely affected joints, but not among intermediate severities).^{2,5} Other schemes, such as those developed by Glasson *et al.*⁶ and the

Osteoarthritis Research Society International (OARSI),⁷ aim to improve upon the Mankin HHGS and evaluation of OA in animal models. In addition, OARSI has recently published recommendations for histological assessments of OA lesions in various species, including the mouse.⁸ The authors recommend using a scheme that is very similar to one developed by

Supplementary material for this article is available on the *Cartilage* website at <http://cart.sagepub.com/supplemental>.

¹Department of Veterinary Population Medicine, University of Minnesota College of Veterinary Medicine, St. Paul, MN, USA

²Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA

³Biostatistical Design & Analysis Center, University of Minnesota, Minneapolis, MN, USA

⁴Department of Orthopaedic Surgery, Wake Forest University School of Medicine, Winston-Salem, NC, USA

Corresponding Author:

Cathy S. Carlson, Department of Veterinary Population Medicine, University of Minnesota College of Veterinary Medicine, 1988 Fitch Avenue, 435H AnSci/VM, St. Paul, MN 55108
Email: carls099@umn.edu

Chambers *et al.*⁹ and is also almost identical to that used by Glasson *et al.*⁶ This scheme evaluates the structural integrity of the articular cartilage (range, grades 0-6) in sections cut at 70- to 80- μ m intervals through the entire joint.⁸ This scheme is designed to simply and rapidly identify lesions within the 4 compartments (medial and lateral tibial plateaus and femoral condyles) and therefore is useful for identifying the site within the joint that contains the most severe lesions and for identifying general changes that are present within the joint; however, it does not provide in-depth information regarding changes within the lesions. Finally, the OARSI Osteoarthritis Cartilage Histopathology Assessment System (OOCHAS), which was developed for use in large joints, evaluates changes in articular cartilage, chondrocytes, and surrounding bone and groups these changes into grades (range, 0-6 [including subgrades]) and stages (range, 0-4), where grade includes the depth of the most severe lesion and stage includes the horizontal extent of the most severe lesion.⁷ The grade and stage are then combined to generate a final score, which represents the overall OA severity. Both the Mankin HHGS and the OARSI scheme focus on changes in articular cartilage and do not include changes of OA in other joint tissues (subchondral and periarticular bone, synovium, meniscus) that may be important in the pathogenesis of OA.¹⁰⁻¹³ The OOCAS scheme evaluates changes in chondrocytes, articular cartilage, and surrounding bone; however, these changes are combined into a single grade, even though the severity of changes in each of these tissues may not be sequentially uniform.

Histomorphometric analysis of bony tissues commonly utilizes measurements of parameters such as changes in trabecular number, spacing, and width; cortical bone porosity and thickness; and various indirect measurements of cellular activity (osteoblasts and osteoclasts).¹⁴ These methods are widely used and accepted in osteoporosis research. A previous study developed a method to quantitate histological changes in murine stifle joints using image analysis on a single mid-coronal section per joint and found that measurements of articular cartilage degradation and osteophyte size were reproducible.¹⁵ Cellular parameters (e.g., chondrocyte numbers) and changes in subchondral bone were not evaluated.

The purposes of the present study were to develop a histological grading scheme that includes changes that are consistently observed in multiple tissues during the development of OA in mice and to evaluate these changes using both semi-quantitative grades and continuous measurements. The purpose of this scheme is not to identify the location in the joint of the most severe OA lesions in a particular disease model (which can more easily be done using the Mankin HHGS or the Glasson scheme) but to comprehensively characterize the lesions that are present once the site of maximum severity has been determined. We hypothesized that the inclusion of continuous measurements of joint changes would allow this new histological grading scheme to be more reproducible and reliable and that

Table 1. Summary of 5 Mouse Studies⁶

Study	Age	Number of animals	Surgical intervention	Age at sacrifice
1	10 wk	11	DMM	18 wk
2	14 wk	15	DMM	22 wk
3	12 mo	17	None	17 mo
	18 mo	13	None	23 mo
4	N/A	16	None	12.8 mo (mean)
5	10 mo	33	None	16 mo

Note: DMM = destabilized medial meniscus model of osteoarthritis; N/A = not applicable.

the addition of meniscal and bone evaluations would provide a more sensitive assessment of early OA changes.

Materials and Methods

Mouse Models of Osteoarthritis

In order to develop a scheme that would be broadly applicable, mice from 5 separate OA studies in which the disease either occurred naturally or was surgically induced were evaluated (**Table 1**).

Histological Preparation and Development of Scheme

Intact stifle joints ($n = 158$ joints from 105 mice; both hindlimbs from 56 animals, 1 hindlimb from 49 animals; of these 3 hindlimbs were not evaluated due to sectioning difficulties) were routinely fixed in 10% formalin where the joint angle was approximately 120°, decalcified in 10% EDTA, and processed. Joints were embedded intact into paraffin with the patella down and the femur and tibia forming equal angles to the rim of the embedding cassette, ensuring that the joint was not rotated medially or laterally. Joints were then serially sectioned in a coronal plane, and 4 to 5 representative midcoronal 4- μ m-thick sections that spanned approximately 100 μ m were selected and stained as previously described¹⁵ with hematoxylin & eosin (H&E) and Safranin-O.

Sections ($n = 4-6$) from each stifle were randomized and relabeled, all sections were carefully examined histologically, and all changes were recorded. These included abnormalities in chondrocytes (e.g., number, viability, morphology), abnormalities in the matrix of articular cartilage and menisci (e.g., fibrillation, clefting, variability in Safranin-O staining), changes in synovial membrane (e.g., inflammation, fibrosis), and abnormalities in subchondral and periarticular bone (e.g., subchondral bone thickness, osteophytes). Those changes that were consistent among stifles (vs. a change involving one or

Table 2. Summary of 15 Parameters That Were Included in the Histological Grading Scheme

Parameter	Abbreviation	Definition
Articular cartilage area, μm^2	AC area	Total area of articular cartilage
Articular cartilage thickness, μm	AC thick	Average articular cartilage thickness
Subchondral bone area, μm^2	SCB area	Total area of subchondral bone
Subchondral bone thickness, μm	SCB thick	Average subchondral bone thickness
Chondrocyte cell death area, μm^2	CCD	Total area of chondrocyte cell death
Percentage of chondrocyte cell death	CCD%	Percentage of chondrocyte cell death within the articular cartilage
Number of chondrocytes	#chond	Total number of viable chondrocytes
Total articular cartilage area per viable chondrocyte, μm^2	AC/chond	Articular cartilage area divided by number of viable chondrocytes
Viable articular cartilage area per viable chondrocyte, μm^2	VAC/chond	Total area of articular cartilage subtracted by chondrocyte cell death area, divided by total number of viable chondrocytes
Size of axial osteophytes, μm^2	Ax OP	Total area of axial osteophytes (if present)
Size of abaxial osteophytes, μm^2	Abax OP	Total area of abaxial osteophytes (if present)
Meniscal area, μm^2	Men	Total area of weightbearing meniscus
Meniscal cell death area, μm^2	Men CCD%	Percentage of area of chondrocyte cell death within the meniscus
Articular cartilage structure score (0-12)	ACS	Semiquantitative score of articular cartilage integrity
Safranin-O staining score (0-12)	Saf-O	Semiquantitative score for loss of Safranin-O staining

a very low number of stifles) were further characterized with measurements using an OsteoMeasure bone histomorphometry system (OsteoMetrics, Atlanta, GA) or were assigned semiquantitative grades according to severity (**Table 2**).

Two representative midcoronal sections from each stifle, one stained with H&E and one with Safranin-O, were selected for final evaluation. Medial and lateral tibial plateaus from all joints were evaluated using the finalized scheme (**Table 2**). Measurements of articular cartilage and subchondral perimeters^{16,17} and areas of chondrocyte cell death were taken at a magnification of 20x using a 700- μm \times 800- μm grid that was centered on the tibial plateau (**Fig. 1**) and included 75% to 100% of the width of each plateau. To evaluate osteophytes and menisci, the grid for these measurements was centered over the area of interest (**Fig. 1**). Articular cartilage (AC) area was defined as the area between the superficial surface of articular cartilage and the most superficial tidemark; calcified cartilage area was the area between the most superficial tidemark and the calcified cartilage–subchondral bone junction; and subchondral bone (SCB) area was the area between the calcified cartilage–subchondral bone junction and the most superficial boundary of the marrow spaces (**Fig. 1A**). Because the deep surface of subchondral bone in mice is not as well delineated as in other species, several rules were generated to provide consistency among measurements. First, adjacent marrow spaces were connected by tracing horizontally across the bone trabeculae at the point where the width of the trabecula was constant or began to narrow. Second, marrow spaces were excluded if the size of the marrow space was less than the vertical distance between 2 adjacent marrow spaces.

In instances where no marrow spaces were present, the upper margin of the growth plate was considered as the deep face of the subchondral bone plate. The area of weightbearing meniscus (Men area) was defined by drawing a straight line connecting the most medial or lateral aspect of the tibial plateau and femoral condyle and outlining the meniscus axial to that line (**Fig. 1B**). Total area composed of dead chondrocytes within the articular cartilage and meniscus was measured by tracing all areas within which 2 or more chondrocytes were dead, as determined by visual assessment of loss of nuclear staining with H&E (**Fig. 1C**), and summing these areas. Articular cartilage, calcified cartilage, and subchondral bone width and area were calculated from perimeter measurements. Percentages of chondrocyte cell death area relative to total articular cartilage area (CCD%) and the total area of cartilage in the weightbearing meniscus (Men CCD%) also were calculated. The number of viable chondrocytes (#chond) in the articular cartilage was counted. In cases where there was complete loss of articular cartilage, CCD% were recorded as 100%, and the remainder of the articular cartilage and chondrocyte measurements were recorded as “0”. Abaxial (**Fig. 1D**) and axial osteophytes (Abax OP and Ax OP, respectively) were measured as previously described for nonhuman primates.²⁰ In addition to these measurements, articular cartilage was evaluated using 2 semiquantitative parameters, the Articular Cartilage Structure score (ACS) (range, 0-12) (**Suppl. Fig. S1; Table 3**) and the Safranin-O staining score (Saf-O) (range, 0-12) (**Suppl. Fig. S2; Table 3**). Similar semiquantitative assessments have been used previously in other species.^{16,18,19}

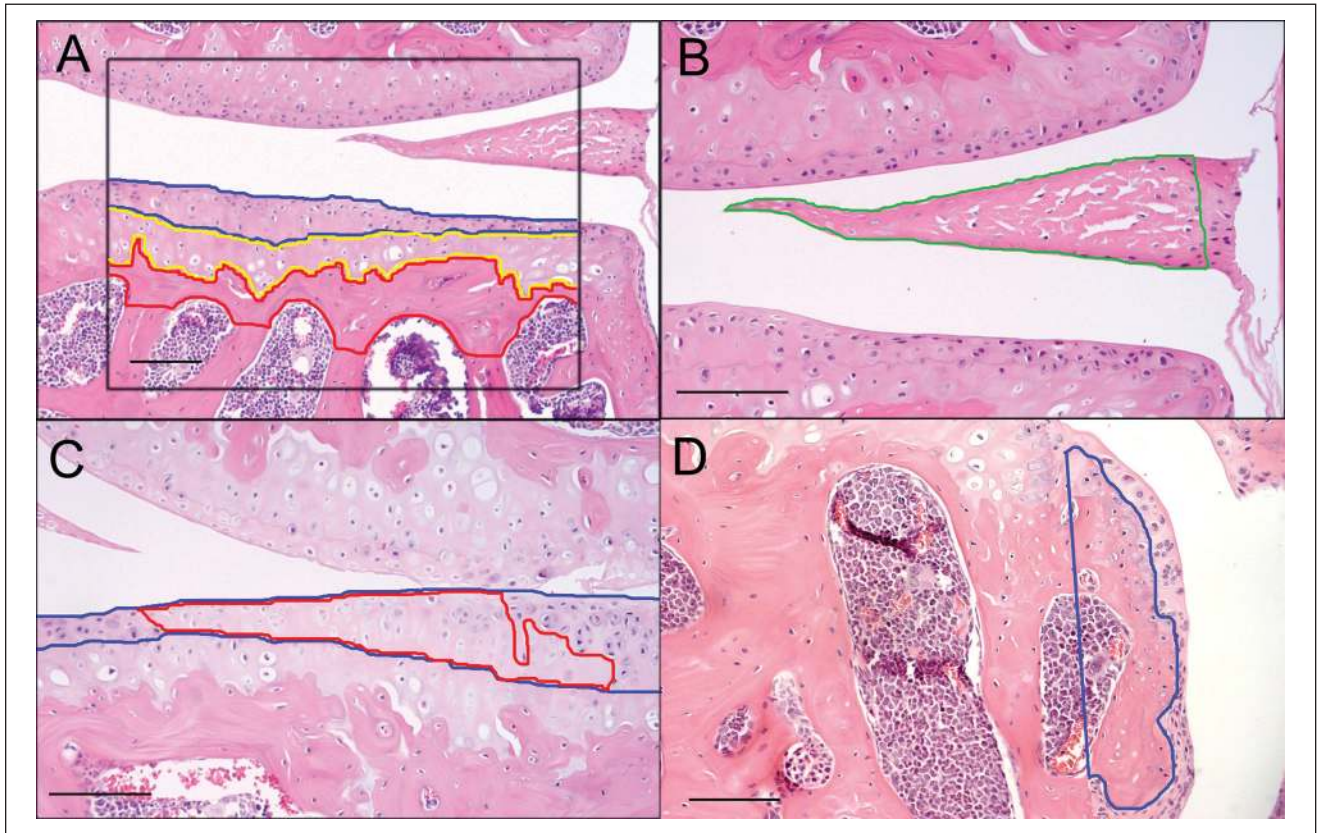


Figure 1. Continuous measurements performed using the OsteoMeasure bone histomorphometry system. **(A)** A box ($800\ \mu\text{m} \times 700\ \mu\text{m}$) was centered on the tibial plateau, and the superficial and deep surfaces of articular cartilage (blue), calcified cartilage (yellow), and subchondral bone (red) were traced. **(B)** Weightbearing meniscal area (green) was outlined. **(C)** Areas of chondrocyte cell death (red) were traced within the articular cartilage (blue). **(D)** Abaxial osteophyte area outlined in blue. Bar = $100\ \mu\text{m}$.

Samples Selected for Evaluation by Multiple Observers

A representative subset of 30 tibial plateaus (lateral, $n = 12$; medial, $n = 18$) from 21 joints was selected for evaluation by multiple observers. Each tibial plateau was independently assessed and assigned an OA severity level based on the semi-quantitative ACS score (**Table 3**). Five levels of OA severity were created based on these ACS scores, and tibial plateaus were chosen to represent these severities: 0 = “normal” ($n = 8$); 1-3 = “minimal” ($n = 8$); 4-6 = “mild” ($n = 6$); 7-9 = “moderate” ($n = 3$); and 10-12 = “severe” ($n = 5$). The newly developed grading scheme (**Table 2**) was applied to one H&E- and one Safranin-O–stained midcoronal section from each plateau ($n = 30$) by each of 4 independent observers (observers 1-4). Two of the observers had extensive experience in evaluating mouse OA and in using an OsteoMeasure histomorphometry system (observers 1 and 4) (OsteoMetrics), one observer had moderate experience (observer 2), and one observer had little to no experience (observer 3). Both slides for each tibial plateau were blinded and randomized to prevent observer bias. Each observer was given written definitions for both schemes and a figure outlining how measurements should be taken. In addition, observers were

given a tutorial (approximately 20 minutes) on how to use the OsteoMeasure system (OsteoMetrics). All measurements and histological grades were performed on the H&E-stained sections with the exception of the Safranin-O semiquantitative grades, for which a corresponding Safranin-O–stained section was used. Observers evaluated all 30 sections over a 2-day period (time 1), after which the slides were rerandomized and re-evaluated by all 4 observers 2 weeks later (time 2).

Statistics

Parameters from the medial joint compartments of all 158 joints were included in a principal components analysis (PCA) to reduce the large number of correlated parameters to a smaller number of orthogonal linear combinations (factors) of these parameters. Parameters with factor loadings >0.40 were considered to contribute substantially to the factors. The final results of PCA are standardized factor scores for each joint. These standardized factor scores have an approximately normal distribution that allows for their use in parametrical statistical methods such as ANOVA, linear regression, and Student t tests. SAS version 9.1 (Cary, NC) was used for all statistical analyses.

Table 3. Description of 2 Semiquantitative Grades Included in New Histological Grading Scheme

Histological parameter	Description
Articular cartilage structure score (ACS)	
Grade 0	Articular surface smooth and intact
Grade 1-3	Fibrillation and/or clefts and/or loss of cartilage involving one fourth or less of articular cartilage thickness involving \leq one third, $>$ one third and \leq two thirds, or $>$ two thirds of the plateau or condyle, respectively
Grade 4-6	Fibrillation and/or clefts and/or loss of cartilage involving one half or less of articular cartilage thickness involving \leq one third, $>$ one third and \leq two thirds, or $>$ two thirds of the plateau or condyle, respectively
Grade 7-9	Fibrillation and/or clefts and/or loss of cartilage involving $>$ one half of articular cartilage but less than full thickness involving \leq one third, $>$ one third and \leq two thirds, or $>$ two thirds of the plateau or condyle, respectively
Grade 10-12	Fibrillation and/or clefts and/or loss of cartilage involving the full thickness of articular cartilage involving \leq one third, $>$ one third and \leq two thirds, or $>$ two thirds of the plateau or condyle, respectively
Safranin-O staining score (Saf-O)	
Grade 0	Uniform staining throughout the articular cartilage
Grade 1-3	Loss of staining in the matrix (but not cells) in \leq one half of the articular cartilage and involving \leq one third, $>$ one third and \leq two thirds, or $>$ two thirds of the plateau or condyle, respectively
Grade 4-6	Loss of staining in the matrix (but not cells) in $>$ one half of the articular cartilage and involving \leq one third, $>$ one third and \leq two thirds, or $>$ two thirds of the plateau or condyle, respectively
Grade 7-9	Complete loss of staining in \leq one half of the articular cartilage involving \leq one third, $>$ one third and \leq two thirds, or $>$ two thirds of the plateau or condyle, respectively
Grade 10-12	Complete loss of staining in $>$ one half of the articular cartilage and involving \leq one third, $>$ one third and \leq two thirds, or $>$ two thirds of the plateau or condyle, respectively

To calculate interobserver and intraobserver reproducibilities in the subset of 30 tibial plateaus that was evaluated, the PCA-derived scores for these joints were calculated for each of the 8 evaluations in this study (4 observers, 2 time points) by 1) standardizing each measurement using means and standard deviations from the original PCA, 2) multiplying these standardized measures by the PCA-derived factor score loadings from the original PCA, and 3) summing the weighted standardized measures into PCA-derived scores for this study.

Reliability. Intraobserver and interobserver reliability coefficients were estimated for each of the PCA-derived scores using an intraclass correlation coefficient (ICC). To estimate intraobserver reliability coefficients, ICCs were calculated for each evaluator using repeated measures of calculated PCA-derived scores. An interobserver reliability analysis was completed using data from all 4 observers, and a separate analysis was completed for the 2 experienced observers (observers 1 and 4). ICCs of 0.90 or greater are considered to be acceptable diagnostic or intervention purposes.²¹

Validity. The accuracy of each PCA-derived score in identifying levels of OA severity was estimated by the area under the receiver operating characteristic (ROC) curve (AUC) for each of the 5 OA severity levels: normal, minimal, mild, moderate, and severe. Combined data from each evaluator from the first time point were used in separate logistic regression models (for each score of the PCA-derived scores) as a

predictor of each level of assigned OA severity (5 levels). The AUC ranges from 0.5 (no discrimination) to 1.0 (perfect discrimination), with values ranging from 0.70 to 0.9, indicating moderately accurate classification, and values $>$ 0.90, representing high classification accuracy.²²

Results

Histological evaluation of these sections (both from all 158 sections and from the subset of 30 sections) revealed a wide range of lesion severities. In the majority of the sections, the lesions were more severe in the medial tibial plateau than any other site (lateral tibial plateau, medial and lateral femoral condyles) within the joint.

Principal Components Analysis (PCA) for All 158 Joints

Because lesions were more severe in the medial tibial plateau in the great majority of the sections and because fewer significant correlations among the 15 parameters were identified in the lateral tibial plateaus (data not shown), PCA was only completed on data from the medial tibial plateaus. PCA resulted in retention of 5 factors that together accounted for 74% of the total variance in the data (Table 4). For clarity, each factor was named according to the major parameters it described (Table 4); the 5 factors included one each that described changes in

Table 4. Components of Factors Retained by Principal Components Analysis (PCA) Using Data from the Medial Tibial Plateaus (factor loadings in parentheses), Including Percentage of Total Variance Accounted for by Each Factor (bottom row)

Factor 1: articular cartilage integrity	Factor 2: chondrocyte viability	Factor 3: subchondral bone	Factor 4: meniscus	Factor 5: periarticular bone
AC area (0.942)	AC area (0.082)	AC area (−0.005)	AC area (−0.074)	AC area (−0.107)
AC thick (0.898)	AC thick (0.017)	AC thick (−0.151)	AC thick (−0.094)	AC thick (0.010)
SCB area (−0.195)	SCB area (0.184)	SCB area (0.711)	SCB area (−0.282)	SCB area (0.191)
SCB thick (−0.086)	SCB thick (0.063)	SCB thick (0.864)	SCB thick (0.076)	SCB thick (0.027)
CCD area (0.219)	CCD area (0.858)	CCD area (0.149)	CCD area (−0.054)	CCD area (0.015)
CCD% (−0.527)	CCD% (0.689)	CCD% (0.285)	CCD% (−0.170)	CCD% (0.149)
#chond (0.604)	#chond (−0.692)	#chond (−0.042)	#chond (0.051)	#chond (−0.161)
AC/chond (0.115)	AC/chond (0.902)	AC/chond (0.036)	AC/chond (−0.051)	AC/chond (−0.058)
VAC/chond (0.707)	VAC/chond (0.445)	VAC/chond (−0.259)	VAC/chond (−0.013)	VAC/chond (−0.058)
Ax OP (−0.021)	Ax OP (−0.155)	Ax OP (0.141)	Ax OP (0.064)	Ax OP (0.837)
Abax OP (−0.159)	Abax OP (0.205)	Abax OP (0.031)	Abax OP (−0.174)	Abax OP (0.716)
Men (−0.044)	Men (−0.041)	Men (0.138)	Men (0.866)	Men (−0.069)
Men CCD% (0.042)	Men CCD% (0.046)	Men CCD% (0.253)	Men CCD% (−0.656)	Men CCD% (−0.002)
ACS (−0.482)	ACS (0.315)	ACS (0.275)	ACS (−0.371)	ACS (0.207)
Saf-O (−0.492)	Saf-O (0.334)	Saf-O (−0.114)	Saf-O (−0.350)	Saf-O (0.290)
30.3%	18.8%	9.9%	8.5%	7%

Note: Parameters with factor loadings >0.40 are highlighted in bold and italics.

Table 5. Intraobserver Reliability Coefficients for Principal Components Analysis (PCA)–Derived Factor Scores

	Articular cartilage integrity	Chondrocyte viability	Subchondral bone	Meniscus	Osteophytes
Observer 1	0.946	0.962	0.973	0.863	−0.090
Observer 2	0.955	0.529	0.821	0.813	0.503
Observer 3	0.978	0.880	0.846	0.845	0.465
Observer 4	0.900	0.924	0.878	0.911	0.685

Note: Coefficients ≥ 0.90 are recommended for diagnostic purposes, and these are indicated in bold. Coefficients ≥ 0.80 are deemed acceptable; these are indicated in italics.

articular cartilage, chondrocytes, subchondral bone, meniscus, and, periarticular bone (i.e., osteophytes), respectively.

Reliability and Validity Using Multiple Observers

Depending on their level of experience, observers took between 2 and 6 hours to complete the histological evaluations of the 30 tibial plateaus. This time decreased as observers became more comfortable with the scheme and the Osteo-Measure system (OsteoMetrics).

Reliability. Intraobserver reliability coefficients were acceptable (≥ 0.90) for all evaluators for the articular cartilage integrity factor score (Table 5), and 2 evaluators (observers 1 and 4) had acceptable coefficients for the chondrocyte viability score. Only one evaluator had acceptable coefficients for the subchondral bone (observer 1) and meniscal (observer 4) factor scores, and no evaluators had acceptable coefficients for the osteophyte factor score.

The ICCs for a single evaluator using data from all 4 observers identified only one factor score, articular cartilage integrity, which had acceptable interobserver reliability (0.95) (Table 6). The same ICCs for a single evaluator were re-estimated using only data from the 2 most experienced observers (observers 1 and 4); in this case, 3 factor scores had acceptable interobserver reliability: articular cartilage integrity (0.94), chondrocyte viability (0.95), and subchondral bone (0.90). When the ICC was estimated for the mean of 4 evaluators using data from all 4 observers, 4 of the 5 factor scores (all except the osteophyte factor score) had acceptable interobserver reliability coefficients. When the ICC was estimated for the mean of 2 evaluators using data from the 2 most experienced observers (observers 1 and 4), 3 factor scores (articular cartilage integrity, 0.97; chondrocyte viability, 0.97; and subchondral bone, 0.94) had acceptable interobserver reliability coefficients.

Validity. The articular cartilage integrity factor score was moderately accurate (AUC = 0.82) in discriminating severe OA from normal, minimal, mild, and moderate disease. The

Table 6. Interrater Reliability Coefficients for Principal Components Analysis (PCA)–Derived Factor Scores

	ICC for single evaluator (4 evaluators)	ICC for mean of 4 evaluators	ICC for single evaluator (2 evaluators)	ICC for mean of 2 evaluators
Articular cartilage integrity	0.95	0.99	0.94	0.97
Chondrocyte viability	0.89	0.97	0.95	0.97
Subchondral bone	0.82	0.95	0.90	0.94
Meniscus	0.78	0.93	0.75	0.86
Osteophytes	-0.01	-0.02	-0.08	-0.18

Note: Reliability coefficients for a single evaluator and mean of 4 evaluators were estimated using data from all 4 observers (first 2 columns). Reliability coefficients for a single evaluator and mean of 2 evaluators were estimated using data from the 2 most experienced observers (observers 1 and 4; second 2 columns). Coefficients ≥ 0.90 were determined to be within the acceptable range for diagnostic purposes, noted in bold and italics. ICC = intraclass correlation coefficient.

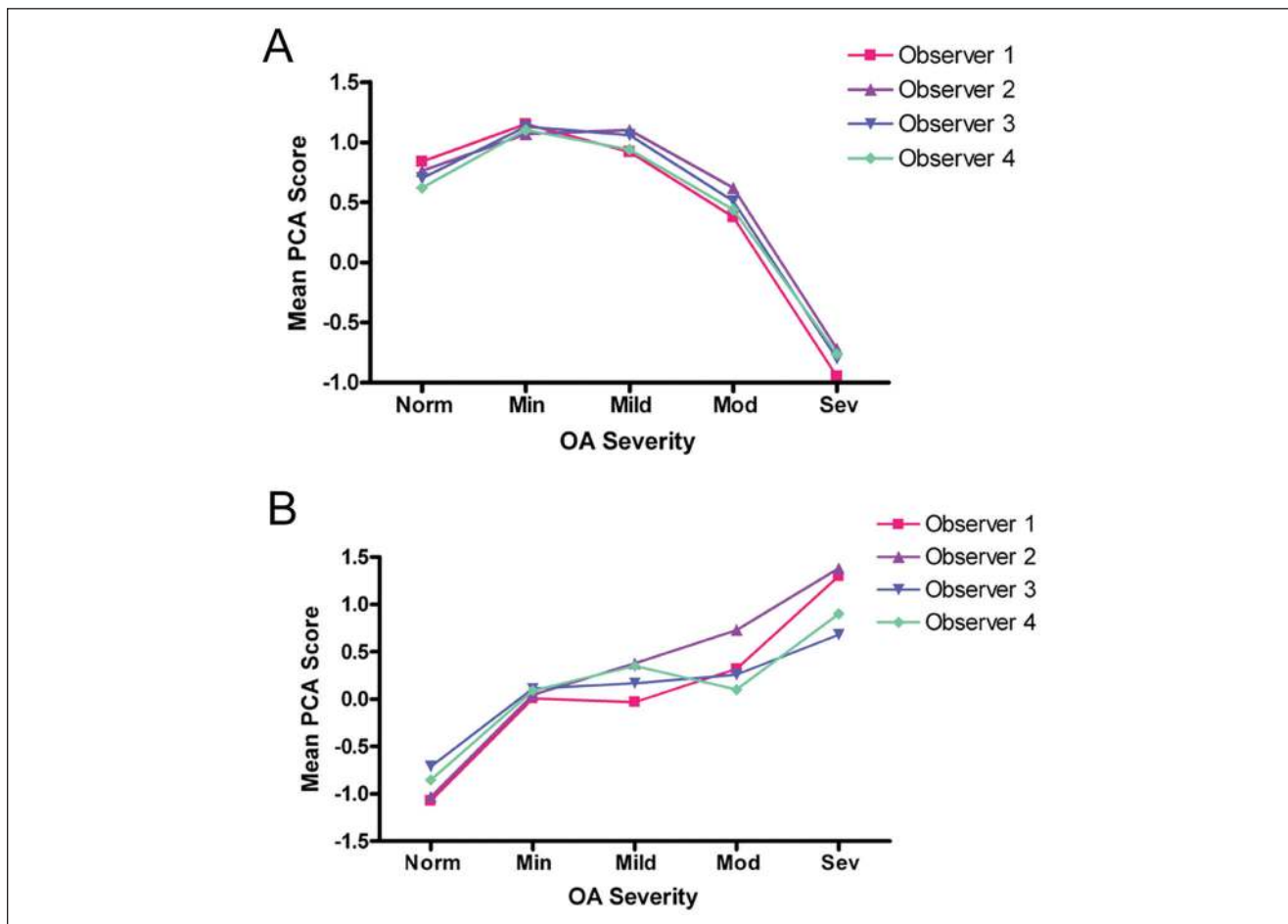


Figure 2. Mean principal components analysis (PCA) score for articular cartilage integrity and chondrocyte viability by osteoarthritis (OA) severity level for each evaluator at time 1. (A) Articular cartilage integrity factor scores. (B) Chondrocyte viability factor scores.

chondrocyte viability factor score was moderately accurate (AUC = 0.81) in distinguishing normal cartilage from all levels of osteoarthritic cartilage and moderately accurate (AUC = 0.74) in distinguishing severe OA from normal, minimal, mild, and moderate disease. These results are consistent with the plots of OA severity level by mean PCA score stratified by evaluator at time 1 (Fig. 2).

Discussion

A comprehensive histological grading scheme was developed that characterizes, in detail, the changes occurring in murine stifle joints in surgically induced and spontaneously occurring OA. The scheme that has been developed is based on a thorough evaluation of one midcoronal section that was

meticulously chosen, as opposed to a less rigorous evaluation focused on articular cartilage in multiple sections. Due to the comprehensive nature of this scheme, it is best utilized in models where the locations of the most severe lesions are already known. If the locations of lesions in a particular model have not been characterized, a more rapid method of identifying OA lesions, such as that developed for use in mice by Glasson *et al.*,⁶ the scheme recommended by OARSI,⁸ or the scheme developed by van der Kraan *et al.*,²³ would be more appropriate. Once the locations of the most severe lesions have been identified, this detailed scheme may then be used to accurately evaluate changes in lesion severity in multiple joint tissues to determine, for example, the effects of a compound on subchondral bone, meniscus, and/or articular cartilage.

In the present study, histological changes that were observed infrequently were not included in the final scheme primarily because PCA does not work effectively when the majority of the scores for a particular parameter are zero. For example, thickening of the synovial membrane was noted only in severely affected joints. Tidemark clefts²⁴ also were not included because these also were present primarily in the most severely affected joints. However, because tidemark clefts result in full-thickness loss of articular cartilage and are associated with chondrocyte cell death, their presence likely was accounted for, albeit indirectly, by parameters such as articular cartilage thickness and chondrocyte cell death measurements. Conversely, some lesions, including vessels crossing the tidemark and chondrocyte clones, were expected but were not identified in this large sample of joints. These features are common in other species; however, their absence in mice provides yet another reason why grading schemes for OA need to be individualized by species. Finally, because many sections were identified in which chondrocyte cell death was a feature, this parameter was included, whereas it has not been included in previous schemes for OA evaluation. Chondrocyte death was considered to be a real (not artifactual) lesion, evidenced by the common association of this change with other lesions of OA including articular cartilage degradation and the characteristic appearance of the affected cells.²⁵

Our finding that lesions were more severe in the medial tibial plateau in the aged models of OA concurs with the results of many studies in numerous species, including humans. It contradicts a previous finding by van der Kraan *et al.*, however, that showed that OA lesions in 1- to 2-year-old C57/Bl6 mice were more severe in the lateral tibial plateau in 63% of the animals.²³ However, the lesions in that study were assessed using a simplified 4-level scoring system (none, slight, moderate, severe) and using only 14 to 16 animals in 2 age groups.

The OsteoMeasure system (OsteoMetrics) is not a requirement for using this grading scheme, as any histomorphometric system that can be used to measure areas and thicknesses can

generate similar data. The OsteoMeasure system (OsteoMetrics) was utilized in this study mainly because it has been developed for and is commonly used to quantitate changes in bony structures and was easily adapted to evaluate changes in joint tissues. The use of PCA is also not necessary, as the majority of the parameters are represented by continuous data that can be evaluated individually by parametric statistical methods. A benefit of using PCA is the ability to generate orthogonal (uncorrelated) factor scores (composed of and weighted by multiple variables) for each individual joint, which are standardized and are approximately normally distributed. Prior studies have successfully used PCA to summarize data from OA lesions in other species,^{18,26} as this technique is particularly useful for combining continuous (e.g., measurements) and discrete (e.g., grades) data.

Intraobserver consistency is critical in creating reproducible results when using any grading scheme. Care should be taken to set out strict definitions of the changes identified, which may be dependent on the species that is being evaluated. For example, the low intraobserver reliability coefficients for the osteophyte factor score most likely occurred because the definitions used by the new scheme to identify osteophytes in mice were originally developed for use in monkeys,²⁰ which, as realized later, did not apply perfectly to mice. In the present study, all observers had at least a general knowledge of the tissues within the joint and changes that were to be evaluated, but only moderately detailed instructions were given for identifying changes within each tissue. Despite this limitation, interobserver and intraobserver reliability coefficients for the new scheme were acceptable for several factors and could be improved upon with more complete instructions and definitions. The extremely low intraobserver reliability coefficient score for observer 1 for the osteophyte factor score (-0.090) likely occurred because the method of evaluating osteophytes during the second evaluation was changed after the observer determined that calcified cartilage margins provided a more reliable method to determine osteophytes. In addition, the relatively low intraobserver reliability coefficient (0.529) for observer 2 for the chondrocyte viability factor score can be attributed to a change in the way in which this observer identified dead chondrocytes, relying upon the OsteoMeasure (OsteoMetrics) computer screen for the first evaluation and using the microscope and the computer image to identify dead chondrocytes during the second evaluation. These types of changes in the way measurements are taken obviously will have an effect on the results and can be corrected in future studies.

For interobserver reliability scores, the ICC model utilized in this study allows for the interobserver reliability estimates to be generalized to the larger population of evaluators rather than limited to the fixed set of evaluators in this study. The interobserver reliability results indicate that having multiple observers evaluating the data using the newly devised scheme

will yield the most reproducible results, even if some of the observers are inexperienced. However, experience in evaluating OA severity and/or using the OsteoMeasure system (OsteoMetrics) also is important, as interobserver reliability coefficients were higher when only data from observers 1 and 4 (the 2 most experienced observers) were included in the ICC analysis. Under normal circumstances, however, it is not a reasonable expectation for multiple individuals to evaluate OA severity in a single study or to expect those individuals to be experts in evaluating OA. The results showed that a single inexperienced observer can produce very reliable results for the articular cartilage integrity factor scores and that a single experienced observer is sufficient to generate reliable results for a minimum of 3 of the 5 factors using the newly developed scheme.

This scheme also was able to distinguish normal from severely affected joints but did not distinguish either normal or severely affected joints from mildly or moderately affected joints; however, we identified a wide range of severity in each individual parameter. Evaluation of the validity of this scheme, as well as other schemes, is complicated by the lack of a “gold standard” by which levels of OA severity may be identified. Ostergaard *et al.*² used the Collins and McElligott scale, which is a macroscopic evaluation scheme,²⁷ to create levels of OA severity among the samples in their study; however, it is not possible to apply this scheme to a mouse stifle joint, which is too small to allow accurate gross assessment. For this reason, a histological method of categorizing OA severity was utilized in this study. However, it was redundant in that it used an initial independent evaluation with the ACS score for each joint in order to place it within the proper OA severity category, and this same ACS score was included in the newly developed grading scheme.

In summary, the scheme developed in this study evaluates changes associated with OA that occur not only within the articular cartilage but also in periarticular bone and meniscus. In addition, factor scores generated by PCA provide a global summary of changes occurring within the joint as well as an indication of which parameters are highly correlated with each other. Finally, the combined results of this study show that the newly developed histological grading scheme demonstrates good intraobserver and interobserver reliability, particularly when used by experienced evaluators.

Acknowledgments and Funding

This work was funded by National Institutes of Health (NIH) Musculoskeletal Training grant T32 AR050938 (CSC, MAM), Arthritis Foundation Innovative Research (RFL), and NIH grant AG16697 (RFL).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

References

1. Mankin HJ, Dorfman H, Lippiello L, Zarins A. Biochemical and metabolic abnormalities in articular cartilage from osteoarthritic human hips, II: correlation of morphology with biochemical and metabolic data. *J Bone Joint Surg Am.* 1971;53(3):523-37.
2. Ostergaard K, Andersen CB, Petersen J, Bendtzen K, Salter DM. Validity of histopathological grading of articular cartilage from osteoarthritic knee joints. *Ann Rheum Dis.* 1999;58(4):208-13.
3. 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(10):1766-71.
4. van der Sluijs JA, Geesink RG, van der Linden AJ, Bulstra SK, Kuyer R, Drukker J. The reliability of the Mankin score for osteoarthritis. *J Orthop Res.* 1992;10(1):58-61.
5. Moussavi-Harami SF, Pedersen DR, Martin JA, Hillis SL, Brown TD. Automated objective scoring of histologically apparent cartilage degeneration using a custom image analysis program. *J Orthop Res.* 2009;27(4):522-8.
6. Glasson SS, Blanchet TJ, Morris EA. The surgical destabilization of the medial meniscus (DMM) model of osteoarthritis in the 129/SvEv mouse. *Osteoarthritis Cartilage.* 2007;15(9):1061-9.
7. Pritzker KP, Gay S, Jimenez SA, Ostergaard K, Pelletier JP, Revell PA, *et al.* Osteoarthritis cartilage histopathology: grading and staging. *Osteoarthritis Cartilage.* 2006;14(1):13-29.
8. Glasson SS, Chambers MG, Van Den Berg WB, Little CB. The OARSI histopathology initiative: recommendations for histological assessments of osteoarthritis in the mouse. *Osteoarthritis Cartilage.* 2010;18 Suppl 3:S17-23.
9. Chambers MG, Bayliss MT, Mason RM. Chondrocyte cytokine and growth factor expression in murine osteoarthritis. *Osteoarthritis Cartilage.* 1997;5(5):301-8.
10. Ding C, Cicuttini F, Jones G. Tibial subchondral bone size and knee cartilage defects: relevance to knee osteoarthritis. *Osteoarthritis Cartilage.* 2007;15(5):479-86.
11. Dore D, Ding C, Jones G. A pilot study of the reproducibility and validity of measuring knee subchondral bone density in the tibia. *Osteoarthritis Cartilage.* 2008;16(12):1539-44.
12. Englund M. The role of the meniscus in osteoarthritis genesis. *Med Clin North Am.* 2009;93(1):37-43.
13. Goldring SR. The role of bone in osteoarthritis pathogenesis. *Rheum Dis Clin North Am.* 2008;34(3):561-71.
14. Parfitt AM. Bone histomorphometry: proposed system for standardization of nomenclature, symbols, and units. *Calcif Tissue Int.* 1988;42(5):284-6.
15. van Valburg AA, van Osch GJ, van der Kraan PM, van den Berg WB. Quantification of morphometric changes in murine experimental osteoarthritis using image analysis. *Rheumatol Int.* 1996;15(5):181-7.

16. Olson EJ, Wentorf FA, McNulty MA, Parker JB, Carlson CS, LaPrade RF. Assessment of a goat model of posterolateral knee instability. *J Orthop Res.* 2008;26(5):651-9.
17. Pastoureau PC, Hunziker EB, Pelletier JP. Cartilage, bone and synovial histomorphometry in animal models of osteoarthritis. *Osteoarthritis Cartilage.* 2010;18 Suppl 3:S106-12.
18. Ham KD, Loeser RF, Lindgren BR, Carlson CS. Effects of long-term estrogen replacement therapy on osteoarthritis severity in cynomolgus monkeys. *Arthritis Rheum.* 2002;46(7): 1956-64.
19. LaPrade RF, Wentorf FA, Olson EJ, Carlson CS. An in vivo injury model of posterolateral knee instability. *Am J Sports Med.* 2006;34(8):1313-21.
20. Olson EJ, Lindgren BR, Carlson CS. Effects of long-term estrogen replacement therapy on bone turnover in periarticular tibial osteophytes in surgically postmenopausal cynomolgus monkeys. *Bone.* 2008;42(5):907-13.
21. Weiner IB, Freedheim DK, Schinka JA, Velicer WF. *Handbook of psychology.* New York: Wiley; 2003.
22. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* 1988;240(4857):1285-93.
23. van der Kraan PM, Stoop R, Meijers TH, Poole AR, van den Berg WB. Expression of type X collagen in young and old C57Bl/6 and Balb/c mice: relation with articular cartilage degeneration. *Osteoarthritis Cartilage.* 2001;9(2):92-100.
24. Otterness IG, Chang M, Burkhardt JE, Sweeney FJ, Milici AJ. Histology and tissue chemistry of tidemark separation in hamsters. *Vet Pathol.* 1999;36(2):138-45.
25. Yttrhus B, Carlson CS, Ekman S. Etiology and pathogenesis of osteochondrosis. *Vet Pathol.* 2007;44(4):429-48.
26. Carlson CS, Loeser RF, Purser CB, Gardin JF, Jerome CP. Osteoarthritis in cynomolgus macaques, III: effects of age, gender, and subchondral bone thickness on the severity of disease. *J Bone Miner Res.* 1996;11(9):1209-17.
27. Collins DH, McElligott TF. Sulphate ($^{35}\text{SO}_4$) uptake by chondrocytes in relation to histological changes in osteoarthritic human articular cartilage. *Ann Rheum Dis.* 1960;19:318-30.