



Published in final edited form as:

*J Orthop Res.* 2011 August ; 29(8): 1145–1151. doi:10.1002/jor.21368.

## Establishment of an Index with Increased Sensitivity for Assessing Murine Arthritis

Erik R. Sampson, PhD<sup>1</sup>, Christopher A. Beck, PhD<sup>2</sup>, John Ketz, MD<sup>1</sup>, Krista L. Canary, MS<sup>1</sup>, Matthew J. Hilton, PhD<sup>1</sup>, Hani Awad<sup>3</sup>, Edward M. Schwarz, PhD<sup>1</sup>, Di Chen PhD<sup>1</sup>, Regis J. O’Keefe, MD, PhD<sup>1</sup>, Randy N. Rosier, MD, PhD<sup>1</sup>, and Michael J. Zuscik, PhD<sup>1</sup>

<sup>1</sup> Department of Orthopaedics, Center for Musculoskeletal Research, University of Rochester Medical Center, 601 Elmwood Avenue, Box 665, Rochester, NY 14642

<sup>2</sup> Department of Biostatistics & Computational Biology, Center for Musculoskeletal Research, University of Rochester Medical Center, 601 Elmwood Avenue, Box 665, Rochester, NY 14642

<sup>3</sup> Department of Biomedical Engineering, Center for Musculoskeletal Research, University of Rochester Medical Center, 601 Elmwood Avenue, Box 665, Rochester, NY 14642

### Abstract

**Purpose**—Establish quantitative outcomes for assessing murine knee arthritis and develop an Arthritis Index that incorporates multiple outcomes into a single calculation that provides enhanced sensitivity.

**Methods**—Using an accepted model of meniscal/ligamentous injury (MLI)-induced osteoarthritis (OA), we assessed mouse knee arthritis using several approaches. Histology-based methods were performed to visualize joint tissues including articular cartilage and subchondral bone. Accepted histologic scoring methods and histomorphometry were performed to grade cartilage degeneration and determine articular cartilage area, respectively. MicroCT was used to visualize and quantify the bony structures of the joint including osteophytes and joint bone volume. A statistical algorithm was then developed that combined histologic scores and cartilage areas into a single Arthritis Index.

**Results**—MLI induced progressive, OA-like articular cartilage degeneration characterized by increasing (worsening) histologic score and decreasing cartilage area. MicroCT revealed osteophytes and increased joint bone volume between the femoral and tibial physes following MLI. Lastly, an Arthritis Index calculation was established, which incorporated histologic scoring and cartilage area. The Arthritis Index provided enhanced quantitative sensitivity in assessing the level of joint degeneration compared to either histologic scoring or cartilage area determination alone; when using the Index, between 29% and 43% fewer samples are needed to establish statistical significance in studies of murine arthritis.

**Conclusions**—Arthritis in the mouse knee can be quantitatively assessed by histologic scoring, measuring cartilage area and determining joint bone volume. Enhanced sensitivity can be achieved by performing the Arthritis Index calculation, a novel method for quantitatively assessing mouse knee arthritis.

### Keywords

Osteoarthritis; Articular cartilage; Meniscal injury; Micro-computed tomography; Arthritis index

## Introduction

Arthritis is the number one cause of disability in the United States and is projected to afflict 59.4 million Americans by 2020<sup>1-2</sup>. OA, the most common form of arthritis, is a degenerative joint disease characterized by dysfunction of articular chondrocytes, articular cartilage degradation, osteophyte formation and subchondral sclerosis<sup>3</sup>. Despite the increasing population of afflicted individuals, only recently have there been advances in understanding the seminal molecular/cellular/tissue events that cause degeneration. These advances have emerged from work in animal models, ranging from injury-induced models in large animals and rodents to genetic models in mice.

While models of OA have been developed that involve knee joint injury (meniscus and ligaments) in the rabbit<sup>4</sup>, dog<sup>5</sup> and rat<sup>6</sup>, only recently have methods been developed for a reproducible mouse model. This advancement permits the characterization of trauma-induced OA on defined genetic backgrounds. The murine injury model, which commonly involves transecting the medial collateral ligament coupled with disruption of the meniscus from its antero-medial attachment, leads to OA by 3–4 months<sup>7</sup>. Joint degeneration is either accelerated or decelerated based on which structures are disrupted and the injury-disease relationship has previously been summarized<sup>8</sup>.

Recent studies have used these mouse injury models of OA to delineate molecular mechanisms, including the role of the aggrecans ADAMTS4 and ADAMTS5<sup>9-10</sup> and the transcription factor Runx2. While these and other studies established the utility of mouse injury-induced OA models for studying molecular mechanism(s) of disease, these studies typically rely on non-quantitative histologic and radiographic techniques and/or semi-quantitative grading to evaluate arthritis severity; quantitative analyses of the joints are generally lacking. Thus, the conclusions of these studies do not stand up well to the rigors of statistical testing. It is this gap in the field that leads us to identify quantitative methods to assess joint degeneration under various genetic or experimental conditions as well as to develop an index incorporating multiple outcomes for increased statistical sensitivity compared to any single outcome.

In this report, we describe a series of outcomes that inform us about murine knee joint OA in a manner that supports statistical testing. Using an established mouse meniscal-ligamentous injury (MLI) model<sup>8</sup>, we assessed articular cartilage degeneration using both conventional semi-quantitative histologic scoring methods<sup>6, 9, 11-12</sup> and the novel cartilage area determination method (histomorphometry). We also employed microCT to quantitatively analyze joint bone parameters associated with OA (sclerosis and osteophyte formation). Finally, we hypothesized that incorporation of multiple semi-quantitative and quantitative outcomes into a single calculation would provide enhanced sensitivity relative to any individual outcome. Our results indicate that fusion of conventional semi-quantitative histologic grading and quantitative histomorphometry into an Arthritis Index calculation provides better sensitivity and statistical power than either measure alone. We propose that use of the Arthritis Index provides a sensitive measure of joint degeneration at early and late time points in the disease process, making it a candidate primary outcome for studying disease, in particular when statistical analysis of a candidate therapeutic is required.

## Materials and Methods

Twelve week old male C57/BL6 mice were administered either MLI or sham surgery as previously described<sup>8</sup>, to semi-quantitatively and quantitatively assess progressive injury-induced knee joint degeneration over a 20 week time period. There were 5 MLI and 5 sham

surgery mice for each of 5 time points: 4, 8, 12, 16, & 20 weeks (post-surgery). Mouse surgical procedures were performed with the approval of the Institutional Animal Care & Use Committee at the University of Rochester Medical Center. Following anesthesia, a 5 mm incision was made on the medial aspect of the joint. The medial collateral ligament was transected, the joint space was opened slightly and a 25 gauge needle was used to detach the medial meniscus from its anterior tibial attachment. The sham group involved a similar incision, but tissues were not manipulated. The skin was closed with 4.0 silk sutures applied in an interrupted pattern. Post surgery, mice were provided analgesia (IP injection of 0.5 mg/kg buprenorphine) every 12 hrs for 72 hrs and the sutures were removed after 10 days.

A systematic approach to preparation, sectioning and visualizing articular cartilage was employed for all tissue-based assays<sup>13</sup>. Mice were sacrificed via an AMVA-approved method and the surgically-manipulated knee joints were dissected with the femur and tibia intact to maintain the structural integrity of the joint. The tissue was fixed at 23°C in 10% neutral buffered formalin for 72 hours, de-calcified in 10% w/v EDTA for 21 days and embedded in paraffin. Tissue blocks were then serially sectioned in the midsagittal plane through the medial compartment of the joint. Sections (3 microns thick) were cut at 3 levels within the medial compartment, each level being 50 microns from the previous level (Figure 1). These cut sections were mounted on positively-charged glass slides, baked at 60°C for 30 min, de-paraffinized in xylene and re-hydrated in decreasing concentrations of ethanol. Sections were stained with alcian blue/orange G for histologic grading and histomorphometry.

One stained section from each of the 3 histologic levels in each joint (described above) was assessed via histologic grading. Three sections (i.e. levels) per joint were evaluated to better visualize the OA status of the entire medial compartment. Grading of articular cartilage (including both the femoral chondyle and tibial plateau) was performed by two blinded observers based on two scoring systems. In the first system, described by Chambers *et al*<sup>9, 11</sup>, 0 = normal cartilage, i.e. lack of superficial zone fibrillation or clefting; 1 = mild superficial fibrillation; 2 = fibrillation and/or clefting extending below the superficial zone; 3 = mild (<20%) loss of non-calcified cartilage; 4 = moderate (20%–80%) loss of non-calcified cartilage; 5 = loss of cartilage to the calcified zone; and 6 = Severe (>80%) loss of non-calcified cartilage. The second system that was employed was a minor modification of the scoring scheme described by Bendele *et al*<sup>6, 12</sup>. Briefly, 0 = normal cartilage; 1 = minimal, superficial zone only; 2 = mild, extends into the upper middle zone; 3 = moderate, well into the middle zone; 4 = marked, into the deep zone but not reaching the tide mark; 5 = severe, full thickness degeneration to the tidemark. The assigned grade in this system is then multiplied by a weighting factor to account for the extent of the degeneration. Specifically, the score is multiplied by 1 if degeneration encompasses less than 1/3 of the articular surface, 2 if it encompasses between 1/3 and 2/3, and 3 if the degeneration encompasses more than 2/3<sup>6, 12</sup>. Agreement between the two blinded observers was confirmed for both scoring methods using simple and weighted kappa. The observer agreement for the Chambers scoring was 0.48 and 0.88 respectively. For the Bendele scoring, observer agreement was 0.48 and 0.93 respectively. The 2 observer scores were averaged for each section. These average scores from all sections in a given experimental group were then combined to calculate an overall average score. Significance between the injured and sham groups was assessed using the Mann-Whitney-Wilcoxon test, whereby a *p*-value of less than 0.05 was taken to indicate significance.

We have established a simple histomorphometric approach to quantify articular cartilage area using sections obtained from mouse knee joints<sup>13</sup>. Using alcian blue/orange G stained sections, the Osteometrics system (OsteoMetrics) was utilized to quantify articular cartilage area on 1 tissue section at each of 3 levels (50 microns apart) in the medial compartment of

every joint. Using the OsteoMetrics stylus, projected images of the articular cartilage, that were obtained using an Olympus microscope (4x objective) outfitted with a video camera, were outlined on the femoral condyle and tibial plateau in an area defined by the anterior and posterior margins of the meniscus (Figure 3C, green outline). Then, using an area-calculating algorithm in the Osteomeasure software, the area of collected regions of interest (ROI) was quantified for every section. Area values for every section from a given joint were then averaged. All histomorphometry data were normalized to the average of the sham-operated control joints from the 4 week time point and significant differences between groups were identified by ANOVA, with  $p$ -values  $<0.05$  taken as significant.

Prior to histologic processing, harvested knee joints were evaluated via microCT using a Scanco vivaCT40 scanner as we have previously described<sup>13</sup>. Joints were scanned at a resolution of 12  $\mu\text{m}$  with a slice increment of 10  $\mu\text{m}$  from mid-femur to mid-tibia. Images from each group were reconstructed at identical thresholds and analysis of bone volume was performed on selected regions between the femoral and tibial physes. Significance differences between groups were identified by ANOVA, with  $p$ -values  $<0.05$  taken as significant.

There are numerous advantages to using a composite index<sup>14–16</sup>, including increased statistical efficiency and a simplified analysis and interpretation. Combining several outcomes into a single composite index avoids the arbitrary selection of a single primary outcome variable, and it makes Type I error adjustments for multiple co-primary outcome variables unnecessary.

In order to create an arthritis index incorporating multiple outcome measures, each measurement must first be converted to a common scale. This can be accomplished by the use of t-score transformations<sup>17–18</sup>, resulting in measurements on a unitless, standard deviation (SD) scale:

$$t_{CA} = \frac{CA - m_{CA}}{s_{CA}}, \quad t_{CS} = \frac{CS - m_{CS}}{s_{CS}}.$$

Here CA and CS represent an individual mouse's cartilage area and cartilage score (Chambers or Bendele),  $m_{CA}$  and  $m_{CS}$  are the mean cartilage area and cartilage score over all sham-operated mice, and  $s_{CA}$  and  $s_{CS}$  are the SD of cartilage area and cartilage score over all sham-operated mice. As a consequence, these t-scores have a zero mean and a unit variance among sham-operated mice. In comparison, MLI mice tend to have smaller t-scores for cartilage area (Figure 5C) and larger t-scores for either Chambers or Bendele cartilage score (Figure 5A and 5B, respectively).

The simplest way to combine these t-scores into a single composite index is to average or sum them<sup>17–18</sup>. However, an adverse effect is in opposite directions for cartilage area and the cartilage score. In order to prevent the t-scores from cancelling when summing, the sign of one of the t-scores must be reversed. Our simple index is then the difference in t-scores:  $t_{CS} - t_{CA}$  (Figure 6).

Ultimately our goal is to have a composite index to use as a primary outcome variable in a longitudinal study comparing treatments among MLI and sham-operated mice. An efficient statistical analysis in such a study would incorporate all measurements over time in comparing treatments. When progression over time is linear, comparing slopes (rates of change over time) is an efficient and valid approach to summarizing treatment efficacy.

Combining the t-scores such that the linearity of the resulting index over time is maximal is therefore desirable.

We considered several alternatives to induce longitudinal linearity of our composite index, including linear and quadratic functions of the t-scores. Regression models were used to estimate coefficients for the t-scores in these functions. Coefficients were chosen such that the correlation of the composite index with time was maximized among the MLI mice. These coefficients were then normalized so that all indices have the same variance among sham-operated mice.

## Results

MLI induced progressive OA-like cartilage degeneration of the mouse knee over a 20 week period (Figure 2), with increasing Chambers and Bendele scores that were statistically significant at all time points (Figure 3A and 3B). The most severely degenerated joints from the MLI group (20 weeks) had a score that was approximately 14-fold higher than the control group for Chambers score and 18-fold for the Bendele score. Next, the Osteometrics system was used to quantify articular cartilage areas and as expected, MLI joints showed a progressive reduction of cartilage area that became statistically significant by 8 weeks post injury (Figure 3D). The cartilage area decrease was maximal at 20 weeks post-injury, when the sham group had 3.1-fold more cartilage than the MLI group (Figure 3D).

Osteophyte formation and subchondral sclerosis are also common features of OA<sup>3</sup>. MicroCT scans were collected at 16 and 20 weeks post-surgery and bone volume in the region between the distal femoral and proximal tibial physes was quantified. There was no evidence of osteophytes or subchondral sclerosis in MLI joints at 16 weeks (data not shown). However, scans of 20 week MLI joints revealed areas of ectopic bone formation at joint margins resembling osteophytes (Figure 4A, yellow arrows) and increased bone volume (Figure 4B) potentially reflecting both osteophyte formation and subchondral sclerosis. As differences in bone parameters between MLI and sham groups only emerged during end-stage OA following the cartilage degeneration phase, we excluded microCT measurements from the Arthritis Index.

As a first step toward the development of an Arthritis Index, raw Chambers and Bendele scores and cartilage areas were transformed into respective t-scores for each outcome to remove units and permit combining two different outcome measures into a single formula to calculate the Index. Data from the plots presented in Figure 3 were transformed and new plots were generated depicting the t-Chambers score (Figure 5A), the t-Bendele score (Figure 5B) and the t-cartilage area (Figure 5C). The simple Arthritis Index involves subtraction of the t-cartilage area from the t-cartilage score (Chambers or Bendele) to arrive at an Index; plots of this transformation and best fit calculations for each dataset are presented in Figures 6A and 6B. A series of formulae were developed to generate curves that represent an Arthritis Index calculation at each time point for each experimental group. Table 1 presents results for each t-score and four of these potential Arthritis Indices, including the simple index ( $t_{CS} - t_{CA}$ ) using either Chambers or Bendele scores, a weighted index using Chambers scores ( $1.2 \cdot t_{CS} - 0.6 \cdot t_{CA}$ ), and a weighted index using Bendele scores ( $0.5 \cdot t_{CS} - 1.2 \cdot t_{CA}$ ). We considered other indices, including linear and quadratic functions of the t-scores, but results were similar (data not shown). Test statistics and *p*-values for comparing observed data from the experimental groups (sham vs. MLI), along with effect size estimates (Cohen's *d*) and sample sizes required to achieve desired significance levels and power for future studies, are listed in Table 1 for each time point and across all time points (slopes). Although the t-Chambers and t-Bendele scores perform better than the indices at a few time points (larger test statistic and effect size, smaller *p*-value and sample

size), overall the indices are far more efficient. This is evident by comparing the results for slopes (rate of change over time), where the indices perform comparably, with both indices having more desirable properties than either individual t-score (i.e. t-Chambers, t-Bendele or t-cartilage area) alone. By defining relative efficiency as the ratio of required sample sizes, the t-Chambers and t-Bendele scores have 57% efficiency and the t-cartilage area has 71% efficiency compared to either index.

## Discussion

The aims of this study were to identify quantitative methods to assess arthritis in murine models and to develop an Arthritis Index with increased statistical sensitivity compared to any single semi-quantitative or quantitative method. While numerous genetic and injury models of OA have been described over the past decade, the assessment of the arthritic phenotype has generally involved non- or semi-quantitative outcomes that fail to support statistical testing. The need for statistics becomes imperative when evaluating the efficacy of candidate therapeutic strategies. We therefore employed an established MLI method to induce a degenerative OA-like process in the mouse knee (Figure 2). Histological evidence of fibrillation and clefting at earlier time points and later erosion to subchondral bone and eburnation were consistent with results from published reports using analogous injury-induced OA models<sup>8-10, 19-21</sup>. We therefore utilized this model to establish quantitative imaging and histology-based methods as well as a novel Arthritis Index that possesses enhanced sensitivity in quantifying disease progression.

The current standard for staging of the disease process in the mouse is the semi-quantitative assessment of joint tissue by blinded observers. Numerous mouse studies have employed histologic scoring methods to gauge the severity of degeneration. Examples include the Chambers and Bendele scoring methods for the assessment of disease in tissue sections<sup>6, 9-12, 19-20</sup>. These scoring protocols effectively distinguishes between early, mid and late disease as exemplified in Figures 3A and 3B, where MLI joints were assigned a higher score than sham-operated joints, even at the 4 week time point.

Since the central tissue phenotype present in a degenerating joint is loss of cartilage, quantitative assessment of cartilage loss should strongly correlate with severity/progression of the OA disease process. Our group recently published a method to quantify cartilage area in serial sections from knee joints that utilizes the Osteometrics system to perform cartilage histomorphometry<sup>13</sup>. Use of this method revealed progressive cartilage loss in MLI joints over the 20 week time course (Figure 3D). Significant reductions in cartilage were first detectable at 8 weeks post-surgery, and by 20 weeks, nearly 70% of the cartilage was lost.

In addition to cartilage degeneration, it is widely accepted that active OA involves the development of subchondral sclerosis and osteophytes<sup>3</sup>. Since microCT represents the state-of-the-art approach to quantifying various bone parameters, several recent studies have used this method to examine the presence of these phenotypes in mouse models of OA<sup>13, 22-23</sup>. OA caused by MLI is also associated with increased bone volume (relative to sham control joints) in the region between the femoral and tibial physes (see Figure 4). This increase could be due to subchondral sclerosis, the deposition of mineral associated with the formation of osteophytes, or a combination of both. This quantification of bone volume distinguishes arthritic from healthy joints at 20 weeks post surgery, suggesting bone volume determination represents another quantitative outcome that can identify joint changes in severely degenerated joints. Since bone volume measurements made at earlier time points in the MLI-induced disease process were not increased relative to control joints (data not shown), there is no correlation between stage of disease and joint bone volume. Thus, although the quantitation of bone volume is valuable for identifying end-stage disease, it is

not useful for comparing diseased and normal joints during the cartilage degeneration phase which precedes total joint destruction and eburnation. Since the phenotypes microCT can measure do not vary in intensity in a manner that correlates with the apparent severity of joint degeneration, we did not include microCT bone measurements in the Arthritis Index.

While both semi-quantitative histologic scoring and quantitative cartilage area determination methods allow for effective staging of arthritic processes in the mouse, we hypothesized that combining these methods in a single calculation would represent a novel approach providing enhanced sensitivity compared to either method alone. Since the raw data derived from the histologic scores (Chambers or Bendele) are unitless and the area values have units ( $\mu\text{m}^2$ ), their direct incorporation into a single equation is not possible. Thus, t-scores were calculated, and then a series of 4 algorithms were developed to calculate an index for MLI and sham-operated joints at each time point. The relative ability of each algorithm to distinguish between MLI and sham-operated joints was compared, with all 4 calculations essentially equally capable of distinguishing between the two experimental groups. As a result, the Arthritis Index was defined to be the Simple calculation, which is the t-score for the cartilage area (t-Cartilage Area, Figure 5C) subtracted from the t-score for the cartilage score (t-Chambers or t-Bendele, Figures 5A and 5B, respectively). Figure 6 depicts the linear fits of Arthritis Index calculations for MLI and control groups incorporating either the t-Chambers (6A) or t-Bendele (6B). Based on the magnitude of the detected difference between MLI and control groups over all time points, the Chambers or Bendele scores alone were only 57% efficient and the cartilage area alone was only 71% efficient as compared to the respective Arthritis Index. This means that future studies using the Arthritis Index as the primary outcome measure would require only 57% – 71% of the sample size required by studies using cartilage scores or cartilage area calculations alone as the primary outcome variable. It should be noted that the Arthritis Index is approximately equally efficient whether t-Chambers or t-Bendele score are used in the calculation, suggesting that other histologic scoring methods would be compatible with the general approach taken here to calculate the Index. Thus, it is based on this enhanced sensitivity and versatility that we propose use of the Arthritis Index as a primary quantitative outcome measure for the assessment of injury- or genetic-based murine arthritis. This enhanced sensitivity would be particularly valuable in studies aimed at testing the efficacy of a candidate therapeutic strategy.

## Acknowledgments

The authors wish to thank Erica Dussmann and Nathan Miller for technical assistance. This work was supported by NIH/NIAMS P50 AR054041 (RNR), NIH/NIAMS RO1 AR045700 (RNR), and an Arthritis Foundation Arthritis Investigator Award (MJZ). ERS was supported by NIH/NIAMS T32 AR053459.

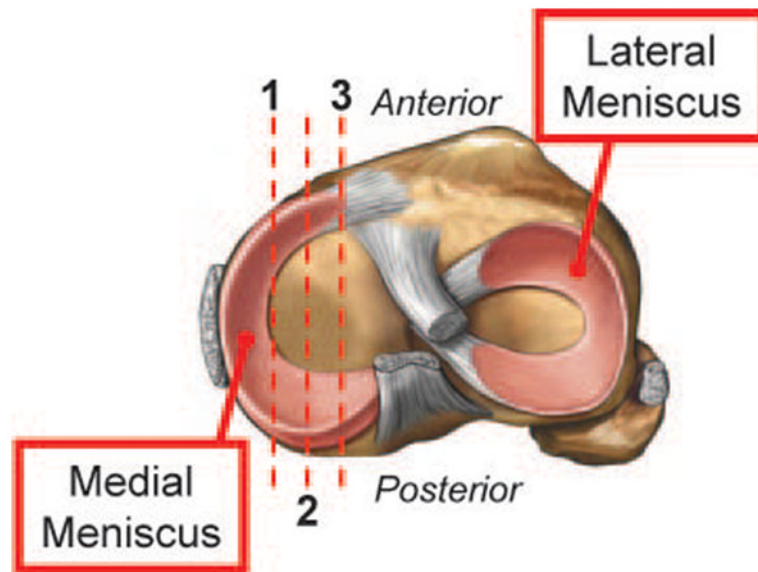
## References

1. CDC. Prevalence of disabilities and associated health conditions among adults--United States, 1999. *MMWR Morb Mortal Wkly Rep.* 2001; 50(7):120–125. [PubMed: 11393491]
2. Lawrence RC, Helmick CG, Arnett FC, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis and Rheumatism.* 1998; 41(5):778–799. [PubMed: 9588729]
3. Buckwalter JA, Mankin HJ, Grodzinsky AJ. Articular cartilage and osteoarthritis. *Instr Course Lect.* 2005; 54:465–480. [PubMed: 15952258]
4. Suzuki Y, Takeuchi N, Sagehashi Y, Yamaguchi T, Itoh H, Iwata H. Effects of hyaluronic acid on meniscal injury in rabbits. *Arch Orthop Trauma Surg.* 1998; 117(6–7):303–306. [PubMed: 9709840]

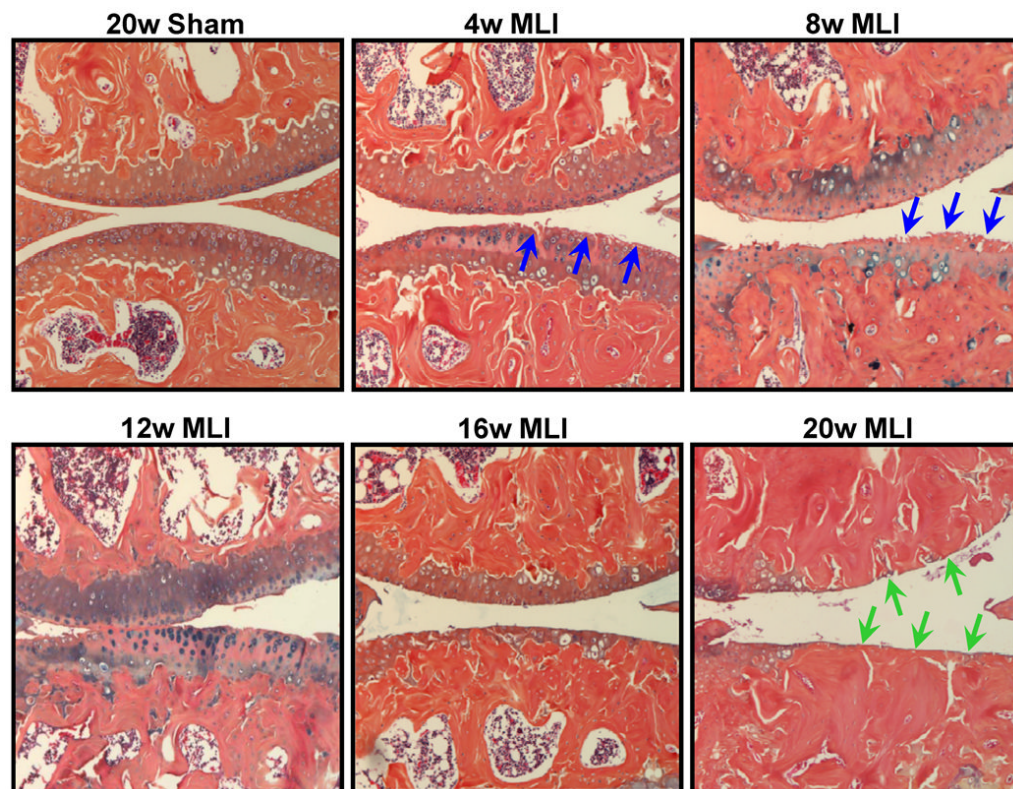
5. Goto H, Shuler FD, Niyibizi C, Fu FH, Robbins PD, Evans CH. Gene therapy for meniscal injury: enhanced synthesis of proteoglycan and collagen by meniscal cells transduced with a TGFbeta(1)gene. *Osteoarthritis Cartilage*. 2000; 8(4):266–271. [PubMed: 10903880]
6. Janusz MJ, Bendele AM, Brown KK, Taiwo YO, Hsieh L, Heitmeyer SA. Induction of osteoarthritis in the rat by surgical tear of the meniscus: Inhibition of joint damage by a matrix metalloproteinase inhibitor. *Osteoarthritis Cartilage*. 2002; 10(10):785–791. [PubMed: 12359164]
7. Clements KM, Price JS, Chambers MG, Visco DM, Poole AR, Mason RM. Gene deletion of either interleukin-1beta, interleukin-1beta-converting enzyme, inducible nitric oxide synthase, or stromelysin 1 accelerates the development of knee osteoarthritis in mice after surgical transection of the medial collateral ligament and partial medial meniscectomy. *Arthritis and Rheumatism*. 2003; 48(12):3452–3463. [PubMed: 14673996]
8. Kamekura S, Hoshi K, Shimoaka T, et al. Osteoarthritis development in novel experimental mouse models induced by knee joint instability. *Osteoarthritis Cartilage*. 2005; 13(7):632–641. [PubMed: 15896985]
9. Glasson SS, Askew R, Sheppard B, et al. Deletion of active ADAMTS5 prevents cartilage degradation in a murine model of osteoarthritis. *Nature*. 2005; 434(7033):644–648. [PubMed: 15800624]
10. Majumdar MK, Askew R, Schelling S, et al. Double-knockout of ADAMTS-4 and ADAMTS-5 in mice results in physiologically normal animals and prevents the progression of osteoarthritis. *Arthritis and Rheumatism*. 2007; 56(11):3670–3674. [PubMed: 17968948]
11. Chambers MG, Cox L, Chong L, et al. Matrix metalloproteinases and aggrecanases cleave aggrecan in different zones of normal cartilage but colocalize in the development of osteoarthritic lesions in STR/ort mice. *Arthritis and Rheumatism*. 2001; 44(6):1455–1465. [PubMed: 11407708]
12. Bendele AM, Hulman JF. Effects of body weight restriction on the development and progression of spontaneous osteoarthritis in guinea pigs. *Arthritis Rheum*. Sep; 1991 34(9):1180–1184. [PubMed: 1930336]
13. Wu Q, Kim KO, Sampson ER, et al. Induction of an osteoarthritis-like phenotype and degradation of phosphorylated Smad3 by Smurf2 in transgenic mice. *Arthritis and Rheumatism*. 2008; 58(10): 3132–3144. [PubMed: 18821706]
14. Ferreira-Gonzalez I, Permanyer-Miralda G, Busse JW, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol*. 2007; 60(7):651–657. [PubMed: 17573977]
15. Freemantle N, Calvert M. Weighing the pros and cons for composite outcomes in clinical trials. *Journal of Clinical Epidemiology*. 2007; 60:658–659.
16. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA*. 2003; 289(19):2554–2559. [PubMed: 12759327]
17. Smythe HA, Helewa A, Goldsmith CH. “Independent assessor” and “pooled index” as techniques for measuring treatment effects in rheumatoid arthritis. *Journal of Rheumatology*. 1977; 4(2):144–152. [PubMed: 881691]
18. Smythe HA, Helewa A, Goldsmith CH. Selection and combination of outcome measures. *Journal of Rheumatology*. 1982; 9(5):770–774. [PubMed: 7175850]
19. Glasson SS, Askew R, Sheppard B, et al. Characterization of and osteoarthritis susceptibility in ADAMTS-4-knockout mice. *Arthritis and Rheumatism*. 2004; 50(8):2547–2558. [PubMed: 15334469]
20. 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):1061–1069. [PubMed: 17470400]
21. Kamekura S, Kawasaki Y, Hoshi K, et al. Contribution of runt-related transcription factor 2 to the pathogenesis of osteoarthritis in mice after induction of knee joint instability. *Arthritis and Rheumatism*. 2006; 54(8):2462–2470. [PubMed: 16868966]
22. Naruse K, Urabe K, Jiang SX, et al. Osteoarthritic changes of the patellofemoral joint in STR/OrtCrlj mice are the earliest detectable changes and may be caused by internal tibial torsion. *Connect Tissue Res*. 2009; 50(4):243–255. [PubMed: 19637060]



23. Blair-Levy JM, Watts CE, Fiorentino NM, Dimitriadis EK, Marini JC, Lipsky PE. A type I collagen defect leads to rapidly progressive osteoarthritis in a mouse model. *Arthritis and Rheumatism*. 2008; 58(4):1096–1106. [PubMed: 18383364]

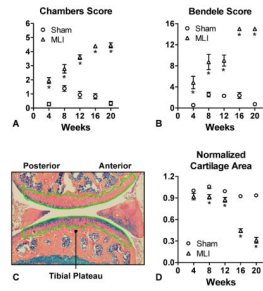


**Figure 1.** Schematic summarizing preparation of sections for histologic analysis of murine knee joints. Red lines represent serial midsagittal sections cut every 50 microns from the medial compartment of the knee joint, resulting in three sections per joint.



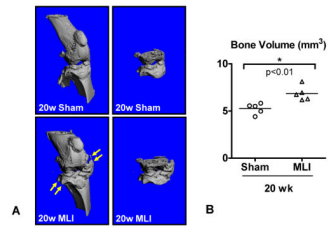
**Figure 2.**

MLI induces progressive degeneration of articular cartilage. Compared to sham-operated joints at 20 weeks post-surgery (20w Sham), joints from mice administered MLI showed progressive cartilage degeneration that was initially discernable 4 weeks post-surgery (4w MLI) and culminated in eburnation 20 weeks out (20w MLI). Evidence of fibrillation and clefting is denoted with blue arrows and erosion to subchondral bone is denoted with green arrows. Presented alcian blue/orange G histology is representative of groups of 5 or more mice for each of the time points.



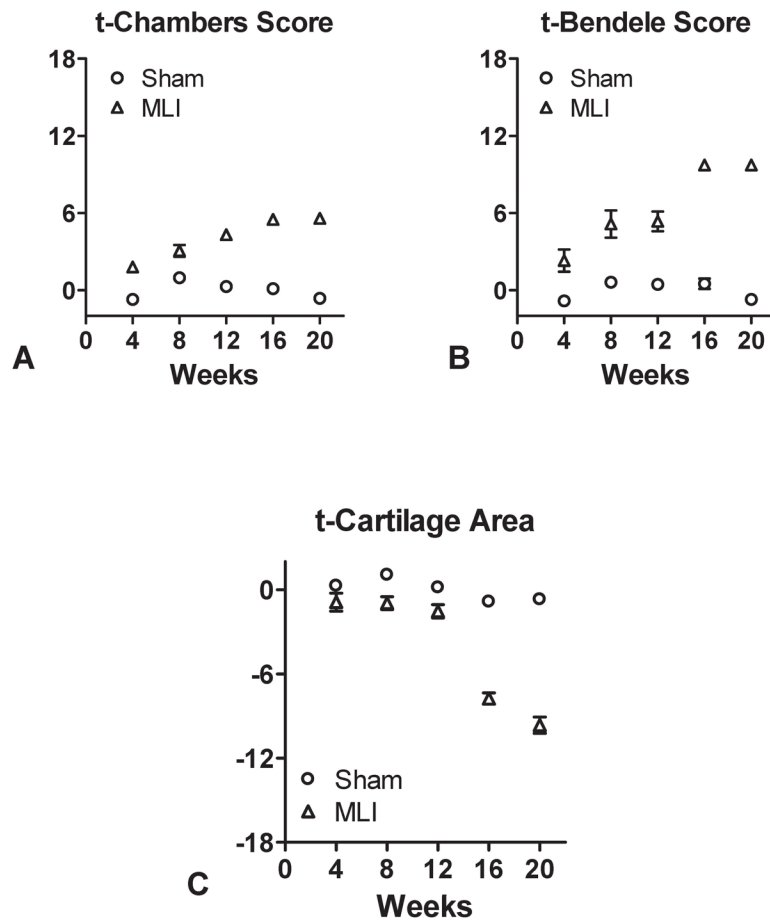
**Figure 3.**

MLI-induced degeneration is significant based on histologic scoring and cartilage area determination. Three sections from each experimental joint (MLI or sham-operated) were analyzed by 2 histologic scoring methods (Chambers and Bendele). For grading, 2 blinded observers assigned a Chambers (A) and Bendele (B) score to each of the 3 sections from a given joint. Representative outlining of articular cartilage to quantitate cartilage area using Osteometrics is depicted in (C), denoted by the green tracing. Cartilage areas were determined and averaged for each of the three sections harvested from a given joint (D). All groups were normalized to the Sham group harvested at 4 weeks. Symbols represent the average Chambers score (A), Bendele score (B) or normalized area (D) for all joints in the same experimental group and the error bars represent SEM (N=3). Asterisks denote statistically significant differences ( $p < 0.5$ ) from the sham operated controls as determined by the Mann-Whitney-Wilcoxon test (A and B) or ANOVA (D).

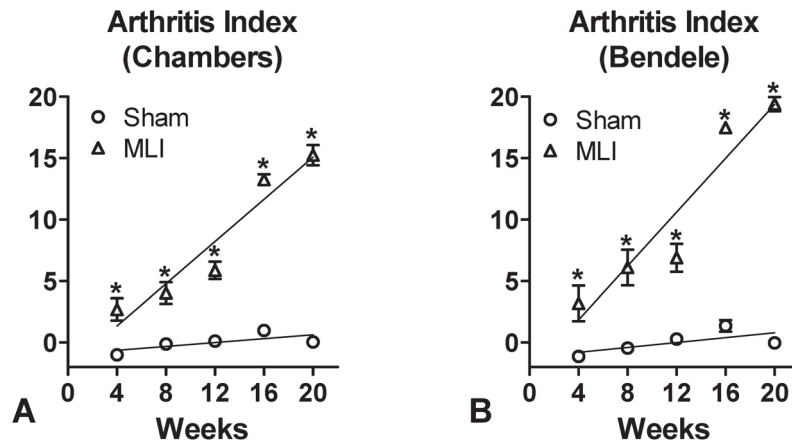


**Figure 4.**

Bone volume is increased in joints administered MLI. High resolution micro CT scans were performed on MLI and sham-operated joints 20 weeks post-surgery. Representative reconstructions are shown in (A), with the region between the distal femoral and proximal tibial physes depicted in the images on the left. Yellow arrows identify periarticular mineralized areas reminiscent of osteophytes. (B) The average joint bone volume for the 2 experimental groups is depicted as a horizontal line with each calculated value depicted with a symbol in a scatter plot (N=5). The asterisk denotes a statistically significant difference between the groups as determined by ANOVA ( $p < 0.05$ ).



**Figure 5.** Calculation of t-Chambers scores, t-Bendele scores and t-cartilage areas. The cartilage scores (Chambers and Bendele) and cartilage area data presented in Figure 3 were converted into t-scores and plotted individually as a prelude to calculation of the Arthritis Index (**A**, **B** and **C**).



**Figure 6.** An Arthritis Index as a quantitative measure of arthritic disease in the mouse knee. The t-Cartilage area (Figure 5C) was subtracted from either the t-Chambers score (Figure 5A) or the t-Bendele score (Figure 5B) and the results are plotted in (A) and (B). Also depicted are the linear best fits for the MLI and sham-operated datasets using either scoring method (Chambers versus Bendele). Symbols in (A) and (B) represent the average Arthritis Index for each experimental group with error bars representing the SEM (N=3). Asterisks denote significant differences between the groups as determined by ANOVA ( $p < 0.05$ ).

**Table 1**

Comparison of t-scores with the Simple and Weighted Indices.

<i>Using the Chambers Grade:</i>												
Time	Grade t-score			CA t-score			Simple Index			Weighted Index		
	t-statistic*	p-value	Effect Size#	N†	t-statistic*	p-value	Effect Size#	N†	t-statistic*	p-value	Effect Size#	N†
4	6.84	1.3E-05	3.22	14	1.75	1.1E-01	0.82	148	3.97	3.5E-03	1.87	32
8	3.83	1.8E-03	1.81	34	3.55	3.0E-03	1.67	38	4.48	1.4E-03	2.11	26
12	9.24	5.9E-07	4.58	10	3.31	7.2E-03	1.54	46	7.11	1.0E-05	3.33	14
16	16.39	4.4E-10	7.73	6	15.47	1.9E-09	7.29	8	20.89	5.4E-13	9.85	6
20	18.08	8.5E-12	8.52	6	14.11	2.2E-08	6.65	8	17.00	6.0E-09	8.01	6
Slope:												
	7.32	2.5E-13	1.59	42	8.34	7.5E-17	1.79	34	10.28	8.3E-25	2.21	24
<i>Using the Bendele Grade:</i>												
Time	Grade t-score			CA t-score			Simple Index			Weighted Index		
	t-statistic*	p-value	Effect Size#	N†	t-statistic*	p-value	Effect Size#	N†	t-statistic*	p-value	Effect Size#	N†
4	3.56	6.7E-03	1.68	38	1.75	1.1E-01	0.82	148	2.94	1.8E-02	1.39	54
8	4.09	2.5E-03	1.93	30	3.55	3.0E-03	1.67	38	4.41	1.8E-03	2.08	26
12	6.23	1.6E-04	2.86	16	3.31	7.2E-03	1.54	46	5.70	3.8E-04	2.61	18
16	22.87	1.4E-08	10.78	6	15.47	1.9E-09	7.29	8	26.28	3.0E-14	12.39	6
20	64.25	3.8E-12	30.29	4	14.11	2.2E-08	6.65	8	30.22	3.3E-12	14.25	6
Slope:												
	7.54	4.7E-14	1.62	42	8.34	7.5E-17	1.79	34	9.50	2.1E-21	2.04	28

\* Welch t-statistic for comparing means (sham vs. MLL) at each time-point; Wald statistic for comparing slopes (sham vs. MLL) over all time-points.

# Maximum Likelihood Estimate of Cohen's d for comparing means (sham vs. MLL) at each time-point and slopes (sham vs. MLL) over all time-points.

† Total sample size required to provide 99% power to detect observed effect sizes at a two-sided 1% significance level.