Abstract

**Objective:** Sheep and goats are commonly used large animal species for studying pathogenesis and treatment of osteoarthritis (OA). This review focuses on the macroscopic and microscopic criteria for assessing OA in sheep and goats and recommends particular assessment criteria to assist standardization in the conduct and reporting of preclinical trials of OA.

**Methods:** A review was conducted of all published OA studies using sheep and goats and the most common macroscopic, microscopic, or ultrastructural scoring systems were summarized. General recommendations regarding methods of OA assessment in the sheep and goat have been made and a preliminary study of their reliability and utility undertaken.

**Results:** The modified Mankin scoring system is recommended for semiquantitative histological assessment of OA in due to its already widespread adoption, ease of use, similarity to scoring systems used for OA in humans, its achievable inter-rater reliability. Specific recommendations are also provided for histological scoring of synovitis and scoring of macroscopic lesions of OA.

**Conclusions:** The proposed system for assessment of sheep and goat articular tissues appears to provide a useful versatile method to quantify OA change. It is hoped that by adopting more standardised quantitative outcome measures, better comparison between different studies and arthritis models will be possible. The suggested scoring systems can be modified in the future as our knowledge of disease pathophysiology advances.
Introduction

Sheep and to a lesser extent goats have been commonly used for the study of OA (see collected references and Table 1) as well as diseases of other musculoskeletal tissues such as intervertebral disc (reviewed in (1)), tendon and ligament (e.g. (2)), and bone (e.g. fracture repair, osteoporosis etc: (3, 4)). These two species offer the advantage of being large animals with their orthopaedic biomechanics perhaps being more comparable to humans than smaller species, despite being quadrupeds (e.g. (5)). As with other larger species, the size of the joints enables many potentially useful OA outcome measures to be used including routine arthrocentesis and synovial fluid collection, radiography/CT/MRI, gait analysis, arthroscopic joint examination, biomechanical evaluation of cartilage, topographical evaluation within a single joint and ample tissue for histological, biochemical and molecular analysis. While a variety of strains/breeds of sheep and goats have been used, the prolonged breeding and selection of these species for meat, wool or milk production means that genetic and/or biological diversity within each breed is reduced which can be advantageous for decreasing biological variability in experiments. Additionally, these species are generally placid in nature and easily managed, and can potentially be more cheaply housed at pasture (depending on the country), although goats can prove more difficult to house than sheep. The disadvantage of both species is that they are ruminants rather than monogastrics, and so the bioavailability of oral therapies may require additional validation. In addition, the full genome sequences are not available for either species, and therefore design of molecular biology reagents (e.g. primers for RT-PCR, siRNA) is more difficult and large-scale microarray chips are not available.

Spontaneous OA has not been reported as a viable model in either sheep or goats. While spontaneous OA appears to be rare in sheep, proteoglycan loss, cartilage matrix atrophy and erosion of the cartilage consistent with OA change occurs spontaneously in the unprotected region of the medial tibial plateau in goats as young as 2 years of age (Murphy et al manuscript in preparation). In addition, goats are susceptible to caprine arthritis/encephalitis caused by a lentivirus, which leads to chronic profound synovitis and joint degeneration, particularly but not exclusively in the carpus and tarsus. There is one report where decreased cartilage proteoglycan was detected in goat knee cartilage 24 hours after various doses of intra-articular papain, but long term follow up was not performed (6). The remaining OA models in sheep and goats are all surgically induced, with the most common involving partial or total meniscectomy (uni-or bilateral, medial or lateral, with or without other ligament transactions). In contrast with other species (e.g. dog, rabbit), anterior (or cranial) cruciate ligament transection (ACLT) has not been reported commonly as a model of OA in sheep and goats. ACLT alone appears to induce very limited/mild cartilage damage in these species (7-10). It is important therefore that studies of ACL reconstruction using sheep and goats should always be compared with a non-repaired group if modulation of OA is an outcome measure (11).
Table 1. Osteoarthritis models in sheep and goats. Note only surgically-induced models of OA have been reported in sheep and goats i.e. no spontaneous or chemically-induced models

<table>
<thead>
<tr>
<th>Species</th>
<th>Surgery</th>
<th>Post-op time</th>
<th>Findings</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>sheep</td>
<td>unilateral radial meniscal tear</td>
<td>6 &amp; 12 mth</td>
<td>Retropatellar OA with or without suture of meniscus</td>
<td>(12)</td>
</tr>
<tr>
<td>sheep</td>
<td>unilateral caudal pole hemi-meniscectomy</td>
<td>3 mth</td>
<td>No difference in pathology between hemi- and total meniscectomy</td>
<td>(13)</td>
</tr>
<tr>
<td>sheep</td>
<td>unilateral medial meniscectomy</td>
<td>6 mth</td>
<td>Early OA features (AC fissuring/fibrillation, chondrocyte enlargement, matrix proliferation, marginal osteophytes) more marked in exercised group</td>
<td>(14-16)</td>
</tr>
<tr>
<td>sheep</td>
<td>unilateral lateral meniscectomy</td>
<td>6 mth</td>
<td>Lateral meniscectomy induced greater depletion of PG than medial meniscectomy; Along with unicompartmental histological cartilage damage, meniscectomy increased osteoid volume &amp; surfaces with increased labeling of subchondral bone</td>
<td>(17-23)</td>
</tr>
<tr>
<td>sheep</td>
<td>bilateral lateral meniscectomy</td>
<td>2 wk - 6 mth</td>
<td>Higher AC lesion scores and lower AC PG content compared with unilateral meniscectomy.</td>
<td>(24-27)</td>
</tr>
<tr>
<td>sheep</td>
<td>unilateral ACLT</td>
<td>20 wk</td>
<td>Significant increase in joint laxity with ACLT at 20 weeks but OA changes very mild.</td>
<td>(7)</td>
</tr>
<tr>
<td>sheep</td>
<td>unilateral ACLT + MCLT</td>
<td>20 wk</td>
<td>Joints laxity at 20 wks; no OA change reported but change in PCL and LCL</td>
<td>(28)</td>
</tr>
<tr>
<td>sheep</td>
<td>unilateral ACLT + medial meniscectomy</td>
<td>6 mth &amp; 1 yr</td>
<td>Joints with meniscectomy alone or in combination had worse OA than ACLT alone</td>
<td>(10)</td>
</tr>
<tr>
<td>goat</td>
<td>unilateral medial meniscectomy</td>
<td>4 &amp; 8 mth</td>
<td>Decreased AC contact area and increased contact stress diminished with time post operatively.</td>
<td>(29)</td>
</tr>
<tr>
<td>goat</td>
<td>unilateral collateral ligament plus meniscal transection and meniscal scarification</td>
<td>14 wk</td>
<td>Minimal cartilage degeneration other than that caused directly by scarification.</td>
<td>(30)</td>
</tr>
<tr>
<td>goat</td>
<td>unilateral ACLT</td>
<td>4 wk - 8 mth</td>
<td>Knee joints unstable but variable/minimal signs of synovitis, inflammation or AC degeneration</td>
<td>(8, 9)</td>
</tr>
<tr>
<td>goat</td>
<td>unilateral ACLT plus medial meniscectomy</td>
<td>6 mth</td>
<td>OA with cartilage erosion; some modulation of OA with stem cells</td>
<td>(31)</td>
</tr>
</tbody>
</table>

Table 2: Table listing previously proposed grading systems for sheep and goat joint degeneration

<table>
<thead>
<tr>
<th>Comments</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sheep: Gross pathology score for cartilage fissuring/erosion and for osteophytosis</td>
<td>(14, 32)</td>
</tr>
<tr>
<td>Sheep: Depth of cartilage structural damage</td>
<td>(25)</td>
</tr>
<tr>
<td>Sheep: Synovial inflammation and fibrosis</td>
<td>(33)</td>
</tr>
<tr>
<td>Sheep: Mankin cartilage histopathology</td>
<td>(34)</td>
</tr>
<tr>
<td>Sheep: Various modifications of Mankin cartilage histopathology</td>
<td>(14, 22)</td>
</tr>
<tr>
<td>Goat: histological cartilage structural damage and toluidine blue staining, subchondral bone thickening and osteophyte maturity.</td>
<td>(31)</td>
</tr>
</tbody>
</table>

1. Anatomy of the sheep and goat knee (stifle) joint.

The predominant appendicular joint OA models described in the literature in sheep and goats have used the stifle/knee joint and are surgically induced (summarised in Table 1). Numerous reports on induced temporo-mandibular joint arthritis in sheep have been published but are not the focus of the present paper. The anatomy of the sheep and goat
stifle is grossly similar that of the human (Fig 1) (35). The articular cartilage is generally thicker in goats compared with sheep, and thus closer to that found in man, which may offer advantages particularly for cartilage repair studies (36). Both sheep and goats have an intra-articular long digital extensor tendon which crosses the joint in the cranio(anterolateral) compartment and there is an associated groove on the anterior aspect of the lateral tibial plateau to accommodate this tendon (Fig 1A&D). The trochlea of the distal femur is much deeper and more well defined in sheep and goats (as it is in most quadrupeds) compared with humans (Fig 1A&B). The medial tibial condyle is smaller than the lateral and the medial meniscus is similarly smaller and more circumferential compared with the larger ovoid shaped lateral meniscus (Fig 1C). The tendon of the popliteus muscle courses diagonally across the caudo(postero)lateral aspect of the sheep knee, deep to the lateral collateral ligament from its origin craniodistal of the lateral epicondyle (Fig 1D). Following lateral meniscectomy in sheep, the popliteal tendon falls into the articulation and undergoes meniscoid structural and biochemical changes and may act to partially protect the cartilage of the posterior tibial plateau (Cake et al, manuscript in preparation).

Figure 1: Anatomy of the sheep knee/stifle joint. Abbreviations: LCL and MCL = lateral and medial collateral ligament; ACL and PCL = anterior and posterior cruciate ligaments; LDE = long digital extensor; PT = popliteal tendon; MM and LM = medial and lateral meniscus; MFC and LFC = medial and lateral femoral condyle; TG = trochlear groove of distal femur.

While sheep and goats of various ages (6 months to 7 years) at the time of surgery have been studied, it is recommended that “skeletally mature” animals be used, which is likely >/= 2 years of age in most sheep and goat breeds, when longitudinal growth of long bones has ceased even though growth plates may not be radiographically fused. Surgically-induced OA in sheep and goats has been evaluated in males (castrated and intact) and females, and the effect of ovariectomy and estrogen replacement has been further examined in the latter. Cartilage changes have been reported as a result of ovariectomy in sheep (37, 38) and OA worsens with time post surgery and with exercise following meniscectomy in sheep (14-16, 39-41), thus the sex and exercise of animals used for OA studies should be standardised.

II. Scoring of Alterations in Joint Structures

IIa. Macroscopic scoring (Staging of the disease process)

Scoring systems for macroscopic grading/staging of cartilage damage (Table 3A) and osteophyte development (Table 3B) have been described in OA models in sheep (14, 25, 32, 34). In general these analyses have been performed on unstained specimens or digitized images, although India ink staining to demarcate fibrillated/fissured/eroded cartilage has been reported in the sheep (34). No previous macroscopic scoring system for synovium has been described in sheep or goats, however a similar system to that reported for dogs (see article in this journal – Table 3C) could be used. Representative images of joints demonstrating the different macroscopic scores for cartilage damage, osteophytosis, and synovial pathology are shown in Figures 2, 3 and 4, respectively. Very mild cartilage and osteophyte changes (Table 3A & 3B - score 1) are difficult to appreciate on the digital images but can be more easily observed in vivo. The authors have not observed macroscopic synovial scores greater than 2 (Table 3C) in untreated sheep OA joints.
### Table 3A: Macroscopic scoring of cartilage

**Gross articular damage** - score each area separately

<table>
<thead>
<tr>
<th>Assessment of central cartilage of</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>- medial tibial condyle</td>
<td></td>
</tr>
<tr>
<td>- lateral tibial condyle</td>
<td></td>
</tr>
<tr>
<td>- medial femoral condyle</td>
<td></td>
</tr>
<tr>
<td>- lateral femoral condyle</td>
<td></td>
</tr>
</tbody>
</table>

| Normal                             | 0     |
| Surface roughening                 | 1     |
| Fibrillation and fissures          | 2     |
| Small erosions down to subchondral bone (<5mm diameter) | 3     |
| Larger erosions down to subchondral bone (>5mm diameter) | 4     |

Add to give lesion score 0 - 16

### Table 3B: Macroscopic scoring of osteophytes

**Osteophyte development** - score each area separately

<table>
<thead>
<tr>
<th>Assessment of central cartilage of</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>- medial tibial condyle</td>
<td></td>
</tr>
<tr>
<td>- lateral tibial condyle</td>
<td></td>
</tr>
<tr>
<td>- medial femoral condyle</td>
<td></td>
</tr>
<tr>
<td>- lateral femoral condyle</td>
<td></td>
</tr>
</tbody>
</table>

| Normal                             | 0     |
| Mild osteophyte development (<2mm outgrowth or < 20% of joint margin) | 1     |
| Moderate osteophyte development (2-4 mm outgrowth or 20-50% of joint margin) | 2     |
| Large osteophyte development (>4mm outgrowth or >50% of joint margin) | 3     |

Add to give osteophyte score 0 - 12

### Table 3C: Macroscopic scoring of synovium (as suggested for use in dog)

**Gross Characteristics**

<table>
<thead>
<tr>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal – opal white, semitranslucent, smooth, with sparse well defined blood vessels</td>
</tr>
<tr>
<td>Slight – focal involvement, slight discoloration, visible fibrillation/thickening, notable increase in vascularity</td>
</tr>
<tr>
<td>Mild – diffuse involvement, slight discoloration, visible fibrillation/thickening, notable increase in vascularity</td>
</tr>
<tr>
<td>Moderate – diffuse involvement, severe discoloration, consistent notable fibrillation/thickening, moderate vascularity</td>
</tr>
<tr>
<td>Marked – diffuse involvement, severe discoloration, consistent and marked fibrillation/thickening, marked synovial proliferation with diffuse hypervascularity</td>
</tr>
<tr>
<td>Severe – diffuse involvement, severe discoloration, consistent and severe fibrillation, thickening to the point of fibrosis, severe proliferation and hypervascularity</td>
</tr>
</tbody>
</table>
Figure 2: Macroscopic scoring of cartilage pathology in the tibia and femoral condyles of sheep following meniscectomy-induced OA. The score based on table 3A for the cartilage pathology of the lateral and medial (left and right sides, respectively in each image) tibial or femoral condyle is shown in each image.

Tibia

Femur
Figure 3: Macroscopic scoring of osteophytosis in the tibial and femoral condyles of sheep following meniscectomy-induced OA. The score based on table 3B for the osteophyte development in the medial and lateral (left and right sides, respectively in each image) tibial or femoral condyle is shown in each image.
Figure 4: Macroscopic scoring of synovial pathology in the tibial plateau and femoral condyle of sheep following meniscectomy-induced OA. The score based on the table reported for scoring synovial pathology in dogs.

IIb. Microscopic scoring of cartilage alterations (Grading of cartilage degeneration)

Histopathology, particularly of cartilage, has been the predominant outcome measure used to quantify OA and response to therapy in sheep and goat models. The majority of the histopathology scoring systems used have been based on that originally reported by Mankin et al where cartilage structure, cellularity, chondrocyte cloning, proteoglycan staining and tidemark integrity are scored (42). Variations in this original scoring system with fewer or more parameters have been suggested (see Table 2). It is recognised that modified Mankin type systems are deficient in only evaluating cartilage and not other OA-affected tissues such as synovium, meniscus, ligaments and bone, and therefore additional scoring systems must be developed and used for these (see later sections). Furthermore, the relative importance of the different cartilage pathology parameters scored is not taken into account, although it may be argued that our understanding of which changes are in fact the most important in pathophysiology, progression or treatment of OA is also lacking. Further issues arise in histological analysis, as even within a single section e.g. across one tibial or femoral condyle, the severity of cartilage damage will be variable and the percentage area affected by the different degrees of damage is not measured with Mankin-based systems. These issues lead to the development of a histopathological (OARSI) scoring system described by Pritzker et al. (43) where only structural damage of cartilage is evaluated, but a measure of both maximal severity and area involved is included. This method has been evaluated in goats following hemi-arthroplasty (44) and in sheep following meniscectomy or meniscal destabilization (Cake et al manuscript in preparation) and appears to be quite applicable to sheep and goat OA models. Nevertheless, until more comparative studies are published, we have continued to recommend use of the modified Mankin system outlined in this paper (see below), although subsequent or simultaneous analysis using the OARSI system may be useful and can in the authors experience be rapidly done.

As change in cartilage is often very focal, particularly in surgically induced OA models, all histopathological analyses are plagued by issues and questions of sampling – i.e. the number of joint regions and sections within each region that should be evaluated. Whether sections for histological comparison in a therapeutic trial should be cut from the same anatomical location or through the most severe lesion in the joint regardless of location is problematic. The former allows for the best comparison given the pre-existing topographical differences in morphology, biochemistry and gene expression that exist in normal joints (22, 23, 27, 45-48). However, if a therapy under investigation resulted in the maximal cartilage lesion being in a different anatomical location (such as may occur if joint mechanics was affected by the treatment), then false positive or negative results may arise from comparing the same topographical region. In the latter scenario additional osteochondral slabs should therefore be cut to enable comparison of histopathology from the site of the maximal lesion but also the same topographic region of the comparison joint. The number of joint regions to be evaluated histologically may be driven by the additional analyses that are to be simultaneously undertaken, such as cartilage biomechanics, biochemistry, gene expression etc. Topographical histopathology of up to 12 regions in a single medial tibial condyle with accompanying biomechanics, biochemistry ± gene expression analysis has been undertaken.
in sheep (27) and humans (47). These painstaking and time consuming analyses may enable more subtle regional differences in OA and effects of therapy to be examined, however that are not recommended for routine studies.

In light of the above discussion, the following recommendations for sampling, fixation, sectioning, staining and scoring are suggested. After gross analysis and digital photography during which cartilage should be kept moist with saline, coronal 3-4mm thick osteochondral slabs which cover the entire width of the tibial (medial and lateral separate) and femoral condyles (medial and lateral separate) are cut with a fine-toothed band saw. These coronal slabs can be cut at various depths from the anterior joint margin to span the complete joint depending on other analyses to be undertaken as per the discussion above. Osteochondral slabs should be immediately fixed in 10% neutral buffered formalin or other cross-linking fixative for 48 hours (non-cross linking fixatives are not recommended as significant leaching of proteoglycan can occur during fixation and subsequent processing steps (49)). Following fixation, samples should be transferred to 70% ethanol for storage or further processing as prolonged formalin fixation may make subsequent immunohistological analysis more difficult and it hardens the tissues. For routine analysis we recommend decalcification and paraffin embedding although plastic embedding and non-decalcified sections have been described (31). For decalcification, 10% formic acid/5% formalin for 8 days with agitation and 4-5 changes of decalcification solution has routinely been used and this method still allows excellent immunohistology on osteochondral sections (22, 50-53). EDTA can also be used but extended decalcification times are needed due to the density of the sheep bone.

After paraffin embedding, 5-micron sections are cut and mounted on superfrost ultraplus positively charged slides, which have proven particularly useful for cartilage adhesion which can be problematic. We routinely heat slides to 85 degrees for 30 minutes and then 55 degrees overnight to maximise adhesion of sections and minimize wrinkles in the cartilage, which can be a problem with large curved osteochondral sections such as a complete tibial or femoral condyle. While this method has been used without apparently compromising immunohistological studies (51, 54), readers should ensure that this protocol works for the particular antibodies they intend to use for IHC.

Sections are de-paraffinised with xylene and graded ethanols and then stained with toluidine blue or safranin O to enable the proteoglycan content of the cartilage to be evaluated. We routinely use toluidine blue (10 minutes, 0.04% toluidine blue in 0.1M sodium acetate ph 4.0) with a fast green counter stain (3 minutes 0.1% fast green in Milli Q water). This method originally described by Getzy el al is a robust and reproducible staining procedure (55). For histological scoring at least 2 observers blinded to the treatment should score each section using the system outlined in Table 4A. In each coronoal slab multiple sections at different depths can be scored, although we have found minimal variation in histopathology scoring through the depth of a 3mm osteochondral slab in sheep following meniscectomy (unpublished observation). Various topographically defined joint locations within the one section can also be defined and scored separately and a composite score can be then generated by averaging or totalling scores from different regions in each section and in different osteochondral slabs. We routinely score inner, middle and outer 1/3rd of the tibial condyle (representing regions normally uncovered, intermittently covered or continuously covered by the meniscus) in a single coronal section from each osteochondral slab of each joint. The most severe lesion as well as the % surface area affected in each topographical region (third) is scored as described in Table 4A and 4B, respectively. It is important to avoid marginal osteophytes when scoring the outer and inner thirds of the joint.

In addition to the histomorphological changes scored in the modified Mankin system, the extent of structural damage across the entire width of the tibial or femoral condyle is scored (Table 4B). This score evaluates the percentage of the joint surface that is affected by structural damage over and above that defined by the score of 3 in Table 4A (i.e. severe surface irregularities (disruption, fissuring to < 10% depth)). We have found that both the degree of cartilage damage (modified Mankin score) and % of surface area affected, increases with time post OA induction. However, if a treatment is instituted some time after OA-induction, focal maximal or near maximal modified Mankin score may already exist and thus cannot be modified, but spread of further joint surface involvement may be noted.
Table 4A: Microscopic scoring of cartilage. Representative examples of each score are shown in figure 5 and instructions on how to score each parameter follow the table.

<table>
<thead>
<tr>
<th>A. Structure (score the worst area in field of view)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
</tr>
<tr>
<td>Slight surface irregularities (surface barely disturbed)</td>
<td>1</td>
</tr>
<tr>
<td>Moderate surface irregularities (surface roughened)</td>
<td>2</td>
</tr>
<tr>
<td>Severe surface irregularities (disruption, fissuring/fibrillation to &lt; 10% depth)</td>
<td>3</td>
</tr>
<tr>
<td>Fissures to transitional zone (1/3 depth)</td>
<td>4</td>
</tr>
<tr>
<td>Fissures to radial zone (2/3 depth)</td>
<td>5</td>
</tr>
<tr>
<td>Fissures to calcified zone (full depth)</td>
<td>6</td>
</tr>
<tr>
<td>Erosion or severe fibrillation to mid zone (1/3 depth)</td>
<td>7</td>
</tr>
<tr>
<td>Erosion or severe fibrillation to deep zone (2/3 depth)</td>
<td>8</td>
</tr>
<tr>
<td>Erosion or severe fibrillation to calcified zone (full depth)</td>
<td>9</td>
</tr>
<tr>
<td>Erosion or severe fibrillation to subchondral bone</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Chondrocyte density (“average” score for whole field of view in non-calcified cartilage)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
</tr>
<tr>
<td>Increase or slight decrease</td>
<td>1</td>
</tr>
<tr>
<td>Moderate decrease</td>
<td>2</td>
</tr>
<tr>
<td>Severe decrease</td>
<td>3</td>
</tr>
<tr>
<td>No cells</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. Cell cloning (score the whole field of view)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
</tr>
<tr>
<td>Several doublets</td>
<td>1</td>
</tr>
<tr>
<td>Many doublets</td>
<td>2</td>
</tr>
<tr>
<td>Doublets and triplets</td>
<td>3</td>
</tr>
<tr>
<td>Multiple cell nests or No cells in section</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D. Interterritorial Toluidine blue (score the worst area in field of view working from AC surface down)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
</tr>
<tr>
<td>Decreased staining to mid zone (1/3 depth)</td>
<td>1</td>
</tr>
<tr>
<td>Decreased staining to deep zone (2/3 depth)</td>
<td>2</td>
</tr>
<tr>
<td>Decreased staining to calcified zone (full depth)</td>
<td>3</td>
</tr>
<tr>
<td>No staining</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E. Tidemark/calcified cartilage/subchondral bone (score the worst area in field of view)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intact subchondral bone plate + single tidemark</td>
<td>0</td>
</tr>
<tr>
<td>Intact subchondral bone plate + duplicated tidemark</td>
<td>1</td>
</tr>
<tr>
<td>Blood vessels penetrate through subchondral bone plate to calcified cartilage</td>
<td>2</td>
</tr>
<tr>
<td>Tidemark penetrated by blood vessels</td>
<td>3</td>
</tr>
</tbody>
</table>

INSTRUCTIONS FOR SCORING DIFFERENT PARAMETERS:

**Structure** – score the worst area seen in the field of view
- surface abnormalities are considered to NOT extend below the upper 10% of the non-calcified cartilage depth. Disruption of the matrix extending below this level should be considered either as fissures or fibrillation.
- fissures may be simple or complex but it could be imagined that by “pushing” the cartilage back together the defect could be closed i.e. no cartilage missing or lost.
- “fibrillation” is included with erosion of cartilage and is defined as “appearing like the frayed end of a piece of material” where it is not possible to imagine that by pushing the cartilage from any direction could the defect be closed, and or it is not possible to readily conclude that no cartilage has been lost. Note that fibrillation confined only to the very surface 10% is included in surface abnormalities

**Chondrocyte density** – “average” score for the field of view of the non-calcified cartilage. Comparison needs to be made to a similar topographical location in a normal joint as cell density varies across a joint and with cartilage depth. Where cell clusters are present these are counted as a single “cell focus” i.e. five clones each with 10 nuclei is considered moderate to severe decrease compared with 100 single nuclei evenly distributed throughout the cartilage.

**Cell Cloning** – score the entire field of view.
- if no cells are present then the maximal score (4) is given

**Interterritorial Toluidine blue** – score the worst area seen in the field of view.
- Do not include the pericellular matrix in evaluation of toluidine blue staining.
- decreased staining does not have to be complete loss of toluidine blue but rather a diminution compared to the same region in a normal joint.
- begin at the cartilage surface and work down to find the depth to which staining is decreased. If a decrease of staining was observed only in deep tissue but not at the upper surface then a score of zero is given.

**Tide mark** – score the worst area seen in the field of view.

**Table 4B: Microscopic scoring of the extent of surface area affected by structural damage greater than score 3.**

<table>
<thead>
<tr>
<th>% surface area affected across the entire femoral or tibial condyle</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>No structural damage extending below 10% depth</td>
<td>0</td>
</tr>
<tr>
<td>&lt; 10% surface area with structural damage extending below 10% depth</td>
<td>1</td>
</tr>
<tr>
<td>10 to 25% surface area with structural damage extending below 10% depth</td>
<td>2</td>
</tr>
<tr>
<td>26 to 50% surface area with structural damage extending below 10% depth</td>
<td>3</td>
</tr>
<tr>
<td>51 to 75% surface area with structural damage extending below 10% depth</td>
<td>4</td>
</tr>
<tr>
<td>&gt; 75% surface area with structural damage extending below 10% depth</td>
<td>5</td>
</tr>
</tbody>
</table>
Figure 5: Representative toluidine blue/fast green stained sections of cartilage from sheep with meniscectomy-induced OA, to demonstrate the features associated with the pathology scores for the different parameters outlined in Table 4A. A higher magnification image is included for cell cloning to demonstrate doublets, triplets and multiple cell nests (clusters).
It is important that the synovium from the same anatomical location in the joint is evaluated as the morphology varies considerably within normal joints. We routinely sample synovium from the suprapatellar fold that normally has an adipose-areolar subsynovial tissue and thus readily enables sub-intimal fibrosis to be scored. Synovial specimens are fixed in 10% (v/v) neutral buffered formalin for 24 hours, transferred to 70% (v/v) ethanol, routinely processed through ascending grades of ethanol (70 – 100% (v/v)), cleared in chloroform, and then infiltrated and embedded in paraffin wax. Sections (4μm) are stained with haematoxylin and eosin (H&E) and scored using the system outlined in Table 5A with accompanying examples in Figure 6. This scoring system allows an “average” score of the entire section of synovium to be calculated. Additional quantitative parameters (Table 5B) can be determined using an eyepiece graticule of 1 cm². Areas counted are required to have a straight edge for 250μm, be at least 250 μm deep and not cover any normal fibrous synovial tissue. Five randomly selected areas are counted per section at each observation. These synovial scoring parameters have been used in sheep to measure changes in synovium with OA and modulation by intra-articular hyaluronan therapy (33).

### Table 5A: Microscopic grading of synovial changes

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Score</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intimal hyperplasia</td>
<td>0</td>
<td>Normal (1 cell deep only)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Mild, focal (2 to 4 cells deep, and &lt;20% area)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Mild diffuse (2 to 4 cells deep, and &gt;20% area) or Moderate focal (5 or more cells deep, and &lt;20% area)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Moderate diffuse (5 or more cells deep, and &gt;20% area)</td>
</tr>
<tr>
<td>Inflammatory cell (lymphocytic/plasmocytic) infiltration</td>
<td>0</td>
<td>Normal (occasional cell)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Mild – focal infiltration, no lymphoid aggregates</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Moderate - diffuse infiltration, no lymphoid aggregates</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Marked – discreet lymphoid aggregates</td>
</tr>
<tr>
<td>Subintimal fibrosis (loose connective tissue areas only)</td>
<td>0</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Light, focal (&lt; 20% area) collagen staining</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Heavy focal (&lt;20% area) or slight diffuse collagen staining</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Heavy diffuse collagenous staining</td>
</tr>
<tr>
<td>Vascularity</td>
<td>0</td>
<td>0 to 2 vascular elements per 100x field</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3 to 4 vascular elements per 100x field</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5 to 8 vascular elements per 100x field</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>More than 8 vascular elements per 100x field</td>
</tr>
<tr>
<td>Aggregate score (joint)</td>
<td>0 - 12</td>
<td>Sum of the scores obtained for the 4 criteria above</td>
</tr>
</tbody>
</table>

### Table 5B: Quantitative microscopic parameters of synovial pathology

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellularity</td>
<td>Number of cell nuclei present along a 250μm strip of intima to a depth of 50μm. Multiplied by 4 to give per mm.</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>Depth of subintimal fibrosis in μm</td>
</tr>
<tr>
<td>Vascularity</td>
<td>Number of blood vessels greater than 10μm in diameter in a 250 x 250μm square abutting the synovial intima</td>
</tr>
</tbody>
</table>
Figure 6: Representative H&E stained sections of synovium from sheep with meniscectomy-induced OA to demonstrate the features associated with the pathology scores for the different parameters outlined in Table 5A. The scores for intimal hyperplasia, lymphocytic/plasmacytic cellular infiltrate, sub-intimal fibrosis and vascularity are shown below each image and a magnified region showing detail of the synovial lining and sub-intima is shown in the bottom right of each section.
IId. Microscopic scoring of bone alterations

Very little has been published in sheep or goats for microscopic scoring of bone in OA models. Murphy et al used a simple 0-4 (none to severe) score for subchondral bone thickening in goats following ACLT and meniscectomy (31). In sheep, changes in subchondral bone thickness and bone mineral density following meniscectomy have been measured histomorphometrically or with an absorptiometer, respectively (26, 32, 56). A simple and validated scoring system for bone change in sheep and goat OA models requires development.

IIf. Scoring of other alterations - menisci, and ligaments

While changes in biochemistry, gene expression, cell number, hydration and collagen fibril diameter have been reported in ligaments and menisci in OA joints in sheep (28, 57-59), there have been no publications using a gross or histopathological scoring system for these tissues in sheep or goats. Scoring systems similar to those reported for other species such as the dog (see article in this journal) could be adapted for use in sheep and goats but would need to be validated in these species.

III. Reproducibility study

Expert (n=5) and non-expert (n=5) scorers were asked to score 29 histopathological images representing a broad range of cartilage pathology, including normal cartilage. Experts were orthopaedic researchers with previous experience in using similar histopathological scoring systems in evaluating an animal model of osteoarthritis although not necessarily in sheep and goats. Non-experts were researchers in other aspects of orthopaedics without previous experience in this or other histopathology scoring methods. Experts were simply sent the images and the written instructions for scoring as outlined in this paper. Non-experts received the same images and instructions but had a 10-minute instruction session on the scoring method by one of the experts (MMS). Reproducibility was calculated as standard difference, Sdiff (standard deviation of inter-observer differences), and also as the percentage of scores within ±0.5 or ±1 of the mean score for each sample image. Inter-observer agreement was calculated for each ordinal variable using Fleiss’s adaptation of Cohen’s kappa coefficient (κ) for multiple observer agreement, as determined by the MacKappa software module (60). To assess any bias between expert and non-expert scorers, a paired t-test was used to test the null hypothesis that the means derived from the two scorer cohorts did not differ at P<0.05 (Table 6).

Table 6: Standard difference, interobserver agreement (Fleiss’s kappa), and precision for each nominal variable as scored by expert and non-expert scorers (n=5 each group)

<table>
<thead>
<tr>
<th>Structure</th>
<th>Cellularity</th>
<th>Cloning</th>
<th>Staining</th>
<th>Tidemark</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sdiff</td>
<td>1.56</td>
<td>0.65</td>
<td>0.95</td>
<td>0.68</td>
<td>0.61</td>
</tr>
<tr>
<td>kappa (±95% C.I.)</td>
<td>0.44 (0.39-0.49)</td>
<td>0.43 (0.36-0.51)</td>
<td>0.37 (0.30-0.44)</td>
<td>0.52 (0.45-0.58)</td>
<td>0.41 (0.25-0.56)</td>
</tr>
<tr>
<td>% mean ±0.5</td>
<td>52</td>
<td>74</td>
<td>66</td>
<td>73</td>
<td>80</td>
</tr>
<tr>
<td>% mean ±1</td>
<td>87</td>
<td>100</td>
<td>92</td>
<td>97</td>
<td>99</td>
</tr>
<tr>
<td>Non-experts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sdiff</td>
<td>1.73</td>
<td>0.84</td>
<td>1.13</td>
<td>0.64</td>
<td>0.75</td>
</tr>
<tr>
<td>kappa (±95% C.I.)</td>
<td>0.45 (0.40-0.50)</td>
<td>0.40 (0.33-0.46)</td>
<td>0.35 (0.28-0.41)</td>
<td>0.58 (0.52-0.64)</td>
<td>0.28 (0.16-0.40)</td>
</tr>
<tr>
<td>% mean ±0.5</td>
<td>57</td>
<td>68</td>
<td>60</td>
<td>78</td>
<td>71</td>
</tr>
<tr>
<td>% mean ±1</td>
<td>71</td>
<td>94</td>
<td>90</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>Experts vs. non-experts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference in meansa</td>
<td>0.23</td>
<td>0.003</td>
<td>0.03</td>
<td>0.32</td>
<td>0.10</td>
</tr>
<tr>
<td>P value</td>
<td>0.02</td>
<td>0.57</td>
<td>0.87</td>
<td>&lt;0.0001</td>
<td>0.26</td>
</tr>
</tbody>
</table>

a = Expert mean – non-expert mean

Reproducibility and inter-observer agreement (Fleiss’s kappa) was moderate and did not vary greatly between variables or between scorer cohorts, with the exception of tidemark changes that were more reliably scored by experts (P<0.0001). However a general trend was evident for greater reproducibility and agreement in the expert group.
Comparison of the means scores of the two groups showed that non-experts significantly underestimated scores for structure and toluidine blue staining, with total scores on average nearly one score lower than expert scorers (Table 6). Overall, these preliminary results demonstrate that the proposed scoring system is reproducible, particularly amongst experienced histopathologists. Novice scorers should receive training and additional feedback after initial scoring and this would be expected to further improve reproducibility and decrease inter-observer variability.

**Discussion**

In this paper we have summarised existing literature and experience in grading models of arthritis in sheep and goats. These species offer significant advantages because of their size, enabling use of numerous outcome measures commonly utilized in man. The “gold standard” for evaluation of arthritis models in all species remains histopathological evaluation of the cartilage ± synovium and bone. The pitfalls of histological evaluation of cartilage damage given the focal nature of diseases such as osteoarthritis has been discussed in the preceding sections. Furthermore, it is unclear whether histology, particularly when limited primarily to cartilage, fully captures the pathological process of osteoarthritis and there is little or no information on the correlation of histopathological changes and clinical signs of disease, particularly joint pain. Our studies using the meniscectomy model of osteoarthritis in sheep have shown a poor relationship between gait abnormalities and histopathology or synovial fluid rheology (61). Further research should be conducted into non-destructive measurements that correlate well with histopathological changes and/or joint pain and disease progression, such as or mechanical properties of cartilage, biomarkers, and different imaging modalities (25, 47, 62-65). One of the problems with the scoring systems recommended for gross and histopathological evaluation of arthritis in sheep and goats as well as other species, is that they use ordinal as opposed to continuous variables generated from actual measurements. This means that non-parametric statistical analyses such as Kruskal-Wallis and Mann-Whitney U should be used rather than the more common parametric tests. We recommend using the Benjamini-Hochberg method for post-hoc correction to reduce false positives with multiple comparisons, which can be used with parametric or non-parametric tests (25). Despite these drawbacks, the advantage of scoring rather than measurement systems is that they can be easily and rapidly adopted by investigators with little training and minimal specialized equipment, as our preliminary validation studies demonstrate. It is hoped that by adopting more standardised quantitative outcome measures, better comparison between different studies and arthritis models will be possible. The suggested scoring systems can be modified in the future as our knowledge of disease pathophysiology advances.

**Acknowledgments:**

“Non-expert” scorers – Dr. James Melrose, Dr. Richard Appleyard, Dr. Ashutosh Barai, Dr. Kenshi Kikukawa, Mr. Daniel Burkhardt. Thanks to Ms Susan Smith for preparing all histopathology presented, and Mr Daniel Burkhardt for preparation of gross pathology images.

**Conflict of interest statement:**
None of the authors have any conflict of interest related to this work.

**Disclosures:**

Drs Little, Smith, Cake and Read have received research grant funding from various pharmaceutical companies where the meniscectomy-induced OA model in sheep was used. These studies were conducted under contractual agreement with the companies and the respective Universities or Area Health Services employing the researchers. Drs Murphy and Barry conducted research using surgically-induced OA in goats at Osiris Therapeutics, and are shareholders in Osiris Therapeutics Inc., Baltimore MD, USA.
References

45. Little C, Ghosh P. Variation in proteoglycan metabolism by articular chondrocytes in different joint regions is determined by post-natal mechanical loading. Osteoarthritis Cartilage 1997;5:49-62.


