Comparison of Cartilage Histopathology Assessment Systems on Human Knee Joints at All Stages of Osteoarthritis Development

Pauli, C; Whiteside, C R; Las Heras, F; Nesic, D; Koziol, J; Grogan, S P; Matyas, J; Pritzker, K P H; D’Lima, D D; Lotz, M K

Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA, USA, Shiley Center for Orthopaedic and Education at Scripps Clinic, La Jolla, CA, USA, Arthroteq Preclinical Services, Guelph, Canada, Department of Pathology, University of Chile, Clinical Hospital, Santiago, Chile, Osteoarticular Research Group, Department of Clinical Research, University of Bern, Bern, Switzerland, McCaig Institute for Bone and Joint Health, Department of Comparative Biology & Experimental Medicine, Faculty of Veterinary Medicine, University of Calgary, Canada, Pathology and Laboratory Medicine, Mount Sinai Hospital, University of Toronto, Canada

e-mail: miotz@scripps.edu

Introduction
The Osteoarthritis Cartilage Histopathology Assessment System (OARSI system) was developed to address perceived shortfalls of the Histologic/ histochemical Grading System (MANKIN system). However, its validity in comparison to the MANKIN system should be further established. In this study, we compared the two scoring systems using human articular cartilage from a large number of donors across the adult age spectrum representing all levels of cartilage degradation. We aimed for a more detailed data analysis to identify strengths and limitations of both systems.

Methods
Human knees (n=125 from 65 donors; age range 23-92) were obtained from tissue banks. Osteochondral slabs representing the entire central regions (sagittal) of both femoral condyles, both tibial plateaus (coronal), and the patella (transversal) were resected, fixed and decalcified. Each slab was cut into tissue blocks at defined locations and processed for Safranin O – Fast Green staining (Fig. 1). From a total collection of approximately 1600 slides, a set of 300 slides was selected that represented normal, aged, and OA tissue at all grades and subgrades. These slides were scanned and electronic images were scored online by five observers familiar with the grading of articular cartilage, but with different levels of experience using the MANKIN and the OARSI system. Statistical analysis was performed for inter- and intra-observer variability, reproducibility and reliability.

Results
The inter-observer variability among five observers for the MANKIN system showed a good intra-class coefficient (ICC >0.81). Among the different parameters, all readers achieved the best agreement for the scoring of changes in the cartilage surface representing lesion depth. For the OARSI system, the same observers showed a good but slightly lower agreement (ICC >0.78). Three of the five observers scored the 300 slides twice, at least three weeks apart, with both grading systems showing a very good agreement (ICC >0.94). Both systems showed a high reproducibility among four of the five readers. There was a good positive correlation between both scoring systems as indicated by the Spearman’s rho value. A summary plot of average MANKIN system versus average OARSI system showed a strong monotonic, roughly linear, relationship between the two systems in our study (Fig. 2).

Fig. 1. Harvesting protocol: (A) For the femoral condyles medial and lateral sagittal osteochondral slabs were harvested as indicated by the red lines. (B) One coronal central slab of each tibia plateau was harvested as indicated. (C) From each patella a central transverse slab was harvested.

Fig. 3. Summary plot of average MANKIN versus average OARSI system: Summary plot comprising 95% confidence ellipse, and marginal histograms of the frequency distribution. The graph shows a strong monotonic relationship between the two scoring systems, reflective of the Spearman rho value of 0.96. The relationship is roughly, but not perfectly, linear: Given a particular MANKIN [or OARSI] score, there is a fair amount of scatter around the corresponding OARSI [or MANKIN] scores.

Conclusion
Both scoring systems were reliable but appeared too complex and time consuming if used to assess lesion severity and thus not very useful to determine a global grade of cartilage damage within a knee. Surface changes representing lesion depth is the parameter where all of the readers showed an excellent agreement in the MANKIN system. Other parameters such as cellularity, Safranin O stain intensity and tidemark had greater inter-reader disagreement. For the purpose of rapidly assessing severity of cartilage degradation, we propose to develop a more simplified system for lesion volume as measured by lesion depth and width. Furthermore, we believe that a standard library of illustrations of grading systems would greatly assist observers with consistency and to find a better consensus. The development of online training programs for less experienced observer based on a library of images would help standardize agreement on vocabulary and disease severity.

Significance
Various cartilage grading systems are being used for the analysis of specimens from humans and experimental animal models. This diversity in methods is a major limitation in comparing results from different studies. Based on the present study, the two major systems in use appear too complex for the main purpose of assessing OA severity. The proposed simplified system would generate more consistent and reproducible results.

Acknowledgments
National Institutes of Health (P01 AG007996) and by Donald and Darlene Shiley.

References