

# MRI SCORING OF OSTEOARTHRITIS OF THE ANKLE

Dr Sharief Aboelmagd

A thesis submitted for the degree of Doctor of Medicine

University of East Anglia

Norwich Medical School

December 2016

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

## Abstract

The aim of this project was to develop a semi-quantitative MRI based scoring system for osteoarthritis of the ankle, and determine its inter and intra-observer reliability.

A systematic search was performed to identify current MRI scoring systems and review the methods and included features of these systems. Based on this, a Delphi survey of a group of experts was undertaken to determine the features to be taken forward for reliability testing.

A retrospective sample of 50 patients who had undergone MRI and plain radiographs were included. Anatomical division of the ankle was based on existing published systems. MR examinations were graded using the proposed system by two consultant radiologists and two radiology trainees. Radiographs were graded using a published Kellgren-Lawrence ankle score. Inter and intra-observer reliability were examined using a weighted kappa statistic ( $k_w$ ), and association between MRI and radiographic severity using Spearman's Rho.

Inter-reader reliability was "almost perfect" for all features for total joint scores ( $k_w$  0.88–0.97) except osteophyte scoring for the trainees that was "substantial" (kappa 0.64). Zonal based assessment of features demonstrated "substantial" agreement between the trainees ( $k_w$  0.63–0.75), and "substantial" or "almost perfect" agreement for the consultants ( $k_w$  0.73–0.92). Inter-reader reliability was "almost perfect" for surface extent of cartilage loss across all readers. Intra-reader reproducibility was "substantial" or "almost perfect" for total joint scores and "moderate" to "almost perfect" for the zonal approach. There was strong positive correlation between all features and radiographic severity (rho 0.75–0.85) except cysts which demonstrated "weak" correlation (rho 0.35).

This new grading system demonstrates "substantial" to "almost perfect" inter and intra-observer reproducibility and may be of use in longitudinal studies. Further research and development will include assessment of validity and sensitivity to change.

# List of Contents

Abstract .....	2
List of abbreviations .....	8
1. Introduction .....	12
2. Background .....	14
Anatomy the ankle .....	14
Kinematics .....	20
Osteoarthritis .....	22
Ankle osteoarthritis .....	30
Management.....	31
Imaging in Osteoarthritis.....	34
Conventional Radiography.....	34
Ultrasound .....	35
Computed Tomography .....	36
Radionuclide Scintigraphy.....	37
MRI .....	38
3. Grading systems .....	43
Radiographic grading.....	43
Radiographic grading of OA of the ankle .....	44
Existing MRI systems .....	47
MRI scoring systems for osteoarthritis: ankle joint .....	47
MRI scoring systems for osteoarthritis: all joints .....	48
WORMS .....	50
KOSS.....	52
BLOKS .....	53
MOAKS.....	56
Park et al.....	57
HOAMS .....	58
SHOMRI .....	60

4. Delphi survey.....	63
Background and methods .....	63
Results .....	65
Tibiotalar joint.....	65
Subtalar Joint .....	68
Discussion .....	70
5. Reliability Study.....	72
Materials and Methods .....	72
Study Design.....	72
Ethics .....	72
Patient selection.....	72
MR Imaging ankle protocol .....	73
Sample Size .....	74
Method .....	74
Justification .....	74
Patient demographics .....	77
Readers.....	77
Radiographic Scoring .....	77
MRI Scoring .....	78
Zones .....	78
Method .....	78
Justification .....	80
Osteophytes .....	82
Method .....	82
Justification .....	82
Bone marrow lesions, Bone marrow oedema and subarticular cysts .....	82
Method .....	82
Justification .....	83
Cartilage integrity and osteochondral defects.....	83
Method .....	83
Justification .....	83

Bone attrition .....	86
Statistical Analysis .....	87
Methods.....	87
Statistical Methods.....	87
Justification .....	87
The kappa statistic and observer agreement .....	87
Interpretation of kappa.....	88
Weighted kappa .....	89
Problems with kappa .....	91
Correlation .....	91
Results .....	94
Radiographic Scoring.....	94
Modified Kellgren-Lawrence score .....	94
Osteophytes.....	95
Joint space narrowing .....	96
Total Kraus score and modified Kellgren-Lawrence Correlation .....	97
MRI Scoring .....	99
Osteophytes.....	99
Bone marrow signal abnormality.....	100
Cartilage .....	102
Distribution of disease .....	109
Correlation .....	110
Discussion .....	118
6. Conclusion .....	122
Appendices.....	125
References.....	146

## List of Tables

Table 1. Kellgren Lawrence grading scale .....	43
Table 2. Components of the Park scoring system.....	57
Table 3. Delphi survey results for the Tibiotalar joint .....	65
Table 4. Delphi survey results for the Subtalar joint .....	68
Table 5. Patient demographics .....	77
Table 6. Cartilage grading system used in reliability study.....	83
Table 7. Modified Noyes grading system.....	84
Table 8. Modified Outerbridge grading system .....	85
Table 9. Suggested interpretations of kappa .....	88
Table 10. Suggested interpretation of kappa as outlined by McHugh .....	89
Table 11. Modified Kellgren-Lawrence Score .....	94
Table 12. Radiographic Osteophyte scoring .....	95
Table 13. Joint space narrowing .....	96
Table 14. Correlation for Kraus total score and Kellgren-Lawrence score .....	97
Table 15. Osteophytes inter-rater results.....	99
Table 16. Osteophytes intra-rater results.....	99
Table 17. Bone marrow lesion zonal and total joint inter-rater results .....	100
Table 18. Bone marrow lesion zonal and total joint intra-rater results .....	101
Table 19. Cartilage inter-rater zonal and total joint results .....	102
Table 20. Cartilage intra-rater zonal and total joints results .....	103
Table 21. Talar Dome Inter-rater results .....	103
Table 22. Zones 10–16 Inter-rater results.....	104
Table 23. Cartilage grading versions .....	106
Table 24. Inter-rater results for different versions of cartilage scoring .....	106
Table 25. Inter-rater results for cartilage surface extent grade .....	108
Table 26. Intra-rater results for cartilage surface extent grade .....	108
Table 27. Distribution of MRI disease severity .....	109
Table 28. Bone marrow lesion and plain radiograph correlation .....	110
Table 29. Cartilage scores and plain radiograph correlation .....	110

## List of Figures

Figure 1. Medial Collateral Ligament complex .....	17
Figure 2. Morphology of Cartilage. ....	26
Figure 3. Severe radiographic OA.....	34
Figure 4. MRI Bone marrow oedema. ....	39
Figure 5. MRI subchondral cyst.....	40
Figure 6. MRI normal cartilage.....	42
Figure 7. MRI full thickness cartilage loss. ....	42
Figure 8. Search for Ankle OA MRI grading systems.....	47
Figure 9. Search for knee and hip OA MRI grading systems .....	49
Figure 10. Multiplot Delphi results for the Tibiotalar joint.....	66
Figure 11. Multiplot Delphi results for Subtalar joint .....	69
Figure 12. Zones 1–9. ....	79
Figure 13. Zones 10–16 .....	79
Figure 14. Consensus Kellgren-Lawrence vs total Kraus score scatter plot.....	98
Figure 15. Bone marrow lesion score versus Kellgren-Lawrence grade.....	112
Figure 16. Bone marrow oedema score versus Kellgren-Lawrence grade .....	113
Figure 17. Cyst score versus Kellgren-Lawrence grade.....	114
Figure 18. Total joint cartilage score versus Kellgren-Lawrence grade .....	115
Figure 19. Cartilage Score 1 versus Kellgren-Lawrence grade.....	116
Figure 20. Cartilage Score 2 versus Kellgren-Lawrence grade.....	117

## List of abbreviations

AIDA	Ankle images digital analysis
AITFL	Anterior inferior tibiofibular ligament
ATFL	Anterior Talofibular ligament
BLOKS	Boston Leeds Osteoarthritis Knee Score
BML	Bone Marrow Lesion
BMO	Bone Marrow Oedema
CFL	Calcaneofibular ligament
CT	Computed Tomography
ECM	Extracellular matrix
HOAMS	Hip Osteoarthritis MRI Scoring System
ICC	Intraclass Correlation Coefficient
ICRS	International Cartilage Repair Society
IGF	Insulin like Growth Factor
IL-1	Interleukin-1
ITFJ	Inferior tibiofibular ligament
KL	Kellgren-Lawrence
KOSS	Knee Osteoarthritis Scoring System
MCL	Medical Collateral Ligament
MOAKS	MRI Osteoarthritis Knee Score
MRI	Magnetic Resonance Imaging
NNUH	Norfolk and Norwich University Hospital
OA	Osteoarthritis
OARSI	Osteoarthritis Research Society International
PACS	Picture archiving and communication system
PD	Proton Density
PET	Positron Emission Tomography
PITFL	Posterior inferior tibiofibular ligament
SHOMRI	Scoring Hip Osteoarthritis with MRI
STIR	Short Tau Inversion Recovery



T1W	T1 weighted
T2W	T2 weighted
TIMP	Tissue inhibitor of metalloproteinase
TNF- $\alpha$	Tumour Necrosis Factor alpha
US	Ultrasound
WORMS	Whole Organ Magnetic Resonance Imaging Score

## Acknowledgements

I would like to thank my supervisor Andoni Toms for his help and support, not only throughout this project but also in my career so far. He has given his time, expertise and encouragement and I am eternally grateful. Without his guidance, I would not have reached this point.

I owe thanks to my colleagues in the radiology department at the Norfolk and Norwich University Hospital for their unfailing support throughout this period of research and write up.

To my son Samir, thank you for always smiling, no matter how many bedtimes I missed. Finally, I would like to thank my wife Billie, for her patience, understanding and continuous encouragement. Thank you for putting up with me. None of this would have been possible without you.

## Statement of contribution

The idea for this project followed conversations between Professor Toms and myself regarding existing MRI scoring systems.

Professor Toms is a consultant radiologist based at the Norfolk and Norwich University Hospital and the University of East Anglia. Professor Toms acted as an expert reader for MRI scoring as well as being the primary supervisor for this MD degree.

Dr Samantha Low is a radiology trainee based at the Norfolk and Norwich University Hospital. Dr Low consensus read radiographs for inclusion in the study as well as reading radiographs and MRI examination for reliability testing.

Dr John Cahir is a consultant musculoskeletal radiologist based at the Norfolk and Norwich University Hospital. Dr Cahir acted as an expert reader for MRI scoring.

Mr David Loveday is a consultant foot and ankle surgeon based at the Norfolk and Norwich university Hospital who took part in the Delphi survey and advised on some component of the project.

Mr George Smith is a consultant foot and ankle surgeon based at the Norfolk and Norwich university Hospital who took part in the Delphi Survey.

Dr James Teh is a consultant musculoskeletal radiologist based at the Nuffield orthopaedic centre in Oxford who took part in the Delphi survey.

I read radiographs for consensus reading and reliability testing as well as being a novice reader for MRI reliability testing. All work conducted in thesis is my own including the statistical analyses unless otherwise stated.

## 1. Introduction

An understanding of the severity of a disease is essential in modern medicine. It helps clinicians and patients decide on the most appropriate treatments as well as giving an insight into the prognosis and natural progression of the disease based on available evidence. For some diseases, the change in one point on a disease severity scale can mean the difference between radical and potentially curative treatment and palliation. The need for reliable and reproducible grading systems is vital in order to provide the best possible care for patients, as well as attempting to help both clinicians and their patients understand the potential for cure or natural disease progression.

Plain radiographs have traditionally been used to grade the severity of osteoarthritis. They are able to depict bone changes and use joint space narrowing as a surrogate for cartilage loss. However, current concepts in osteoarthritis treat the whole joint as an organ with multiple components that can be implicated in osteoarthritis, many of which cannot be appreciated on plain radiographs. MRI has become increasingly used to evaluate osteoarthritis, especially in the research setting, due to its ability to demonstrate many of the features implicated in osteoarthritis. Large-scale epidemiological studies in osteoarthritis use MRI data to identify factors associated with disease progression and association with clinical features. In order to do this they need reliable and reproducible MRI scoring systems. Most of the research in this area has focused on the knee joint with many systems described, some of which have undergone various iterations and unpublished modifications. More recently two similar scoring systems for the hip have been described.

In the published literature to date there is no semi-quantitative scoring system for evaluating osteoarthritis of the ankle. This may in part be due to the relatively limited therapeutic options for ankle osteoarthritis when compared

to the knee and hip joints. With a resurgence in ankle arthroplasty using the current third generation prostheses an MRI tool that can evaluate ankle osteoarthritis may be useful. The aim of this project is to determine the features that should be included in such a system and test the reliability and reproducibility of these features.

The background of this thesis addresses the relevant anatomy and kinematics of the ankle joint. Current concepts in the pathogenesis of ankle osteoarthritis and management are described. Existing radiographic grading systems are then reviewed, including those specific to the ankle joint. The results of a systematic search of existing MRI scoring systems is presented followed by a narrative review of these systems. The features to be included in an MRI ankle scoring system were determined by a Delphi survey of experts who are currently involved in managing patients with ankle osteoarthritis. These features were taken forward for reliability testing on a sample of patient with varying severities of ankle osteoarthritis. Inter-rater agreement was assessed between two radiology trainees and two radiology consultants. Intra-rater reliability was also examined.

The culmination of this project is the first described MRI scoring system for ankle osteoarthritis: the Norwich Osteoarthritis of the Ankle MRI Score.

## 2. Background

This chapter reviews the anatomy and kinematics of the ankle joint in order to identify structures that may be important in the evolution or distribution of osteoarthritis (OA). Factors implicated in the pathogenesis of OA are explored followed by an overview of the current management strategies specific to ankle OA. Different imaging techniques that can be employed in the evaluation of OA will be described.

### ANATOMY THE ANKLE

The ankle joint complex is made up of two joints; the subtalar joint and the true ankle joint, or tibiotalar joint. The tibiotalar joint, also sometimes referred to as the talocrural joint, is a synovial joint consisting of a fork shaped dome formed by the distal tibia and fibula, and the talus. The ankle mortise is the concave surface formed by the tibia and fibula. The subtalar joint can refer to one or two articulations. The anatomic subtalar joint is the talocalcaneal joint or posterior subtalar joint. The talocalcaneonavicular joint is the anterior subtalar joint, sometimes termed the clinical subtalar joint (1).

#### Tibiotalar Joint

The distal tibia has two cartilage covered articular surfaces. The dorsal surface of the talus is convex in the anterior-posterior dimension while the tibia is concave (1). The talus is slightly concave from medial to lateral while the tibia is mildly convex. The fibrous joint capsule attaches to the inferior and anterior ridge of the distal tibia and the malleoli.

The lateral surface of the medial malleolus articulates with the medial articular facet of the talus. Unlike the rigid medial ankle mortise, the lateral mortise is more flexible (1). The lateral talus and medial facet of the lateral malleolus form an articulation. The lateral surface of the distal tibia and distal fibula are attached by the syndesmotic ligament complex, interosseous membrane and interosseous ligament.

### Lateral Ligament Complex

The lateral collateral ligament complex consists of the anterior and posterior talofibular ligaments and the calcaneofibular ligament. Injuries to the lateral ligaments are one of the most common contributing factors in the development of ankle OA and correlate with varus malalignment (2,3). The anterior talofibular ligament (ATFL) runs anteromedially to its insertion on the talar body, anterior to the joint surface for the lateral malleolus, at a 45° angle to the frontal plane (4). It plays a role in limiting anterior translation of the talus and plantar flexion of the ankle as well as external tibial rotation (4). The ligament is closely related to the joint capsule of the ankle and although there are numerous descriptions it is most commonly comprised of a double-band morphology (5). Arterial branches from the peroneal artery and anastomoses with the lateral malleolar artery separate the bands (5). The cranial band is the thickest and strongest (1). In the neutral position the ligament runs almost horizontally to the ankle but moves upwards in dorsiflexion and downwards in plantar flexion. In the plantar flexed position, the superior band of the ligament becomes taut whilst the inferior band becomes taut in dorsiflexion. The ligament is most susceptible to injury when the ankle is flexed, especially if the foot is inverted (5).

The calcaneofibular ligament (CFL) originates from the anterior portion of the lateral malleolus, just below the lower band of the ATFL and fibres connecting

these ligaments can be observed in dissection. When the ankle is in the neutral position the ligament passes obliquely inferiorly and posteriorly to attach to the posterior region of the lateral calcaneal surface, being the only lateral ligament to bridge both the talocrural and subtalar joint (5). Superficially, the ligament is crossed by the peroneal tendons. It remains under tension throughout the arc of movement with tension highest during dorsiflexion where inversion resistance is most effective (4). The ligament is relaxed in the valgus position, and taut in the varus position, explaining the potential for injury without flexion-extension of the ankle. CFL and ATFL run at an angle of  $105^{\circ}$  to each other allowing the two ligaments to act synergistically (4).

The posterior talofibular ligament is a broad (PTFL), flat, triangular ligament that originates from the malleolar fossa on the medial aspect of the lateral malleolus, passing almost horizontally to insert into the posterolateral talus (5). The ligament is taut during dorsiflexion and relaxed in the neutral and plantar flexed positions. It is a multifasciculated ligament with fibres inserting into the posterior surface of the talus and the lateral talar process and, if present, os trigonum (5). The lateral ligament complex tends to fail in a predictable sequence with inversion injury; ATFL, CFL and PTFL (4). If ATFL ruptures, there is an increase in internal hind-foot rotation predisposing to further ligamentous injury (4).

### Medial Ligament Complex

The stabilisers of the ankle medially are the deltoid ligament, also known as the medial collateral ligament (MCL), and the medial malleolus. The MCL is the strongest stabilising ligament (4). Insufficiency of the MCL may lead to instability and resultant OA (6). Sectioning of the MCL has been shown to decrease bony contact area by 15–20% leading to an increase in force per unit area (3). Six different bands have been described for the MCL: four superficial bands (tibiospring ligament, superficial posterior tibiotalar, tibiocalcaneal and



tibionavicular ligament) where only the tibiospring and tibionavicular ligament are constant, and two deep bands (deep anterior and posterior tibiotalar ligaments) of which only the deep posterior tibiotalar ligament is constant (7). Descriptions of the anatomy of the MCL vary but it is generally accepted that it is composed of two layers; the deep and superficial components (4–7). The deep component of the MCL is a broad fan shaped ligament composed of anterior and posterior tibiotalar fibres that originate from the posterior border of the anterior colliculus, intercollicular groove and posterior colliculus and pass transversely to insert into the non-articular surface of the medial talus (1,4). It is a synovial lined intra-articular ligament (1).

The superficial portion is a broad, flat, triangular fibrous component that courses from the medial malleolus to the navicular, calcaneus and spring ligament (6). The superficial portion acts mainly to prevent hindfoot inversion (4).

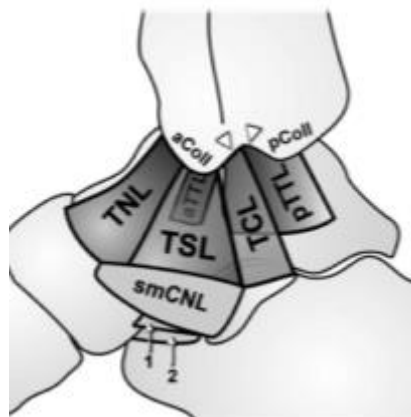


Figure 1. Medial Collateral Ligament complex; deep and superficial layers (6). TNL tibionavicular ligament, TSL tibiospring ligament, TCL tibiocalcaneal, aTTL Anterior Tibiotalar ligament, pTTL Posterior Tibiotalar ligament. Superomedial calcaneonavicular ligament (smCNL), medioplantar oblique calcaneonavicular ligament (1), and inferoplantar longitudinal calcaneonavicular ligament (2). Intercollicular groove (arrowheads), aColl anterior colliculus, pColl posterior colliculus. (Permission from publisher and author to use this figure. Appendix 1).

### Inferior Tibiofibular Joint

The inferior tibiofibular joint (ITFJ) is a syndesmosis with three main supporting ligaments: the anterior inferior tibiofibular ligament (AITFL), posterior inferior tibiofibular ligament (PITFL), and interosseous ligament. The inferior segment of the interosseous membrane also aids in stabilising the tibiofibular syndesmosis. The syndesmotic ligament complex ensures stability between the distal fibula and tibia, resisting rotational, axial and translational forces (5). Inferior to the insertion site of the ligament the remaining anterior surface represents the tibiofibular synovial recess of the ankle, and at the posterior surface there is the fatty synovial fringe (1,5). This structure has been implicated as a cause of chronic pain following ankle sprain in anterolateral impingement syndrome (5).

The AITFL is a flat, multifasciculated, fibrous band that takes its origin from the anterior tubercle of the tibia with fibres that run distally and laterally to insert on the anterior margin of the lateral malleolus. At the fibular insertion the caudal fibres interdigitate with ATFL (1,4,5).

The PITFL is composed of a deep and superficial component. The superficial portion is a broad fan-shaped ligament which originates from the posterior edge of the lateral malleolus and runs medially and proximally to insert on the posterior tibial tubercle. The deep component runs from the posterior margin of the tibia to the posteromedial aspect of the distal fibula. The deep fibres are twisted around each other forming a strong thick fibrous band (1).

The interosseous tibiofibular ligament is a thickening of the interosseous membrane and is more flexible than the other tibiofibular ligaments allowing subtle diastasis of the tibia and fibula during dorsiflexion (4). This functions as a buffer neutralising forces during, for instance, the heel strike in walking (8). It is a short, dense band that extends from the lateral surface of the tibia to the medial surface of the fibula (5).

### Subtalar Joint

The subtalar joint allows supination and pronation and is made up of two independent synovial lined articulations: the anterior and posterior subtalar joints or, talocalcaneonavicular and talocalcaneal joints (1,4). The anterior subtalar joint is formed by the convex talar head and the concave proximal body of the navicular bone. Anterolaterally, the plantar surface of the head of the talus articulates with the superior surface of the calcaneus and posteriorly with the sustentaculum tali of the calcaneus. The calcaneonavicular ligament (part of the spring ligament) and the deltoid ligament support the medial and plantar articulations. The spring ligament is a support structure of the longitudinal arch of the foot that is made up of three bands (superomedial, medioplantar oblique and inferoplantar) and supports the head of talus at the anterior and middle calcaneal facets (1). The posterior subtalar joint is formed where the posterior calcaneal facet of the talus articulates with the posterior talar facet of the calcaneus. The anterior and posterior subtalar joints are separated by the sinus tarsi and have separate joint capsules although share a similar axis of rotation.

The sinus tarsi is a bony canal with a medial apex and lateral outlet. The anterior and middle talar facets are separated from the posterior talar facet by the sinus tarsi (1). It contains fat and branches from the posterior tibial artery and nerve, peroneal artery and nerve and supporting ligaments.

The three main supporting ligaments of the subtalar joint are the cervical ligament, inferior extensor retinaculum and the interosseous talocalcaneal ligament (1,4). The cervical ligament runs from the cervical tubercle of the calcaneus, forwards and medially to the talar neck. The interosseous ligaments lie posterior to the cervical ligaments and run upwards and medially. It is a broad, flat, oblique ligament originating from the calcaneus just anterior to the posterior joint capsule and inserts into the talar neck. It is thought that the

interosseous talocalcaneal ligament, together with the cervical ligament, may play an important role in subtalar instability (9).

### KINEMATICS

Normal kinematics minimises stress on the bones, joints and soft tissues across the joint (10). Gait describes the set of motions in walking and running between the heel strike of one step and the heel strike of the same foot on a subsequent step (11). The cycle is divided into a stance phase, where the foot is on the ground, and a swing phase, where the foot is off the ground (11). The stance phase is further divided into periods of double and single limb support. The stance phase can be further divided into three intervals (10,11). The first interval is from the initial heel strike to the foot lying flat on the floor, the second interval occurs as the body passes over the foot, and the third interval extends from ankle joint flexion, as the heel rises from the floor, to toe lift-off. During the walking cycle one foot is always in contact with the floor. As gait speed increases, there is incorporation of a float phase during which both feet are off the ground. During running there is no period of double limb support as in walking but instead this float phase (11). As gait speed increases, the stance phase decreases as a percentage of the gait cycle and time.

The movements that can occur at the ankle joint are dorsiflexion and plantarflexion, as well as internal and external rotation when flexed.

Dorsiflexion is best assessed when the ankle is in the neutral position. The ankle allows about 20° of dorsiflexion and 50° of plantarflexion about the axis of the ankle that runs between the tips of the malleoli (10,11). The trochlear surface of the talus rotates around this axis, as would a section of a cone whose apex is based medially (11).

Two series of dorsiflexion and plantarflexion occur in the walking cycle. At the heel strike the ankle is dorsiflexed after which point it rapidly plantarflexes resulting in the foot being placed flat on the floor. Following this there is progressive dorsiflexion. Plantarflexion begins as the heel rises and continues until toe-off beyond which dorsiflexion again begins during the swing phase. Forces applied across the ankle vary greatly depending on activity with the force transferred through the joint in walking being 1.2 times body weight compared to 2 to 2.5 times body-weight in running and higher for extreme push off activities such as jumping (9–11). The nature of these forces also vary from vertical force, torque forces and side-to-side shear forces (11).

The movements that occur at the subtalar joint are that of inversion and eversion. Inversion is the action of turning the heel inwards while eversion is the action of turning the heel outwards. The range of motion includes inversion of about 30 ° and eversion of about 10 ° although there is variation between individuals (10). In normal gait, eversion occurs at initial ground contact after which progressive inversion occurs until toe-off. The amount of eversion varies depending on whether the patient has a normal foot or a flat foot (greater degree of eversion). Normal function of the subtalar joint requires normal function of the talonavicular and calcaneocuboid joints and if normal motion is not permitted at these joints the motion at the subtalar joint is restricted (10). Similarly, if subtalar joint movement is impeded then abnormal stress is placed across the ankle joint which if longstanding may lead to osteoarthritis (10).

### OSTEOARTHRITIS

Osteoarthritis is the clinical and pathological outcome of a range of disorders that result in structural and functional failure of synovial joints (12). It is the most prevalent form of arthritis and is an increasing cause of social and economic burden to the ageing society (13). It is a multifactorial process in which mechanical factors have a central role and is characterised by change in structure and function of the whole joint.

The Osteoarthritis Research Society International (OARSI)<sup>1</sup> provide the following definition of osteoarthritis.

*Osteoarthritis is a disorder involving movable joints characterized by cell stress and extracellular matrix degradation initiated by micro- and macro-injury that activates maladaptive repair responses including pro-inflammatory pathways of innate immunity. The disease manifests first as a molecular derangement (abnormal joint tissue metabolism) followed by anatomic, and/or physiologic derangements (characterised by cartilage degradation, bone remodelling, osteophyte formation, joint inflammation and loss of normal joint function), that can culminate in illness. (14)*

Although traditionally considered a disease of the articular cartilage, current concepts accept that the disease process affects not only the articular cartilage but the entire joint organ, including ligaments, capsule, subchondral bone, synovial membrane and periarticular muscles (15,16). The synovial joint is

---

<sup>1</sup> OARSI is a non-profit scientific organisation whose mission is to promote and advance research for the prevention and treatment of osteoarthritis. [www.oarsi.org](http://www.oarsi.org)

considered an organ and OA represents failure of that organ. OA can be initiated by multiple factors including developmental, genetic, traumatic and metabolic. Both mechanical and biological events affect normal degradation and synthesis of articular cartilage. The disease process can be initiated by abnormalities arising from any of its constituent tissues with no common physiological pathway but a common end-stage (16,17).

OA is commonly misnamed degenerative joint disease, as the cells of cartilage and bone are normal and if the inciting mechanism is reduced it is possible for the damaged tissue to be restored to normal (16,17). At the cellular level, OA occurs when the equilibrium between the repair and breakdown of synovial joint tissues is mismatched (17).

### Cartilage

Unlike most tissues, articular hyaline cartilage does not have blood vessels, lymphatics or nerves. It is composed of chondrocytes embedded within an extracellular matrix (ECM). The ECM is primarily composed of water, collagen and proteoglycans along with other proteins and glycoproteins in lesser quantities. These components help retain water within the ECM which is important to maintain its hydrostatic properties (18).

Articular cartilage is composed of various zones, each with distinct features. The superficial zone is thin making up 10–20% of articular cartilage thickness and acts to protect the deeper zones from shear stresses. The collagen fibres of this zone are primarily type II and IX and are packed tightly parallel to the articular surface (18). The superficial zone is in contact with synovial fluid and is responsible for much of the tensile properties of cartilage in resisting sheer and compressive forces. The integrity of this layer is vital in protecting the deeper layers. The transitional zone (middle zone) provides a bridge between the superficial and deep zones and lies immediately below the superficial zone.

Collagen fibres in this layer are less organised but typically oblique in orientation to the surface. The deep basal layer contains chondrocytes which are arranged vertically, perpendicular to the surface. It is responsible for the greatest resistance to compressive forces. The deep zone contains the highest proteoglycan content and the lowest water concentration. The tidemark distinguishes the deep zone from the calcified zone that is a remnant of the cartilage analogue, which participated in endochondral ossification during growth in childhood (Figure 2).

Chondrocytes regulate the balance between matrix synthesis and breakdown. This process is influenced by multiple factors including the matrix composition, injury, ageing, loading and local growth factors (18). Type II collagen is the major collagen of articular cartilage representing 90–95% of total collagen. Types IX and XI are the most abundant minor types.

In normal physiological conditions, collagen metabolism is slow. In disease states however, turnover can increase markedly, exceeding the capacity of chondrocytes to produce a well organised matrix. The disorganised matrix may undergo more rapid mechanical failure leading to degeneration and ultimately arthritis.

Compared with normal cartilage, cartilage from osteoarthritic joints deforms more readily in response to the same load, and more fluid is lost during loading (19). The alteration in cartilage biomechanics may change perception of normal mechanical load into pathologically larger loads of greater duration which are damaging to cartilage. A sufficient degree of loading appears to be imperative for cartilage health and changes in articular cartilage similar to those found in OA have been demonstrated experimentally in disuse atrophy (20). These changes include reduced proteoglycan content and synthesis, increased metalloproteinase levels with decreased levels of tissue inhibitor of metalloproteinase (TIMP) (21).



In a normal joint, chondrocytes are subject to static and dynamic compressive forces and shear stresses. This is associated with changes in the expression of genes for collagen, matrix metalloproteinases, aggrecan, cytokines and growth factors that are transduced into metabolic responses (22). In vitro, injurious compression of cartilage has been demonstrated to result in increase in expression of the genes for stromelysin, aggrecanase and TIMP in magnitudes that may lead to breakdown of the articular cartilage matrix (23). Breakdown of cartilage in OA is mediated by matrix metalloproteinases whose production is stimulated by interleukin-1 and tumor necrosis factor. These are produced by chondrocytes in response to mechanical loads. It is considered that this process may be driven by abnormal mechanical stresses on the joints. Changes seen in vitro are not identical to what is seen in an osteoarthritic joint in vivo. Before there is obvious damage to articular cartilage, excessive loading of the joint leads to fracture of the bone (16). Pharmacological inhibition of metalloproteinases has not been wholly successful in halting the progression of OA, and although influencing mechanical integrity of the articular cartilage, mechanical factors would also seem to play a major role.

Much of the loss of articular cartilage in OA appears to be attributable to breaking off of enzymatically weakened segments of the joint surface. Fibrillation and vertical splitting of the cartilage can persist without fragmentation of the joint surface, although the thinned cartilage is also subject to deep horizontal splits (16). When these join the vertical splits, breaking off of cartilage shards occurs which may be detected in synovial fluid before they incorporate into the synovial membrane where they incite an inflammatory response. In addition, reactivation of the secondary centre of ossification results in endochondral ossification which advances the tidemark, gradually thinning the cartilage from below (16,17). The majority of articular cartilage loss in OA seems to be the result of these two processes (16).

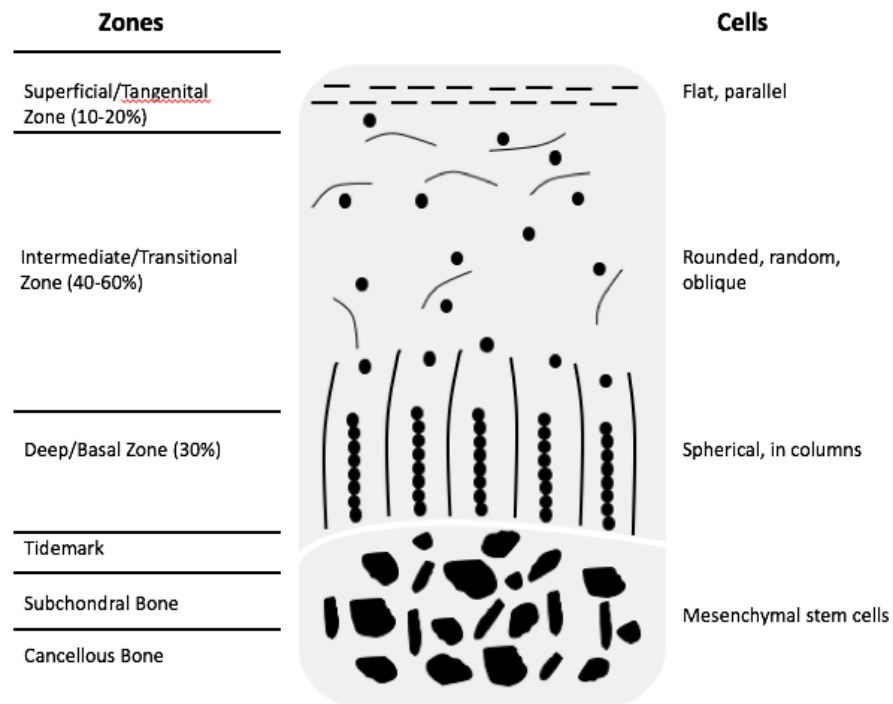


Figure 2. Morphology of Cartilage.

### Subchondral Bone

Under load, cartilage on both sides of the joint deforms maximising the contact area and minimising stress within the cartilage. As the load increases cartilage deformation is insufficient, in isolation, and it is necessary for deformation of the underlying bone to also occur. Joint damage from excessive loading is related to not only the magnitude of the load but also the rate of loading due to the limitation is the chondroprotective effect of the subchondral bone.

Subchondral bone is viscoelastic; it deforms less when load is applied more rapidly than when is loaded more gradually. Rapid loading does not allow the time necessary for the flow of interstitial fluid out of the bone, which would aid in absorbing the energy transmitted and protect the cartilage matrix. If deformation under impulsive loads is restricted the cartilage cannot adequately deform. The size of contact area becomes restricted, generating high stresses in the cartilage matrix (16). The subchondral cancellous bone, although stiffer

than cartilage is softer than cortical bone and serves as a major shock absorber.

In the normal joint the subchondral bone attenuates load more than the articular cartilage or surrounding soft tissues (16). In an osteoarthritic joint the subchondral sclerotic bone is less able to absorb and dissipate energy, thus increasing the force which is transmitted through the joint. The total volume of subchondral trabecular bone increases in OA predominantly due to thickening of the trabeculae, reduction in trabecular separation and some increase in number of trabeculae. This is demonstrated as subchondral sclerosis, a characteristic features of OA on radiographs, representing an apparent increase in density compared with normal bone. As the subchondral bone in OA is actively remodelling in response to increased stress, the newly formed bone does not have time to fully mineralise. This bone is less stiff than in normal controls (24). In order to retain a normal degree of stiffness the volume of subchondral bone in the osteoarthritic joint increases markedly (25). Although stiffening of the subchondral bone may play a key part in OA, it alone is not likely to account for the destruction of articular cartilage. Thinning of the cartilage as subchondral bone thickens and migrates towards the joint space eventually causes the cartilage to fragment.

### Metalloproteinases and cytokines

The metalloproteinases are a group of zinc containing enzymes involved in the degradation of cartilage and are felt to be key elements in the development of OA (26). They are inhibited by TIMP. When the normal ratio of the two groups becomes mismatched, catabolic effects of the metalloproteinases predominate resulting in cartilage destruction (27).

The role of cytokines in OA is complex. Some such as interleukin-1 (IL1) are predominantly catabolic whilst others such as insulin-like growth factor (IGF)–1

are anabolic. Other cytokines have been implicated but their roles are less clear. IL-1 is involved in cartilage degradation. Synthesised by chondrocytes, it acts in multiple pathways to suppress synthesis of type 2 cartilage and promote the formation of type 1 cartilage. It also induces catabolic enzymes such as the metalloproteinases. The levels of IL-1 are increased in OA compared to those without (28). Furthermore, chondrocytes from osteoarthritic joints are more sensitive to IL-1 than non-arthritic chondrocytes. Tumour Necrosis Factor alpha (TNF- $\alpha$ ) is a cytokine with similar effects to IL-1. Although likely to be involved in the pathogenesis of osteoarthritis, its role is less clear than in rheumatoid arthritis. IGF-1 is an anabolic cytokine found in decreased levels in the serum of patients with OA (29) and if deficient has been demonstrated to lead to cartilage loss in animal models (30).

### Synovium

Synovial responses to OA include synovial hyperplasia, fibrosis, activated synoviocytes and lymphocytic infiltration (17). Synovitis in OA may be due to phagocytosis of wear particles from bone and cartilage from abraded joint surfaces or release from cartilage of soluble matrix macromolecules (16). Earlier in the course of disease synovium from patients with full thickness articular cartilage ulceration may be histologically normal suggesting that in those cases pain is not attributable to synovitis (31). Conversely, in patients who have OA but no joint pain the severity of articular cartilage damage and synovitis may be as great as those who have pain (16).

### Periarticular muscles and alignment

Periarticular muscles absorb a large amount of energy in addition to providing joint function. Sensorimotor muscle dysfunction has been implicated in the pathogenesis of osteoarthritis by adversely affecting protective mechanisms

that prevent detrimental abnormal joint motion, damage and pain (32). Abnormalities in joint alignment may have a variety of underlying causes including congenital, post-traumatic, or related to an inflammatory joint disease. This results in abnormal mechanical stresses across the joint. Varus and valgus malalignment in the knee joint, for example, has been demonstrated to lead to increases in the progression of medial and lateral osteoarthritis (33,34).

### Obesity

Obesity is a further recognised risk factor for OA. In addition to the increased loading on the joint, it has been demonstrated that increase in body fat specifically as opposed to merely an increase in body mass, is more closely related to symptoms. Adipose tissue is able to release cytokines, growth factors and adipokines and obese individuals have greater concentrations of inflammatory markers; all of which may be involved in the progression of osteoarthritis and affect muscles function and pain thresholds (17,19).

### Ageing

Ageing is considered to contribute to the pathogenesis of OA via several methods. During ageing, structural changes occur in the components of the ECM. There are changes in the distribution of chondrocytes with deep cartilage layers demonstrating an increase whilst the superficial layers have diminished numbers (35). There is increasing dehydration of the matrix with ageing with increase in compressive stiffness which may have implications on the subchondral bone. There is also a decrease in size and structural organisation of aggrecan, the major cartilage proteoglycan, further contributing the changes in the ECM (36).

### **Ankle osteoarthritis**

In the UK, the incidence of symptomatic osteoarthritis of the ankle has been estimated at 47.7 per 100,000 (37). The most common underlying aetiological factor in ankle OA development is previous trauma whether it be fracture or ankle sprain (38). The incidence of these influences is increasing and ankle osteoarthritis is likely to become an increasing health burden (37,38). In a study of patients presenting with end stage ankle osteoarthritis 70% were determined to be due to previous trauma with previous fracture found to be the most common cause, followed by ligamentous injury (39). Another study had similar conclusions with 78% of end stage ankle arthritis identified due to previous trauma with other less prevalent causes including rheumatoid arthritis, haemochromatosis, haemophilia, talar dome avascular necrosis and previous septic arthritis (40).

The tibiotalar joint experiences the greatest contact force per unit area of any major joint making it susceptible to degeneration from minor changes to joint loading (41). Changes in the mechanical axis of the joint can result in increased loading per unit area leading to asymmetric wear. When the mechanical load falls on the medial or lateral part of the ankle joint varus and valgus ankle arthritis occurs respectively (41). Pathological loading can occur from a variety of underlying factors including malpositioned knee arthroplasties, tibial osteotomies and tibial malunions, as well as underlying congenital abnormalities such as tarsal coalition. Ankle trauma, resulting in ligamentous disruption, can lead to chronic instability, which again adversely affects biomechanics and transmission of forces through the joint.

Inflammatory arthritis such as rheumatoid arthritis and psoriatic arthritis can also lead to ankle arthritis. Excessive synovitis can lead to destruction of the cartilage surface. Ligamentous and capsular laxity as well as osseous degeneration in inflammatory arthritis can result in alterations in foot

biomechanics and alignment (41). Septic arthritis results in a global insult to the joint with articular cartilage being sensitive to infection (41). Other less common aetiologies such as neuropathic, haemorrhagic and neoplastic invariably insult the joint by either inflammation, altered biomechanics or both.

### **Management**

Management of ankle osteoarthritis can broadly be divided into non-operative and operative. Non-operative treatment ranges from modification of footwear and bracing to oral and intra-articular medications. Bracing aims to minimise motion across the joint by providing support above and below the joint. The most commonly used pharmacological method addressing the symptoms of ankle osteoarthritis is the use of nonsteroidal anti-inflammatory drugs. These are not without risk and side effects mean careful prescribing and use are needed. Corticosteroid injection into the tibiotalar joint is the final nonsurgical option offered after failing anti-inflammatory drugs and activity modification (3).

If a patient fails conservative management options, surgical options may be considered. The most commonly used surgical options include arthroscopy, osteotomies, distraction arthroplasty, ankle arthrodesis and total ankle replacement (TAR).

Patients with loose bodies, early osteoarthritis and osteochondral lesions may be suitable candidates for arthroscopy as well as those with impinging osteophytes (3). Corrective osteotomies address loading problems caused by malalignment of the lower limb which may contribute to ankle arthritis and symptoms. Distraction arthroplasty is a treatment option for advanced ankle arthritis which involves the use of an external fixator to forcibly separate the opposing surfaces of articular cartilage. It has been shown to help with pain,

function and clinical scores (42). In a small retrospective case series 91% of patients reported improvement in pain scores and only two of the 25 patients in the study requiring ankle fusion at 30 month follow up (43). However, a systematic review of evidence relating to distraction arthroplasty concluded that there is insufficient high quality evidence to support or refute its use for post-traumatic ankle arthritis.

Tibiotalar arthrodesis, or ankle fusion, has been regarded as the most well established operative management for end stage ankle arthritis (42). The main indication for arthrodesis is failed conservative therapy in patients with intractable pain or deformity (3). It can be performed via both open or arthroscopic techniques and using internal or external fixators with one of its main advantages being the reliability of good pain relief (3). Outcomes vary dependent on the technique used but is generally regarded to provide excellent pain relief but the resultant lack of motion at the joint leads to accelerated arthrosis at the subtalar joint and joints of the foot (42). This accelerated arthritis of the surrounding joints has been shown to lead to deficits in functional outcomes and limitations in activities of daily living (44,45)

Although arthrodesis is the main form of treatment for end-stage ankle osteoarthritis, total ankle replacement is being increasingly used as an alternative (38). The difficulty with TAR is identifying the most appropriate patients who will benefit in the short and long term (3). There are a number of prerequisites to be fulfilled and multiple contraindications to TAR. Despite this, TAR has seen a resurgence in the last decade with the current third generation of implants that are in use (3,42). The treatment for ankle arthritis is evolving rapidly with options that avoid fusion becoming more common in an attempt to preserve more joint mobility (42). The scope of indications for TAR is being expanded and evidence demonstrates that the use of TAR has long term positive impact on patients' lives (38). A systematic review exploring the outcomes of total ankle replacement demonstrated there was a statistically



significant improvement in pain and function scores at a mean follow up of 8 years with an implant survival reported at 89% at ten years with an annual failure rate of 1.2% (38). The most common reasons for failure of TAR are malalignment, aseptic loosening and infection which together account for approximately 50% of failures (3). The use of ankle allograft transplantation and viscosupplementation injection are also being explored (42).

## IMAGING IN OSTEOARTHRITIS

### Conventional Radiography

Despite advancements in medical imaging, radiographs remain the most commonly used and easily accessible investigation in the assessment of OA. They are used to establish a diagnosis of OA and to monitor disease progression. Acquisition is relatively inexpensive, readily available, and technically simple. They provide excellent visualisation of the pathological changes of osseous structures. They are able to depict the bony features such as marginal osteophytes, subchondral cysts, and subchondral sclerosis. As the disease progresses the joint space width is used as a surrogate for the integrity of cartilage. Radiographic assessment relies predominantly on the evaluation of joint space narrowing and osteophytes. Osteophytes often develop at an earlier stage of disease than joint space narrowing and are the most widely applied criterion for defining the presence of OA, whilst severity assessment relies mainly on joint space narrowing and abnormalities of the subchondral bone (46).

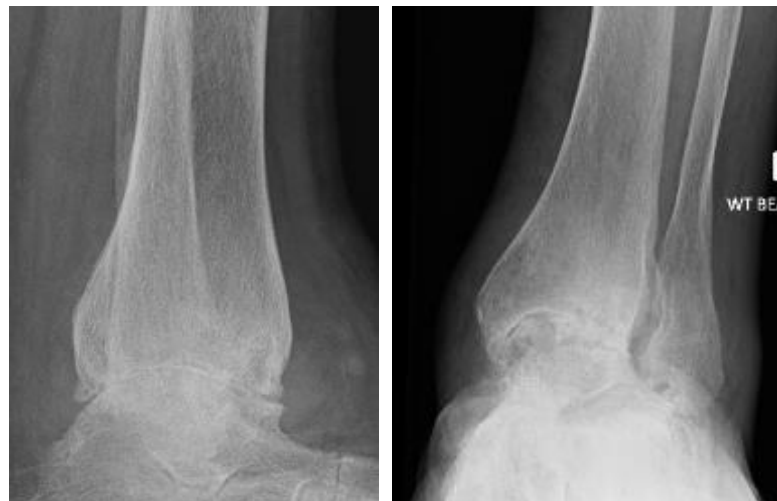


Figure 3. Severe radiographic OA.

Frontal and lateral radiographs demonstrating marked joint space loss, osteophytes, subchondral sclerosis and cysts.

Conventional radiography has several significant limitations despite usually being the primary investigation in OA. Changes within the cartilage early in disease precede any radiographically detectable abnormality. The inability of conventional radiography to distinguish different soft tissue structures is a major drawback if attempting to assess the whole organ. Progression of early stages of OA are therefore not well characterised (47). Changes in the position in the x-ray source and detector will affect the perception of joint space width that is a further limiting factor. Although radiographs are commonly used to assess joint space width it has been demonstrated that cartilage loss can be demonstrated on MRI without radiographic progression of disease or joint space narrowing (48). Conventional radiography also suffers from the two-dimensional representation of a three-dimensional structure with superimposed bone. Radiographs are also less sensitive than both MRI and CT when assessing for osteophytes (49).

### **Ultrasound**

Ultrasound (US) allows multiplanar, real time imaging with no ionising radiation at relatively low cost. It allows imaging of soft tissues without the need for contrast administration. The main limitations of US are in the inability to assess deeper articular structures and the subchondral bone as well as being operator dependent with a long learning curve (47,50). Much of the work regarding US has focused on inflammatory arthritis where it is more sensitive than MRI in assessing joint and tendon sheath effusions and more sensitive than radiography in detecting chondral lesions (51). Although US may be sensitive to bony erosive changes over time (52), subchondral bone changes are not visualised as the soft tissue–bone interface impedes sound waves penetrating the cortex, preventing the visualisation of subchondral pathology such as cysts. Despite the fact that cartilage can be visualised with US, in practice,

visualisation of cartilage in large, load bearing joints is difficult and the clinical relevance is questionable (50).

Doppler US is also useful in the depiction of synovial vascularity in large joint OA. A distinct role for US in imaging large joint OA is uncertain with MRI having many advantages. It does however provide a useful tool in differentiating inflammatory arthritis from OA and is a useful aid to joint injections.

### **Computed Tomography**

Computed Tomography (CT) is an imaging modality that can be used in the evaluation of OA, especially osseous change or when pre-surgical planning is required (50). Multi-planar reconstruction is also possible. Depiction of soft tissue calcification and cortical bone is better assessed using CT compared with MRI (53). CT is also useful in the detection of intra-articular osseous loose bodies. Its main limitation is the low soft tissue contrast.

Visualisation of cartilage with conventional radiography or CT is not possible due to not being radio-opaque. Indirect visualisation of cartilage with CT arthrography provides an alternative when MRI is contraindicated or not available. Penetration of contrast within layers of the cartilage surface is indicative of a defect of the chondral surface. High spatial resolution and the differing attenuation between cartilage and contrast media within the joint allows demonstration of focal changes. The limitations of CT arthrography include the fact it is an invasive procedure with factors such as patient pain and risk of infection requiring consideration. A further limitation is the insensitivity to changes of deep cartilage layers in the absence of surface alterations (50).

### **Radionuclide Scintigraphy**

Bone scintigraphy uses radiopharmaceuticals to demonstrate skeletal metabolism with technetium 99-m being the most commonly used radioisotope (47). It has the advantage over conventional radiography, CT, MRI and US (which provide structural assessment) in that it provides a physiological assessment of bone turnover. In OA of the hand scintigraphy demonstrates increased activity in the subchondral region preceding changes visible on radiography (54). Its value has also been illustrated as a predictor of disease progression in OA of the knee (55). A major disadvantage of bone scintigraphy is that the images are planar with superposition of a 3D array into 2D with loss of resolution for complex joints (47). Indications for performing a bone scan specifically in OA are limited, being used clinically to aid differentiation between pathologies (47).

SPECT (single photon emission computed tomography) improves contrast and localisation over conventional bone scintigraphy by separating sequential tomographic planes. Technical advancements have allowed SPECT to be fused with CT in the region of interest combining high levels of structural information with highly sensitive functional information (47). The main disadvantage being the radiation dose to the patient of combining SPECT and CT.

In Positron Emission Tomography (PET), 18-fluorodeoxyglucose (18-FDG) acts as a glucose analogue, taken up by body cells that have a high glucose need. This includes cardiac tissue, brain tissue and cells with a high metabolic activity. With improved resolution and localisation of tracer, resultant energy from positrons can be used to create 3D functional images (47). PET does demonstrate increased uptake in OA but this is not specific to OA and the value of PET in the clinical setting with respect to OA is yet to be demonstrated (56). The limited availability, cost and radiation dose are further generic limitations.

### **MRI**

MRI offers many advantages for OA imaging. It enables a cross-sectional perspective of anatomy without the projectional limitations of plain radiography and allows manipulation of image contrast to highlight different tissue types (50,57,58). MRI allows relevant features of the whole organ joint to be evaluated such as articular cartilage integrity, subarticular cysts, subchondral bone marrow abnormalities, subarticular bone attrition, osteophytes and joint specific features such as ligament and meniscal integrity (50,53). In addition to morphological data that is acquired in clinical practice, advanced MRI techniques used in the research setting that obtain biochemical data are providing further insight to the pathogenesis of OA (47,53,57,58).

In the imaging of specific OA features, conventional MR sequences may be employed. There are also many advanced sequences available, some of which have been developed for the sole purpose of imaging a particular feature of the joint. Some of the common conventional sequences used in everyday clinical practice in musculoskeletal imaging include fat saturated proton density (PD), T1 weighed and T2 weighted (and T2 with fat suppression) (59). Despite the many advantages of MRI, it does have some limitations. Image acquisition time is long compared to radiography and CT. With the exception of PET it is expensive compared to the other modalities. It is also contraindicated in some patients with certain medical implants.

### **Subchondral bone**

Subchondral bone is clearly important in the progression of OA and is well visualised using MRI. Bone marrow lesions (BMLs), sometimes termed bone marrow oedema-like lesions (BMO), subchondral cyst-like lesions and subchondral bone attrition are particular features of interest.

BMLs are degenerative lesions comprised of bone marrow necrosis, oedema and fibrosis. They are often seen alongside and correlated with cartilage damage (60). BMLs are best demonstrated on fat suppressed PD-weighted sequences, T2W weighted sequences and STIR sequences. They appear as areas of low signal intensity on T1-weighted spin echo sequences. The use of gradient echo sequences is discouraged since they are insensitive to marrow abnormalities even with fat suppression (61).

Subchondral cyst-like lesions appear as areas of well-demarcated fluid-like signal intensity on non-enhanced imaging sequences. They are a common finding in patients with OA although the exact pathogenesis is unknown. It is thought they result either from synovial fluid intrusion secondary to elevated intra-articular pressure or from post-traumatic bone necrosis following impact of articular surfaces (58).

Subchondral bone attrition is demonstrated in patients with advanced OA. It may be secondary to altered mechanical loading and is associated with the presence of bone marrow lesions (62). On MRI bone attrition appears as flattening or depression of the subchondral surface.

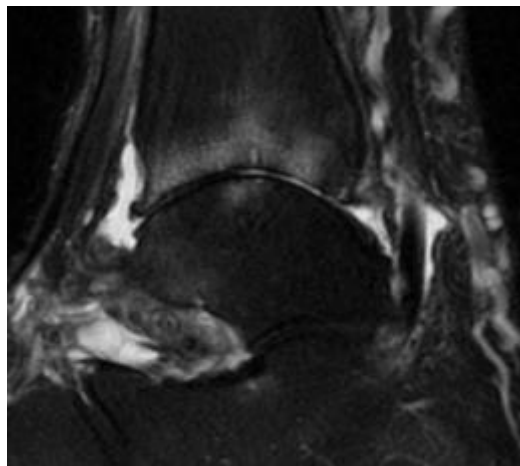


Figure 4. Ill-defined abnormal high signal in the subchondral distal tibia in representing bone marrow oedema.



Figure 5. Coronal STIR image depicting a well circumscribed high signal in the distal tibia representing a cyst

### Synovium

Synovitis is regarded as a key feature in OA pathogenesis and although it is possible to demonstrate synovitis on US, MRI is the imaging modality of choice for large joints where US may be limited (53,58). In contrast, for small joints, US is commonly used in assessing synovitis. Synovitis can be assessed using both non-contrast enhanced and post-contrast MRI. On non-contrast enhanced MR hypointense T1 signal and hyperintense signal on T2-weighted or proton density weighted sequences are used as a surrogate for synovitis (58). Contrast enhanced MRI better differentiates synovitis from effusion. Synovium with inflammatory activity enhances while effusions remains hypointense (58).



### Cartilage

MRI of articular cartilage is of particular importance in OA imaging as cartilage degeneration is often regarded as a structural hallmark of OA progression (63). MRI is the only imaging modality that allows visualisation of cartilage with sufficient contrast (64). Morphological changes such as tissue loss and degradation are thought to be preceded by biochemical changes, highlighting the importance of both morphological and physiological evaluation. Commonly used morphological MRI techniques include spin echo and gradient echo sequences. In addition to the choice of MRI sequence used, the strength of the MRI machine (1.5T vs 3T) can also play a factor with 3T demonstrating improved diagnostic performance for evaluating cartilage (64,65). There are many systems used to grade the severity of cartilage damage with respect to depth, most of these are modifications of previously described arthroscopy grading systems. The Outerbridge, Noyes and ICRS (International Cartilage Repair Society) scores are perhaps the most well-known (64). There are a number of advanced physiological techniques described in the literature that provide information on the composition of cartilage but these are beyond the remit of thesis. These include T2 mapping, T1rho mapping, sodium MRI, diffusion weighted imaging, and delayed gadolinium enhanced MR imaging of cartilage (dGEMRIC) (57).



Figure 6. PD sequence with normal articular cartilage at the tibiotalar joint.

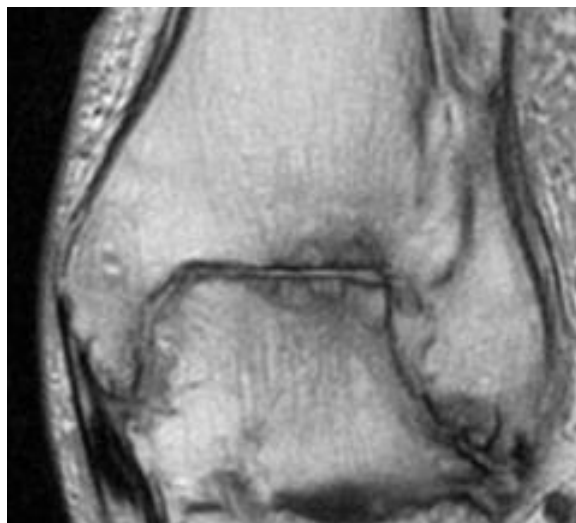


Figure 7. PD sequence with full thickness cartilage loss at the tibiotalar joint with underlying bony irregularity.

### 3. Grading systems

#### RADIOGRAPHIC GRADING

The most widely used system for grading the severity of OA on radiographs is the Kellgren-Lawrence (KL) grading system (66). The system is outlined in the table below. Grade 2 is generally regarded as the threshold level for the presence of osteoarthritis (67).

KL Grade	Description
0	No features of OA
1	Possible osteophytic lipping and possible JSN*.
2	Definite osteophytes with possible JSN
3	Moderate multiple osteophytes, definite JSN and sclerosis
4	Large osteophytes, marked JSN, severe sclerosis and deformity of the bone ends
* Joint space narrowing	

Table 1. Kellgren Lawrence grading scale

The KL grading scale is limited by the mixing of multiple distinct components of disease into one scale (osteophytes, joint space narrowing, subchondral sclerosis bone end deformation) (68). A further criticism of the system is that it emphasises osteophytes over joint space narrowing. A patient with definite joint space narrowing and cartilage degeneration but no osteophytes will therefore be classified as not having osteoarthritis (67). In contrast, the Osteoarthritis Research Society International (OARSI) atlas uses different individual scores for osteophytes and joint space narrowing, specifically for the joints of the hand, knee and hip (69). In addition, there is the option of scoring additional features separately such as the presence or absence of subchondral cysts. Sclerosis is scored separately for all the joints whereas it is not included in the KL score until grade 3. As the individual features are graded individually as

opposed to being grouped as in the KL system it is more sensitive to change and therefore more adapt for use in longitudinal studies (69). However, compared to the KL scoring system it is a more involved scoring system that is not routinely used outside the research setting. Despite its limitations, the KL scale is the most widely used radiological classification to identify and grade osteoarthritis (67).

## RADIOGRAPHIC GRADING OF OA OF THE ANKLE

The OARSI atlas does not include the ankle joint and the KL system did not include the ankle joint in original descriptions. There are radiographic ankle scores, mostly based on the KL system, described in the literature.

Moon et al evaluated the reliability of previously described grading systems for radiographic tibio-talar joint OA and compared this to cartilage damage on arthroscopy (70). The three systems included for grading were the KL scale and two other systems that had been used specifically to grade OA of the ankle prior to an orthopaedic intervention described by Takakura and van Dijk (71,72). The systems were modified by adjusting for the presence of talar tilt (defined as an angle of greater than two degrees measured between the tibial plafond and the upper surface of the talus on weight bearing antero-posterior radiographs of the ankle). This was based on the principle that tilting of the talus in the ankle mortise will result in asymmetric overload of focal areas of cartilage medially. Inter-rater reliability ranged from 0.58–0.89. The modified KL score demonstrated the greatest correlation coefficient compared to arthroscopic cartilage damage (0.53 vs 0.42 vs 0.52). The sensitivity and specificity of radiographs to predict arthroscopic cartilage damage was increased when talar tilt is considered. Lateral OA was not included in the study based on the premise that the incidence of lateral OA is rare in comparison and

the authors not being aware of any reports of isolated OA confined to the lateral malleolus and talus.

AIDA (Ankle Images Digital Analysis) is a semi-automatic digital technique that can be used to measure joint space width and provide quantitative data regarding the degree of subchondral sclerosis (73). The difference in joint space width between using AIDA and standard technique (measuring with a ruler), was shown to be statistically different with AIDA producing smaller measurements. AIDA also enables quantification of the degree of sclerosis which was demonstrated to be 'highly reliable' between observers. It relies on specific software for image analysis and although may be useful in the research setting this limits use in the clinical setting.

Holzer et al applied a modified KL grade to the ankle joint demonstrating good inter and intra-observer reliability (ICC 0.61 and 0.75) (74). KL grade 3 was subclassified into grade 3a and 3b, without and with talar tilt respectively. Joint space width was assessed quantitatively with AIDA. The presence of talar tilt was associated with significantly higher pain levels.

The use of the KL system has been shown to not be reliable for the subtalar joint. A study assessing the peritalar joints (subtalar and talonavicular) both before and 5 years after total ankle replacement, using the KL grading system demonstrated inadequate inter and intra-reader reliability (kappa values 0.37–0.43) (75).

#### Kraus scoring system

Kraus et al outlined the first radiographic atlas of osteoarthritis for the ankle and subtalar joint (76). Osteophytes and joint space narrowing are scored separately at multiple defined locations at the ankle and subtalar joint enabling calculations for a total joint osteophyte score and joint space narrowing score.

There are no discrete measurements given for the size or number of osteophytes or a quantitative measure of joint space narrowing but rather a comprehensive atlas is provided with examples of each grade of said feature at each location. The Kraus score also provides a separate modified KL score for both the ankle and subtalar joint with the ankle joint scored separately on anterior and lateral radiographic projections and the subtalar joint scored on the lateral projection only. The main modification of the Kraus KL score is the removal of subchondral sclerosis. This is not included as it was felt that it can be seen commonly at the ankle joint with 'mild' OA but with the original KL system, it leads to a higher grade for the affected joint. Grade 2 in the Kraus modified score allows mild joint space narrowing with definite osteophytes where with definite joint space narrowing in the original KL system, however mild, means a grade 3 or 4.

The reliability of the system was tested on a sample of 30 ankle joints. Results for experienced and trained readers spanned the range of "moderate" to "very good" based on the interpretation of kappa outlined by Landis and Koch (77). Inter-rater agreement for the modified KL score was at the upper limits of the "weak" range with a kappa value of 0.40. A weakness of the Kraus system was the use of a simple kappa statistic in reliability testing as opposed to a weighted kappa given that the study concerns ordinal data. This underestimates the true value of kappa with the possibility that actual reliability and reproducibility values would have been higher. The system was also only tested on a random sample of 30 subjects with no demonstration of the spread of disease across this sample although there is a mention of relatively fewer grade 3 and 4 examinations included and some features occurring infrequently (e.g. posterior talar osteophytes). This will also adversely affect the kappa values for inter and intra-reader agreement. Despite some of these shortcomings, it is the first comprehensive atlas of radiographic OA at the ankle and subtalar joint that has conducted reliability testing.

## EXISTING MRI SYSTEMS

**MRI scoring systems for osteoarthritis: ankle joint**

A systematic search of the Medline and Embase databases was undertaken to identify any studies that have described whole organ MRI grading of OA of the ankle. Search terms used included; Osteoarthritis (exploded), Magnetic Resonance Imaging (exploded), grading, scoring, score, grade and ankle. The search terms were kept broad to attempt to include as many results as possible as from prior reading no studies had been identified. The limits were human studies, adult subjects and published in the last 15 years.

(Figure 8). The search did not identify any described whole organ MRI based grading systems for ankle OA.

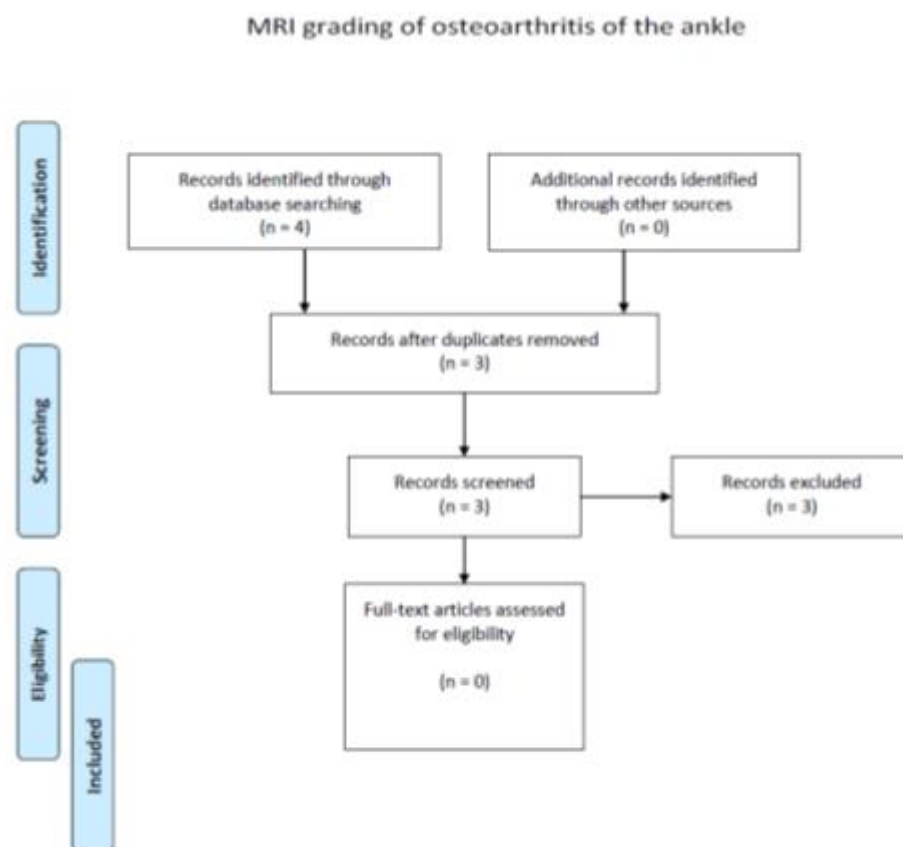


Figure 8. Search for Ankle OA MRI grading systems

#### **MRI scoring systems for osteoarthritis: all joints**

The majority of research regarding MRI imaging of OA has focused on the knee, likely due to the high prevalence of knee OA and un-complicated image acquisition compared with other joints (78). There are some knee based MRI whole organ scoring systems that exist. The three most commonly discussed MRI scoring systems in the literature include the WOMBS, KOSS and BLOKS systems and more recently the newer MOAKS scoring system (79–82). A search was undertaken to identify other systems that may exist for the large weight bearing joints.

A systematic search of the Medline and Embase databases was undertaken to identify any studies that have described whole organ MRI grading of OA in weight bearing joints. Search terms used included; Osteoarthritis (exploded), Magnetic Resonance Imaging (exploded), grading, scoring, score and grade. The limits were human studies, adult subjects and published in the last 15 years. (Figure 9). Seven studies were identified which take a whole organ approach to grading OA on MRI. These include 2 hip based scoring systems and 5 knee based scoring systems.



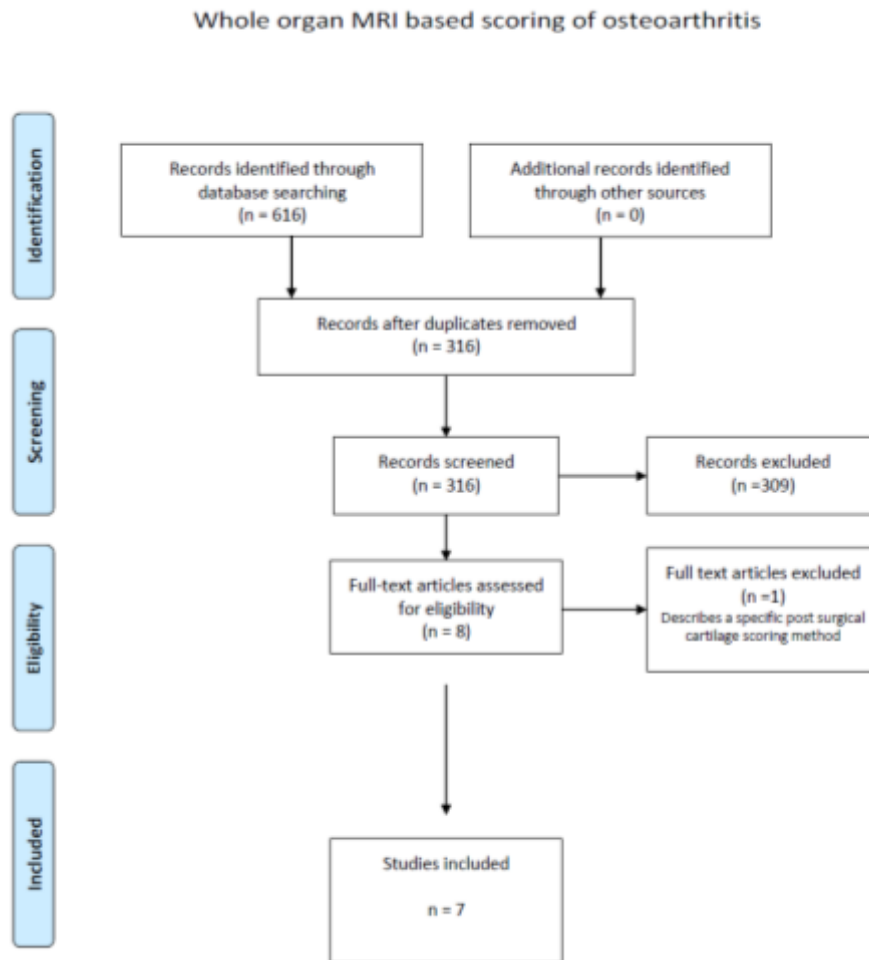


Figure 9. Search for knee and hip OA MRI grading systems

Due to the differing components of the MRI scoring systems along with multiple different scoring methods for a given component a narrative review of the identified systems will follow.

## WORMS

The Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis was published in 2004 and was the earliest scoring system identified in the search (79). WORMS uses a system for zonal division of knee compartments: the patella is divided into medial and lateral facets. The tibial plateau and medial and lateral femur are subdivided into anterior, central and posterior subregions. The spinous process is classed alone as a separate subregion. Multiple features are included in the WORMS system although there is no justification for these provided. The features included are cartilage integrity, subchondral bone marrow lesions, subchondral cysts, osteophytes, meniscal status, osteophytes, effusion and synovitis score, and integrity of the cruciate and collateral ligaments. Periarticular features such as popliteal cysts and loose bodies are also evaluated. A comparative view of the scoring of all features for all the scoring systems is outlined in appendix 2.

The WORMS system was tested on a sample of 19 knees that had KL grade 2 or 3 osteoarthritis on plain radiograph. Nearly all features demonstrated high intraclass correlation coefficients (ICC) of greater than 0.81 with ICC values for the cartilage score and osteophyte exceeding 0.9 (79).

The WORMS cartilage score is the most complex of all the MRI scoring systems comprising of eight different grades with the authors stating the system aims to capture different patterns of regional cartilage loss and give more information about the extent of surface involvement. Having so many increments makes the cartilage score sensitive to change. Despite the complexity, the score demonstrated near perfect ICC values for total joint scores and subregional scores; 0.99 and 0.97–0.99 respectively (79).

The scoring of osteophytes was also the most complex of all the systems using an eight-point scale. The osteophyte score was based on the method outlined in the Osteoarthritis Research Society International (OARSI) Atlas for radiographic grading of osteophytes in the knee but was expanded from four grades to the eight-point system. This also demonstrated high ICC values at 0.99 for total joint score and 0.93–0.98 for the subregion scores (79).

WORMS is the only system that grades severity of bone attrition, defined as flattening of the articular surface, and only one of two system that records its presence at all. Bone attrition demonstrated the weakest agreement with an ICC value of 0.61 but due to the low prevalence of this feature the ICC in this instance was regarded to be unreliable (79).

The complex subregional approach used in WORMS contrasts the more lesion-based approach of some of the other systems, notably for BML and cartilage scoring. Using the lesional approach may lead to difficulties in determining the exact number of lesions as they may be directly adjacent to each other or even merging (78). There are some limitations to WORMS. The sample size is relatively small with no justification for the number of knees analysed. There is no clear justification given for the complex subregional division used. Similarly, there is no justification provided for the features that are included in the system and some features demonstrated a low prevalence giving ICC values that were deemed to be unreliable (bone attrition). Although inter-rater reliability on the whole gave high ICC values there was no intra-rater analysis. Furthermore, the readers in WORMS were expert musculoskeletal radiologists with experience in MRI based osteoarthritis imaging which may limit generalisability. Although the conclusion of the proposed system is that of a reliable multi-feature tool for OA of the knee, it is stated that WORMS was not intended as a 'definite solution' but an initial step in the process of development in whole-organ evaluation of OA.

## KOSS

The Knee Osteoarthritis Scoring System (KOSS), introduced by Kornaat, covers similar OA features as the WOMBS system (81). There are however some differences. Cartilage integrity, cysts and subchondral BMLs are scored for each subregion, with scores differentiated by size of the lesion. Osteophytes are differentiated into marginal, intercondylar and central. Medial and lateral menisci are reviewed for the presence of meniscal tears, subluxation, intra-substance degeneration or absence of a meniscal portion. Other features included are effusion, synovitis and Baker's cyst. Unlike many of the other scoring systems, KOSS not only accounts for the presence of osteochondral defects as part of the cartilage score, but there is a separate sub-score within the cartilage score for the presence and depth of osteochondral lesions. There is a further score to grade the surface extent of the lesion giving three separate scores for cartilage and osteochondral integrity. Comparative view of the scoring of all features for all the scoring systems is outlined in appendix 2.

There are nine subregions in the KOSS system differentiating the medial and lateral patellar facet, patellar crest, medial and lateral trochlear articular facet, medial and lateral femoral condyle, and medial and lateral tibial plateau.

The system was tested on a sample of 25 knees with confirmed radiographic OA of KL grade 2 and 3. Inter-rater ICC values were generally high at 0.77 covering all features. BMO had the highest ICC value at 0.91 whereas the cartilage score had the lowest at 0.64. Unlike some of the other scoring systems, intra-rater agreement testing was performed which demonstrated ICC scores ranging from 0.76–0.96.

As with WOMBS, the sample size used to test KOSS was relatively small with no justification given. No justification was given for the features included in the system. Compared with WOMBS, KOSS demonstrated lower ICC values for

osteophytes and cartilage scoring although higher ICC values for BMO. A potential limitation of the KOSS cartilage score is that it is more simplistic than many the cartilage scores of other scoring systems with no inclusion of abnormal signal change within the cartilage without a tear. The authors state this was not included as the system was designed to test for all grades of OA and not early OA change. Nevertheless, this has the potential to under-score cartilage that is abnormal but not torn. The system was only tested on OA grades 2 and 3 so although the reliability values are generally high this is not applicable to KL grade 1 or severe, KL grade 4.

## BLOKS

The Boston Leeds Osteoarthritis Knee Score (BLOKS) was designed with the aim of developing a comprehensive, semi-quantitative scoring system specific for knee OA (80). In addition, the validity of the BML score component was examined. As a result of a collaborative program with two international centres, an initial meeting addressed items and scoring to be included in BLOKS based on MRI literature with many of the items selected based on likely relevance to pain and structural damage or progression of OA. The scoring system put forward included cartilage integrity, attrition, BMLs and cysts, osteophytes, ligaments, meniscus and synovitis in addition to other items which may were felt may warrant further attention in OA such as meniscal displacement, collateral ligament contours, osteophyte signal, synovitis separate from effusion, subchondral plate signal and thickness, limb alignment and muscle quality.

Compared to the WOMBS system, BLOKS focused on the weight bearing components of the tibiofemoral joint and in addition the patellofemoral joint in a similar approach to that in KOSS. In contrast to WOMBS there are three components to the BLOKS BML score to account for the overall size of the

lesion, percentage of surface area adjacent to the subchondral plate and finally the percentage of the BML that is comprised of a cyst. Cartilage grading is composed of two scores components. The first grades the presence of any cartilage loss in a subregion and the percentage of full thickness loss in a subregion. The second component grades cartilage from 0–2 at eleven specific locations (not subregions). The scoring system and subregional description are outlined in full in appendix 2.

A series of three reliability exercises were undertaken with 10 MRI knee examinations scored in each session followed by an adjudication session. The focus of this exercise was to refine the features included and remove ‘redundant items’ to develop a more user friendly tool. Features with a weighted kappa value of less than 0.2 were removed from the scoring system. The inter-rater reliability results of the third and final round (n=10) were presented for the final features that are displayed in appendix 3. The inter-rater agreement (weighted kappa) values ranged from 0.51–0.79 across the different features with scores relating to BMLs and cartilage demonstrating the highest values being in the range of “good” agreement (Landis and Koch kappa interpretation (77)). Features including synovitis, effusion and meniscal extrusion demonstrated “good” and “moderate” agreement but confidence intervals were wide extending into the “poor” agreement range (77).

The validity of BLOKS was examined compared with WORMS as well as determining correlation between components of both systems. This was performed using data obtained from the Boston Osteoarthritis of the Knee Study (BOKS) (83). A sample of 71 patients who were part of BOKS were included in the analysis who had baseline, 15 month and 30 month MRI scans as well as completing knee pain questionnaires. The WORMS cartilage score was collapsed from the originally described eight-point scale to a five-point scale (80).

In the medial and lateral knee compartment Spearman's rho correlation values were 0.63 and 0.79 respectively. There was a significant association with increasing BML grade and increasing pain using the BLOKS system but this was not significant when using WORMS. Higher baseline BML scores were associated with more severe cartilage loss using both the BLOKS and WORMS methods ( $p < 0.001$ ) (80).

Potential limitations of the BLOKS system is the complexity of the instrument, something the authors acknowledge. As with WORMS, the authors (some of who were also involved with developing the WORMS system) are expert musculoskeletal radiologists with experience in MRI based OA imaging which may again limit generalisability. Although the cartilage and scoring components demonstrated "good" inter-rater agreement, some of the other features included had less favourable levels of agreement with wide confidence intervals. This may be in part due to the relatively small sample size.

Two related studies comparing BLOKS and WORMS have been performed (84,85). They concluded that little extra information was added by using the more complex BLOKS BML score with both system giving equivalent results for both extent and number of BML (84). The WORMS BML score also better predicted cartilage loss and was subjectively easier to use (85). The BLOKS system was however more sensitive for full thickness cartilage defects than the WORMS cartilage score (85).

## MOAKS

The MRI Osteoarthritis Knee Score (MOAKS) was developed based on limitations of both WOMBS and BLOKS which since their inception had undergone multiple unpublished iterations (82). Expert readers met to establish the limitations of the current scoring systems. BMLs and meniscal abnormalities were the main features noted for revision with refining of BML scoring to include subregional assessment and refine elements of meniscal morphology evaluation. In addition, redundancies in cartilage scoring were removed. The grading system is outlined in appendix 2. Many of the features are very similar to BLOKS or WOMBS with the addition of elements to meniscal scoring and refinement of cartilage and BML scoring.

Reliability testing was undertaken on a sample of 20 MRI knee examinations with an equal number of left and right knee and KL grades 2 or 3 examined. Inter and intra-rater reliability testing was undertaken on all 20 examinations. With the exception of a few features in certain subregions, inter and intra-rater reliability was “substantial” or “almost perfect”. Features such as cartilage area at the tibia demonstrated “weak” levels of agreement despite percentage agreement being 70% and this was felt to be due to relatively low frequency of occurrence adversely influencing the kappa values.

The MOAKS system refines limitations of the WOMBS and BLOKS systems and again, many of the authors of MOAKS were involved in development of the previous two systems. Limitations of MOAKS include the relatively small sample size that may have led to the less favourable kappa values despite high levels of percentage agreement. The system also examined reliability on a sample of knee MRI examinations that were all performed on 3T MRI machines, limiting generalisability of using the system in many departments that do not have access to 3T scanners. It is however the most recent and updated semi-quantitative tool for evaluation of knee designed for use in longitudinal studies.



## PARK ET AL

A study published in 2013 by Park et al set out to evaluate reliability and association with radiographic scores of an MRI grading system for knee OA that does not require complicated calculations (86). Due to the structure of the grading system it is outlined in the table below and not included in appendix 2.

MR Osteoarthritis Grade	Description
0	No cartilage injury with no or minimal osteophyte (<5mm)
1	Cartilage injury grade 1 and at least one of the following: Osteophytes >5mm, BMO >10mm, subchondral cyst >10mm
2	Cartilage injury grade 2 and at least one of the following: Osteophytes >5mm, BMO >10mm, subchondral cyst >10mm
3	Cartilage injury grade 3 and at least one of the following: Osteophytes >5mm, BMO >10mm, subchondral cyst >10mm
4	Cartilage injury grade 3 and complex meniscal tear

Table 2. Components of the Park scoring system (86)

Unlike the other MRI scoring systems that have individual scores for each feature, the Park system aims to mirror the KL scale by means of grades running from zero through four with multiple features included in each grade. The main determining feature is the cartilage component with cartilage graded from zero to three. 0 represents normal cartilage, 1 represents altered signal intensity in the cartilage only, 2 represents cartilage defect less than 99% and 3 represents 100% defect with or without bony ulceration. The choice of size limits for BMO and subchondral cysts were based on evidence suggesting that lesions over 1cm in diameter are increasingly common with increasing KL (87).

The system was tested on a retrospective sample of 105 patients. Both inter and intra-observer agreement was “almost perfect” ranging from 0.82–0.84 and 0.91–0.94 respectively. Correlation between MR grade and KL grade was also high with Spearman’s rho ranging from 0.90–0.97.

Compared with the other MRI grading systems, the Park system is less comprehensive in terms of providing separate scores and reliability information for each OA component included. The system does display high inter and intra-observer agreement and is more suited for MRI grading of osteoarthritis in routine clinical practice compared to the time consuming and complex methods of some of the other systems.

## HOAMS

The Hip Osteoarthritis MRI Scoring System (HOAMS) was the first semi-quantitative whole organ scoring system for the hip joint (88). Thirteen articular features are assessed in the system which include cartilage morphology, BMLs, subchondral cysts, labral lesions, synovitis, effusions, loose bodies, attrition, herniation pits, paralabral cysts, labral hypertrophy and trochanteric bursitis. The subregional divisions used in HOAMS are complex with a total of 15 subregions used with eight subregions for the femoral head and seven for the acetabulum. A full description of the system is outlined in appendix 2.

Although 52 hip examinations were scored with the new system only 15 cases were used for inter and intra-rater reliability testing. Values for inter and intra-rater agreement for cartilage and BMLs were in the range of “substantial” and “almost perfect”. Agreement levels for some features such as cysts were only “slight”. Although some features such as bone attrition demonstrated perfect agreement the prevalence of this, and many other features in the sample tested was very low ( $n=0-3$ ). The confidence levels for many features were

wide extending to levels that would significantly alter the interpretation of the presented reliability values. Lower limits of the confidence intervals for cartilage for example extended to the lower boundaries of the “fair” agreement range. Some confidence intervals such as those for cysts extended below zero.

Nearly all features (except labral tear, labral hypertrophy and herniation pits) demonstrated statistically significant association with radiographic KL grades. Validity of the scoring system was assessed against hip outcome score questionnaires that included both pain and functional components although the association between the MRI scores and the patient rated outcomes were not statistically significant.

Despite some features demonstrating favourable levels of agreement there are limitations to HOAMS. No justification is given for the features included in the system or the divisions of the subregions. Synovitis is assessed on contrast enhanced images. Intravenous contrast media is not routinely administered for MRI scans of the hip for osteoarthritis limiting the generalisability of this scored feature. Many of the confidence limits extended into low levels of agreements. For many features the level of agreement assessed with a simple percentage agreement was high as opposed to the weighted kappa. Cysts for example had an inter-rater percentage agreement of 95% with a corresponding weighted kappa of 0.15 and negative confidence intervals. This may be due to the small sample size of only 15 studies that were used for reliability testing. Further evidence of the limitations due to sample size is the very low or absence of many features which led to interpretation of “perfect” agreement of some features where raters did agree but only on the single occurrence of a feature. Despite these limitations HOAMS is the first system of its kind for hip OA scoring and demonstrated “satisfactory” reliability overall. The authors conclude that HOAMS was not intended to be a “definite solution” but rather an “initial step” in development and improvement for a semi-quantitative whole organ approach to scoring hip OA.

## SHOMRI

The Scoring Hip Osteoarthritis with MRI (SHOMRI) system set out to develop and test a practical semi-quantitative MRI based scoring system for OA, test reproducibility and correlate with radiographic and clinical scores (89). The subregional divisions used in SHOMRI were simpler than those used in HOAMS with only 10 subregions that are based on those outlined by the Arthroscopy Society of North America. The features included were based on a literature review of hip OA findings and were articular cartilage loss, BMO, subchondral cysts, labral abnormalities, paralabral cysts, intra-articular loose bodies, effusion, and ligamentum teres abnormality. The system is outlined in appendix 2. Features including osteophytes and bone attrition were excluded from SHOMRI as were felt to be better visualised on plain radiographs than MRI. Clinical assessment included range of hip motion measurements and patient self-reported functional and pain assessment.

The study involved a sample of 98 patients although only 23 of these had radiographically evidence OA (KL grade  $\geq 2$ ). ICC values were “excellent” ranging from 0.91–0.97. Inter-rater agreement assessed with weighted kappa was within the range of “moderate” or “good” for all features ranging from 0.55 to 0.79. Intra-rater agreement was “good” with values from 0.65–0.79.

Correlations between MRI scores and radiographic scores (KL and OARSI radiographic scores) were statistically significant for all MRI features with correlation strongest for cartilage loss and subchondral cysts where Spearman’s rho was 0.52 and 0.49 respectively. MRI correlation with clinical scores were statistically significant for BMO and cysts with relation to pain and impact on activities of daily living although strength of correlation was only “weak” to “moderate” (0.27–0.44)(90).

Limitations of the SHOMRI system include the simplicity of grading some features such as cartilage that used a three-point scale versus the six-point scale used in HOAMS which may affect the sensitivity of the system to interval change. The system was tested on a sample of patients who had mild to moderate OA radiographically with no severe grades (KL grade 4) included and the majority of subjects having KL grades of 0 or 1 which limits generalisability. This may also have impacted the levels of agreement as features including BMO and cartilage were only recorded as having an abnormality in 31% and 12% of subjects respectively and despite demonstrating high percentage agreement (75–97%) kappa values were only in the “moderate” range of agreement.

Compared to HOAMS, SHOMRI is a more user friendly system for grading of hip OA with fewer subregions and less complex grading of the features included. It is based on standard MRI protocols for imaging the hip with no requirements for intravenous contrast as required for HOAMS.

A summary of the main inter-reader reliability results of the difference scoring systems is outlined in appendix 3.

#### Summary

The current published MRI scoring systems all score similar features of OA with BMLs and cartilage integrity having specific focus. They all used a zonal approach to scoring although how zones were chosen did vary widely and was not always clearly justified. Some such as BLOKS and SHOMRI also assessed MRI findings with relation to clinical findings although most focused purely on reliability testing. There are some limitations that applied to many of the systems that should be considered in designing our study. There was no clear justification for sample size in any of them with sample sizes varying widely. The underlying prevalence of disease severity also varied with only MOAKS attempting to be representative across severity with some not including any patients classified as having 'severe' radiographic OA. These systems are still evolving, drawing on potential limitations identified in studies beyond their initial description as demonstrated in the MOAKS system.

## 4. Delphi survey

### BACKGROUND AND METHODS

There are many features of OA that can be demonstrated on MRI as previously discussed. Not all the features are included in every scoring system of other joints. Furthermore, some features are merely recorded as present or absent by other scoring systems whilst other features have numerous grades assigned to them such as the eight-point grading scale for cartilage integrity in WORMS.

In order to determine what features to include in the ankle MRI score a Delphi survey was conducted to ascertain a consensus opinion from a group of experts regarding both the presence or absence of the feature as well as if a consideration of severity should be included.

The Delphi technique is a structured communication process used to collect and rank data to reach a consensus from a group of people without requiring face-to-face contact (91). It is composed of sequential questionnaires with feedback to participants. Responses to the first questionnaire are summarised and used to develop the second questionnaire that aims to seek agreement. The process can be repeated as desired.

A Delphi survey is a well-suited method to try and determine a consensus opinion in this situation and has a number of advantages over other consensus techniques such as a focus group or nominal group method. It allows for analyses, ranking and priority setting and can clearly determine if there is consensus or not (92). It allows anonymity of respondents and does not require participants to meet physically reducing costs and avoiding issues relating to participant reticence. The remote nature of the process avoids negative group influences such as dominating members that can potentially be problematic in other consensus methods. Potential disadvantages of this process are the time

consuming nature of it and potential for low response rate and participant drop out between questionnaires.

Following review of the survey protocol, the R&D department at the NNUH approved the survey to be held at the institution confirming that formal ethical approval and local NHS permission was not required (appendix 4).

A group of experts who either report or interpret the findings from MRI examinations of the ankle in OA were selected as participants. This comprised of four consultant radiologists who have a subspecialty interest in musculoskeletal radiology, two consultant rheumatologists and two consultant orthopaedic foot and ankle surgeons. One of the orthopaedic surgeons was based at a local district general hospital but had recently completed fellowship training in foot and ankle surgery at a tertiary centre. One musculoskeletal radiologist was based at the Nuffield Orthopaedic Centre in Oxford, a tertiary musculoskeletal unit. All other participants held substantive posts at the NNUH.

The survey was hosted on an online platform ([www.surveymonkey.com](http://www.surveymonkey.com)). It comprised of a five point Likert scale concerning each component in question. Participants were asked to score from one to five the perceived importance of each component listed to be included in the report of an MRI of the ankle. These included the presence and extent of subchondral marrow oedema, the presence and number of osteophytes, the degree of cartilage degeneration, integrity of the supporting ligaments, the presence of osteochondral defects, presence of subarticular cysts, presence and severity of subchondral bone attrition, presence of a joint effusion, presence of synovial hypertrophy and the presence of intra-articular loose bodies. These features were assessed for both the tibiotalar and the subtalar joint. The components included in the survey were based on the components of other existing scoring systems with only joint specific components removed e.g. menisci for the knee. An outline of the questionnaire can be found in appendix 5.



## RESULTS

The online link to the first-round survey was sent to participants by email and open for four weeks. All participants responded. All components with a median score  $<3$  were excluded from round two. The second-round survey link was open for four weeks with all participants responding. Features with a median score of  $\geq 4$  were taken forward to the MRI scoring system.

### Tibiotalar joint

The results for both the round one and two surveys for the tibiotalar joint are displayed in table 3. The results are also displayed graphically in Figure 10 which highlights differences in responses between the specialties.

	Round 1			Round 2		
	Mean	Median	SD	Mean	Median	SD
Presence of BMO	4	4	0.71	4.13	4	0.6
Extent of BMO	3.75	3.5	0.83	4.13	4	0.78
Presence of osteophytes	3.75	4	1.09	3.88	4	1.17
Number of osteophytes	2.88	3	1.05	2.75	3	0.97
Cartilage integrity	4.63	5	0.48	4.63	5	0.48
Ligament integrity	3	3	1	2.88	3	1.05
Presence of OCD	4.13	4.5	1.05	4.25	4	0.66
Presence of cysts	4.25	4	0.66	4.38	4	0.48
Presence of bone attrition	4.25	4	0.66	4.38	4	0.48
Severity of bone attrition	4.38	4.5	0.7	4.38	4	0.48
Presence of joint effusion	2.75	3	0.66	3.13	3	1.05
Presence of synovitis	2.88	3	0.6	2.88	2.5	1.05
Presence of loose bodies	2.5	2.5	0.5	-	-	-
BMO; BMO. OCD; Osteochondral defect. SD; Standard deviation						

Table 3. Delphi survey results for the Tibiotalar joint

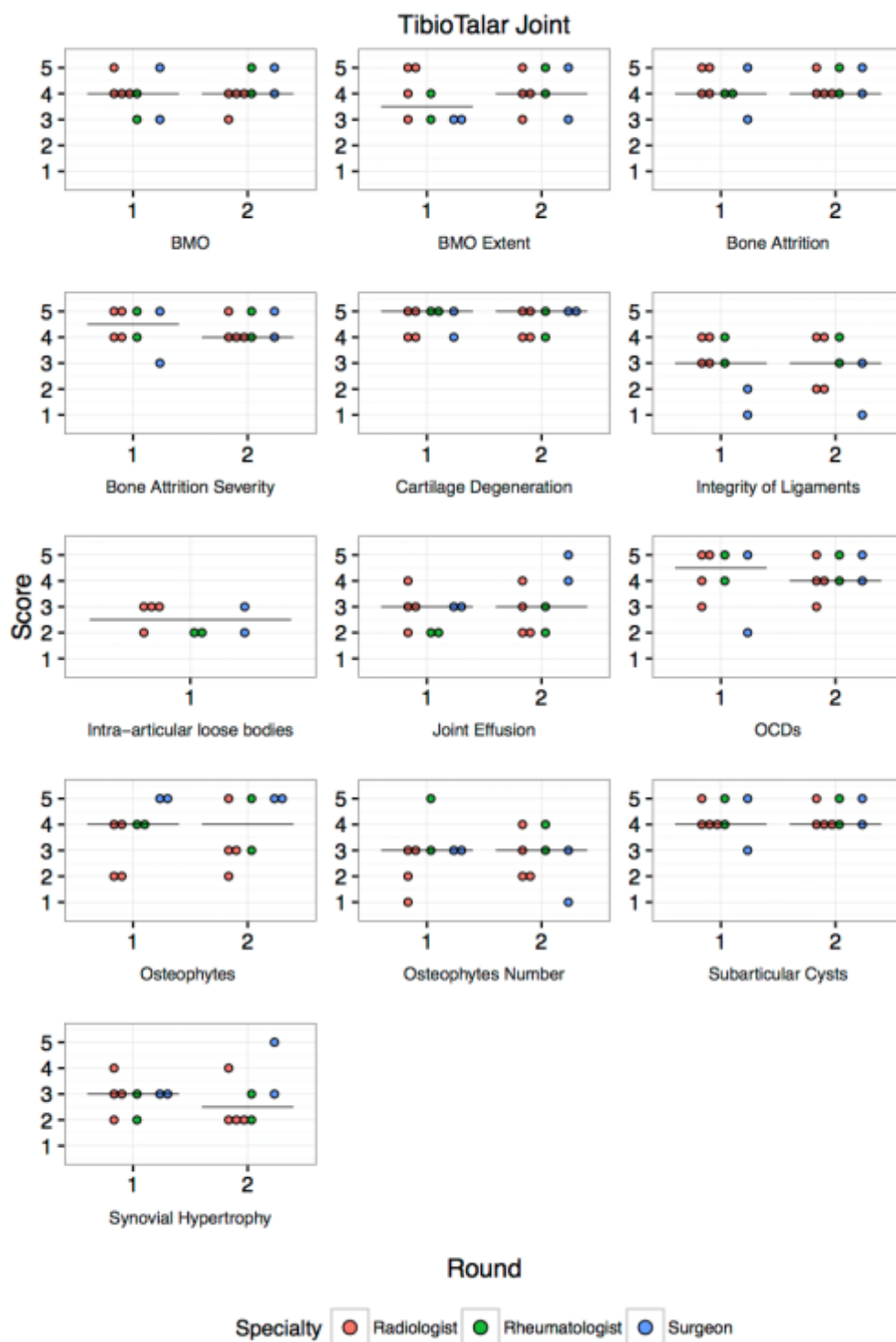


Figure 10. Multiplot outlining the Delphi results for the Tibiotalar joint for each feature by clinical specialty. The horizontal bar denotes the median score.

All components except the presence of intra-articular loose bodies made the threshold median score of three or more to be included in the second round. The standard deviation decreased for most features between round one and round two indicating an improvement in consensus. There was a slight increase in standard deviation for osteophytes (1.09 to 1.17). The presence of joint effusion and the presence of synovitis both demonstrated an increase in standard deviation between rounds but neither of these features made it past the second cull to the MRI score. Figure 10 demonstrates that for both of these features the second round score from the surgeons increased whilst the view of the other specialists remained relatively unchanged, causing the increase in the standard deviation. Although the purpose of the Delphi survey is to determine the consensus opinion it is interesting to observe some of the differences in perceived importance between specialties. In general the surgeons scored the importance of ligament integrity quite low compared to both radiologists and rheumatologists whilst the presence of osteophytes was considered very important by the surgeons whilst the radiologists scored this feature much lower.

The features that made it through the first round cut and had a resultant median score of four or five in the second round were the presence and severity of BMO, degree of cartilage degeneration, presence of chondral defects, presence of subarticular cysts, presence and severity of bone attrition and the presence of osteophytes.

## Subtalar Joint

The results for both the round one and two surveys for the subtalar joint are displayed in table 4. The results are also displayed graphically in Figure 11 which highlights differences in responses between the specialties.

	Round 1			Round 2		
	Mean	Median	SD	Mean	Median	SD
Presence of BMO	3.63	4	1.2	3.63	4	1.11
Extent of BMO	3.5	4	1.1	3.63	4	1.22
Presence of osteophytes	3	3.5	1.2	2.75	2.5	1.2
Number of osteophytes	2.5	2.5	1.2	-	-	-
Cartilage integrity	4.13	4.5	1.2	3.75	4	1.2
Ligament integrity	2.25	2	1.2	-	-	-
Presence of OCD	3.13	3	1.4	3.63	4	1.11
Presence of cysts	3.25	3.5	1.2	3.63	4	1.32
Presence of bone attrition	3	3	1.5	3.13	3	1.17
Severity of bone attrition	3.13	3.5	1.5	3.13	3	1.17
Presence of joint effusion	2.13	2	0.7	-	-	-
Presence of synovitis	2.13	2	0.7	-	-	-
Presence of loose bodies	1.75	2	0.6	-	-	-
BMO; BMO. OCD; Osteochondral defect. SD; Standard deviation						

Table 4. Delphi survey results for the Subtalar joint

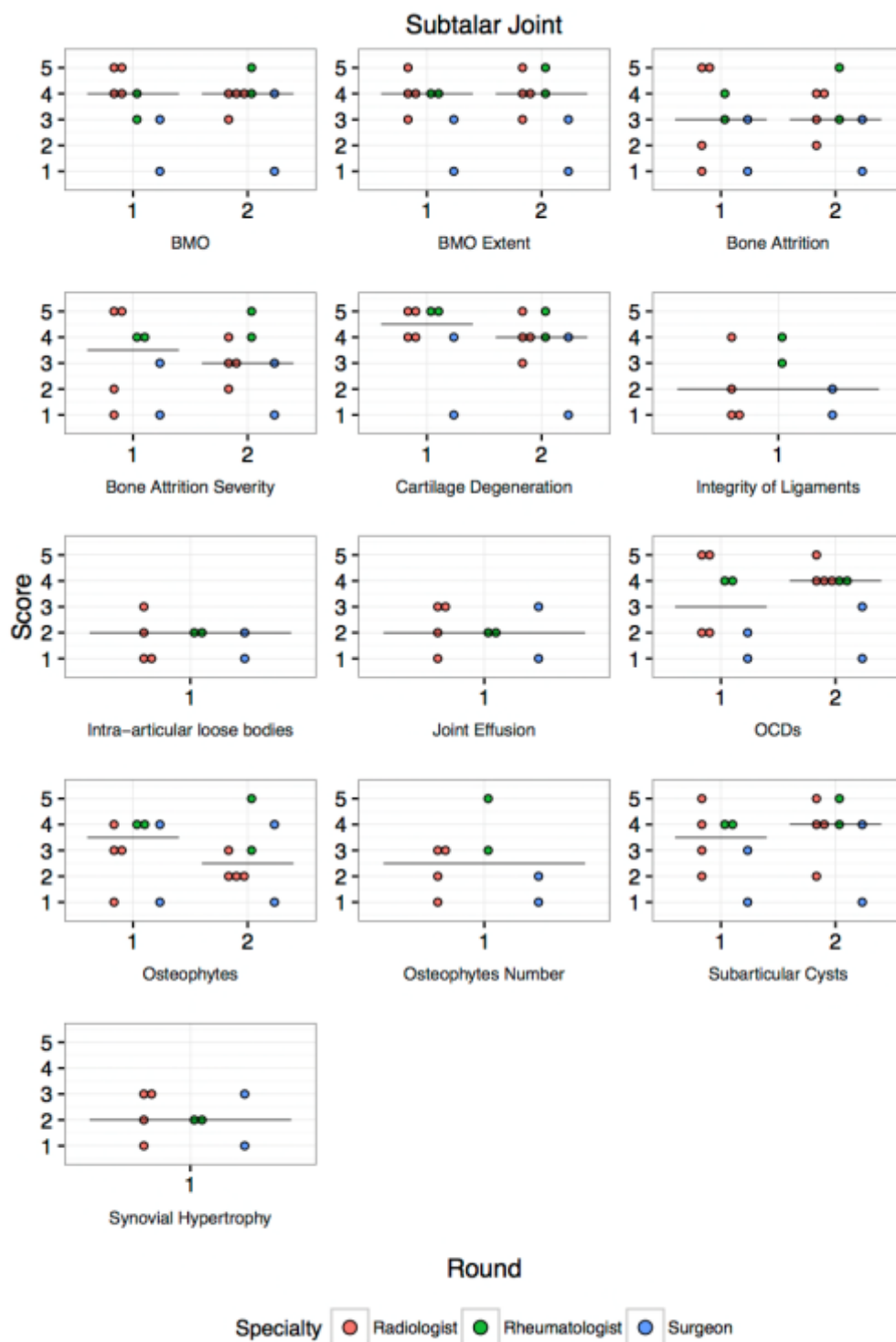


Figure 11. Multiplot outlining the Delphi results for Subtalar joint for each feature by clinical specialty. Horizontal bar denotes median score

Number of osteophytes, ligament integrity, presence of joint effusion, presence of synovitis and presence of intra-articular loose bodies were cut after the first round due to low median scores ( $\leq 2$ ). In the second round only presence and severity of BMO, presence of subarticular cysts, cartilage integrity and presence of chondral defects had a median score of at least four. Of the features that were included in the second round all features except the extent of BMO and presence of cysts demonstrated a reduction in standard deviation indicating an improved consensus. The level of consensus for the subtalar joint was less than the tibiotalar joint with higher standard deviations for all features across both rounds of the survey.

## DISCUSSION

Other than for the features already mentioned that demonstrated an increase in standard deviation, nearly all features demonstrated a decrease suggesting a gain in consensus between the two rounds. It was felt that the features to be excluded were quite clear when studying both the tables and the multiplot charts. It is possible that a third round may have helped in gaining further consensus but as all features that were above the threshold score of 4, except presence of osteophytes at the tibiotalar joint, demonstrated reduction or static standard deviation the results from the two rounds was accepted. From a practical point of view, the first two survey rounds had taken longer than initially planned for all responses to be gathered and conducting a third round may have been difficult. Given that the features to be included appeared clear from the two rounds no third round was undertaken.

A limitation of the Delphi survey with relation to osteophytes is the exact phrasing of the questions. Both the presence and severity of osteophytes were features included in the Delphi survey. In retrospect, scoring osteophytes as merely present or absent for the whole joint appears rather limited compared

to other systems. Other MRI scoring systems recorded osteophytes at pre-defined locations as well as osteophyte severity that was assessed on size criteria. Although the severity of osteophytes was not deemed to be important based on the Delphi results, the questions did not address if the location of osteophytes was important. The use of a scoring system in a longitudinal study that records the mere presence or absence of osteophytes is rather insensitive compared to a system that records their presence at different sites and possibly even their size. Having said this, the SHOMRI system excluded osteophytes completely and stated that this was done as these are better assessed and followed up on radiographs. A further possible limitation is in the choice of threshold boundaries set for progression to round 2 and beyond. These were arbitrary and the choice for progression from round 2 to reliability testing is in retrospect stringent. Decreasing this threshold level to 3 would have resulted in the inclusion of number of osteophytes, effusion and ligament integrity.

Although one expert was from another institution the remainder were all from the NNUH. Despite the mix of specialties included this potentially introduced an institutional bias to the survey.

Although the Delphi survey concerned both the tibiotalar joint and subtalar joint the aim of this project centred on the ankle joint proper and the remainder of this thesis will focus on the tibiotalar joint. The features taken forward for consideration of inclusion in the MRI scoring system based on the Delphi results are presence and severity of BMO, degree of cartilage degeneration, presence of chondral defects, presence of subarticular cysts, presence and severity of bone attrition and the presence of osteophytes.

## 5. Reliability Study

### MATERIALS AND METHODS

#### STUDY DESIGN

This was a retrospective reliability study of a new MRI scoring system for osteoarthritis of the ankle conducted at the Norfolk and Norwich University Hospital.

#### ETHICS

The project protocol was submitted to an ethics review panel for proportionate review. The Lancaster Research Ethic Committee granted ethical approval on the 22<sup>nd</sup> of February 2016. IRAS project ID 198605, REC reference 16/NW/0152 (appendix 6). The project was granted full NHS permission for research at the NNUH by the research and development department on the 29<sup>th</sup> of April 2016 (appendix 7).

#### PATIENT SELECTION

Cases were eligible for inclusion if there was an MRI ankle examination on the NNUH Picture Archiving and Communication System (PACS) between the 1<sup>st</sup> of January 2012 and 31<sup>st</sup> of December 2015, there was a preceding ankle radiograph within 4 months of the MRI examination and patients were over the age of 18. Exclusion criteria were similar to other MRI scoring systems including; inflammatory arthritis, previous surgery to the ankle, recent trauma, bone tumour in that limb, haemoglobinopathy, haemachromatosis or any neurological condition limiting function e.g. hemiplegia following stroke.



Consecutive cases of ankle MRI examinations between the 1<sup>st</sup> of January 2012 and 31<sup>st</sup> of December 2015 were screened and assessed against inclusion and exclusion criteria. Ankle radiographs were consensus scored by two radiology trainees with two and four years' experience in reporting appendicular radiographs in the role of a radiologist (SHA and SL). The modified KL score as outlined by Kraus was used. Cases that met the inclusion criteria were included until there were 10 examinations in each of the five KL groups (n=50). Cases were then assigned a unique identifier code, anonymised and sent back to PACS in the anonymised format with only the unique identifier code present on each examination for both plain radiograph and MRI although note that the MRI and radiograph were not grouped in the same packet, i.e. a scorer could not access the radiograph if scoring the MRI. Patients were not included or excluded based on which of the MRI scanners they were examined on and therefore could have been scanned on a 1.5T or 3T MRI machine.

### MR IMAGING ANKLE PROTOCOL

At the Norfolk and Norwich University hospital the standard protocol for imaging the ankle joint includes a T1 weighted TSE sagittal sequence, T2 weighted TSE fat suppressed sagittal sequence, proton density weighted coronal sequence, T2 weighted STIR coronal sequence and T2 weighted TSE axial sequence. Other sequences, such as post contrast imaging may be used if the clinical concern merits use e.g. post-contrast imaging.

Ankle MRI examinations are usually performed on one of two 1.5T GE MRI machines or a 3T GE machine. Patients were only eligible for inclusion if the full protocol was completed. The full protocol parameters for each machine are outlined in appendix 8.

## SAMPLE SIZE

### Method

A sample size of 50 was selected based on sample size calculations, considerations regarding underlying marginal prevalence of disease and feasibility. Tables outlined by Sim et al and nomograms outlined by Hong et al were used with the assumption of an underlying equal marginal prevalence of disease (93,94). The sample of 50 allowed an equal number of 10 examinations for each KL grade zero to four to be taken forward for MRI scoring.

### Justification

In the scoring systems previously discussed the sample sizes varied significantly from  $n=19$  in WORMS and  $n=20$  in MOAKS, to  $n=109$  in the system outlined by Park. No justification was given in testing these systems for the sample size used. Sample size calculations for reliability studies are not routinely dealt with in core medical statistic texts but there is information within the statistics research literature that deals with these issues. The mathematical justification in these papers is beyond the remit of this thesis but fortunately, tables are provided. They involve conducting a hypothesis test and predefining a minimum level of accepted kappa and a kappa to detect;  $H_0$  and  $H_1$ . Note that these values can differ between variables. In addition, an estimate of the underlying marginal prevalence of disease is also needed. Sim et al provide tables for determining sample size for a dichotomous variable with varying option for underlying marginal variance and required power (93). As in the case of the radiographic scoring and cartilage scoring where there are multiple grades this is a little more complex but Hong et al has outlined nomograms again enabling selection of sample size dependent on varying marginal prevalence and value of  $H_0$  and  $H_1$  (94).

Although in theory the sample size should be determined solely by the sample size calculation, in reality it is also important to be aware of what is feasible. For example, to detect a kappa of 0.5 where  $H_0=0.4$  and a power of 90% with equal marginal prevalence for a dichotomous variable requires a sample size nearing 700 while to detect a kappa of 0.8 where  $H_0=0.4$  for a five-point rating scale with equal marginal prevalence at a power of 80% requires a sample of 13. There is clearly a trade-off between the level of kappa needed to detect and the difference permitted between  $H_0$  and  $H_1$ ; the smaller the difference leading to increasingly large sample sizes.

At this point, it is worth noting that the number of ankles taken forward for use in the study does not correspond directly to the sample size. For osteophytes where they are only scored once per ankle the number of ankle radiographs and MRI examinations included equals the number of ratings for reliability assessment between readers. This also applies to total joint scores for BMLs, subchondral cysts, BMO and cartilage totals. For individual scoring of these components on the zonal basis there will be 16 times the number of ankle joints included owing to the number of zones. As mentioned and discussed in the statistical methods section, the determination of  $H_0$  and  $H_1$  may not necessarily be the same for all items and depends on what value of kappa to detect is considered appropriate for that individual item as well as what null hypothesis is appropriate.

A sample size of 50 ankle examinations would give 50 data points for testing for osteophytes and total joint scores and 800 data points for zonal based assessment of variable.

For osteophytes with a dichotomous outcome of present or absent, a sample size of 31 is required to detect a kappa of 0.61 or above where  $H_0=0.2$  assuming equal marginal prevalence and a power of 80%. The reasoning behind this choice is that 0.61 and above is classed as “substantial”, “good” and

“moderate” agreement as per Landis and Koch, Altman and Hugh respectively (95,96).

The analysis and sample size determination of cartilage integrity, BMLs, BMO and subchondral cysts was based on an assumption of equal marginal prevalence of disease. For zonal based scoring of BML, BMO and cysts the kappa to detect was defined as 0.61 with  $H_0=0.4$  and a power of 80%. This gives a sample size of 53. Recall that for these items a sample of 50 ankle gives 800 ranks between observers per item. If the underlying marginal prevalence was heavily skewed despite efforts to control it, this sample size more than accounts for this.

Criteria were the strictest for total joint scores with  $H_0=0.6$  and a kappa to detect defined at 0.81. This is stringent with a null hypothesis rejecting anything classed as even “moderate” agreement by Landis and Koch and to detect a kappa defined as “strong” even by Hugh (almost perfect by Landis and Koch). The reasoning behind this is that inter-rater reliability is likely to be higher for total joint scores than for the zonal based approach. If rater A classes a variable in a zone whilst rater B classes it in the adjacent zone this will be by kappa using the zonal based approach. In the total joint score this will be treated as equal. This therefore allows a more stringent approach.

### Summary

For osteophytes, to detect a kappa of 0.61 with  $H_0$  set at 0.2 a sample of  $n=31$  is required. For the zonal assessment of BML, BMO, cysts and cartilage a sample of  $n=43$  is required to detect a kappa of 0.61 with  $H_0$  set at 0.4. For total joint scores of BML, BMO, cysts and cartilage a sample size of  $n=41$  is required to detect a kappa of 0.81 with  $H_0$  set at 0.6. A sample size of 50 examinations was therefore considered appropriate. Although it would have been possible to use 41 examinations, a sample of 50 would allow attempting to account for equal marginal prevalence of disease across baseline KL scores 0–4.

## PATIENT DEMOGRAPHICS

Patient demographic information is displayed in the table below for the total sample of 50 patients as well as a breakdown of demographic information for different KL grades.

	Mean age	Age Range	M:F
Total sample	54	23–83	27:23
Grade 0	35	23–58	4:6
Grade 1	49	34–66	8:2
Grade 2	61	33–82	3:7
Grade 3	57	33–75	6:4
Grade 4	69	50–83	6:4

Table 5. Patient demographics by KL grades

## READERS

Readers 1 and 2 are two radiology trainees (SHA and SL) who have four and two years' experience respectively as radiologists. Readers 3 and 4 are two consultant musculoskeletal radiologists who each have in excess of ten years' experience in reporting musculoskeletal examinations at consultant level (AT and JC).

## RADIOGRAPHIC SCORING

The Kraus radiographic scoring system including the modified KL score and full zonal based approach was retested in full using a weighted kappa statistic. The radiographs were graded by readers 1 and 2 at a time no sooner than 8 weeks after initial consensus grading for patient selection. Intra-rater reliability was assessed on a sample of 10 radiographs reflecting equal spread across the initial consensus determined KL grades. This was done 8 weeks after individual scoring.

### MRI SCORING

All readers performed the MRI scoring. Inter-rater reliability was assessed between the two experienced radiology consultants and between the two radiology trainees. Intra-rater reliability was tested on a sample of ten MRI examinations that reflected an equal spread across KL grades and were randomly selected for rescoring. Rescoring for this purpose was performed no sooner than 4 weeks before the end of initial scoring for each rater.

### ZONES

#### **Method**

The ankle joint was divided into 16 zones with each variable, except osteophytes, scored in each subregion.

The talar dome was divided into nine equal zones by way of a three-column by three-row grid as outlined by Raiken (97). The nine equal zones were assigned numerical identifiers from one to nine beginning with the most anterior and medial region, proceeding laterally, then posteriorly. Zone 1 was therefore the most anterior and medial, zone 3 the most lateral and anterior, zone 5 represents the middle region of the talar dome and zone 9 represents the most posterolateral of the posterior three zones.

The medial and lateral aspects of the talus represent zones 10 and 11 respectively. There was no further subdivision from anterior to posterior of these zones. The distal tibial articulation is divided into three zones from medial to lateral representing zones 12, 13 and 14. Zone 12 therefore articulates with zones 1, 4 and 7 of the talus. Zone 13 articulates with zones 2, 5 and 8. Zone 14 articulates with zones 3, 6 and 9. The medial malleolus represents zone 15, adjacent to zone 10 of the talus. The medial aspect of the distal fibula represents zone 16, adjacent to zone 11 of the talus.

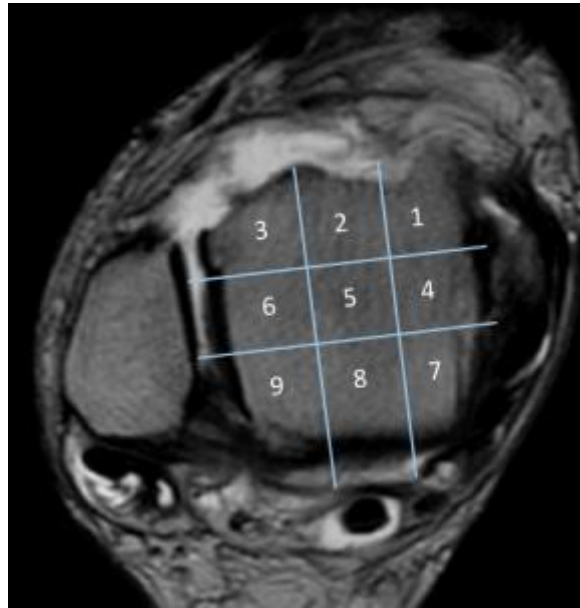


Figure 12. Axial MRI section through a normal ankle outlining the talar dome divisions from zones 1–9.

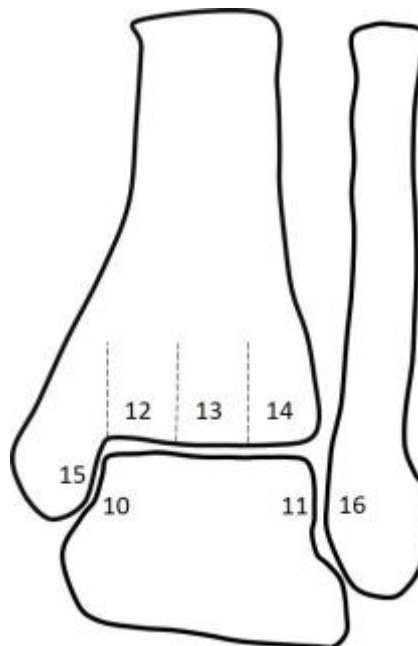


Figure 13. Diagram outlining the positions of zones 10–16

### Justification

Previous described scoring systems divide the joint of interest into zones that are scored individually, dependent on the feature being scored. These subdivisions can appear complex and in many of them there appears to be no justification given for the division used.

In the SHOMRI system, the zones were based on geographic zones outlined by the Arthroscopic Society of North America with a maximum of 10 zones covering the femur and acetabulum, although some features were only scored at some sites. In the HOAMS system, there were 9 zones for assessment of cartilage and 15 zones for subchondral bone marrow assessment although there is no clear justification described for this.

For the knee joint the WORMS system divided the knee into 15 zones and the MOAKS system 14 zones whilst BLOKS and KOSS systems simplified this somewhat dividing into a maximum of 9. Park et al had a more simplified approach dividing the knee into 3 regions, medial, lateral and patellofemoral compartments.

The radiographic grading system outlined by Kraus et al divides the tibiotalar joint in to medial and superior for assessing joint space narrowing on the anterior view, and anterior and posterior on the lateral radiographic view. Subtalar joint space narrowing is assessed on the lateral view only and the subtalar joint is not subdivided. Joint space narrowing at the talofibular joint was also assessed. Osteophytes are assessed at a number of set locations around the ankle joint: medial and lateral tibia, medial and lateral talus, distal fibula, anterior and posterior tibia, anterior and posterior talus, and subtalar joint.



With respect to osteochondral lesions of the talar dome some authors describe the location simply as medial, lateral or central (98). In a study to evaluate the incidence of osteochondral lesions of the talar dome by location, Raiken et al divided the talar dome further into a three by three grid to attempt to better characterise lesion location with zone 1 representing the most anteromedial location and zone 9 representing the most posterolateral location (97). The medial talar dome was more frequently involved than the lateral dome. In the anteroposterior plane, the mid dome was more frequently involved than the anterior or posterior dome. Overall zone 4 was the most frequently involved zone followed by zone 6. This means that if more simplistic four square grid was employed most lesions could occur at the junction of the anterior and posterior quadrants, possibly causing difficulty accurately placing the lesion reliably. Although a three by three grid could therefore be advocated for classifying osteochondral lesions, for other features such as cartilage integrity the zones may prove too small with these lesions potentially covering multiple zones. In attempting to grade BMLs in particular, such small zones may be particularly problematic as the other grading systems stratify severity by extent of subregion involved. The alternative would be to record a positive or negative for every subregion the BML involves with the more zones involved indicating more severe disease. On discussion with the foot and ankle surgeons at the NNUH a more simplistic division of the talar dome into medial, central and lateral zones was preferred (99). Furthermore, whatever position the ankle is in the medial, central and lateral zones of the talar dome will always be opposed, giving further merit for this way of division. If data were recorded in a three by three grid system for the talar dome it would be possible to adjust at a later stage to compute reliability for a simplified three-zone system (medial, central, lateral). This will allow comparison of the different methods of division to see if there is any increase in reliability of the simplified system versus the more detail given by the nine zone approach.

For the MRI scoring, a modification on the radiographic division by Kraus was proposed. The tibiotalar joint was divided into medial, superior and talofibular components for scoring all articular components with further subdivision of the talar dome into a 9-zone grid system. The distal tibial articulation with the talar dome was divided into corresponding medial, central and lateral zones. This gave 16 separate zones at the ankle to be used in the proposed scoring system as detailed above.

### OSTEOPHYTES

#### **Method**

Osteophytes were recorded as a binary outcome of present or absent for the examination.

#### **Justification**

The Delphi survey did not specifically address the perceived importance of location or size of osteophytes. Therefore, osteophytes were only recorded as present or absent for each examination.

### BONE MARROW LESIONS, BONE MARROW OEDEMA AND SUBARTICULAR CYSTS

#### **Method**

BMLs, BMO and subarticular cysts were recorded as present or absent in each of the 16 zones. BMLs were defined as any subchondral signal abnormality and therefore included both subchondral cysts and BMO. BMO, as in the previously described scoring systems, was defined as ill-defined subchondral high signal on fluid sensitive sequence. Subchondral cysts were defined as well defined areas of high signal on fluid a sensitive sequence.

## Justification

Although the Delphi survey concluded that BMO was important to include for both its presence and severity it did not address the size or degree of involvement for subchondral cysts. Recording in this way allowed testing of reliability between observers for the presence of BMLs, BMO and cysts as well as the severity of involvement by way of calculating a total score for each joint for each item giving a possible maximum score of 16 for a joint.

## CARTILAGE INTEGRITY AND OSTEOCHONDRAL DEFECTS.

### Method

Cartilage integrity was graded on a six-point scale with a score recorded for each zone. If a cartilage lesion spanned multiple zones a score was recorded for each zone for the degree of involvement in that specific zone. The system used for grading cartilage integrity is outlined below.

Grade	Description
0	Normal cartilage
1	Abnormal signal of morphologically normal cartilage
2A	Superficial partial thickness cartilage defect <50% of total articular thickness.
2B	Superficial partial thickness cartilage defect >50% of total articular thickness.
3	Full thickness cartilage defect
4	Chondral injury with a bony component

Table 6. Cartilage grading system used in reliability study

It represents a modified Noyes system of cartilage grading for MRI (65,100). Versions of the Noyes system were also used in the Park and KOSS systems (81,86).

### Justification

Cartilage scoring in existing scoring systems have varying complexity from eight grades in WOMBS to the far more simplistic three-point scale in SHOMRI.

Although there are grading systems proposed for grading cartilage integrity on MRI, many are adapted from arthroscopy grading systems, none have been for specifically tested for reliability and reproducibility at the ankle joint. In Saifuddin's core musculoskeletal radiology text, the modified Noyes system is outlined as a proposed method of grading cartilage integrity in the ankle (65,101). This is an MRI cartilage grade based initially on arthroscopy, adapted for MRI by Kijowski et al, and is outlined below (65). Although Saifuddin's text recommends the modified Noyes system it has not been tested on the ankle joint and the evidence cited is a modified Noyes system for assessing cartilage at the knee (65).

Grade	Description
0	Normal cartilage
1	Abnormal signal of morphologically normal cartilage
2A	Superficial partial thickness cartilage defect <50% of total articular surface thickness.
2B	Superficial partial thickness cartilage defect >50% of total articular surface thickness.
3	Full thickness cartilage defect

Table 7. Modified Noyes grading system

Park et al used a modified Noyes classification for cartilage grading and the grading system used in KOSS is very similar with only the exclusion of grade 1. This version of the score was actually the initial MRI modification of the Noyes arthroscopic score outlined by Recht et al which did not include a grade 1 (instead jumping from grade zero to 2A) which reflected softened but intact articular cartilage at arthroscopy (100). It did include grades 3a and 3b which referred to full thickness chondral loss normal underlying bone contour and full thickness chondral loss with bony injury which is not included in the version outlined by Kijowski. The version of the Noyes score therefore used in this reliability study is a combination of that initially outlined by Recht with the addition of a grade 1 score as proposed by Kijowski (65,100).

The other main MRI cartilage scoring system, that is similar to the modified Noyes classification, is the modified Outerbridge scale. This is also initially based on arthroscopy findings and is outlined below.

Grade	Description
I	Focal areas of hyperintensity with normal contour
II	Blister-like swelling/fraying of articular cartilage extending to surface
III	Partial thickness cartilage loss with focal ulceration
IV	Full thickness cartilage loss with underlying bone reactive changes

Table 8. Modified Outerbridge grading system

In addition to the depth of lesion, some of the existing scoring systems grade the surface extent of cartilage loss in a specific subregion, as either an additional score or incorporating surface extent of damage into the cartilage grade itself. BLOKS for example scores depth of cartilage damage as merely absent, partial loss of complete loss with additional scoring for the size of the lesion and another score for the size of lesion that is full thickness. A similar system is used in the MOAKS score. This method of scoring would classify Noyes grades 1, 2A and 2B into a score of 1, with additional scoring for percentage of subregion affected overall and percentage subregion suffering from a full thickness loss. Out of all the scoring systems the BLOKS system along with SHOMRI have the simplest approach to classifying depth of cartilage involvement but BLOKS also accounts for area of subregion involved.

Osteochondral defects are focal areas of chondral damage that extend to involve injury of the adjacent subchondral bone. Only the KOSS scoring system incorporates a specific component for any grading of osteochondral lesion. The Park scoring system classes a grade 3 Noyes as a full thickness cartilage defect with bony involvement therefore including a bony component or osteochondral injury.

For simplicity, the presence of bony involvement was classified at the most severe end of the scale of cartilage involvement and this was classified as a grade 4 in the Noyes system. The ICRS score (International Cartilage Repair Society) classifies bony involvement in a similar way with the most severe score, grade 4, representing an osteochondral lesion (102,103).

### BONE ATTRITION

Bone attrition was only graded in the WORMS system and scored as only present or absent in the HOAMS system. In WORMS it demonstrated the weakest reliability values of all included features and was subsequently excluded from the MOAKS system. None of the other scoring systems graded bone attrition. On discussion with musculoskeletal radiologists at the NNUH there was some uncertainty regarding the term bone attrition, despite the fact it had passed through the Delphi survey. Although often mentioned when discussing OA in the research literature, the distinction between full thickness chondral loss with bony involvement and bone attrition was not clear and bone attrition was not felt to be a term used routinely in everyday clinical practice. On discussion with two experienced professors of musculoskeletal radiology, it was felt that bone attrition should not be included separately in the scoring system as it was not a term used routinely outside a research setting and demonstrated poor reliability results in the systems in which it was included (104).

## STATISTICAL ANALYSIS

### Methods

The main measure of inter and intra-rater reliability presented in this study is the weighted kappa statistic and, where appropriate, the unweighted kappa statistic. Kendall's Tau and Spearman's rho correlation values are also presented when comparing radiographic and MRI scores. A description and justification including the benefits and drawbacks of these statistics is presented in the statistical methods section.

All statistical analyses in this thesis was performed using the R programming language within the R environment for statistical computing using the base package with the additional "irr", "psych", and "ggplot2" packages (105–108) .

## STATISTICAL METHODS

### Justification

#### The kappa statistic and observer agreement

The most straightforward method of assessing agreement between observers or observations is to calculate the percentage agreement ( $P_0$ ). This is the percentage of scores or subjects observed as the same on two occasions by two observers. The main drawback of this method is that it does not account for the level of agreement that is expected by chance ( $P_e$ ).

For categorical data, Cohen's Kappa ( $\kappa$ ) is a measure of agreement that adjusts for agreement that would be expected by chance. Cohens Kappa was initially created for the study of agreement between two equally skilled observers. It is

the observed agreement in excess of the chance agreement. The proportion of units which would be expected to agree by chance is denoted  $P_e$ .

Kappa is the proportion of agreement between two raters following removal of chance agreement.

$$\kappa = \frac{P_0 - P_e}{1 - P_e}$$

When  $P_0 = 1$ , kappa has a maximum value of one denoting perfect agreement. When  $P_e = P_0$ , kappa is zero denoting the level of agreement is no better than expected by chance. Negative kappa values demonstrate agreement is worse than expected by chance. Only value between 0 and 1 have useful meaning (96).

### Interpretation of kappa

There are no objective criteria for judging intermediate values. Probably the most widely used interpretation of kappa are values suggested by Landis and Koch (77). These are slightly adapted by Altman. Both are listed below (96).

Value of kappa	Landis and Koch (77)	Altman (96)
<0.00	Poor	-
0.00–0.20	Slight	Poor
0.21–0.40	Fair	Fair
0.41–0.60	Moderate	Moderate
0.61–0.80	Substantial	Good
0.81–1.00	Almost perfect	Very good

Table 9. Suggested interpretations of kappa

A criticism of this interpretation of kappa is that relatively low values of kappa can be deemed to have moderate agreement. In a clinical setting, such interpretation may be inappropriate. With a percentage agreement for



example of 55%, it can be seen that 45% of the data is 'faulty'. 45% disagreement in a clinical setting is likely to be unsuitable. Although the corresponding kappa value will be lower, categorizing 0.41–0.60 as moderate agreement implies 0.41 itself is classified as moderate agreement. McHugh suggests a stricter interpretation of kappa where values below 0.6 are classified as weak at best (95). However, unless otherwise stated, the Landis & Koch criteria will be referred to in this as it is the most widely used criteria.

Value of kappa	Level of agreement
0–0.20	None
0.21–0.39	Minimal
0.40–0.59	Weak
0.60–0.79	Moderate
0.80–0.90	Strong
Above 0.90	Almost perfect

Table 10. Suggested interpretation of kappa as outlined by McHugh (95)

### Weighted kappa

With ordinal categorical data kappa does not distinguish between the magnitude of disagreement between observers. For example, on a scale of disease severity of mild, moderate and severe the difference in disagreement between a mild and a severe rating is greater than between a mild and a moderate rating. Disagreement by two points on the scale is more serious than by one point on the scale.

To account for the amount of disagreement weightings can be attached to kappa. Weighted kappa penalises disagreement dependent on magnitude while unweighted kappa treats all disagreement equally. Unweighted kappa is therefore not suitable for ordinal scales (93). A number of weighting systems are available with quadratic weighting often used (93) where weights are

proportional to the square of the number of categories apart. Linear weights can also be used which are proportional to the number of categories. It is also possible to weight for agreement rather than disagreement where perfect agreement has a weighting of 1.0 with smaller weights for differing degrees of disagreements, the smaller assigned to the largest disagreement. The determination of weights is issue and scenario dependent, although in practice it is likely that the default weights of the statistical software package are used (96). Having said this the quadratic weighting system is most commonly used and will be used in this thesis.

Kappa can be assessed for statistical significance through hypothesis testing. As a negative kappa does not usually have any meaning a 1-tailed hypothesis test is usually sufficient if  $H_0=0$  (93). In practice specifying a value of zero for kappa in the null hypothesis may not be meaningful as agreement will usually be better than expected by chance in a clinical setting (93). The value of kappa used in the null hypothesis can be set at a higher level e.g. 0.4 or 0.6 with any lower value deemed unacceptable in the given setting. If the null hypothesis is specified at a value greater than zero, a two-tailed hypothesis test is preferable as cannot assume the reliability is necessarily better than the pre-defined threshold for importance. The minimum acceptable value for kappa will depend on the clinical context. Unlike percentage agreement that is a direct measure, kappa is an estimate of inter-rater agreement and confidence intervals are therefore useful.

### Problems with kappa

In addition to the care needed with the descriptive characteristics of kappa strength and the need for weighting in the case of ordinal data, there are some other considerations to make when interpreting kappa.

The kappa value is influenced by the marginal prevalence of the attribute (the trait prevalence in the study population). The kappa statistic alone is appropriate if the marginal totals are relatively balanced. If the prevalence of given responses is very high or very low the resultant value for kappa may be low even when the observer proportion of reliability is high. This is sometimes referred to as the kappa paradox (93,109). Interpretation based solely on the value of kappa with this circumstance may be misleading. Reporting simply the percentage agreement and  $P_e$  is an option. MOAKS reported percentage agreement alongside low kappa values when this paradox was suspected.

### Correlation

The secondary aim of this research is to compare the severity of scores on MRI with that on plain radiographs e.g. the KL score versus total joint cartilage score. The reason to do this is to assess if the severity of disease on plain film correlates with that demonstrated on MRI and if so to what degree. E.g. does a plain radiograph score of 0 or 1 mean no to little BMO versus severe BMO with a radiographic score of 4. Clinically this may give an indication of whether performing an MRI for assessment of OA is worthwhile if the plain radiograph can give an indication of the severity of cartilage damage or severity of BMLs for a given radiographic score.

The data that will be assessed is the radiographic score (0–4 grade) versus the total component score the joint on MRI. Total BML score will have a maximum

value of 16 where total cartilage score will have a maximum value of 80.

Although these numbers, especially in the case of total cartilage score, are quite large, they will be treated as categorical ordinal data as there is a maximum value possible and treating as continuous data for these purposes is not appropriate.

For ranked categorical data as in the case of both the radiographic and MRI score, the Spearman's rank correlation is perhaps the most widely recognised method to assess for correlation in this case. The other measure of correlation to be discussed in this section is Kendall's Tau.

Both Spearman's rank (or rho) and Kendall's Tau are used to assess association based on the ranks of the data. Spearman's rho is based on the deviations in the ranks of the data. It is calculated by

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

where  $d_i$  is the rank difference between the observations and  $n$  is the number of observations.

Differing from Spearman's rho, Kendall's Tau is not affected by the difference between the ranks but only by whether ranks differ between variables. Each pair is assessed to be concordant (ordered the same way), discordant (ordered in opposing ways), or equal/not ordered and therefore tied. Kendall's Tau is the difference between the proportion of concordant and discordant pairs.

Although Spearman's rho is more commonly used, it is more sensitive to errors and discrepancies in the data than Kendall's Tau (110). Kendall's Tau will usually result in slightly smaller coefficient values than those given by the Spearman's rho but in almost all situations the values of Kendall's Tau and Spearman's rho are very close and "invariably lead to the same conclusions" (110).

Given the above, Kendall's Tau will be presented as the main measure of correlation for assessing for association between Kellgren Lawrence score and total joint scores for cartilage, BML, BMO and subchondral cysts. For completeness, and as it is more commonly used the Spearman's rho is also presented.

## RESULTS

### RADIOGRAPHIC SCORING

#### Modified Kellgren-Lawrence score

The inter-rater reliability results for the modified KL score assessed on the mortise and lateral view are detailed in the table below (table 11) in the form of a weighted kappa in addition to the intra-rater results. The subtalar KL results are also listed for completeness as the Kraus system covers both tibiotalar and subtalar. Intra-rater reliability results are also presented.

	Inter-rater†	Intra-rater†	
	Readers 1 & 2	Reader 1	Reader 2
Mortise	0.86 (0.78,0.93)	0.55 (0.04,1.00)	1.00
Lateral	0.84 (0.76,0.92)	0.60 (0.20,1.00)	1.00
Subtalar	0.66 (0.47,0.85)	0.67 (0.46,0.87)	0.89 (0.66,1.00)
†Weighted kappa. Inter-rater p<0.01. Intra-rater p<0.05			

Table 11. Modified KL Score

The Kraus KL scoring system demonstrated “almost perfect” agreement between the raters for the ankle joint proper on the mortise and lateral radiographic projections. The Kraus KL score for the subtalar joint also demonstrates “substantial” agreement. Intra-rater agreement for reader 2 was “perfect” for the ankle joint proper and “almost perfect” for the subtalar joint. Reader 1 had “moderate” reliability for the ankle joint proper and “good” for the subtalar joint.

## Osteophytes

The zones assessed for osteophyte severity on the mortise view all demonstrated “substantial” agreement (table 12). On the lateral view, there was “substantial” or “perfect” agreement for most components. Posterior talar osteophytes only had “moderate” agreement but the confidence interval is wide extending into the range of “fair” agreement.

	Inter-rater	Intra-rater	
	Readers 1 & 2	Reader 1	Reader 2
<i>Mortise View</i>			
Medial Tibial	0.71 (0.57,0.85)	0.73 (0.51,0.96)	0.91 (0.76,1.00)
Lateral Tibial	0.77 (0.62,0.92)	0.83 (0.59,1.00)	0.83 (0.72,0.95)
Medial Talar	0.71 (0.52,0.91)	0.77 (0.54,1.00)	1.00
Lateral Talar	0.71 (0.67,0.91)	0.94 (0.87,1.00)	0.62 (0.28,0.95)
Fibular	0.77 (0.60,0.94)	0.88 (0.77,0.99)	0.95 (0.86,1.00)
<i>Lateral View</i>			
Anterior Tibial	0.83 (0.69,0.97)	0.76 (0.69,0.82)	1.00
Posterior Tibial	0.90 (0.83,0.96)	0.82 (0.55,1.00)	0.84 (0.62,1.00)
Anterior Talar	0.66 (0.52,0.81)	0.67 (0.33,1.00)	0.93 (0.84, 1.00)
Posterior Talar	0.51 (0.23,0.79)	0.91 (0.78,1.00)	0.81 (0.56, 1.00)
Subtalar	0.63 (0.41,0.84)	0.90 (0.80,1.00)	0.83 (0.69,0.97)
Weighted kappa. Inter-rater $p < 0.01$ . Intra-rater $p < 0.05$			

Table 12. Radiographic Osteophyte scoring

Results for intra-rater reliability for osteophytes were “substantial” or “almost perfect” for all zones for both readers with the lower level of the confidence intervals extending in to the “moderate” agreement range for some items and into the “fair” agreement range for the lateral talar zone for reader 2 and the anterior talar zone for reader 1.

## Joint space narrowing

For joint space narrowing, inter-rater agreement was “substantial” or “almost perfect” for all components (table 13). The lower level of the confidence intervals lie within the “substantial” range for all items except anterior talar and subtalar joint space narrowing where they extend only into the “moderate” agreement range.

	Inter-rater	Intra-rater	Intra-rater
	Readers 1 & 2	Reader 1	Reader 2
<i>Mortise View</i>			
Medial Tibiotalar	0.81 (0.70,0.93)	0.90 (0.77,1.00)	0.88 (0.76 1.00)
Superior Tibiotalar	0.81 (0.69,0.93)	0.79 (0.56,1.00)	0.92 (0.83 1.00)
Talofibular	0.80 (0.68,0.92)	0.57 (0.26,0.88)	0.93 (0.80 1.00)
<i>Lateral View</i>			
Anterior Tibiotalar	0.72 (0.58,0.86)	0.90 (0.80,1.00)	0.95 (0.88 1.00)
Posterior Tibiotalar	0.76 (0.61,0.91)	0.69 (0.33,1.00)	0.68 (0.26 1.00)
Subtalar	0.63 (0.41,0.84)	0.84 (0.69,0.99)	0.88 (0.72 1.00)
Weighted kappa. Inter-rater $p < 0.01$ . Intra-rater $p < 0.05$			

Table 13. Joint space narrowing

Intra-rater agreements were “substantial” or “almost perfect” for all items for reader 2 and all items for reader 1 except talofibular joint space narrowing which still demonstrated “moderate” agreement. The confidence intervals for posterior tibiotalar joint space narrowing for both readers and talofibular joint space narrowing for reader 1 were wide with the lower levels extending into the range of “fair” agreement.



### Total Kraus score and modified Kellgren-Lawrence Correlation

Table 14 displays correlation values for the total joint score as per the full Kraus radiographic system against the modified KL score for each reader and the consensus KL score. There is strong positive correlation between the full score and the modified KL score.

Total Kraus score	Reader 1/2 KL		Consensus KL	
	<i>tau</i> <sup>†</sup>	<i>rho</i> <sup>‡</sup>	<i>tau</i> <sup>†</sup>	<i>rho</i> <sup>‡</sup>
Reader 1	0.86 (0.80,0.91)	0.95 (0.90,0.97)	0.86 (0.80,0.91)	0.95 (0.91,0.97)
Reader 2	0.85 (0.79,0.89)	0.94 (0.89,0.96)	0.82 (0.75,0.87)	0.92 (0.85,0.95)

† Kendall's tau. ‡Spearman's rho. p<0.01

Table 14. Correlation for Kraus total score and modified Kellgren-Lawrence score

The corresponding scatter plots for the total Kraus score and Kraus KL score is displayed in figure 14.

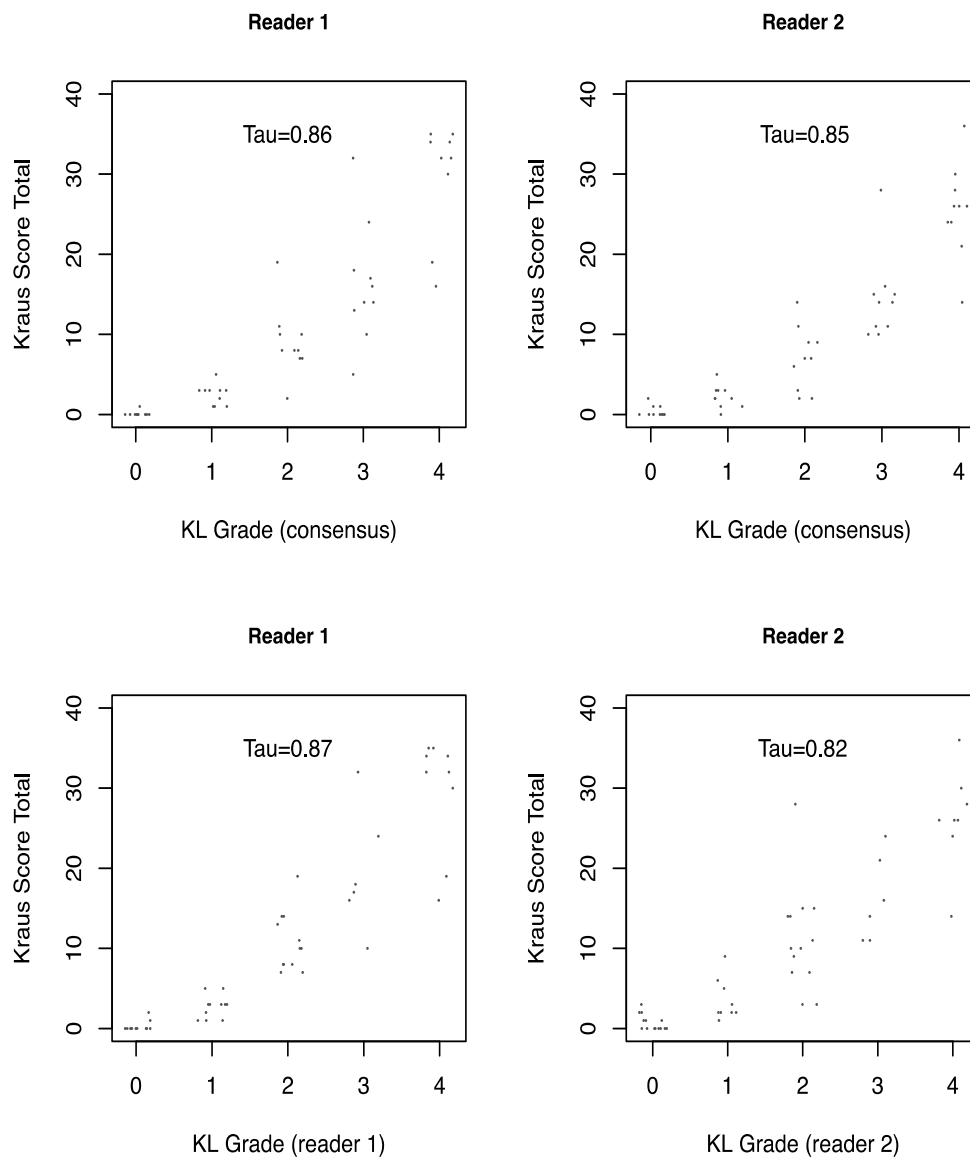


Figure 14. Consensus Kellgren-Lawrence vs total Kraus score scatter plot

## MRI SCORING

### Osteophytes

Amongst the experienced observers there was “almost perfect” agreement on the presence or absence of osteophytes (table 15). The results for the trainee radiologists were not as high but still demonstrated “substantial” agreement although the confidence interval extends into the “moderate” agreement range.

Raters 1 and 2	Raters 3 and 4
0.64 (0.43,0.85)	0.92 (0.81,1.00)
Inter-rater weighted kappa. $p < 0.01$	

Table 15. Osteophytes inter-rater results

Intra-rater reliability was “perfect” for both experienced radiologists and was “substantial” for the junior radiologists although the confidence intervals were wide extending into the fair agreement range (table 16).

Reader 1	Reader 2	Reader 3	Reader 4
0.78 (0.39, 1.00)	0.78 (0.39, 1.00)	1.00	1.00
Intra-rater weighted kappa. $p < 0.05$			

Table 16. Osteophytes intra-rater results

## Bone marrow signal abnormality

### Inter-rater agreement

Inter-rater agreement for the trainee radiologist was “substantial” for zonal based assessment of BMLs, BMO and cysts with confidence intervals extending down to the “moderate” agreement range for cysts but remaining in the “substantial” agreement range for the other two components (table 17).

	Raters 1 and 2	Raters 3 and 4
Bone marrow lesion	0.75 (0.69,0.81)	0.82 (0.77,0.87)
Bone marrow oedema	0.73 (0.67,0.79)	0.81 (0.75,0.86)
Cysts	0.63 (0.48,0.78)	0.73 (0.63,0.83)
Bone marrow lesion Total	0.97 (0.96,0.99)	0.96 (0.93,0.98)
Bone marrow oedema Total	0.97 (0.97,0.99)	0.94 (0.90,0.98)
Cysts Total	0.94 (0.87,1.00)	0.91 (0.86,0.97)
Inter-rater weighted kappa. $p < 0.01$		

Table 17. Bone marrow lesion zonal and total joint inter-rater results

The experienced radiologists demonstrated “almost perfect” agreement for BMLs and BMO with the lower limits of the confidence intervals extending into the “substantial” agreement range. “Substantial” agreement was demonstrated for cysts with the confidence intervals also lying in this agreement range.

Inter-rater agreement for total joint scores for all marrow signal related items agreement was “almost perfect” with confidence intervals also remaining in this range.

### Intra-rater agreement

For zonal based approach the intra-rater agreement was “substantial” or “almost perfect” for all readers for all components except cysts for reader 4 where intra-rater agreement was “moderate” but the confidence intervals was wide with the lower limit extending down to the slight agreement range and below the previously defined null hypothesis value. Although for some of the other items the lower limits of the confidence intervals extended to the “moderate” agreement boundaries they did not extend below the null hypothesis value.

	Reader 1	Reader 2	Reader 3	Reader 4
Bone marrow lesion	0.73 (0.60,0.85)	0.60 (0.45, 0.76)	0.85 (0.75,0.96)	0.70 (0.56, 0.83)
Bone marrow oedema	0.69 (0.56,0.82)	0.62 (0.47,0.77)	0.83 (0.72 0.94)	0.70 (0.56, 0.83)
Cysts	0.89 (0.75, 1.00)	0.79 (0.52, 1.00)	0.81 (0.61, 1.00)	0.51 (0.19, 0.83)
Bone marrow lesion Total	0.97 (0.94,0.99)	0.79 (0.67, 0.92)	0.94 (0.85,1.00)	0.91 (0.78,1.00)
Bone marrow oedema Total	0.97 (0.94, 0.99)	0.93 (0.83, 1.00)	0.92 (0.83, 1.00)	0.91 (0.78, 1.00)
Cysts Total	0.91 (0.77,1.00)	0.88 (0.76, 1.00)	0.97 (0.89, 1.00)	0.77 (0.50,1.00)
Intra-rater. Weighted kappa. $p < 0.05$				

Table 18. Bone marrow lesion zonal and total joint intra-rater results

Intra-rater agreement for all total scores lay in the “almost perfect” range for all items and all readers except BMLs total for reader 2 and cyst total for reader 4 that still demonstrated “substantial” agreement. Again, the lower limits of the confidence intervals for reader 4 for cysts extended into the “slight” agreement range.

## Cartilage

### Inter-rater agreement

The trainee radiologist demonstrated “substantial” agreement for the zonal based cartilage assessment and “almost perfect” agreement for total cartilage joint score with confidence intervals also lying within these categories. The experienced radiologists demonstrated “almost perfect” agreement for zonal based cartilage assessment and total cartilage joint score with confidence intervals not extending below these categories.

	Raters 1 and 2	Raters 3 and 4
Cartilage	0.71 (0.67,0.76)	0.88 (0.85,0.91)
Cartilage Total	0.88 (0.82,0.95)	0.96 (0.92,0.99)
Inter-rater weighted kappa. $p < 0.01$		

Table 19. Cartilage inter-rater zonal and total joint results

### Intra-rater agreement

Intra-rater agreement for cartilage assessment by zones was classed as “almost perfect” for all raters with the lower limits of the confidence intervals extending below to the range of “substantial” agreement. Intra-rater agreement for cartilage whole joint scores was classed as “almost perfect” for readers 1 and 4 with confidence intervals lying within this range. Reader 3 demonstrated “almost perfect” agreement but the lower limit of the confidence interval extended into the range of “moderate” agreement. Reader 2 demonstrated “substantial” agreement but the lower limit of the confidence interval extended into the category of “fair” agreement.

	Reader 1	Reader 2	Reader 3	Reader 4
Cartilage	0.85 (0.70, 0.91)	0.82 (0.75, 0.90)	0.88 (0.81, 0.92)	0.84 (0.77, 0.91)
Cartilage Total	0.95 (0.91, 0.98)	0.76 (0.40, 1.00)	0.81 (0.59, 1.00)	0.94 (0.89, 0.99)
Intra-rater weighted kappa. $p < 0.05$				

Table 20. Cartilage intra-rater zonal and total joints results

### Inter-rater agreement Talar Dome

The table below demonstrates results for inter-rater agreement using the 9-zone grid for the talar dome and the 3-zone division approach for the radiology trainees and the experienced radiologists.

	Readers 1 & 2	Readers 3 & 4
Talar Dome 9 zone	0.74 (0.68, 0.80)	0.89 (0.86, 0.92)
Talar Dome 3 zone	0.79 (0.72, 0.86)	0.85 (0.79, 0.91)
Inter-rater weighted kappa. $p < 0.01$		

Table 21. Talar Dome Inter-rater results

The radiology trainees demonstrated “substantial” agreement using the 9-zone approach with confidence intervals also lying in this range. The 3-zone approach did result in an increase in the kappa value but the interpretation of the kappa value did not change with the result still within the range of “substantial” agreement.

The experienced radiologists demonstrated “almost perfect” agreement using the 9-zone approach with the confidence interval also included in this range. Interestingly the kappa value using a simplified three zone approach actually decreased slightly and although remained within the category of “almost

perfect” agreement the lower limit of the confidence interval dropped just into the category of “substantial” agreement.

#### Inter-rater agreement zones 10–16

Table 22 outlines the results for inter-rater agreement for zones 10 through to 16 for the radiology trainees and experienced radiologist.

Zone	Raters 1 and 2	Raters 3 and 4
10	0.66 (0.45,0.86)	0.87 (0.79,0.95)
11	0.31 (0.07,0.55)	0.91 (0.84,0.98)
12	0.60 (0.37, 0.84)	0.76 (0.61,0.92)
13	0.77 (0.63,0.91)	0.92 (0.82,1.00)
14	0.85 (0.76,0.94)	0.91 (0.84,0.97)
15	0.45 (0.21,0.69)	0.76 (0.56,0.96)
16	0.45 (0.16,0.74)	0.91 (0.82,0.99)
Inter-rater weighted kappa. $p < 0.01$		

Table 22. Zones 10–16 Inter-rater results

The experienced radiologists demonstrated “substantial” or “almost perfect” agreement for all zones 10–16. The inter-rater agreement between the radiology trainees was more variable. The lower limits of the confidence interval for zones 11 and 16 extend into the only slight agreement category with those for zone 12 and 15 fall into the fair agreement category.

#### Modified cartilage score

With sixteen zones for cartilage scoring it was likely that in all but the most severely diseased joint there would be a reasonable number of zones where both readers score the zone as normal. Although it is important to include these ratings in the overall score for cartilage, for both zonal approach and total



joint score as they still represent agreement, they may increase the kappa value obtained at the expense of discrepancies in the positive ratings i.e. non-normal cartilage scores.

If all data points where both readers grade a zone as normal are removed it leaves 365 data points for the radiology trainees and 320 data points for the experienced radiologists. Recall that 800 zones were assessed across 50 ankles so there is a significant number where both readers score the zone as a zero.

Unsurprisingly when the inter-rater agreement is calculated with removal of all 'normal' zones the kappa values drop. For the experienced radiologists the agreement is still regarded as "substantial" with a weighted kappa of 0.72 (0.67, 0.77). For the radiology trainees the fall in kappa value is greater to 0.50 (0.43, 0.57). Essentially the interpretation in level of agreement is a fall of one agreement class for the for both trainee and experienced radiologist if using the Landis and Koch classification.

To assess if the inter-rater agreement could be improved for cartilage integrity with a simplified system some classes were grouped for analysis with different iterations attempted. Recall that the modified Noyes criteria was used for assessment of cartilage integrity with an additional grade added for bony component as detailed in the methods section (table 6)

Taking forward only the positive ratings (in other words removal of all data points where both readers scored a zone zero) inter-rater agreement was assessed using a number of different groupings in cartilage score as outlined in the table below (Table 23). For example in version 1 the Noyes grade 2A and 2B were combined. The most simplified version is version 4 where the grades originally used are combined to give only 3 grades.

Original	0	1	2A	2B	3	4
Cart v1	0	1	2		3	4
Cart v2	0	1	2		3	
Cart v3		0	1		2	3
Cart v4		0	1		2	
Cart v5	0	1	2A	2B	3	
Original is the modified Noyes score with the addition of grade 4						

Table 23. Cartilage grading versions

The inter-rater agreement values for the different cartilage grade groupings tested are outlined in the table below. The kappa values for the original full zonal based cartilage grading and values after removal of data points where both readers score a zone zero are also detailed in the table for comparison.

	Raters 1 & 2	Raters 3 & 4
All data		
Original (n=800)	0.71 (0.67, 0.76)	0.88 (0.85, 0.91)
Cartilage v1	0.70 (0.65, 0.74)	0.85 (0.82, 0.88)
Cartilage v2	0.70 (0.66, 0.75)	0.85 (0.82, 0.88)
Cartilage v3	0.65 (0.59, 0.71)	0.85 (0.82, 0.88)
Cartilage v4	0.67 (0.61, 0.73)	0.85 (0.82, 0.88)
Cartilage v5	0.72 (0.67, 0.76)	0.88 (0.85, 0.91)
Zero grades removed		
Original (n=365, n=320)	0.50 (0.43, 0.57)	0.72 (0.67, 0.77)
Cartilage v1	0.43 (0.36, 0.50)	0.61 (0.55, 0.67)
Cartilage v2	0.40 (0.33, 0.48)	0.59 (0.53, 0.65)
Cartilage v3	0.52 (0.45, 0.60)	0.73 (0.68, 0.79)
Cartilage v4	0.53 (0.46, 0.61)	0.73 (0.67, 0.78)
Cartilage v5	0.49 (0.41, 0.56)	0.71 (0.66, 0.77)
Inter-rater weighted kappa. p<0.01		

Table 24. Inter-rater results for different versions of cartilage scoring for original data and with zero grades removed

When all 800 data points are included there is actually a slight decrease in weighted kappa values for the consultant radiologists for the modified versions except version 5 that gives the same results and confidence intervals. Kappa values for the radiology trainees are similar with a reduction in values across all versions except version 5 that has the same confidence levels with a tiny increment in the kappa value that makes no difference to the interpretation. Interestingly the version 5 cartilage score corresponds to the ICRS score.

When the agreed zero grades are removed only simplified versions 3 and 4 demonstrate an increase in kappa value with the other versions actually leading to a decrease in the kappa value. Although versions 3 and 4 demonstrated an increase in kappa value, this is minimal with no resultant change in suggested interpretation for either trainee or experienced radiologists.

### Surface extent score

Comparing severity of cartilage damage between cases using a total joint score does have a limitation in that multiple possible patterns of disease can give the same score. For example, a total joint score of 16 could be made up of a cartilage grade of 1 through all sixteen zones, or a score of 4 in four zones with the remainder of the zones having normal cartilage. The BLOKS and MOAKS systems incorporated severity scores that accounted for the surface extent of cartilage damage across the joint as well as the extent of full thickness cartilage loss across the joint.

To provide more information than the total joint cartilage score alone, cartilage damage extent scores can be determined from the data. Extent score 1 is the number of zones that display any cartilage damage. Score 2 is the number of zones that display full thickness cartilage damage. Therefore, for both scores the scores can range from zero to sixteen.

The tables below outline the inter and intra-rater reliability results for both score 1 and score 2.

	Raters 1 and 2	Raters 3 and 4
Score 1	0.94 (0.90,0.98)	0.95 (0.89,1.00)
Score 2	0.93 (0.89,0.97)	0.93 (0.92,0.98)
Inter-rater weighted kappa. $p < 0.01$		

Table 25. Inter-rater results for cartilage surface extent grade

	Reader 1	Reader 2	Reader 3	Reader 4
Score 1	0.94 (0.86,1.00)	0.86 (0.60,1.00)	0.90 (0.79, 1.00)	0.97 (0.96,0.99)
Score 2	0.76 (0.54, 0.97)	1	0.73 (0.40, 1.00)	0.84 (0.66,1.00)
Intra-rater weighted kappa. $p < 0.05$				

Table 26. Intra-rater results for cartilage surface extent grade

Inter-rater reliability for both score 1 and score 2 demonstrates “almost perfect” reliability with the confidence intervals also lying in this range. Intra-rater reliability was “almost perfect” for all readers for score 1 although confidence intervals extended into the “moderate” and “substantial” ranges for reader 2 and 3 respectively. For score 2, intra-rater reliability was “almost perfect” for readers 2 and 4, and in the “substantial” range for readers 1 and 3 although confidence intervals were wide, extending into the “fair” range for reader 3.

## Distribution of disease

The table below outlines the distribution of MRI disease severity for BML total, cartilage extent scores and cartilage total joint score.

Score range	0–3	4–7	8–11	12–16
BML total	64	20	9	9
Cartilage score 1	66	12.5	9	12.5
Cartilage score 2	79	8	5.5	4.5
Score range	0–20	21–40	41–60	61–80
Cartilage total	73	11.5	11.5	4
Figures are percentages				

Table 27. Distribution of MRI disease severity

Despite equal spread of radiographic severity grades being included in reliability testing the MRI scores for the sample tested are distributed towards the lower end of the severity scale with the higher severity scores occurring less frequently in comparison. Note that the score ranges used in the table are arbitrary and used for simplicity of presentation.

## Correlation

Correlation values for the KL score compared to the total joint BML score, BMO score, cyst score, total cartilage score and cartilage extent scores 1 and 2 are detailed in the tables below for each reader. The primary measure of correlation used is Kendall's Tau as discussed in the methods section although the Spearman's rho is presented alongside as this statistic is often more commonly presented.

	BML*	BMO**	Cyst
Reader 1	0.68 (0.57,0.78) † 0.79 (0.63,0.89) ‡	0.67 (0.51,0.76) † 0.77 (0.61,0.88) ‡	0.35 (0.10,0.55) † 0.40 (0.10,0.62) ‡
Reader 2	0.69 (0.57,0.78) † 0.80 (0.70,0.91) ‡	0.67 (0.54,0.77) † 0.77 (0.61,0.88) ‡	0.31 (0.11,0.52) † 0.35 (0.05,0.56) ‡
Reader 3	0.65 (0.49,0.77) † 0.75 (0.62,0.88) ‡	0.65 (0.47,0.78) † 0.75 (0.55,0.88) ‡	0.47 (0.27,0.61) † 0.54 (0.36,0.71) ‡
Reader 4	0.70 (0.58,0.81) † 0.80 (0.67,0.90) ‡	0.70 (0.55,0.81) † 0.80 (0.65,0.91) ‡	0.52 (0.37,0.66) † 0.61 (0.45,0.75) ‡
*Bone marrow lesion. **Bone marrow oedema. †Kendall's tau. ‡Spearman's rho. p<0.05			

Table 28. Bone marrow lesion and plain radiograph correlation

	Cartilage total	Score1	Score2
Reader 1	0.73 (0.60,0.81) † 0.85 (0.74,0.92) ‡	0.74 (0.62,0.84) † 0.84 (0.71,0.92) ‡	0.66 (0.52,0.75) † 0.75 (0.61,0.85) ‡
Reader 2	0.71 (0.57,0.80) † 0.82 (0.72,0.91) ‡	0.70 (0.54,0.81) † 0.80 (0.60,0.90) ‡	0.67 (0.55,0.79) † 0.77 (0.61,0.88) ‡
Reader 3	0.69 (0.56,0.81) † 0.79 (0.65,0.89) ‡	0.73 (0.60,0.82) † 0.83 (0.71,0.92) ‡	0.71 (0.64,0.80) † 0.80 (0.66,0.88) ‡
Reader 4	0.78 (0.68,0.85) † 0.89 (0.79,0.94) ‡	0.76 (0.62,0.83) † 0.86 (0.76,0.91) ‡	0.72 (0.63,0.80) † 0.81 (0.68,0.89) ‡
†Kendall's tau. ‡Spearman's rho. p<0.05			

Table 29. Cartilage scores and plain radiograph correlation

There is “strong” positive correlation that is similar for BMLs and BMO although not as strong for total cyst score that only demonstrated “weak” positive correlation (90). Cartilage total joint score, score 1 and score 2 all demonstrated “strong” or “very strong” positive correlation across all readers. There is little difference between the cartilage correlations with radiographic severity although score 2 demonstrated lower values than total joint score and score 1 although this does not change the interpretation. As discussed in the methods sections, the Spearman’s rho value is always slightly greater than the value for Kendall’s Tau.

The correlation values alone only demonstrated that there is correlation between the MRI feature and the severity of disease on plain radiograph.

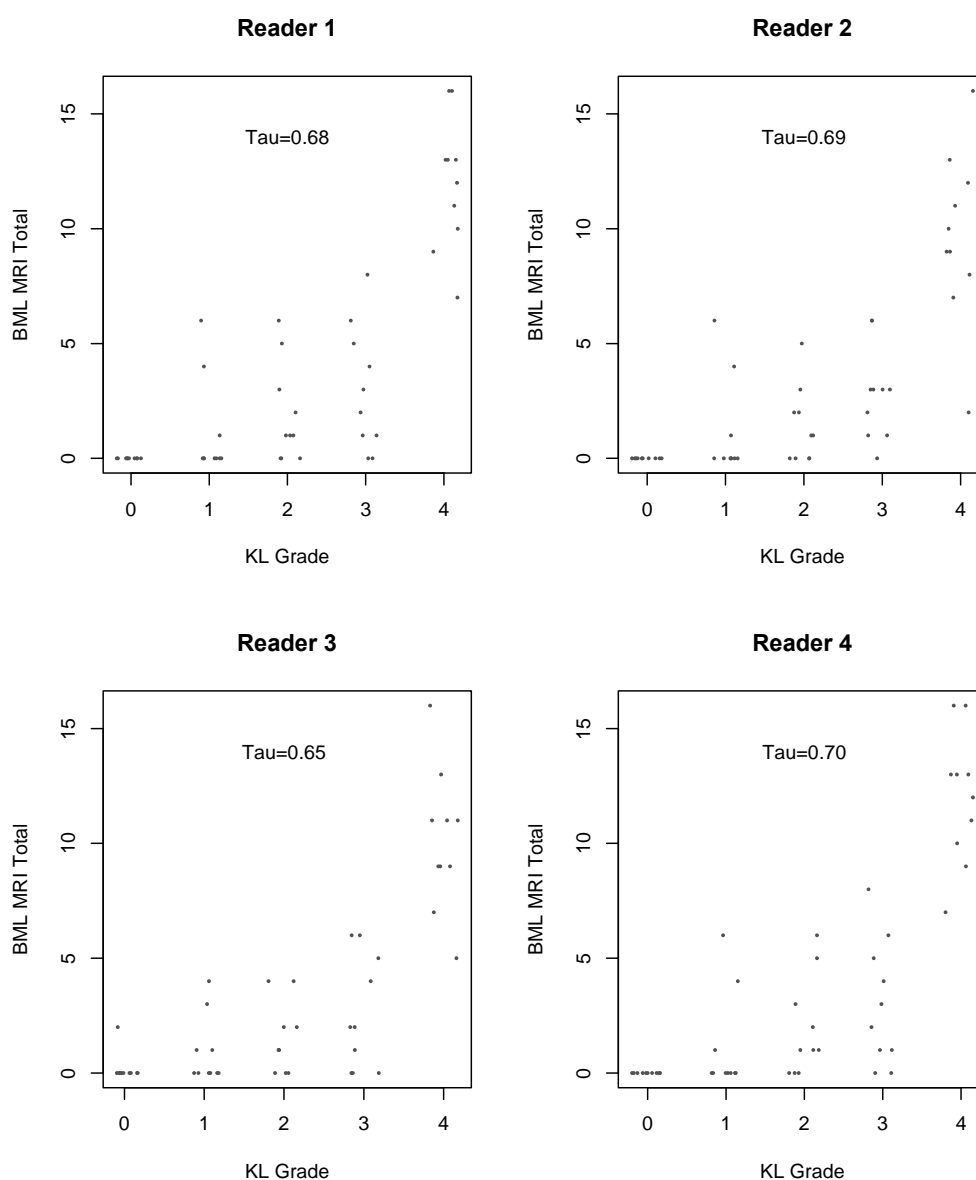


Figure 15. Total bone marrow lesion score versus Kellgren-Lawrence grade

Of the 10 cases that were classified as KL grade 0, only one was graded on MRI as having a BML and for reader 3 only. In contrast, a radiographic score of four related to between five and 16 zones affected other than for a solitary outlier for reader 2. Radiographic scores of one to three demonstrated a wide range of overlapping score including zero and extending to more than five.



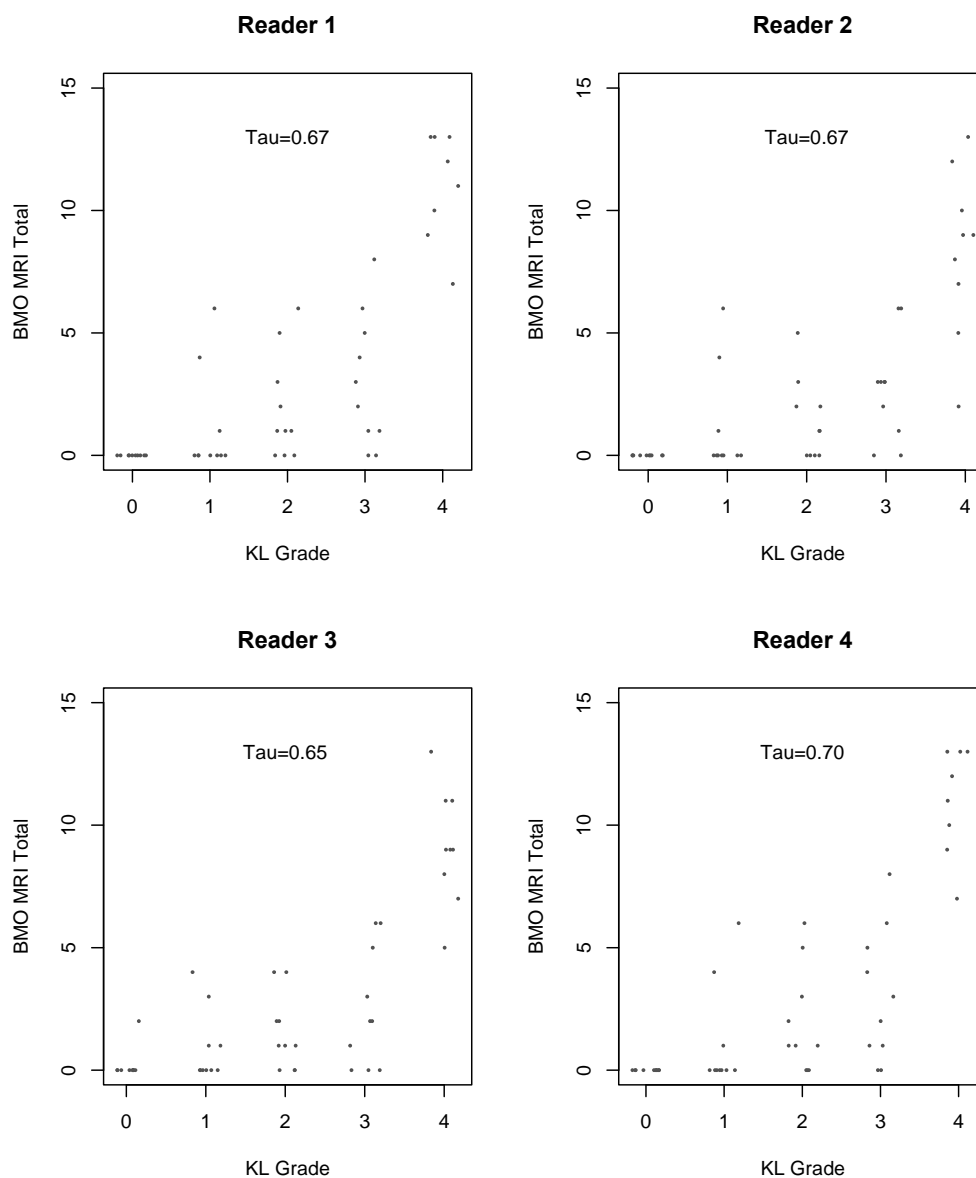


Figure 16. Total bone marrow oedema score versus Kellgren-Lawrence grade

Total joint correlations of BMO with radiographic severity are very similar to BML correlation with a similar interpretation when studying the scatter plots. A radiographic score of zero corresponds to an MRI score of zero (except the same solitary outlier for reader 3) and a radiographic score of four always corresponding to BMO on MRI in this sample. The MRI score related to radiographic scores of one to three are again overlapping and wide ranging.

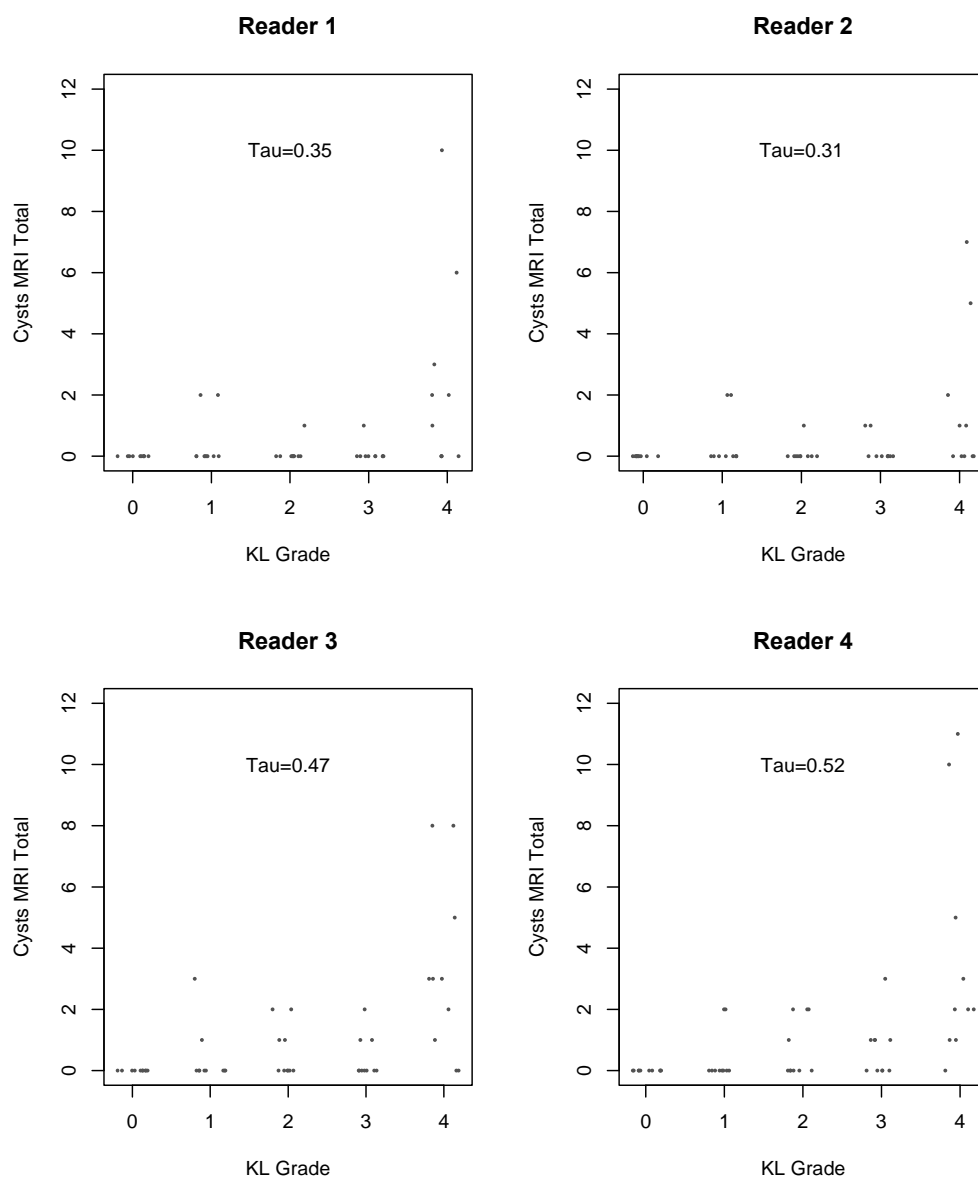


Figure 17. Total joint cyst score versus Kellgren-Lawrence grade

Although there is positive correlation for total joint cyst score related to radiographic score this is not as strong as those demonstrated for the other variables. The most severe cyst scores were associated with a radiographic score of four and a radiographic score of zero corresponds to an MRI cyst score of zero for all subjects in the sample. All radiographic scores, including grade four, demonstrated cyst score of zero.

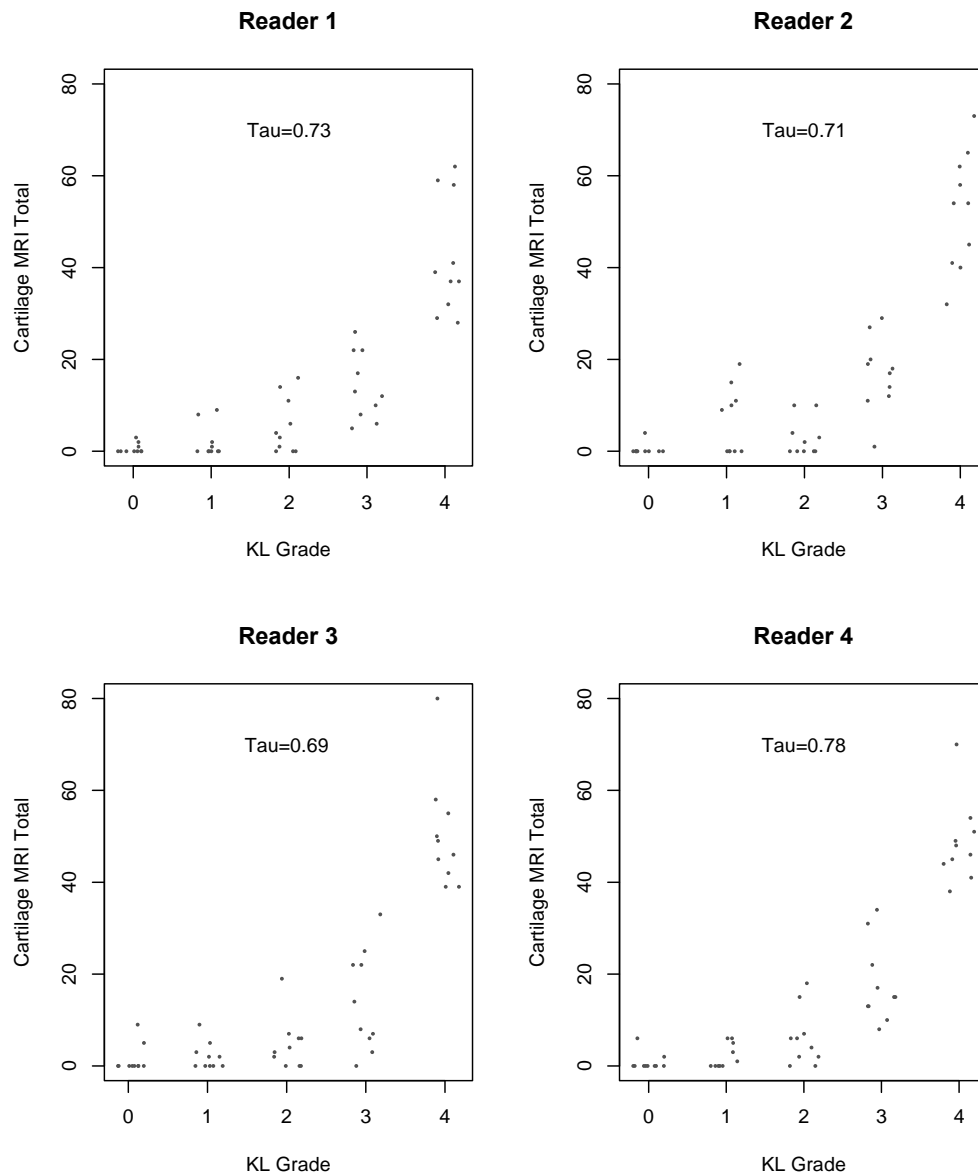


Figure 18. Total joint cartilage score versus Kellgren-Lawrence grade

There is overlap in the range of scores between radiographic grades one to three with radiographic scores of one and two both including MRI scores of zero. All KL grade 4 radiographs demonstrated a total joint cartilage score of at least 30 in this sample population although there is overlap with grade 3 radiographic scores.

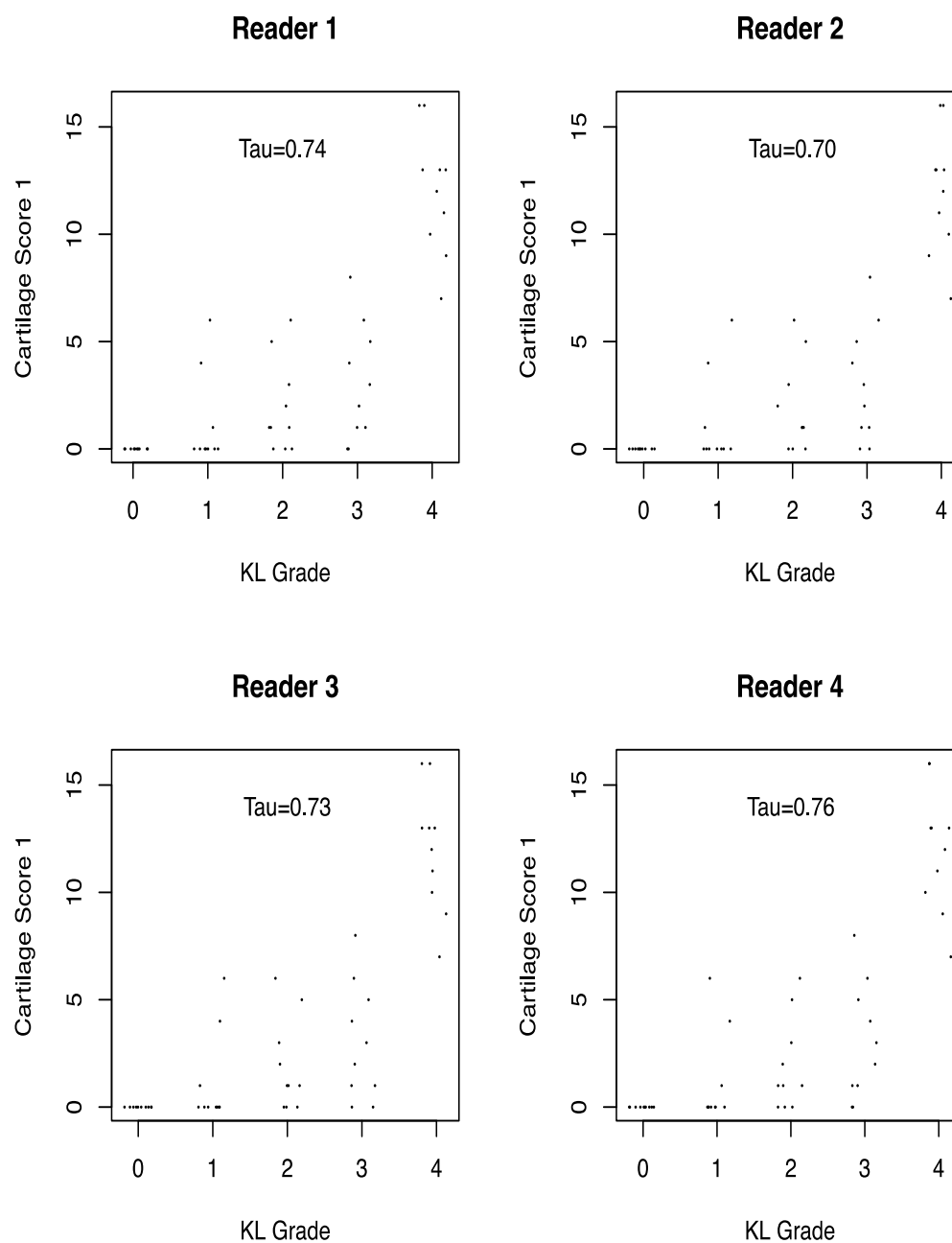


Figure 19. Cartilage Score 1 versus Kellgren-Lawrence grade

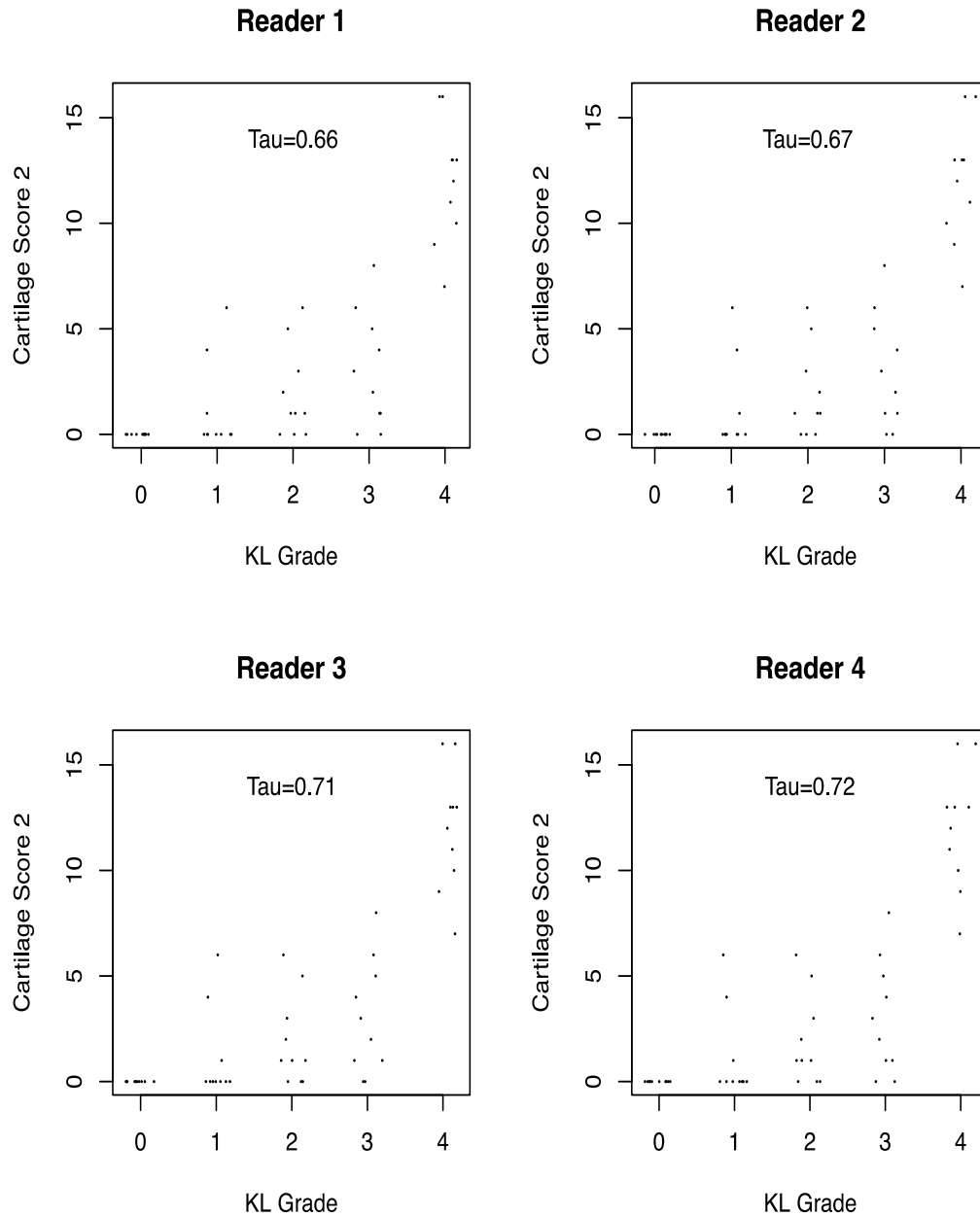


Figure 20. Cartilage Score 2 versus Kellgren-Lawrence grade

The interpretation of the scatter plots for cartilage extent score 1 and score 2 are similar. A radiographic score of zero corresponds to an MRI score of zero but radiographic scores of one to three are wide ranging and overlapping and include zero. Radiographic score of four always corresponding to cartilage damage with both extent scores 1 and 2 in the sample tested.

## DISCUSSION

### Sample

Despite a considered methodology, the current study does have some limitations. The intra-reader reliability was only tested on a sample of 10 patients. The confidence intervals for intra-reader reliability were broader than for inter-reader reliability where a sample of 50 was used. Some intra-reader results demonstrated some disparity such as in the case of mortise view KL score for the radiographic scoring where reader 2 demonstrated perfect agreement but reader 1 only demonstrated “moderate” agreement with confidence intervals extending into the “weak” range. On review, the percentage agreement for reader 1 was 80%. The small sample size used for intra-reader testing is possibly the cause of these suboptimal intra-rater results for a small proportions of features across the radiographic and MRI scoring, especially given the high percentage agreement. In retrospect, rescoring all radiographs and MRI examinations for all features by each observer is desirable to attempt to reduce this effect.

As in the MOAKS system there was an attempt to control the underlying marginal prevalence of disease to avoid some of the issues faced by the other systems where some severities were underrepresented with some features only scored once. Despite this, the spread of MRI OA severity, as detailed in Table 27, was distributed towards to lower end of disease severity. The drawback in some of the other systems is the underrepresentation of severe OA in the reliability testing. Using a random sample may have made the situation worse as in KOSS and SHOMRI where there were no patients included who were defined as having severe radiographic OA.

In selecting studies for use in the reliability study consecutive date ordered radiographs were consensus read until 10 grades for each radiographic grade had been selected. As long as accompanying MRI examinations conformed to the inclusion criteria they were included. Recruiting in this method aimed to minimise any selection bias. The selection of cases for intra-rater assessment were chosen using a random number generator, again to prevent any selection bias (111). The ordering in which the cases were presented for both radiographic and MRI grading was random attempting to minimise any observer bias and readers were blinded to the corresponding plain radiograph and radiographic OA score. One potential source of observer bias is that readers graded all features on one examination and there is the potential that the severity of one feature may influence how they grade the other features. One way to avoid this would have been for readers to grade all 50 cases for cartilage alone and then at a time interval grade for the BMLs and so on. In practice this may have been prohibitively time consuming. Intra-reader assessment was performed at a period of no shorter than four weeks, again to reduce any potential observer bias.

### MRI Protocols

Although not expanded upon in the methodology section, patients were not selected based upon which MRI scanner they were examined with. In fact, 40 of the 50 patients were examined on a 1.5T machine with the remaining 10 patients examined using a 3T machine. It has been demonstrated that cartilage injury is better demonstrated using 3T versus 1.5T (65). This project aims to test the reliability and reproducibility of the system, not the sensitivity of the system using different strength MRI machines so this is a relative limitation. The results of a separate analysis of the 3T group and comparing to the 1.5T group may not be meaningful given the relatively small sample.

### Cartilage score

The score used in the reliability study corresponds to the Noyes score modified for MRI initially outlined by Recht with the addition of a grade 1 score outlined by Kijowski (65,100). Despite multiple cartilage score iterations tested, there was no improvement in reliability by simplifying the cartilage score. Only the version 5 cartilage score demonstrated near equivalent results. The version 5 cartilage score corresponds to the ICRS cartilage score modified for MRI (103). The difference between the ICRS system and the original score used in the reliability study are with respect to the higher grades where the original system scores full thickness chondral loss and chondral loss with a bony component separately and the modified ICRS score groups these together. Despite it representing a simplified score in comparison to the original system, it only demonstrates an increase of 0.01 in the weighted kappa value and for the trainee readers only. The confidence intervals and result for the expert readers were the same.

The cartilage extent scores are complimentary to the total joint cartilage score. The existing knee scoring systems incorporated some form of surface extent assessment in their cartilage scores. These had a particular focus not only on extent of any cartilage damage but also full thickness chondral loss leading to the choice of the two scores in this study, which demonstrated “almost perfect” reliability.

### Radiographic prediction of MRI features

The total joint BML, BMO and cartilage scores demonstrated strong positive correlation with radiographic OA severity while cysts demonstrated positive, although weak correlation (90). In the sample tested the scatter plots give an indication that radiographic severity at each end of the KL scale may give an



indication of MRI score. For example, a radiographic score of four always related to cartilage degeneration and BMLs on MRI in the sample tested. Similarly, except for one data point for one reader a KL score of zero corresponded to no BMLs on MRI. This information may be of use for a clinician when deciding if an MRI of the ankle is warranted if the radiograph demonstrates a KL grade of zero or four. The correlation values and scatter plot distributions does not equate to the predictive power and care must be taken when interpreting these results. This study did not set out to determine the predictive power of plain radiographs for MRI features of OA. In order to determine this a formal diagnostic accuracy study is required to calculate the positive and negative predictive value for each KL grade. Given that there are only ten subjects for each radiographic grade in this reliability study, performing logistic regression and a diagnostic accuracy study on this data may not be meaningful.

## 6. Conclusion

This thesis demonstrates the reliability and reproducibility of a new MRI grading system for OA of the ankle that demonstrates “substantial” to “almost perfect” inter and intra-rater agreement. To the best knowledge this is the first and only proposed MRI grading system for OA of the ankle.

Based on the results of the Delphi survey and reliability study, the “Norwich Osteoarthritis of the Ankle MRI Score” grades the following features of osteoarthritis: bone marrow lesions, BMO, subchondral cysts and cartilage injury depth for all zones with the presence or absence of osteophytes. The full 16-zone division as described in the reliability study is recommended for the expert reader. In addition, summing the scores across zones for cartilage integrity and BMLs gives the feature total joint score. Cartilage damage surface extent (score 1) and full thickness cartilage injury surface extent (score 2) can also be determined. The MRI modification of the Noyes score is the recommended cartilage score.

Given that the 3-zone talar dome score demonstrated similar results to the 9-zone division there is an argument for using the more simplistic approach and method that was preferred on discussion with a foot and ankle surgeon (99). As a research tool, the 9-zone division has the potential to be more sensitive to change. For example, if there is an abnormality in zone 1 at baseline and an additional lesion in zone 7 at an interval study this will be reflected in any increase in severity score. Using a 3-zone approach in the current scoring system would not record an increase in disease severity for BMLs, cartilage total joint score or surface extent score. Whether in practice the 9-zone system is more sensitive to change to change will need to be determined in future research.

When compared to the other described MRI scoring systems for the hip and knee, the number of features included in this score appears limited. This new scoring system grades five components of OA compared to between ten and twelve items for the other systems. The choice of features to be included was justified being the opinion of a group of experts routinely involved in the care of patients with ankle OA. Nevertheless, given that features such as ligament integrity can play a role in the development of ankle OA there is an argument that some other features could have been included.

Comparing reliability figures for each OA feature against the other scoring systems is not particularly meaningful given the different sample sizes and underlying marginal prevalence of disease. In terms of subjective interpretation the features are comparable with the other systems except for perhaps osteophytes. Despite demonstrating almost “perfect agreement” for the expert readers, the osteophyte score in this new system is limited in comparison to the other scores by merely grading for presence or absence for whole joint and not assessing size or location as in the other systems.

The nature of these whole organ MRI scoring systems make them too complex and time consuming for use in everyday clinical practice and more suited as a tool in the research setting. The exception to this is the system outlined by Park et al that mixes multiple constructs to give an overall MRI grade from 0–4, paralleling the radiographic KL score (86). The goal in developing the Park system was for use in routine clinical practice, contrasting the other systems and this study where the primary aim was to develop a research tool. Mixing the different components of this MRI ankle score in a similar fashion to that outlined by Park is a possibility to adapt the system for better clinical utility. The choice of threshold score for each severity level and the number of items on the scale would be arbitrary if applied at this stage to the current system. With cartilage extent and BMLs scores ranging from zero to sixteen while cartilage total joint score extends to 80, the choice of weights of individual

features would also need to be determined. Other than the Park system, the other published systems have not attempted at this stage to give a standalone MRI severity score with the components of each score instead interpreted in isolation.

Certain components of the score may be of use in the clinical setting in isolation. The cartilage score for the talar dome demonstrated “substantial” and “almost perfect” inter-rater reliability for the trainee and expert readers respectively for both three and nine zone divisions of the talar dome. A reproducible and detailed MRI cartilage scoring system for the talar dome that correlates well with arthroscopy findings may be useful to the orthopaedic surgeon. To establish this, further research to assess correlation of talar dome MRI cartilage score and arthroscopic grading of the talar dome is required.


Some of the scoring systems such as BLOKS and MOAKS included analyses of MRI score related to patient reported outcome measures with the results leading to refinements of the MRI scoring system itself. In addition, the MOAKS system compared other clinical components such as range of motion to individual components of the MOAKS system. Comparing the results of a patient reported outcome measure, which considers both pain and function, alongside the described MRI scoring system will allow assessment for possible correlation between individual components of the score and clinical features.

Demonstrating the sensitivity to change of this new MRI score will establish the validity of the tool for use in longitudinal studies. BLOKS assessed components of both the BLOKS and WORMS systems in a sample of patients from a longitudinal cohort study who underwent interval MRI examinations to examine these features (80). Conducting a similar study for this ankle MRI score is desirable prior to possible use in longitudinal studies. This new proposed system may not be the final “solution”, but further research will help establish validity and possibly refine some features when applied in clinical studies.

## Appendices

Appendix 1. Figure 1 permissions.....	126
Appendix 2. Components of MRI Scoring Systems Compared. Table 1 .....	127
Appendix 2. Components of MRI Scoring Systems Compared. Table 2 .....	128
Appendix 2. Components of MRI Scoring Systems Compared. Table 3 .....	129
Appendix 2. Components of MRI Scoring Systems Compared. Table 4 .....	130
Appendix 2. Components of MRI Scoring Systems Compared. Table 5 .....	131
Appendix 2. Components of MRI Scoring Systems Compared. Table 6 .....	132
Appendix 2. Components of MRI Scoring Systems Compared. Table 7 .....	133
Appendix 3. Summary results of MRI Scoring Systems main features.....	134
Appendix 4. R&D Delphi Survey Permission.....	135
Appendix 5. Delphi Survey Questionnaire.....	136
Appendix 6. Ethics Committee favourable opinion .....	137
Appendix 7. R&D Approval .....	141
Appendix 8. Ankle MRI protocols .....	142
Appendix 9. MRI Scoring sheet 1 .....	143
Appendix 9. MRI Scoring sheet 2 .....	144
Appendix 10. Presentations.....	145

## Appendix 1. Figure 1 permissions



**RSNA**  
Radiological Society  
of North America

620 Jane Blvd  
Oak Brook, IL 60521  
Tel: 630-571-2670  
Fax: 630-571-7837  
[RSNA.org](http://RSNA.org)

December 8, 2016

Sharief Aboelmagd  
Norfolk and Norwich University Hospital  
Norwich NR4 7UY  
United Kingdom

Dear Sharief Aboelmagd:

The Radiological Society of North America (RSNA™) is pleased to grant you permission to reproduce the following figure in print format for educational, non-profit use in your research degree thesis, provided you give full credit to the authors of the original publication.

Figure 1  
Mengiardi B, Pfirrmann C W.A., Vienne P, et al. Medial collateral ligament complex of the ankle: MR appearance in asymptomatic subjects. *Radiology* 2007;242:817-824.

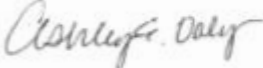
This permission is a one-time, non-exclusive grant for English-language use and is exclusively limited to the usage stated and underlined above. The requestor guarantees to reproduce the material as originally published. Permission is granted under the condition that a full credit line is prominently placed (i.e. author name(s), journal name, copyright year, volume #, inclusive pages and copyright holder).

This permission becomes effective upon receipt of this signed contract. Please sign a copy of this agreement, return a signed copy to me and retain a copy for your files. Thank you for your interest in our publication.

[Print Name]: S. ABOELMAGD

SIGNATURE: S. Aboelmagd Date: 12/12/16

Sincerely,



Ashley E. Daly  
Senior Manager, Journal Rights & Communications  
Publications  
Phone: 630-590-7771

Appendix 2. Components of MRI Scoring Systems Compared. Table 1

	<b>WORMS</b>	<b>BLOKS</b>	<b>MOAKS</b>	<b>KOSS</b>	<b>HOAMS</b>	<b>SHOMRI</b>
<b>Subregional divisions</b>	15 subregions: medial and lateral patella, medial/lateral femur (anterior/central/posterior), medial/lateral tibia (anterior/central/posterior), subspinous tibia	9 subregions: medial and lateral patella, trochlea, weight-bearing femur, weight-bearing tibia, subspinous tibia	15 subregions: medial/lateral patella, medial/lateral femur (trochlea/central/posterior), medial/lateral tibia (anterior/central/posterior), subspinous tibia	9 subregions: medial patella, patellar crest, lateral patella, medial/lateral trochlea, medial/lateral femoral condyle, medial/ lateral tibial plateau	15 subregions: 8 subregions for the femoral head and 7 for the acetabulum.	10 subregions. The femur is divided into 6 subregions and acetabular articular surface is divided into 4 subregions.
<b>Bone marrow lesions</b>	Graded in each subregion from 0 to 3 dependent on volume of subregion involved. 0=none; 1=<25%, 2=25-50%; 3=>50%.  The maximum score for the entire knee is 45	Scoring of individual lesions 3 different aspects of BMLs scored: (A) Size of BML scored from 0 to 3 concerning percentage of subregional bone volume (B) % of surface area adjacent to subchondral plate (C) Percentage of BML that is noncystic.	Summed BML size/volume for subregion from 0 to 3 based on percentage of subregional bone volume Number of BMLs counted Percentage of the volume of each BML that is noncystic is graded from 0 to 3	Scoring of individual lesions from 0 to 3. 0 = absent. 1 = minimal (<5mm). 2 = moderate (5mm–2cm) 3 = large (>2cm)	BMLs and cysts assessed with relation to % involvement of each zone and size. Features are scored together from 0–3. 0 = absent. 1= mild (<33% subregion involved). 2= moderate (33–66% subregion involved). 3= severe (>66% subregion involved).	4-point scale. Scored in each zone.  Each subregion scored separately and then totalled for a BML score.  0 = No lesions present. 1 = equal to or less than 0.5cm 2 = >0.5cm ≤ 1.5cm 3 = >1.5cm

Appendix 2. Components of MRI Scoring Systems Compared. Table 2

	<b>WORMS</b>	<b>BLOKS</b>	<b>MOAKS</b>	<b>KOSS</b>	<b>HOAMS</b>	<b>SHOMRI</b>
<b>Cartilage</b>	Graded in each subregion using from 0-6 on an 8-point scale. 0=Normal 1 = normal thickness but high signal on T2 image 2.0 = partial thickness focal defect<1cm 2.5 = full thickness focal defect<1cm in width. 3 = multiple area of partial thickness focal defect or grade 2.0 = >1cm<75% of subregion. 4 = >75% partial thickness loss 5 = Multiple areas of full thickness loss or grade 2.5>1cm<75% of subregion. 6 = diffuse full thickness loss.	Uses 2 scores Score 1: subregional approach (A) Percentage of any cartilage loss in subregion (B) Percentage of full-thickness cartilage loss in subregion  Score 2: site-specific approach. Scoring of cartilage thickness at 11 specific locations (not subregions) from 0 (none) to 2 (full-thickness loss)	Subregional approach: each articular cartilage region is graded from 0 to 3 for size of any cartilage loss as a percentage of surface area of each individual region surface, and percentage in this subregion that is full-thickness loss	Subregional approach: Depth of lesion is scored from 0 to 3 0=normal 1=<50% thickness 2= >50% thickness 3= Full thickness defect.  Osseous component depth graded 0–3. 0= absent 1=<2mm 2= 2–5mm 3= >5mm  Surface extent/diameter graded 0–3. 0= absent 1=<5mm 2= 5–10 mm 3= >10 mm	Scored from 0–4 0= normal cartilage 1= focal partial thickness defect. 2= focal full thickness defect. 3= several partial thickness defects or single large superficial defect. 4= several large full thickness defects or single full thickness defect.  Scored in 9 of the 15 subregions.	Articular cartilage loss scored in each of the 10 subregions with a 3-point scale: 0 = no loss, 1= partial thickness loss, and 2= full thickness loss



Appendix 2. Components of MRI Scoring Systems Compared. Table 3

	<b>WORMS</b>	<b>BLOKS</b>	<b>MOAKS</b>	<b>KOSS</b>	<b>HOAMS</b>	<b>SHOMRI</b>
<b>Subchondral cysts</b>	Graded in each subregion 0 to 3. 0=none. 1=<25% of subregion. 2=25–50% subregion. 3=>50% of subregion	Scored together with BMLs	Scored together with BMLs	Scoring of individual lesions from 0 to 3. 0 = absent. 1 = minimal (<3mm). 2 = moderate (3mm–5mm) 3 = large (>5mm)	See BMLs.	3–point scale in each subregion. Maximum diameter measured. 0 = absent 1 = ≤ 0.5cm 2 = >0.5cm
<b>Ligaments</b>	Cruciate ligaments and collateral ligaments scored as intact or torn	Cruciate ligaments scored as normal or complete tear. Associated insertional BMLs are scored in tibia and in femur. Collateral ligaments not scored.	Same as BLOKS	Not scored	Not scored	Ligamentum teres abnormalities graded as 0 for normal, 1 for signal abnormalities or fraying, 2 for partial tear and 3 for complete tear.
<b>Loose Bodies</b>	Scored from 0 to 3 depending on number of loose bodies	Scored as present or absent	Same as BLOKS	Not scored	Loose intra-articular bodies are scored from 0–3. 0= None. 1= single loose body. 2= two loose bodies. 3= three or more	Scored as 1 or 0 for present of absent

Appendix 2. Components of MRI Scoring Systems Compared. Table 4

	<b>WORMS</b>	<b>BLOKS</b>	<b>MOAKS</b>	<b>KOSS</b>	<b>HOAMS</b>	<b>SHOMRI</b>
<b>Meniscal status</b> (Knee scores)	Anterior horn, body, posterior horn scored separately in medial/lateral meniscus from 0 to 4:	Intrasubstance signal changes in anterior horn, body, posterior horn scored separately in medial/lateral meniscus.	Same as BLOKS, plus additional scoring for meniscal hypertrophy, partial maceration, and progressive partial maceration	No subregional division of meniscus described. Presence or absence of tears:	Assessed at defined locations in different imaging planes and scored from 0 to 3. Labral hypertrophy and paralabral cysts scores in addition as present or absent.	Scored on 3 planes in 4 different subregions: anterior and posterior on axial plane imaging, anterosuperior on sagittal plane, and superior on the coronal plane.
<b>Labrum</b> (Hip Scores)	<ol style="list-style-type: none"> <li>1. Minor radial or parrot beak tear</li> <li>2. Nondisplaced tear or prior surgical repair</li> <li>3. Displaced tear or partial resection</li> <li>4. Complete maceration or destruction or complete resection</li> </ol> <p>Meniscal extrusion: Not scored</p>	<p>Presence/absence scored for the following:</p> <p>Intrameniscal signal, tear (vertical, horizontal, complex, root), maceration, meniscal cyst.</p> <p>Extrusion: Scored 0–3: medial and lateral extrusion on coronal image, and anterior extrusion for medial or lateral meniscus on sagittal image.</p>	<p>Meniscal extrusion: Scored as medial and lateral extrusion on coronal image, and anterior extrusion for medial or lateral meniscus on sagittal image from 0 to 3</p>	<p>Horizontal tear, vertical tear, radial tear, complex tear, bucket-handle tear.</p> <p>Meniscal intrasubstance degeneration scored from 0 to 3</p> <p>Meniscal extrusion: Scored on coronal image from 0 to 3</p>	<p>0 = No signal changes 1= Intralabral signal alteration 2= Definite labral tear 3= Partial or complete labral maceration</p>	<p>Graded as 0 for normal or normal variant. 1 for abnormal signal or fraying. 2 for a simple tear, 3 for labrocartilaginous separation, 4 for complex tear and 5 for maceration.</p>

Appendix 2. Components of MRI Scoring Systems Compared. Table 5						
	<b>WORMS</b>	<b>BLOKS</b>	<b>MOAKS</b>	<b>KOSS</b>	<b>HOAMS</b>	<b>SHOMRI</b>
<b>Osteophytes</b>	Scored from 0 to 7 at 14 sites along margin of the knee. 0: None 1: Equivocal 2: Small 3: Small–moderate 4: Moderate 5: Moderate–large 6: Large 7: Very large	Scored from 0 to 3 at 12 sites	Same as BLOKS: scored from 0 to 3 at 12 sites	Differentiated in to marginal, intercondylar and central  Scored from 0 to 3, measured from base to tip.  0 = absent 1 = <3mm 2 = 3–5mm 3 = >5mm	Osteophytes are assessed in five locations and scored from 0–4 representing absent, equivocal, small beak like definite osteophyte, intermediate size osteophyte and proliferative large osteophyte.	Not scored
<b>Bone Attrition</b>	Scored from 0 to 3 in 14 subregions. Normal, mild, moderate, severe.	Not scored	Not scored	Not scored	Flattening of the femoral head/ with presence of attrition	Not scored

Appendix 2. Components of MRI Scoring Systems Compared. Table 6

	<b>WORMS</b>	<b>BLOKS</b>	<b>MOAKS</b>	<b>KOSS</b>	<b>HOAMS</b>	<b>SHOMRI</b>
<b>Effusion</b>	Scored from 0 to 3	Scored from 0 to 3	Scored from 0 to 3 (termed effusion synovitis)	Scored from 0 to 3 Absent, small, moderate or massive effusion based on descriptive definitions (no measurements).	Joint effusion was graded from 0–2 on coronal and axial images according to the degree of capsular distension	Joint effusion scored as a surrogate of synovitis and the present of fluid signal at the femoral neck region greater than 0.7cm in thickness scored as 1.
<b>Synovitis</b>	Combined effusion/synovitis score	(A) Scoring of size of signal changes in Hoffa fat pad (B) Five additional sites scored as present or absent	Scored from 0 to 3 (called Hoffa synovitis)	Synovial thickening recorded as present or absent.	Scored according to thickness of the synovium at defined locations in axial and coronal planes wherever contrast enhanced imaging was used only. Scored from 0 to 2 dependent on synovial thickness. 2–4mm=Grade 1 ≥ 4mm = Grade 2	See effusion score

Appendix 2. Components of MRI Scoring Systems Compared. Table 7

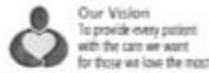
	<b>WORMS</b>	<b>BLOKS</b>	<b>MOAKS</b>	<b>KOSS</b>	<b>HOAMS</b>	<b>SHOMRI</b>
<b>Periarticular features</b>	Popliteal cysts, anserine bursitis, semimembranosus bursa, meniscal cyst, infrapatellar bursitis, prepatellar bursitis, tibiofibular cyst scored from 0 to 3	Features scored as present or absent: Patellar tendon signal, pes anserine bursitis, iliotibial band signal, popliteal cyst, infrapatellar bursa, prepatellar bursa, ganglion cysts of the tibiofibular joint, meniscus, anterior and posterior cruciate ligaments, semimembranosus, semitendinosus, other	Same as BLOKS	Only bursal cysts. Noted as absent, minimal, moderate or severe. Subjective with no measurement or descriptive definition given.	Dysplasia scored as present or absent. Trochanteric bursitis scored as present or absent. Herniation pits scored as present or absent	

Appendix 3. Summary results of MRI Scoring Systems main features.

Joint Feature	WORMS <sup>†</sup>	BLOKS <sup>‡</sup>	MOAKS <sup>‡</sup>	KOSS <sup>†/‡</sup>	HOAMS <sup>‡</sup>	SHOMRI <sup>‡</sup>
BML	0.74	0.72 (0.58–0.87)	0.82–0.96 (0.69–0.95)	0.91 (0.88–0.93)/0.88	0.85 0.67 1.00	0.55 (0.46–0.64)
BML % area and % of lesion BML (BLOKS only). Number of BML (MOAKS)	N/A	0.69 (0.55–0.82) 0.72 (0.58–0.87)	0.65–1.00 (0.40–1.00) 0.72–0.93 (0.51–1.00)	N/A	N/A	N/A
Osteophytes	0.97	0.65 (0.52–0.77)	0.44–0.80 (0.24–1.00)	N/A	0.63 (0.38–0.88)	N/A
Cartilage	0.99	0.72 (0.59–0.85)	0.73–0.85 (0.52–0.98) 0.36–0.71 (0.05–0.99)	0.64 (0.58–0.69)/0.57	0.65 (0.31–1.00)	0.57 (0.52–0.62)
Cartilage 2 (BLOKS only)	N/A	0.73 (0.60–0.85)	N/A	N/A	N/A	N/A
Osteochondral defects	N/A	N/A	N/A	0.63 (0.55–0.70)/0.66	N/A	N/A
Synovitis	0.74	0.62 (0.05–1.00)	0.72 (0.52–0.92)	0.74 (0.58–0.85)	0.60 (0.23–0.97)	0.55 (0.33–0.76)
Effusion	*	0.61 (0.05–0.85)	0.70 (0.47–0.93)	*	0.65 (0.34–0.97)	*
Meniscal subluxation/ Labrum	N/A	0.51 (0.24–0.78)	0.66–0.79 (0.46–0.91)	0.82 (0.75–0.86)/0.82	0.48 (0.15–0.81)	0.65 (0.60–0.71)
Meniscal degeneration	N/A	0.68 (0.44–0.93)	N/A	0.76 (0.66–0.83)/0.56	N/A	N/A
Meniscal Tear	0.87	0.79 (0.40–1.00)	0.95–0.97 (0.86–1.00)	0.78 (0.70–0.83)/0.78	N/A	N/A
Ligaments/ Ligamentum teres	1.00	N/A	N/A	N/A	N/A	0.72 (0.65–0.84)
Subchondral cyst	0.94	Part of BML score	Part of BML score	0.90 (0.87–0.92)/0.87	0.15 (-0.24–0.54)	0.71 (0.60–0.81)
Bakers Cysts/ Paralabral cysts	N/A	N/A	No result	0.96 (0.90–0.98)/0.91	0.58 (0.07–1.00)	0.63 (0.42–0.84)
Loose bodies	None identified	No result	No result	Not scored	1.00	0.79 (0.40–1.00)

<sup>†</sup> Intraclass correlation coefficient. <sup>‡</sup> Weighted kappa with 95% confidence intervals. \* scored with synovitis.

Appendix 4. R&D Delphi Survey Permission



Norfolk and Norwich University Hospitals **NHS**  
NHS Foundation Trust

Research and Development Office  
Level 3 East  
Norfolk and Norwich University Hospital  
Colney Lane  
Norwich NR4 7UY  
direct dial: 01603 286611  
email: [lisa.chalkley@nnuh.nhs.uk](mailto:lisa.chalkley@nnuh.nhs.uk)

Dr Sharief Aboelmagd  
Radiology Registrar  
Norfolk and Norwich University Hospitals NHS Foundation Trust  
Radiology Academy  
Cotman Centre  
Colney Lane  
Norwich  
NR4 7UY

27th July 2015

Dear Dr Aboelmagd,

Thank you for your enquiry regarding the Radiology study entitled:

*"A Delphi Survey to Establish the Importance of Different Components of Ankle Osteoarthritis on MRI".*

I can confirm from the information provided that this project does not meet the criteria for NHS research. This survey includes anonymised data from opinions of NHS staff with relation to their current practice and therefore does not require formal ethical approval or local NHS permission.

Once you have developed your scoring system for ankle osteoarthritis, if you then wish to conduct a study to validate the system then this will be defined as Research and will be require formal NHS Ethics and local R&D approval.

Please do not hesitate to contact the Research and Development office should you require any further information.

Yours sincerely

Lisa Chalkley  
Research Services Manager

## Appendix 5. Delphi Survey Questionnaire

Adapted from the online version. Completed for tibiotalar and subtalar joints separately.

*Clinical Specialty:* Radiologist\_\_\_\_ Surgeon\_\_\_\_ Rheumatologist\_\_\_\_

Please grade on the 5-point likert scale how important you feel the following features of osteoarthritis of the ankle are in reporting the disease severity on MRI where 1 represents not at all important and 5 represents very important

Tibiotalar Joint \_\_\_\_

Subtalar joint \_\_\_\_

Presence of subchondral bone marrow signal abnormality 1—2—3—4—5

Extent of subchondral bone marrow signal abnormality 1—2—3—4—5

The presence of Osteophytes 1—2—3—4—5

The number of osteophytes 1—2—3—4—5

Degree of Cartilage degeneration 1—2—3—4—5

Integrity of the supporting ligaments of the ankle 1—2—3—4—5

The presence of osteochondral defects 1—2—3—4—5

The presence of subarticular cysts 1—2—3—4—5

The presence of subchondral bone attrition  
(depression/flattening of the subchondral bone) 1—2—3—4—5

The severity of subchondral bone attrition 1—2—3—4—5

The presence of a joint effusion 1—2—3—4—5

The presence of synovitis 1—2—3—4—5



## Appendix 6. Ethics Committee favourable opinion

Page 1



### North West - Lancaster Research Ethics Committee

Barlow House  
3rd Floor  
4 Minshull Street  
Manchester  
M1 3DZ

Telephone: 020 71048008

22 February 2016

Professor Andoni Toms  
Consultant Radiologist  
Norfolk and Norwich University Hospital NHS Foundation Trust  
Colney Lane  
Norwich  
NR4 7UY

Dear Professor Toms

<b>Study title:</b>	<b>Osteoarthritis of the Ankle MRI Score</b>
<b>REC reference:</b>	<b>16/NW/0152</b>
<b>IRAS project ID:</b>	<b>198605</b>

The Proportionate Review Sub-committee of the North West - Lancaster Research Ethics Committee reviewed the above application on 23 February 2016.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager Mrs Carol Ebenezzer, [nrescommittee.northwest-lancaster@nhs.net](mailto:nrescommittee.northwest-lancaster@nhs.net). Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

#### Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

#### Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

**You should notify the REC once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. Revised documents should be submitted to the REC electronically from IRAS. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which you can make available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC**

**may cause delay in obtaining permissions.**

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

*Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).*

*Guidance on applying for HRA Approval (England)/ NHS permission for research is available in the Integrated Research Application System, [www.hra.nhs.uk](http://www.hra.nhs.uk) or at <http://www.rdforum.nhs.uk>.*

*Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.*

*For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.*

*Sponsors are not required to notify the Committee of management permissions from host organisations.*

## Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact [hra.studyregistration@nhs.net](mailto:hra.studyregistration@nhs.net). The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

**It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).**

## **Ethical review of research sites**

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion").

## **Summary of discussion at the meeting (if applicable)**

The Committee asked for an explanation as to how you were directly involved in every patient's care that you selected, when it would not have necessarily have been you that reported on those images, and why you were not requesting the PACS manager to select images, anonymise them, and then pass them on to the radiologist.

The Committee would also asked how you would know the clinical history e.g. stroke when all you were reviewing were the images; you needed the case notes for this information as rarely would this be noted on an imaging request card

*You responded that direct care team in the context of your application referred to research team members reporting radiological and specifically MRI ankle examinations as part of their routine clinical work at the Norfolk and Norwich Hospital. No special access was required as all members of the research team had access to this data and the archives ordinarily.*

*The inclusion and exclusion criteria were such that you felt adequate details would actually be included in the request for an examination such as a MRI examination of the ankle at your institution. The paperless system at the Norfolk and Norwich Hospital allowed review of patient data such as the imaging archive, if necessary, without needing to review the patient notes specifically. This was again normal access for the research team in clinical practice with no special access required.*

*You said that you had a robust system in place for anonymizing images. Reviewing and screening requests for the examination for inclusion by a radiologist was preferable than by a PACS team member who was not medically trained and could possibly be a considerable extra workload for them.*

### Approved documents

The documents reviewed and approved were:

Document	Version	Date
Letter from sponsor		
REC Application Form [REC_Form_16022016]		16 February 2016
Research protocol or project proposal	1	01 February 2016
Summary CV for Chief Investigator (CI)		

### Membership of the Proportionate Review Sub-Committee

The members of the Sub-Committee who took part in the review are listed on the attached sheet.

### Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

### After ethical review

#### Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

#### User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: <http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/>

#### HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at <http://www.hra.nhs.uk/hra-training/>

With the Committee's best wishes for the success of this project.

<b>16/NW/0152</b>	<b>Please quote this number on all correspondence</b>
-------------------	---

Yours sincerely



**Dr Lisa Booth**  
Chair

Email: nrescommittee.northwest-lancaster@nhs.net

*Enclosures: List of names and professions of members who took part in the review*

*"After ethical review – guidance for researchers"*

*Copy to: Dr Sharief Aboelmagd  
Karen Baucutt, Research and Development Office*

#### North West - Lancaster Research Ethics Committee

##### Attendance at PRS Sub-Committee of the REC meeting on 22 February 2016

#### Committee Members:

Name	Profession	Present	Notes
Dr Lisa Booth	Senior Lecturer / Chair	Yes	
Mrs Gillian Rimington	Paralegal	Yes	
Professor Jois Stansfield	Professor of Speech Pathology	Yes	

#### Also in attendance:

Name	Position (or reason for attending)
Mrs Carol Ebenezer	REC Manager

## Appendix 7. R&D Approval



Norfolk and Norwich University Hospitals **NHS**  
NHS Foundation Trust

Dr Sharief Aboelmagd  
Radiology Academy  
Norfolk and Norwich University Hospital  
Cotman Centre,  
Colney Lane  
Norwich  
NR4 7UY

Research & Development Office  
Level 3 East  
Norfolk & Norwich University Hospitals NHS Foundation Trust  
Colney Lane  
Norwich  
NR4 7UY  
Direct dial: 01603 287378  
Internal: 3378  
E-mail: [rdoffice@nnuh.nhs.uk](mailto:rdoffice@nnuh.nhs.uk)  
Website: [www.nnuh.nhs.uk](http://www.nnuh.nhs.uk)

29<sup>th</sup> April 2016

Dear Dr Aboelmagd

**Re: IRAS Reference Number:** 198605  
**R&D Reference Number:** 198605 (28-02-16)  
**Project Title:** Osteoarthritis of the Ankle MRI Score  
**Sponsor:** Norfolk & Norwich University Hospitals NHS Foundation Trust

I am pleased to inform you that the above Non CTIMP project has been given full NHS permission for research at Norfolk & Norwich University Hospitals NHS Foundation Trust.

Please note that only NNUH employed staff with current (within 2 years) GCP training are permitted to work on this study.

This NHS permission for research has been granted on the basis described in the application form, protocol and supporting documentation as listed below:

Document	Version / Date	Research Ethics Approval
Protocol	V1 1/2/2016	22/2/2016
REC Approval	16/NW/0152	22/2/2016
Proportionate Review		
IRAS R&D form	198605/928388/14/232	
NHS SSI Form	198605/943175/6/221/311249/344580 22/03/2016	

The agreed total local recruitment target for your study is 45 participants.

To support requirements of the National Institute of Health Research (NIHR) we will be monitoring and publishing outcomes of recruitment into your study. This includes benchmarking against a 70 day period from the time of receipt of a valid research application to this time of recruitment of the first patient for your study.

The date of receipt of a valid application for this study is 22/03/2016 and the benchmark of 70 days to recruit the first patient is 30/05/2016.

The investigator agrees to notify the R&D department when the first patient is enrolled/consented into the study. Wherever the duration exceeds 70 days of the Trust receiving a valid research application, the investigator will be expected to explain the reason for the delay in writing.

Version 1

18/12/2015



Norfolk and Norwich University Hospitals **NHS**  
NHS Foundation Trust

Please note, under the agreed Standard Terms and Conditions for Hosted / Sponsored studies you must inform the R&D department of any proposed changes to this study and submit annual progress reports to the R&D department.

If you have any queries regarding this or any other project please contact Karen Baucutt, Research Facilitator, at the above address. Please note, the reference number for this study is 198605 (28-02-16) and this should be quoted on all correspondence.

Yours sincerely

*pp hchatterjee*

Professor Marcus Flather  
R&D Director

## Appendix 8. Ankle MRI protocols

Two 1.5 GE machines and one 3T GE machine

### GE HDXT 1.5T

AX T2 FRFSE-XL (04:12). FOV: 18. Ph FOV: 0.80. Slice thickness/gap: 4.0 x 1.0. Freq direction: AP TR: 4583. TE: 85. ETL: 16. Matrix: 320x416. NEX: 3. BW: 31:25.  
 Sag T1 FSE-XL (03:24). FOV: 18. Ph FOV: NPW. Slice thickness/gap: 3.0x1.0. Freq direction: SI. TR: 461. TE: Min Full. ETL: 3. Matrix: 320x320. NEX: 2. BW: 31:25  
 Sag T2 FATSAT FRFSE-XL (04:33). FOV: 18. Ph FOV: NPW Slice thickness/gap: 3.0x1.0. Freq direction: SI. TR: 4483. TE: 85. ETL: 16. Matrix: 320x320. NEX: 3. BW: 31:25.  
 Cor PD FRFSE-XL (04:34). FOV: 16. Ph FOV: 0.80. Slice thickness/gap: 3.0x1.0. Freq direction: SI. TR: 2981. TE: 35. ETL: 7. Matrix: 256x384. NEX: 3. BW: 35:71  
 Cor STIR (05:19). FOV: 16. Ph FOV: 0.80. Slice thickness/gap: 3.0x 1.0. Freq direction: SI. TR: 4300. TE: 50. ETL: 10. Matrix: 224x256. NEX: 2. BW: 20:83. TI: 140

### GE HDe 1.5T

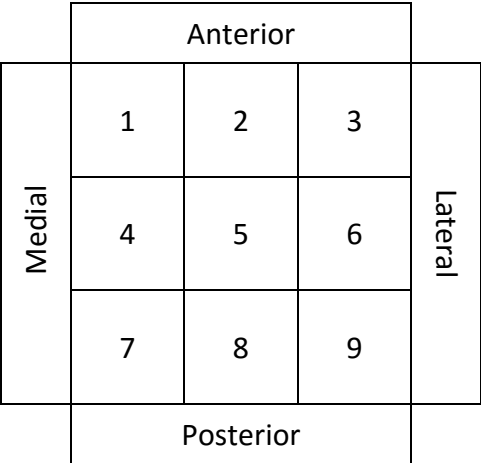
Ax T2 FRFSE (2:07) (20 slices). FOV: 16. Ph FOV: 0.80. 4.0 x 1.0. Freq direction A/P (Ph direct R/L). Auto TR: 4075. TE: 102. 20 slices. ETL: 19. 320x 224. Nex: 3. BW: 25  
 Sag T1 FSE (2:26) (20 slices). FOV: 16.0. Ph FOV: 1.0. 3.0x1.0. Freq dir S/I. (Ph direct A/P). TR: 413. TE: min full. ET: L3. 352x256. NEX 2. BW25  
 Sag T2FS FSE (3:18). Freq FOV 16. Ph FOV 1.0. 3.0x1.0. Freq direct S/I. (Ph direct A/P). TR: 3947. TE: 85. ETL:16. 352x256. NEX: 3. BW: 25  
 Cor PD FSE (2:07) (24 slices). Freq FOV: 16. Ph FOV: 0.70. 3.0x1.0. Freq direction S/I. (Ph direct R/L). TR: 2000. TE: 30. ETL: 6. 352x256. NEX: 2. BW: 25  
 Cor STIR FSE (4:58). Freq FOV: 17. Ph FOV: 0.70. 3.0x 1.0. Freq direct S/I. (Ph direct R/L). TR: 3542. TE: 35. ETL: 8. 320x224. NEX: 2. BW: 12:50

### GE 750W Discovery 3T

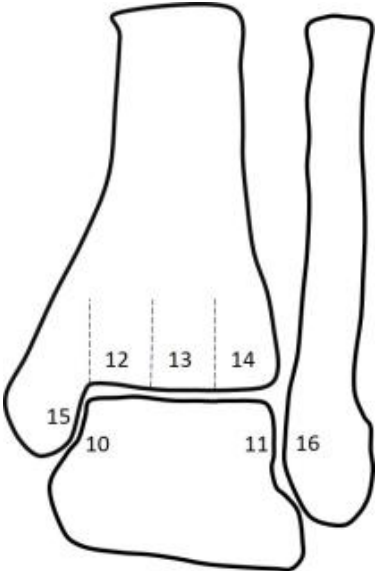
AX T2 FRFSE-XL (05:11). FOV: 14. Ph FOV: NPW. Slice thickness/gap: 4.0 x 1.0. Freq direction: AP. TR: 5088. TE: 70. ETL: 16. Matrix: 320x448. NEX: 3. BW: 50  
 Sag T1 FSE-XL (03:09). FOV: 16. Ph FOV: NPW. Slice thickness/gap: 3.0x1.0. Freq direction: SI. TR: 400. TE: Min Full. ETL: 2. Matrix: 448x512. NEX: 2. BW: 62.5  
 Sag T2 FATSAT FRFSE-XL (03:18). FOV: 16. Ph FOV: NPW Slice thickness/gap: 3.0x1.0. Freq direction: SI. TR: 3947. TE: 85 ETL: 16. Matrix: 320x384 NEX: 3. BW: 31:25  
 Cor PD FRFSE-XL (04:28). FOV: 14. Ph FOV: NPW. Slice thickness/gap: 3.0x1.0. Freq direction: SI. TR: 3264. TE: 30. ETL: 7. Matrix: 256x416. NEX: 2. BW: 62.5  
 Cor STIR (05:18). FOV: 14. Ph FOV: NPW. Slice thickness/gap: 3.0x 1.0. Freq direction: SI. TR: 7750 TE: 40 ETL: 6 Matrix: 224x256. NEX: 1 BW: 31:25 TI: 170

Appendix 9. MRI Scoring sheet 1

Zones



Talar Dome divisions



Zones 10-16

Cartilage Score

Grade	Description
0	Normal cartilage
1	Abnormal signal of morphologically normal cartilage
2A	Superficial partial thickness cartilage defect <50% of total articular thickness.
2B	Superficial partial thickness cartilage defect >50% of total articular thickness.
3	Full thickness cartilage defect exposed bone
4	Chondral injury with a bony component

## Appendix 9. MRI Scoring sheet 2

Bone marrow lesions (BML) refers to any subchondral bone marrow signal abnormality – includes both BMO and subchondral cysts.

Bone marrow oedema (BMO) refers to ill-defined subchondral bone marrow increased signal on fluid sensitive sequences.

Cyst refers to well defined subchondral bone marrow signal increased signal on fluid sensitive sequences.

Osteophyte are recorded as present or absent for the joint.

### Scoring Table

Zone	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Cartilage																
BML																
BMO																
Cyst																

Cartilage Total Score:

BMO Total Score:

Cyst Total Score:

Osteophytes: Present ☐ Absent ☐

Case ID \_\_\_\_\_



## Appendix 10. Presentations

The Delphi survey was presented as an electronic poster presentation at the summer meeting of the European Society of Musculoskeletal Radiology (ESSR) meeting in Zurich, Switzerland. June 9<sup>th</sup>–11<sup>th</sup>.

The radiographic reliability study will be presented as a poster presentation at the European Congress of Radiology in Vienna, Austria. March 1<sup>st</sup>-5<sup>th</sup> 2017.

The MRI reliability study will be presented as an oral presentation at the European Congress of Radiology in Vienna, Austria. March 1<sup>st</sup>-5<sup>th</sup> 2017.

## References

1. Pope T, Bloem HL, Beltran J, Morrison WB, Wilson DJ. Normal Ankle and Foot. In: Imaging of the Musculoskeletal System, 2-Volume Set: Expert Radiology Series, 1e. 1 Har/Cdr edition. Philadelphia, PA: Saunders; 2008. p. 414–30.
2. Valderrabano V, Hintermann B, Horisberger M, Fung TS. Ligamentous posttraumatic ankle osteoarthritis. *Am J Sports Med*. 2006 Apr;34(4):612–20.
3. Grunfeld R, Aydogan U, Juliano P. Ankle Arthritis. *Med Clin North Am*. 2014 Mar;98(2):267–89.
4. Dawe EJC, Davis J. Anatomy and biomechanics of the foot and ankle. *Orthop Trauma*. 2011 Aug 1;25(4):279–86.
5. Golanó P, Vega J, de Leeuw PAJ, Malagelada F, Manzanares MC, Götzens V, et al. Anatomy of the ankle ligaments: a pictorial essay. *Knee Surg Sports Traumatol Arthrosc*. 2010 May;18(5):557–69.
6. Mengiardi B, Pfirrmann CWA, Vienne P, Hodler J, Zanetti M. Medial Collateral Ligament Complex of the Ankle: MR Appearance in Asymptomatic Subjects. *Radiology*. 2007 Mar 1;242(3):817–24.
7. Milner CE, Soames RW. The medial collateral ligaments of the human ankle joint: anatomical variations. *Foot Ankle Int*. 1998;19(5):289–292.
8. Hermans JJ, Beumer A, De Jong TAW, Kleinrensink G-J. Anatomy of the distal tibiofibular syndesmosis in adults: a pictorial essay with a multimodality approach: Anatomy of the distal tibiofibular syndesmosis. *J Anat*. 2010 Dec;217(6):633–45.
9. Snedeker JG, Wirth SH, Espinosa N. Biomechanics of the Normal and Arthritic Ankle Joint. *Foot Ankle Clin*. 2012 Dec;17(4):517–28.

10. Hsu JD, Michael J, Fisk J. Biomechanics of the foot Andrew Haskell and Roger A. Mann. In: AAOS Atlas of Orthoses and Assistive Devices, 4e. 4 edition. Philadelphia: Mosby; 2008. p. 311–24.
11. Miller MD, Thompson SR. Foot and Ankle Biomechanics. In: DeLee & Drez's Orthopaedic Sports Medicine: 2-Volume Set, 4e. 4 edition. Saunders; 2014. p. 1313–21.
12. Nuki G. Osteoarthritis: a problem of joint failure. *Z Für Rheumatol*. 1999;58(3):142–147.
13. Szebenyi B, Hollander AP, Dieppe P, Quilty B, Duddy J, Clarke S, et al. Associations between pain, function, and radiographic features in osteoarthritis of the knee. *Arthritis Rheum*. 2006 Jan;54(1):230–5.
14. Standardization of Osteoarthritis Definitions [Internet]. Osteoarthritis Research Society International. [cited 2016 Nov 25]. Available from: <https://www.oarsi.org/research/standardization-osteoarthritis-definitions>
15. Hunter DJ, Felson DT. Osteoarthritis. *BMJ*. 2006 Mar 16;332(7542):639–42.
16. Brandt KD, Dieppe P, Radin EL. Etiopathogenesis of Osteoarthritis. *Rheum Dis Clin N Am*. 2008 Aug;34(3):531–59.
17. Hunter DJ. Imaging Insights on the Epidemiology and Pathophysiology of Osteoarthritis. *Rheum Dis Clin N Am*. 2009 Aug;35(3):447–63.
18. Ulrich-Vinther M, Maloney MD, Schwarz EM, Rosier R, O'Keefe RJ. Articular cartilage biology. *J Am Acad Orthop Surg*. 2003;11(6):421–430.
19. Kalunian, Kenneth C, Ritter, Susan. Pathogenesis of osteoarthritis [Internet]. Pathogenesis of osteoarthritis. Up to Date. Wolters Kluwer. [cited 2015 May 20]. Available from: <http://www.uptodate.com/contents/pathogenesis-of-osteoarthritis>

20. Buckwalter JA. Osteoarthritis and articular cartilage use, disuse, and abuse: experimental studies. *J Rheumatol Suppl.* 1995 Feb;43:13–5.
21. Grumbles RM, Howell DS, Howard GA, Roos BA, Setton LA, Mow VC, et al. Cartilage metalloproteases in disuse atrophy. *J Rheumatol Suppl.* 1995 Feb;43:146–8.
22. Fitzgerald JB, Jin M, Dean D, Wood DJ, Zheng MH, Grodzinsky AJ. Mechanical Compression of Cartilage Explants Induces Multiple Time-dependent Gene Expression Patterns and Involves Intracellular Calcium and Cyclic AMP. *J Biol Chem.* 2004 May 7;279(19):19502–11.
23. Lee JH, Fitzgerald JB, DiMicco MA, Grodzinsky AJ. Mechanical injury of cartilage explants causes specific time-dependent changes in chondrocyte gene expression. *Arthritis Rheum.* 2005 Aug 1;52(8):2386–95.
24. Li B, Aspden R. Mechanical and material properties of the subchondral bone plate from the femoral head of patients with osteoarthritis or osteoporosis. *Ann Rheum Dis.* 1997 Apr;56(4):247–54.
25. Burr DB. The importance of subchondral bone in the progression of osteoarthritis. *J Rheumatol.* 2004;70:77–80.
26. Monfort J, Garcia-Giralt N, López-Armada MJ, Monllau JC, Bonilla A, Benito P, et al. Decreased metalloproteinase production as a response to mechanical pressure in human cartilage: a mechanism for homeostatic regulation. *Arthritis Res Ther.* 2006 Sep 14;8(5):R149.
27. Lohmander LS, Hoerrner LA, Lark MW. Metalloproteinases, tissue inhibitor, and proteoglycan fragments in knee synovial fluid in human osteoarthritis. *Arthritis Rheum.* 1993 Feb 1;36(2):181–9.

28. Smith MD, Triantafillou S, Parker A, Youssef PP, Coleman M. Synovial membrane inflammation and cytokine production in patients with early osteoarthritis. *J Rheumatol*. 1997 Feb;24(2):365–71.
29. Denko CW, Boja B, Moskowitz RW. Growth promoting peptides in osteoarthritis: insulin, insulin-like growth factor-1, growth hormone. *J Rheumatol*. 1990 Sep;17(9):1217–21.
30. Ekenstedt KJ, Sonntag WE, Loeser RF, Lindgren BR, Carlson CS. Effects of chronic growth hormone and insulin-like growth factor 1 deficiency on osteoarthritis severity in rat knee joints. *Arthritis Rheum*. 2006 Dec 1;54(12):3850–8.
31. Myers SL, Brandt KD, Ehlich JW, Braunstein EM, Shelbourne KD, Heck DA, et al. Synovial inflammation in patients with early osteoarthritis of the knee. *J Rheumatol*. 1990 Dec;17(12):1662–9.
32. Hurley MV. Muscle, exercise and arthritis. *Ann Rheum Dis*. 2002 Aug 1;61(8):673–5.
33. Sharma L, Song J, Felson DT, Cahue S, Shamiyeh E, Dunlop DD. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. *JAMA*. 2001 Jul 11;286(2):188–95.
34. Cerejo R, Dunlop DD, Cahue S, Channin D, Song J, Sharma L. The influence of alignment on risk of knee osteoarthritis progression according to baseline stage of disease. *Arthritis Rheum*. 2002 Oct 1;46(10):2632–6.
35. Sophia Fox AJ, Bedi A, Rodeo SA. The Basic Science of Articular Cartilage. *Sports Health*. 2009 Nov;1(6):461–8.
36. Goldring SR, Goldring MB. Clinical aspects, pathology and pathophysiology of osteoarthritis. *J Musculoskelet Neuronal Interact*. 2006;6(4):376.

37. Goldberg AJ, MacGregor A, Dawson J, Singh D, Cullen N, Sharp RJ, et al. The demand incidence of symptomatic ankle osteoarthritis presenting to foot & ankle surgeons in the United Kingdom. *The Foot*. 2012 Sep;22(3):163–6.
38. Zaidi R, Cro S, Gurusamy K, Siva N, Macgregor A, Henricson A, et al. The outcome of total ankle replacement A systematic review and meta-analysis. *Bone Jt J*. 2013 Nov 1;95–B(11):1500–7.
39. Saltzman CL, Salamon ML, Blanchard GM, Huff T, Hayes A, Buckwalter JA, et al. Epidemiology of Ankle Arthritis. *Iowa Orthop J*. 2005;25:44–6.
40. Valderrabano V, Horisberger M, Russell I, Dougall H, Hintermann B. Etiology of Ankle Osteoarthritis. *Clin Orthop*. 2009 Jul;467(7):1800–6.
41. DiStefano JG, Pinney S. Ankle Arthritis: Etiology and Epidemiology. *Semin Arthroplasty*. 2010 Dec 1;21(4):218–22.
42. American Orthopaedic Foot and Ankle Society. Ankle arthritis. *OrthopaedicsOne* [Internet]. 2011 Jun [cited 2016 Oct 21]; Available from: <http://www.orthopaedicsone.com/display/Main/Ankle+arthritis>
43. Tellisi N, Fragomen AT, Kleinman D, O'Malley MJ, Rozbruch SR. Joint Preservation of the Osteoarthritic Ankle Using Distraction Arthroplasty. *Foot Ankle Int*. 2009 Apr;30(4):318–25.
44. Fuchs S, Sandmann C, Skwara A, Chylarecki C. Quality of life 20 years after arthrodesis of the ankle. *J Bone Jt Surg*. 2003 Sep 1;85(7):994–8.
45. Hintermann B. Ankle arthrodesis. *Total Ankle Arthroplasty Hist Overv Curr Concepts Future Perspect*. 2005;11–24.
46. Roemer FW, Crema MD, Trattnig S, Guermazi A. Advances in imaging of osteoarthritis and cartilage. *Radiology*. 2011;260(2):332–354.

47. Hing CB, Harris MA, Ejindu V, Sofat N. The Application of Imaging in Osteoarthritis. In: Rothschild BM, editor. Principles of Osteoarthritis- Its Definition, Character, Derivation and Modality-Related Recognition [Internet]. InTech; 2012 [cited 2014 Nov 15]. Available from: <http://www.intechopen.com/books/principles-of-osteoarthritis-its-definition-character-derivation-and-modality-related-recognition/the-application-of-imaging-in-osteoarthritis>
48. Amin S, LaValley MP, Guermazi A, Grigoryan M, Hunter DJ, Clancy M, et al. The relationship between cartilage loss on magnetic resonance imaging and radiographic progression in men and women with knee osteoarthritis. *Arthritis Rheum*. 2005 Oct;52(10):3152–9.
49. Chan WP, Lang P, Stevens MP, Sack K, Majumdar S, Stoller DW, et al. Osteoarthritis of the knee: comparison of radiography, CT, and MR imaging to assess extent and severity. *AJR Am J Roentgenol*. 1991 Oct;157(4):799–806.
50. Guermazi A, Hayashi D, Crema MD, Roemer FW. Current Trends in Osteoarthritis Imaging – An Update from a Radiological Viewpoint. *Eur Musculoskelet Rev*. 2010;5:30–5.
51. Hoving JL, Buchbinder R, Hall S, Lawler G, Coombs P, McNealy S, et al. A comparison of magnetic resonance imaging, sonography, and radiography of the hand in patients with early rheumatoid arthritis. *J Rheumatol*. 2004 Apr 1;31(4):663–75.
52. Bajaj S, Lopez-Ben R, Oster R, Alarcón GS. Ultrasound detects rapid progression of erosive disease in early rheumatoid arthritis: a prospective longitudinal study. *Skeletal Radiol*. 2006 Oct 11;36(2):123–8.
53. Guermazi A, Hayashi D, Eckstein F, Hunter DJ, Duryea J, Roemer FW. Imaging of Osteoarthritis. *Rheum Dis Clin N Am*. 2013 Feb;39(1):67–105.

54. Hutton CW, Higgs ER, Jackson PC, Watt I, Dieppe PA. 99mTc HMDP bone scanning in generalised nodal osteoarthritis. II. The four hour bone scan image predicts radiographic change. *Ann Rheum Dis.* 1986 Aug;45(8):622–6.
55. Dieppe P, Cushnaghan J, Young P, Kirwan J. Prediction of the progression of joint space narrowing in osteoarthritis of the knee by bone scintigraphy. *Ann Rheum Dis.* 1993 Aug 1;52(8):557–63.
56. Omoumi P, Mercier GA, Lecouvet F, Simoni P, Vande Berg BC. CT Arthrography, MR Arthrography, PET, and Scintigraphy in Osteoarthritis. *Radiol Clin North Am.* 2009 Jul;47(4):595–615.
57. Crema MD, Roemer FW, Marra MD, Burstein D, Gold GE, Eckstein F, et al. Articular cartilage in the knee: current MR imaging techniques and applications in clinical practice and research. *Radiogr Rev Publ Radiol Soc N Am Inc.* 2011 Feb;31(1):37–61.
58. Braun HJ, Gold GE. Diagnosis of Osteoarthritis: Imaging. *Bone.* 2012 Aug;51(2):278–88.
59. Fischer, Guermazi, Roemer, Carrino, Crema, Grainger, et al. MRI-Essentials. An illustrated atlas of orthopedic MRI. 1st edition. mri-publisher.com; 2014. 640 p.
60. Hunter DJ, Zhang Y, Niu J, Goggins J, Amin S, LaValley MP, et al. Increase in bone marrow lesions associated with cartilage loss: A longitudinal magnetic resonance imaging study of knee osteoarthritis. *Arthritis Rheum.* 2006 May 1;54(5):1529–35.
61. Guermazi A, Roemer FW, Hayashi D. Imaging of osteoarthritis: update from a radiological perspective. *Curr Opin Rheumatol.* 2011 Sep;23(5):484–91.



62. Neogi T, Nevitt M, Niu J, Sharma L, Roemer F, Guermazi A, et al. Subchondral bone attrition may be a reflection of compartment-specific mechanical load: the MOST Study. *Ann Rheum Dis*. 2010 May;69(5):841–4.
63. Shapiro LM, McWalter EJ, Son M-S, Levenston M, Hargreaves BA, Gold GE. Mechanisms of osteoarthritis in the knee: MR imaging appearance. *J Magn Reson Imaging JMRI*. 2014 Jun;39(6):1346–56.
64. Link TM. *Cartilage Imaging: Significance, Techniques, and New Developments*. Springer; 2011. 253 p.
65. Kijowski R, Blankenbaker DG, Davis KW, Shinki K, Kaplan LD, De Smet AA. Comparison of 1.5- and 3.0-T MR Imaging for Evaluating the Articular Cartilage of the Knee Joint. *Radiology*. 2009 Mar 1;250(3):839–48.
66. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *Ann Rheum Dis*. 1957 Dec;16(4):494–502.
67. Schiphof D, Boers M, Bierma-Zeinstra SMA. Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. *Ann Rheum Dis*. 2008 Jul 1;67(7):1034–6.
68. Hunter DJ, Le Graverand M-PH, Eckstein F. Radiologic markers of osteoarthritis progression: *Curr Opin Rheumatol*. 2009 Mar;21(2):110–7.
69. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage*. 2007;15, Supplement 1:A1–56.
70. Moon J-S, Shim J-C, Suh J-S, Lee W-C. Radiographic Predictability of Cartilage Damage in Medial Ankle Osteoarthritis. *Clin Orthop*. 2010 Aug;468(8):2188–97.

71. Takakura Y, Tanaka Y, Kumai T, Tamai S. Low tibial osteotomy for osteoarthritis of the ankle. Results of a new operation in 18 patients. *Bone Jt J*. 1995 Jan 1;77-B(1):50–4.
72. Dijk CN van, Verhagen R a. W, Tol JL. Arthroscopy for Problems After Ankle Fracture. *J Bone Joint Surg Br*. 1997 Mar 1;79-B(2):280–4.
73. Marijnissen ACA, Vincken KL, Viergever MA, van Roy HLAM, Van Roermund PM, Lafeber FPJG, et al. Ankle images digital analysis (AIDA): digital measurement of joint space width and subchondral sclerosis on standard radiographs. *Osteoarthritis Cartilage*. 2001 Apr;9(3):264–72.
74. Holzer N, Salvo D, Marijnissen ACA, Vincken KL, Ahmad AC, Serra E, et al. Radiographic evaluation of posttraumatic osteoarthritis of the ankle: the Kellgren–Lawrence scale is reliable and correlates with clinical symptoms. *Osteoarthritis Cartilage*. 2015 Mar;23(3):363–9.
75. Mayich DJ, Pinsker E, Mayich MS, Mak W, Daniels TR. An Analysis of the Use of the Kellgren and Lawrence Grading System to Evaluate Peritalar Arthritis Following Total Ankle Arthroplasty. *Foot Ankle Int*. 2013 Nov 1;34(11):1508–15.
76. Kraus VB, Kilfoil TM, Hash TW, McDaniel G, Renner JB, Carrino JA, et al. Atlas of radiographic features of osteoarthritis of the ankle and hindfoot. *Osteoarthritis Cartilage*. 2015 Dec;23(12):2059–85.
77. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159–74.
78. Roemer FW, Eckstein F, Guermazi A. Magnetic Resonance Imaging-Based Semiquantitative and Quantitative Assessment in Osteoarthritis. *Rheum Dis Clin N Am*. 2009 Aug;35(3):521–55.

79. Peterfy C., Guermazi A, Zaim S, Tirman PF., Miaux Y, White D, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. *Osteoarthritis Cartilage*. 2004 Mar;12(3):177–90.
80. Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston–Leeds Osteoarthritis Knee Score). *Ann Rheum Dis*. 2008 Feb 1;67(2):206–11.
81. Kornaat PR, Ceulemans RYT, Kroon HM, Riyazi N, Kloppenburg M, Carter WO, et al. MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)?inter-observer and intra-observer reproducibility of a compartment-based scoring system. *Skeletal Radiol*. 2005 Feb;34(2):95–102.
82. Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). *Osteoarthritis Cartilage*. 2011 Aug;19(8):990–1002.
83. Felson DT, Chaisson CE, Hill CL, Totterman SM, Gale ME, Skinner KM, et al. The association of bone marrow lesions with pain in knee osteoarthritis. *Ann Intern Med*. 2001 Apr 3;134(7):541–9.
84. Lynch JA, Roemer FW, Nevitt MC, Felson DT, Niu J, Eaton CB, et al. Comparison of BLOKS and WORMS scoring systems part I. Cross sectional comparison of methods to assess cartilage morphology, meniscal damage and bone marrow lesions on knee MRI: data from the osteoarthritis initiative. *Osteoarthritis Cartilage*. 2010 Nov;18(11):1393–401.
85. Felson DT, Lynch J, Guermazi A, Roemer FW, Niu J, McAlindon T, et al. Comparison of BLOKS and WORMS scoring systems part II. Longitudinal assessment of knee MRIs for osteoarthritis and suggested approach based

- on their performance: data from the Osteoarthritis Initiative.  
Osteoarthritis Cartilage. 2010 Nov;18(11):1402–7.
86. Park H-J, Kim SS, Lee S-Y, Park N-H, Park J-Y, Choi Y-J, et al. A practical MRI grading system for osteoarthritis of the knee: Association with Kellgren–Lawrence radiographic scores. *Eur J Radiol*. 2013 Jan;82(1):112–7.
  87. Hayes CW, Jamadar DA, Welch GW, Jannausch ML, Lachance LL, Capul DC, et al. Osteoarthritis of the Knee: Comparison of MR Imaging Findings with Radiographic Severity Measurements and Pain in Middle-aged Women. *Radiology*. 2005 Dec 1;237(3):998–1007.
  88. Roemer FW, Hunter DJ, Winterstein A, Li L, Kim YJ, Cibere J, et al. Hip Osteoarthritis MRI Scoring System (HOAMS): reliability and associations with radiographic and clinical findings. *Osteoarthritis Cartilage*. 2011 Aug;19(8):946–62.
  89. Lee S, Nardo L, Kumar D, Wyatt CR, Souza RB, Lynch J, et al. Scoring hip osteoarthritis with MRI (SHOMRI): A whole joint osteoarthritis evaluation system. *J Magn Reson Imaging JMRI*. 2014 Aug 20;
  90. Campbell MJ, Swinscow TDV. Correlation and regression. In: *Statistics at Square One*. 11th Revised edition edition. Chichester, UK ; Hoboken, NJ: Wiley-Blackwell; 2009. p. Chapter 9.
  91. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assess Res Eval*. 2007;12(10):1–8.
  92. Cuhls K. The Delphi Method. Delphi Surveys. In: *Teaching material for UNIDO foresight seminars*. 2005. p. 93–112.
  93. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257–268.

94. Hong H, Choi Y, Hahn S, Park SK, Park B-J. Nomogram for sample size calculation on a straightforward basis for the kappa statistic. *Ann Epidemiol*. 2014 Sep;24(9):673–80.
95. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica*. 2012 Oct 15;22(3):276–82.
96. Bland M. *An Introduction to Medical Statistics*. 4 edition. Oxford: Oup Oxford; 2015. 448 p.
97. Raikin SM, Elias I, Zoga AC, Morrison WB, Besser MP, Schweitzer ME. Osteochondral Lesions of the Talus: Localization and Morphologic Data from 424 Patients Using a Novel Anatomical Grid Scheme. *Foot Ankle Int*. 2007 Feb 1;28(2):154–61.
98. McGahan PJ, Pinney SJ. Current Concept Review: Osteochondral Lesions of the Talus. *Foot Ankle Int*. 2010 Jan;31(1):90–101.
99. Loveday D. Personal communication at British Society of Skeletal Radiology refresher course. Norwich. 2016.
100. Recht MP, Piraino DW, Paletta GA, Schils JP, Belhobek GH. Accuracy of fat-suppressed three-dimensional spoiled gradient-echo FLASH MR imaging in the detection of patellofemoral articular cartilage abnormalities. *Radiology*. 1996;198(1):209–212.
101. Saifuddin A. *Musculoskeletal MRI*. 1 edition. London: CRC Press; 2008. 1232 p.
102. Forney M, Subhas N, Donley B, Winalski CS. MR imaging of the articular cartilage of the knee and ankle. *Magn Reson Imaging Clin N Am*. 2011 May;19(2):379–405.
103. Brittberg M, Winalski CS. Evaluation of Cartilage Injuries and Repair. *J Bone Jt Surg Am*. 2003 Apr 1;85(suppl 2):58–69.

104. Zanetti M, Toms A. Personal communication at British Society of Skeletal Radiology refresher course. Norwich. 2016.
105. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
106. Matthias Gamer, Jim Lemon, Ian Fellows Puspendra Singh. (2012). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84. <https://CRAN.R-project.org/package=irr>.
107. Revelle, W. (2016) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <http://CRAN.R-project.org/package=psych> Version = 1.6.6.
108. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.
109. Gwet K. Kappa statistic is not satisfactory for assessing the extent of agreement between raters. Stat Methods Inter-Rater Reliab Assess. 2002;1(6):1–6.
110. Crichton N. Information point. Kendalls Tau. J Clin Nurs. 2001 Sep 15;10(5):715.
111. RANDOM.ORG - True Random Number Service [Internet]. [cited 2016 Dec 14]. Available from: <https://www.random.org/>