Characterising novel imaging biomarkers for use in knee osteoarthritis clinical trials

Bright Dube

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Leeds Institute of Rheumatic and Musculoskeletal Medicine

School of Medicine

October 2020

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The publications are as follows

Dube B, Bowes MA, Kingsbury SR, Hensor EMA, Muzumdar S, Conaghan PG. *Where does meniscal damage progress most rapidly? An analysis using three-dimensional shape models on data from the Osteoarthritis Initiative*. Osteoarthritis and Cartilage 2018; 26: 62-71.

Dube B, Bowes MA, Hensor EMA, Barr A, Kingsbury SR, Conaghan PG. *The relationship between two different measures of osteoarthritis bone pathology, bone marrow lesions and 3D bone shape: data from the Osteoarthritis Initiative*. Osteoarthritis and Cartilage 2018.

Acknowledgements

There are many people I would like to thank for making this thesis a success. Firstly, I would like to acknowledge the guidance and support from my supervisors Professor Philip Conaghan, Dr Sarah Kingsbury and Dr Elizabeth Hensor. They were invaluable in helping me secure the research scholarship to complete my doctorate. They were also available at all times to provide advice, critique, academic oversight and encouragement throughout the PhD process. I am eternally grateful for their support.

Secondly, I would like to thank the University of Leeds for awarding me one of the University of Leeds 110 Anniversary Research Scholarships that funded the research for this PhD. I would also like to thank Dr Michael Bowes for providing the 3D imaging data underpinning all the work in this thesis, without which this work would not have been possible.

For the meniscus study in Chapter 4, I would like to thank Dr Michael Bowes, Professor Conaghan, Dr Sarah Kingsbury, Dr Elizabeth Hensor and Siddhant Muzumdar for helping in the design of the study and the writing of manuscript. For the BML study in Chapter 5, I would like to thank Professor Conaghan, Dr Sarah Kingsbury, Dr Elizabeth Hensor, Dr Andrew Barr and Dr Michael Bowes for assisting in the study design and manuscript writing. For the work in Chapters 6 and 7, I would like to thank Professor Conaghan, Dr Sarah Kingsbury, Dr Elizabeth Hensor and Dr Michael Bowes for assisting in the design of the study and guidance with writing the Chapters.

Finally I would like to thank my wife Precious, for her love, patience and understanding while I undertook this thesis, including very busy times after the birth of our daughter Christine, during my doctorate. Thanks also goes to my son Michael for always asking when each Chapter would be complete which kept me motivated. I would also like to acknowledge my parents, and brothers, Thabo, Tsepo and Mircea, for their encouragement and support throughout.

Co-author contributions:
This research has been carried out by a team with my own contributions and the contributions of the other members explicitly indicated below:

Chapter 4: Where does meniscal damage progress most rapidly? An analysis using three-dimensional shape models on data from the Osteoarthritis Initiative

BD, MB and PGC made substantial contributions to the conception and design of the study, data collection, statistical analysis and interpretation, and manuscript writing. SK, SM and EH helped with the analysis, interpretation of data and manuscript writing. BD wrote the Chapter.

Chapter 5: The relationship between two different measures of osteoarthritis bone pathology, bone marrow lesions and 3D bone shape

BD, PGC and MB made substantial contribution to the conception and design of the study, data acquisition, data interpretation and writing of the manuscript. BD undertook the statistical analysis.EH participated in the design of the study, provided guidance for the statistical analysis and data interpretation. SK made contributions to the design of the study and manuscript writing. AB contributed to the data interpretation. BD wrote the Chapter.

Chapter 6: Determinants of osteoarthritis 3D bone shape and its change in the three knee bones: a latent growth modelling approach on 37,583 MR images from the Osteoarthritis Initiative.

BD made substantial contributions to the conception and design of the study, data acquisition, undertook the statistical analysis, data interpretation and writing of the Chapter.EH provided guidance for the statistical analysis and data interpretation. SK made contributions to the design of the study and Chapter writing. MB and PGC were involved in conception and design of the work, data analysis and interpretation. BD wrote the Chapter.

Chapter 7: Structure-based trajectories in the three knee bones: data from the Osteoarthritis Initiative.
BD made substantial contributions to the conception and design of the study, data acquisition, undertook the statistical analysis, data interpretation and writing of the Chapter.EH provided guidance for the statistical analysis and data interpretation. SK made contributions to the design of the study and Chapter writing. MB and PGC were

involved in conception and design of the work, data analysis and interpretation. BD wrote the Chapter.

List of publications / presentations arising from the thesis

Papers published

Dube B, Bowes MA, Kingsbury SR, Hensor EMA, Muzumdar S, Conaghan PG. *Where does meniscal damage progress most rapidly? An analysis using three-dimensional shape models on data from the Osteoarthritis Initiative*. Osteoarthritis and Cartilage 2018; 26: 62-71.

Dube B, Bowes MA, Hensor EMA, Barr A, Kingsbury SR, Conaghan PG. *The relationship between two different measures of osteoarthritis bone pathology, bone marrow lesions and 3D bone shape: data from the Osteoarthritis Initiative.* Osteoarthritis and Cartilage 2018.

International Conference Poster Presentations

Dube B, Bowes MA, Barr AJ, Hensor EM, Kingsbury SR, Conaghan PG. *The relationship between two different measures of osteoarthritis bone pathology, bone marrow lesions and 3D bone shape: data from the osteoarthritis initiative*. Osteoarthritis and Cartilage 2018; 26: S444.Osteoarthritis Research Society International (OARSI) Annual Conference, Liverpool, 2018

Dube B, Bowes MA, Hensor EM, Kingsbury SR, Conaghan PG. *Determinants of osteoarthritis 3D bone shape and its change in the three knee bones: a latent growth modelling approach on 37 583 MR images from the osteoarthritis initiative*. Osteoarthritis and Cartilage 2019; 27: S391-S392. Osteoarthritis Research Society International (OARSI) Annual Conference, Toronto, 2019

Dube B, Bowes M, Hensor E, Kingsbury S, Conaghan P. *What Is the Relationship Between the 3 Knee Bones in Osteoarthritis? Baseline and Longitudinal Associations Using a Latent Growth Modelling Approach on 37,583 MR Images from the Osteoarthritis Initiative [abstract]. Arthritis Rheumatol.* 2019; 71 (suppl 10). American College of Rheumatology (ACR) Annual Conference, Atlanta 2019

Invited Conference podium presentations

Responsiveness of 3D MRI measures of bone and meniscus in knee OA

Optimising Knee Therapies EPSRC Symposium, Weetwood Hall Hotel, Leeds, January 11, 2018

Determinants of osteoarthritis 3D bone shape and its change in the three knee bones: a latent growth modelling approach on 37 583 MR images from the Osteoarthritis Initiative

Optimising Knee Therapies EPSRC Programme Grant Scientific Meeting, Weetwood Hall Hotel, Leeds, January 9, 2019

Abstract

Knee osteoarthritis (OA) prevalence is increasing globally. Detailed understanding of structural deterioration in OA knees is hampered by lack of responsive, reliable imaging biomarkers. Magnetic resonance imaging (MRI) has shown that OA pathology involves multiple tissues. Machine-learning based 3D image analysis of MR images accurately quantifies individual tissues and their spatial and temporal relationships. This thesis tested the hypothesis that novel 3D quantitative measures would provide valid imaging biomarkers for knee OA in terms of construct validity, reliability and responsiveness. The Osteoarthritis Initiative provided a unique, large, longitudinal database of knee MRIs to enable detailed and novel statistical analyses of novel imaging biomarkers.

A longitudinal study exploring a range of quantitative meniscal measures in 86 patients demonstrated that two exhibited responsiveness comparable to other MRI outcomes, and better than radiographic JSN. Cross-sectional analysis of 600 participants demonstrated a relationship between two potential bone imaging biomarkers, a relatively well studied bone pathology, bone marrow lesions (BMLs) and a novel 3D bone shape measure. Longitudinally, bone shape was more responsive than BMLs. Latent growth modelling on 37,583 knee measurements established that 3D bone shape changed linearly in all three knee bones, with greatest change in the femur, but all three were influenced by clinical covariates in a similar manner. Parallel process growth models showed that onset and rates of structural deterioration were interrelated among the femur, tibia and patella. Latent class growth analysis revealed that distinct trajectories of structural change exist in knee OA. Knee pain, obesity, ethnicity and knee surgery were associated with classification into the fastest trajectory group.

In summary, novel quantitative imaging biomarkers of meniscus and bone shape are valid knee OA imaging biomarkers. The introduction of these measures should improve understanding of OA structural pathogenesis, improve clinical trial sensitivity and potentially enable better stratification for clinical trial inclusion.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAM	Active Appearance Models
ACR	American College of Rheumatology
AIC	Akaike`s Information Criterion
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
BIC	Bayesian Information Criterion
BLOKS	Boston Leeds Osteoarthritis Knee Score
BLRT	Bootstrap Likelihood Ratio Test
BMD	Bone Mineral Density
BMI	Body Mass Index
BML	Bone Marrow Lesion
CES-D	Center for Epidemiologic Studies Depression Scale
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CI	Confidence Interval
CT	Computerised Tomography
DAG	Directed Acyclic Graph
DAMP	Damage-associated Molecular Patterns
DESS	Double Echo Steady State
DF	Degrees of Freedom
DMOAD	Disease-Modifying OA Drugs
DV	Dependent Variable
ES	Effect Size
EULAR	European League Against Rheumatism

FIML	Full Information Maximum Likelihood
FNIH	Foundation for the NIH Biomarkers Consortium
GEE	Generalised Estimating Equations
GEMRIC	Delayed Gadolinium-enhanced MRI of Cartilage
GMM	Growth Mixture Models
GRoLTS	Guidelines for Reporting on Latent Trajectory Studies
IA	Intra-Articular
IACS	Intra-Articular Corticosteroids
ICC	Intra Class Correlation
IQR	Inter Quartile Range
JSN	Joint Space Narrowing
JSW	Joint Space Width
KL	Kellgren Lawrence
KOOS	Knee Injury And Osteoarthritis Outcome Score
KOSS	Knee Osteoarthritis Scoring System
LCA	Latent Class Analysis
LCGA	Latent Class Growth Analysis
LCGM	Latent Curve Growth Modelling
LDA	Linear Discriminant Analysis
LLI	Limb Length Inequality
LMR-LRT	Lo, Mendell, and Rubin Likelihood Ratio Test
MCID	Minimum Clinically Important Difference
MOAKS	MRI OsteoArthritis Knee Score
MOST	Multicenter Osteoarthritis Study
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NFI	Normed Fit Index

NGF	Neuronal Growth Factor
NHANES	National Health and Nutrition Examination Survey
NIH	National Institute of Health
NSAIDS	Non-Steroidal Anti Inflammatory Drugs
OA	Osteoarthritis
OAI	Osteoarthritis Initiative
OARSI	Osteoarthritis Research Society International
OMERACT	Outcome Measures in Rheumatology
OR	Odds ratio
PASE	Physical Activity Scale for the Elderly
PET	Positron Emission Tomography
PROM	Patient Reported Outcome Measure
QOL	Quality Of Life
RMSEA	Root Mean Square Error of Association
ROA	Radiographic Osteoarthritis
ROC	Receiver Operator Curve
SD	Standard Deviation
SDD	Smallest Detectable Difference
SE	Standard Error
SEM	Structural Equation Modelling
SQ	Semi-Quantitative
SRM	Standardised Response Mean
SRMR	Standardised Root Mean Square Residual
SSM	Statistical Shape Modelling
TFJ	Tibio-Femoral Joint
TKR	Total Knee Replacement
TLI	Tucker Lewis Index

TNF	Tumour Necrosis Factor
VAS	Visual Analogue Scale
WOMAC	Western Ontario MacMaster Universities Osteoarthritis Index
WORMS	Whole-ORgan Magnetic Resonance Imaging Score

Chapter 1 Introduction

1.1 Background

Osteoarthritis (OA) is the most common form of arthritis. Its major symptoms are joint pain and stiffness which negatively impact individuals` ability to perform activities of daily living and substantially reduce quality of life. At least 242 million people globally have hip/knee OA [1], and as the world's population continues to age, coupled with increased obesity prevalence worldwide, it is anticipated that the burden of OA will become a major problem for health systems [1].Current treatment of OA is based on treating symptoms, with non-pharmacological treatments having poor uptake and pharmacological therapies limited by both efficacy and toxicity. There are currently no licenced disease-modifying OA drugs (DMOADs), a putative class of agents that target key tissues involved in OA pathophysiologic processes with an aim of preventing structural progression and consequently improving symptoms [2]. Preliminary evidence from recent DMOAD trials suggests some structural improvement is possible, however concomitant symptomatic improvement has not been shown yet.

The pathogenesis of OA is complex and still poorly understood. Modern imaging studies, especially with magnetic resonance imaging (MRI), have confirmed that, at least in established OA, there are multiple tissue pathologies, leading to the concept of OA as a whole-organ disease. Treatment innovation in OA has been slow [3] and the reasons for this are multifactorial. A lack of understanding of where, or in which tissue, pain arises from in an individual OA joint has hampered appropriate target development and resulted in no new effective analgesic therapies being licensed in many years; recent therapies in development have targeted peripheral nociceptive pathways.

In terms of structural modification, there has been a lack of valid, predictive and responsive soluble and imaging biomarkers [4] which has created a cycle in which larger numbers of people and longer follow-up has been required for OA clinical trials in order to detect any changes, resulting in higher costs to pharmaceutical companies[5, 6] who have therefore become reluctant to pursue such trials [4]. Current regulatory requirements for disease progression are based on conventional radiography (X-rays), with the current gold standard for measuring structural benefits in DMOAD development being radiographic joint space narrowing (JSN). Using JSN outcomes, the integrity, thickness and health of hyaline cartilage is inferred [7]. However JSN is not

specific for articular cartilage and reflects changes in other tissues such as meniscal extrusion and other degenerative changes in the meniscus[8]. Historically the focus of OA structure-modification trials has been a single tissue, cartilage; however OA is a whole-joint problem. The poor relationship between joint structural pathology (based on JSN) and symptomatic disease has also contributed to the failure in DMOAD development [3]. Compounding the problems of measuring structure and also structure-pain associations is that trials have used small numbers over short periods not sufficient to demonstrate true change, and the statistical methodology not robust to study OA natural history.

Previously the quantification of OA images has neither been accurate nor specific owing to the (often single) view obtained from X-rays. MRI semi-quantitative and quantitative measures have been developed and have been recommended for clinical trials of OA, on the basis of their preferable validity and responsiveness[9]. Until recently, only cartilage quantification has been possible, using manual segmentation techniques.

Supervised machine learning, an artificial intelligence methodology, removes the need for prior human input into data interpretation by building models based on “training data” in order to make informed predictions. Using machine learning methods, the appearance of a tissue can be learnt, and that learning applied to automatically find and delineate that tissue in new, unseen images.

Applied to MR image analysis, machine-learning enables accurate, reader-independent quantification and we have previously demonstrated it can measure a characteristic OA 3-dimensional bone shape with good precision. Further developments in imaging technology using machine learning techniques have resulted in the development of novel 3D tissue data from segmented MRI-images through a process of statistical shape modelling (SSM)[10, 11]. Use of machine learning techniques has enabled the accurate quantification of MRI images, and enabled rapid analysis of large cohorts in a short time-frame. Previous work using this technique was applied to cartilage and bone and determined that 3D-MRI bone shape provides the much-needed measure of OA status, demonstrating strong associations with the risk of current and future important clinical outcomes.

This thesis explored the application of this new 3D imaging analysis obtained from machine learning in the largest OA cohort to date, the USA-based Osteoarthritis Initiative (OAI) an ongoing longitudinal multicentre observational cohort of patients with, or at risk of OA with MRI data collected annually. The aim was to characterise and provide validity for these novel measures, to improve understanding of OA

pathogenesis and improve OA clinical trial outcomes. Bone and meniscus were the focus of this work, given limited previous studies using quantitative methods, in an effort to validate these novel measures as imaging biomarkers.

Advanced statistical methodology (latent growth curve analysis and growth mixture modelling) was also employed to fully characterise longitudinal changes. A few analyses performed in the OA field previously and termed longitudinal analyses have largely assessed the association between an outcome recorded for example 2 years after study initiation with a variable at baseline, but this has been shown to be inefficient use of the data. Moreover, where covariates change over time, these effects cannot be captured by these type of analyses.

The hypothesis underlying this thesis was that novel machine-learning derived quantitative, imaging biomarkers will provide valid and responsive measures for understanding OA pathogenesis and for use in subsequent knee OA clinical trials.

1.2 Structure of the thesis

The following is an outline of the thesis structure.

1.2.1 Chapter 2: Literature review

This narrative literature review summarises aspects of OA with a focus on the need to assess interrelationships between different knee tissues and applying a whole organ approach for knee OA, with consideration of the role of semi-quantitative and quantitative imaging analysis. The unique resources of this thesis are the Osteoarthritis Initiative (OAI) and quantitative imaging of meniscus and bone.

1.2.2 Chapter 3: Methods

Chapter 3 details the methodologies that were used in studies included in this thesis. To characterise different imaging biomarkers fully, a mixed statistical approach using quantitative methods was adopted. 3D data derived from machine learning techniques used in this thesis was obtained from MRI data from the OAI database as described before, set up in the USA and set up to focus primarily on knee OA.

1.2.3 Chapter 4 :Where does meniscal damage progress most rapidly? An analysis using three-dimensional shape models on data from the Osteoarthritis Initiative

Given the importance of the meniscus in OA, Chapter 4 presents the results of the investigation into the spatial distribution of meniscal change and also assesses which meniscal pathologies (measured quantitatively in a novel way) change more rapidly during a 1-year follow up in a cohort typical to that included in OA clinical trials. Two meniscal measures (medial tibial coverage and thickness) were found to be the most responsive measures of change at follow-up, with change comparable to other MRI outcomes and better than radiographic JSN. The spatial location of damage was predominantly in the posterior subregion of the medial meniscus providing face validity for these novel measures. This study provided preliminary evidence that meniscal measures should add to the discriminatory power in assessment of OA progression. It is also possible that most measures in current use are not as responsive because they show small changes in the midst of large measurement error.

1.2.4 Chapter 5: The relationship between two different measures of osteoarthritis bone pathology, bone marrow lesions and 3D bone shape

Chapter 5 presents the results of the investigation into the relationship between two potential imaging biomarkers, a relatively well studied bone pathology, bone marrow lesions (BMLs) and the novel 3D bone shape. This is a multi-level modelling analysis and responsiveness study using participants from a sub-study of the OAI aimed at establishing the predictive and concurrent validity and responsiveness of biomarkers relevant to knee OA. This study suggested a moderate relationship between bone shape and BMLs, provided some evidence for the temporal order of MRI-detected OA bone pathologies and demonstrated that 3D bone shape had much better responsiveness than semi-quantitatively assessed BMLs over time periods typical of a clinical trial. The study also demonstrated the challenges faced in combining semi-quantitative or ordinal measures in OA.

1.2.5 Chapter 6: Determinants of osteoarthritis 3D bone shape and its change in the three knee bones: a latent growth modelling approach on 37,583 MR images from the Osteoarthritis Initiative.

Chapter 6 presents the results of the investigation into the longitudinal changes seen for the three knee bones, femur, tibia and patella using latent growth curve modelling in 4796 participants. Latent growth curve analysis has provided evidence that the direction of change is linear and the same in all three knee bones over time; they share common clinical or demographic determinants but vary by gender. Also, all three knee bones as measured by bone shape had measurement units that were positively correlated with each other at baseline, with stronger associations between the femur and tibia, than between the tibia and patella or femur and patella, indicating similar structural disease status. This suggests that the three knee bones are part of a single disease process, with the femur providing the greatest amount of change of the three bones. This could have implications on the choice of tissue for biomarker development.

1.2.6 Chapter 7: Structure-based trajectories in the three knee bones: data from the Osteoarthritis Initiative.

Chapter 7 presents the results of the investigation into the trajectories of change seen for the three knee bones (femur, tibia and patella) using the full 4796 OAI cohort and employing growth mixture models, a type of latent class growth analysis (LCGA). While the three knee bones have been previously shown to change linearly, this change has been assumed homogenous among all participants however significant variation in the

growth is thought to exist. Using LCGA, this study identified three distinct trajectories of structural change using bone shape and this was consistent for all bones and in both genders. Obesity, knee pain, non-white ethnicity and history of knee surgery were strongly associated with being in the fastest trajectory group . The study also identified a group of participants that had the fastest rate of change in all three bones. This ability to classify participants more accurately based on structure may be useful in enriching OA clinical trials.

1.2.7 Chapter 8: Discussion, future directions and conclusions

This chapter discusses the results of this thesis, the conclusions that can be drawn, an update on the literature review and the future directions in this field.

Chapter 2 Literature Review

2.1 Introduction

Osteoarthritis (OA) is a major personal and societal problem with few effective analgesic therapies and is the major cause of joint replacement surgery [12]. In the UK, approximately 20% of adults over 45 years old have knee OA [13], with associated direct and indirect healthcare costs approximately £1 billion/year [14]. OA is leading cause of disability worldwide, largely due to pain, the primary symptom of the disease and negatively impacts individuals` ability to perform activities of daily living resulting in reduced quality of life. Currently no proven disease–modifying therapies are available for OA that can slow, alter or reverse the processes of OA [2].

2.2 Defining OA

Clinically and for individual sufferers, OA is a syndrome of joint pain and stiffness that results in functional problems and reduced quality of life. It encompasses a number of problems that share common pathological and radiographic features manifesting themselves as structural abnormalities. The OA disease process in synovial joints may involve individual tissues in early stages, and evolves, by a complex cascade of biomechanical and biochemical pathways, into the typical whole-joint multi-tissue pathologies where all components of the joint are affected. In the later stages the pathological changes are characterised by cartilage degradation, involvement of subchondral bone and synovitis; peri-articular tissues especially muscle may also be involved.

2.2.1 The complexity of defining OA

There is considerable heterogeneity in the description of OA; it may be defined based on duration (early versus late), location (anatomical site), number of joints (monoarticular versus polyarticular), clinical and/or imaging involvement, tissue-based classification (involvement of bone, cartilage or inflammation), and primary or secondary causes [15].

Part of the complexity is that current OA definitions are derived from epidemiological studies and clinical trials where OA is defined using 3 options: clinical findings alone, the presence of imaging pathology or a combination of the two. There is however discordance between symptom and radiographic changes. Only 50% of knees with

radiographic OA (ROA) have knee OA symptoms [16].Likewise some people with clinical signs associated with OA may have a normal radiograph.

Radiographs are used to evaluate osteophyte formation and joint space narrowing (JSN); grading schemes such as the Kellgren-Lawrence (KL) grading scheme [17] and the Osteoarthritis Research Society International (OARSI) classification score have established guidelines for the diagnosis of OA progression [18]. These various scoring systems used to quantify radiographic OA result in different OA definitions depending on the system. Common radiographic features on an osteoarthritic joint include the presence of JSN, bony sclerosis, subchondral cysts, bony spurs (osteophytes) and joint erosions. There is however variability in the reporting which may be attributed to inadequacy in some studies not using radiographic views of all knee compartments; even when all knee compartments have been imaged there is still some discordance with pain [19].

While conventional radiography is feasible and inexpensive, it is limited by its relative insensitivity at detecting key OA structural abnormalities. Although useful for evaluation of JSN, there is evidence that a number of symptomatic patients show cartilage loss on MRI even when JSN or progression is not visible on radiography [20]. In that study, detection of radiographic progression had a sensitivity of 23% and specificity of 91% [20]. A large cohort, the Framingham cohort showed that in 88% of participants aged 50 and above, with no pain and apparently normal radiograph, at least one OA tissue lesion could be seen in them [21].

The classic definition of OA is that of a non-inflammatory disease, due to lack of neutrophils in synovial fluid, a lack of subchondral bone erosions and no evidence of systemic inflammation or features of autoimmunity which have been used to distinguish OA from rheumatoid arthritis (RA). Recent imaging advances such as contrast-enhanced MRI indicate that extensive synovitis is prevalent in most knees with established OA [22, 23]. Using ultrasonography synovial inflammation and effusion have been detected in painful knee OA and found to correlate with clinical parameters suggestive of an inflammatory flare [24].

Symptomatic OA definition, while relevant due to involvement of a person`s symptoms, may not be applied if there are no accompanying radiographic changes and this often excludes people with these signs but not having developed radiographic changes. On the other hand, using only a symptom-based definition of OA for clinical trials could make recruitment more feasible, but suffers from biases such as misclassification of individuals.

A classification criterion for knee OA using clinical and laboratory findings has been developed by the American College of Rheumatology (ACR) (Table 1), and clinical trial inclusion of participants commonly uses these criteria. While useful for recruitment in to trials these classification criteria are less useful for OA diagnosis in clinical practice due to the fact that these criteria were developed using hospital referred patients and a control group of people with inflammatory arthritis [25] .The European League Against Rheumatism (EULAR) also proposed recommendations for diagnosis of knee and hand OA. They suggest that 3 symptoms (persistent knee pain, limited morning stiffness and reduced function) and 3 signs (crepitus, restricted movement and bony enlargement) are most useful in OA diagnosis, and presence of all these was enough to conclude a diagnosis of OA without need for imaging [26].This suggests that a thorough clinical assessment is needed for OA diagnosis, and use of radiographs should be an adjunct if differential diagnosis is considered.

Table 1. Idiopathic OA classification of the Knee - ACR criteria

Clinical and laboratory	**Clinical and radiographic**	**Clinical**
Knee pain + at least 5 of 9	**Knee pain + at least 1 of 3**	**Knee pain + at least 3 of 6**
Age > 50 years	Age > 50 years	Age > 50 years
Stiffness < 30 minutes	Stiffness < 30 minutes	Stiffness < 30 minutes
Crepitus	Crepitus	Crepitus
Bony Tenderness	+ Osteophytes	Bony Tenderness
Bony enlargement		Bony enlargement
No palpable warmth		No palpable warmth
Erythrocyte sedimentation rate<40mm/hr		
Rheumatoid factor <1:40		
92% sensitive	91% sensitive	95% sensitive
75% specific	86% specific	69% specific

The Osteoarthritis Research Society International (OARSI) definition of OA is that: "Osteoarthritis is a disorder involving movable joints characterised 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."

Through the use of magnetic resonance imaging (MRI) we are now able to view knee structures with better accuracy. MRI offers a distinct advantage over knee radiography, since it is tomographic and cartilage can be assessed directly. Several semi-quantitative MRI scoring systems have been developed for evaluation of multiple pathologies in knee joint OA [27]. Ultrasound (US) also has the capacity to acquire 3D images and, for example, has been shown to detect many more osteophytes than CR in small joint (hand) OA [28]. It is limited only by its acoustic window, so it cannot, for example, visualise knee cartilage.

There are therefore many challenges in defining OA and definitions are varied. Modern imaging studies highlight that many of our concepts and definitions of OA have been based on an inaccurate imaging phenotype. There is a need to develop widely accepted criteria for clinical settings, prospective clinical trials and epidemiological studies because the current lack of "gold standard" for defining OA cases has resulted in substantial heterogeneity.

2.3 Epidemiology of OA

The epidemiology of OA is complex owing to the multifactorial aetiology of the disease and the heterogeneity in the disease definition as alluded to previously. All tissues of the joint are involved, but the most prominent features are loss of articular cartilage and corresponding bone changes. Broadly, OA is described clinically based on the symptoms (of which weight-bearing joint pain is the dominant symptom) and the clinical impact being the degree of disability resulting in functional limitation and reduced quality of life [29].

2.3.1 Prevalence

The most common form of arthritis in the world is OA and its prevalence increases with age [30]. With an ageing population and increased obesity worldwide, it is anticipated that the burden of OA will become a major problem for health systems globally [31].

Most frequently affected joints are the knee, hand and hip, spine and foot [29]. The greatest impact at population level is for the hips and knees, since OA is very common at these sites. Additionally, due to pain and stiffness at these sites there are significant problems associated with decreased mobility which manifest as significant activity limitation. As shown in the Framingham study, knee OA was the major contributor of activity limitation [32].

Prevalence estimates vary significantly between radiographic and symptomatic OA, and across studies. Data from population studies show that the prevalence of clinical/symptomatic OA is usually lower than radiographic OA, since joints may remain asymptomatic even in the presence of radiographic OA.

2.3.2 Symptomatic and radiographic OA

The prevalence of OA ranges from 12.3% (self-reported) [33] to 21.6% (physician diagnosed) [34]. Globally, prevalence peaked at the age of 50 with the highest prevalence in the 2010 estimates coming from Asia-Pacific followed by Oceania then North Africa/Middle East, while the lowest prevalence was seen in South and Southeast Asia [31]. In England, approximately 20% of adults over 45 years of age have self-reported knee OA. Prevalence estimates may also vary depending on age-group studied, gender (more likely in women), geographical region and study design [31]. Current information on prevalence is most commonly estimated from population-based radiographic surveys.

2.3.3 Radiographic OA

Data from the USA using the National Health Examination Survey which included only hand and feet examination showed that radiographic OA was rare in participants under 25 but by the ages of 65-74 almost all participants showed evidence of radiographic OA in the hands and about 50% showed similar evidence in the feet. In the US Framingham study, prevalence of radiographic OA in the tibiofemoral compartment of the knee was estimated at about 33%, rates comparable to those in Europe [35]. However, the age standardized prevalence of radiographic knee OA in adults older than 45 was 19.2% in the Framingham Study and 27.8% in the Johnston County OA Project. In the third National Health and Nutrition Examination Survey (NHANES III), close to 40% of participants aged over 60 years had radiographic knee OA [34]. In over 50 year olds, prevalence of radiographic knee OA is estimated at about 25% (data from community surveys) [36].Globally, the prevalence of radiographically confirmed knee OA in 2010 was estimated to be 3.8%, higher in females (mean 4.8% compared to 2.8% in males) when age was not accounted for.

A Dutch population study estimated that the prevalence of OA was as high as 75% in women aged 70+ when the left and right knee were combined, and even at the age of 40 about 10%-20% had evidence of severe radiographic disease. The study also showed that joints with low prevalence of OA in one population tend to have low prevalence in most populations and likewise for joints known to have a high prevalence of OA, due to the hypothesised aetiology being the same. They also suggested the possibility of extrapolating results across population surveys suggesting that differences between estimates could be due to inter-observer variation and lack of methodological standardisation [37].

2.3.4 Symptomatic OA

Data on symptomatic OA prevalence is also varied. The general definition of symptomatic OA is presence of pain, aching or stiffness in a joint with radiographic OA. Age-standardized prevalence of symptomatic OA was reported as about 5% in the Framingham study (participants aged over 25 years) while in the Johnston County OA project this was about 17% (participants aged over 45 years)[38]. Approximately 14% of adults older than 25 years and 34% of adults older than 64 years have clinical OA of one joint or more [39].Symptom definition in the Framingham study was based on medical history which included the question about whether participants had ever had pain in or around the knee on most days for at least a month. In another survey in participants aged 50 or over with evidence of radiographic OA, over half of these had symptomatic OA [36]. The prevalence of symptomatic knee OA in two UK studies ranged from 11 to 19%, and estimates of 5 –15% were noted in surveys undertaken in other countries [36]. This variation in estimates was based on case definition which differed across studies for example pain was defined as “knee pain at any time”, “knee pain for most days in the past month”, “current knee joint problems” (time period unspecified), “knee pain for more than a week in the past month”. Within the knee joint of symptomatic individuals, the most common radiographic osteoarthritis pattern of involvement is combined tibiofemoral and patellofemoral changes [40].

Across joints, the most common site of peripheral joint pain lasting for more than one week in adults 45 years was is in the knee (19%) while the highest prevalence of knee pain was shown to be amongst women aged 75 and over (35%) [41]. Global disability was also highest in those reporting isolated knee pain. In adults aged 50 years and over 23% report severe pain and disability [42].

2.3.5 Limitations of prevalence estimates

The importance of prevalence estimates is to allow accurate planning for the needs and options for delivery of healthcare. There are nonetheless great variations in these estimates due to various reasons for example disease ascertainment and diagnostic reporting. Accurate disease ascertainment is important for prevalence estimates and also useful for planning and aetiological research. Many databases claiming to represent the general population have been shown to contain imperfect and incomplete information and inferences from imperfect databases can sometimes bias prevalence estimates [43]. It is therefore important that prevalence estimates are adjusted for misclassification errors. A Canadian study showed that using only physician-derived diagnosis resulted in an unadjusted OA prevalence of 10.1% but this increased by 50% to 14.5% after robust adjustment for misclassification [43]. Other potential explanations for differences include differences in study populations, definitions of OA, different distribution of risk factors of OA and variation by radiographic readers.

2.4 Incidence

While radiographic changes may be seen more often, symptoms and loss of function are not always present therefore incidence rates may also be influenced by this discordance. These rates have varied depending on epidemiological study and case definition. In a population-based study the age and sex-standardised incidence rate of hand OA was 100 per 100 000 person-years, for hip OA 88 per 100 000 person-years and for knee OA 240 per 100 000 person years. The incidence increased with age and women had higher rates than men after the age of 50. However around the age of 80 these rates levelled off with incident symptomatic knee OA among women being about 1% [44]. Earlier known incidence age- and sex-adjusted rates for knee OA in the USA were 163.8 per 100 000 person-years in 1985 in a population study [45].

A UK study estimated the consultation incidence rate (the rate of new cases presenting to general practice) with a diagnosis of knee OA as 6.5 per 1000 persons aged greater than 45 years old. There was evidence of increased incidence of recorded cases of OA among adults aged 35-44, an increase from 0.3/100 persons in 2003 to 2.0/1000 persons in 2010 [46].

2.5 Traditional risk factors for OA

Risk factors of OA can be generally grouped into person-level factors (age, gender, obesity, genetic and diet) and joint level factors (injury, malalignment and abnormal joint loads) that interact in a complex manner [47]. A conceptual model has been

developed (Figure1) that is useful in understanding the pathogenesis of OA, on the one hand are joint and tissue components (for example cartilage and bone) that are important for movement and maintenance of correct loading while systemic factors that increase overall susceptibility to degradation and local biomechanical factors determine the risk of OA. The complex interplay between these factors is important in development of OA and therefore its progression.

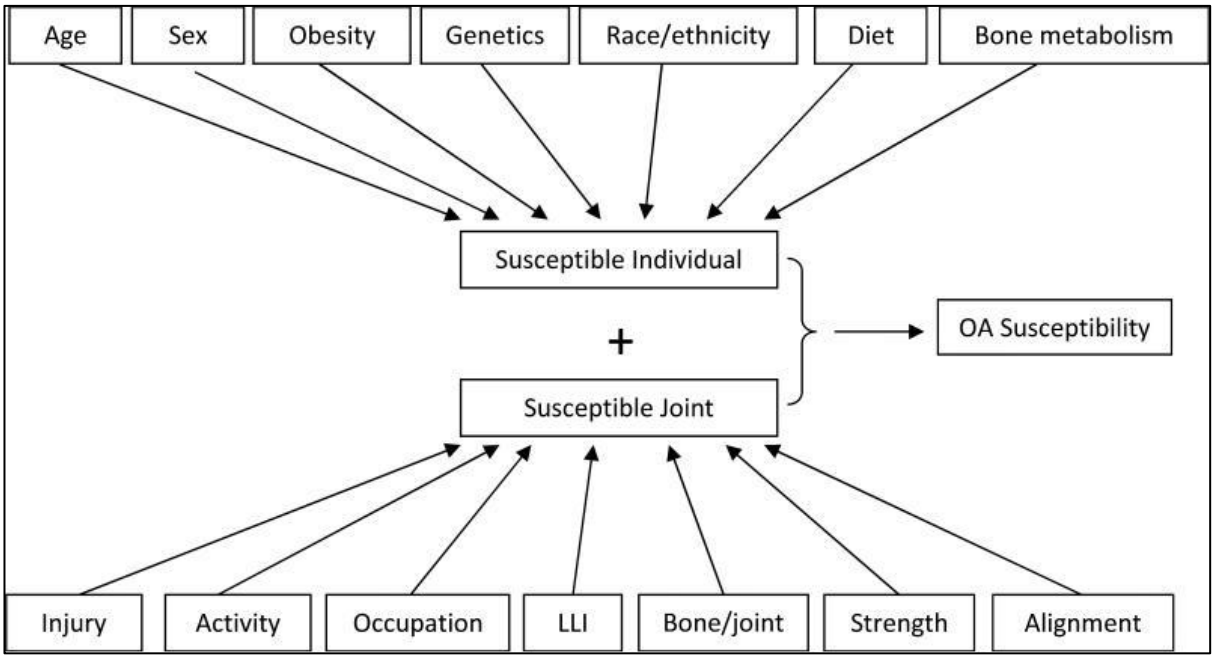


Figure 1.Conceptual model for the pathogenesis of OA
Adapted from [48].

2.5.1 Age

Strong evidence points to an increase in both prevalence and incidence of OA with age [35, 39, 49, 50]. This relationship is more profound in the most commonly affected joints, such as the knee, hip and hands [29]. The direct effect of age is not clearly understood but this relationship between OA and age may be mediated by age-related factors such as increased muscle weakness, ligamentous laxity, decreased proprioception, cartilage thinning accompanied by poor anabolic response to growth factors and loss of chondrocytes which all contribute to joint susceptibility [51, 52].

2.5.2 Gender and sex hormones

Females are at a greater risk of OA for both prevalent and incident OA of the hand and knee especially after the age of 55 years [53]. A recent meta-analysis estimated the odds ratio for the risk of developing symptomatic knee OA as 1.68 for females compared to males, although there was high heterogeneity in those studies [54].

There have been suggestions that increased prevalence of OA in women post-menopause could be linked to sex-hormones primarily oestrogen [55], with lower prevalence of OA seen in oestrogen users [56]. Some evidence also suggests that postmenopausal women using oestrogen also have larger knee cartilage volumes assessed by MRI compared to non-oestrogen users [57]. However conflicting evidence has emerged and thus the effect of hormones is not fully understood and is difficult to interpret [58]. A randomised trial found no difference in knee OA related symptoms between women receiving oestrogen compared to placebo [59]. Due to potential sources of bias and confounding in different studies, this association is not entirely clear and warrants further clinical and biochemical studies.

2.5.3 Race/ethnicity

There are racial and ethnic differences in the prevalence of OA. While the prevalence of hip OA in African American women was similar to Caucasian women in the Johnston County study, higher prevalence was seen in African men when compared to Caucasian men [60].Compared to estimates from the Far East, Caucasians had higher prevalence of hip and hand OA, this was noted by comparing estimates from the Beijing OA study to those in USA [61]. An attempt to determine the impact of racial differences in baseline radiographic features at 6-year follow up had mixed results: while prevalence of radiographic hip OA was similar between Caucasians and African Americans, Caucasians showed more radiographic hip OA progression based on JSN. However, African Americans had more frequent disability and greater progression for symptomatic OA based on range of motion or disability [62]. Although some evidence suggests differences in OA between racial groups could be through mediation via obesity, for example populations in the Far East are less likely to be obese this is not conclusive and there is a need to examine biological, psychological and lifestyle factors that may contribute to these differences.

2.5.4 Genetics

Genetic factors have been found to be strong determinants of OA. There are numerous sources of evidence in the genetic influence of OA, ranging from epidemiological studies of family history, family clustering and twin studies. Twin studies have shown that the influence of genetic factors is between 39% and 65% in radiographic OA of the hand and knee in women, 60% in hip OA and 70% in OA of the spine which indicates that about half the variation in disease susceptibility is explained by genetic variation [63]. Susceptibility to OA can be influenced by a wide number of genetic variations, and as thus may not follow the typical pattern of Mendelian inheritance but rather multiple

gene interactions which have led many to agree on the theory of polygenic inheritance [64, 65]. These genome-wide studies reveal a substantial genetic component to OA comprising multiple contributing variants with small effect sizes [65]. As OA is likely a complex polygenetic problem, gene studies in isolation may not be adequate in stratifying individuals into who may be predicted to develop OA or not because the magnitude of these associations are weak, however innovative findings with respect to the pathophysiology of OA may be useful in developing new targeted therapies [66].

2.5.5 Bone mineral density

In general, increasing bone volume fraction, trabecular number, trabecular thickness, and decreasing trabecular spacing are associated with structural progression and severity of OA of the knee [67]. In the Framingham cohort, radiographic subchondral sclerosis was associated with ipsilateral increase in BMD [68]; greater BMD associated with subsequent joint space narrowing and cartilage defects [69, 70]. It has also been observed that increased severity of ROA is associated with thicker trabeculae and decreased space between trabeculae [71].

The material properties of bone may influence OA susceptibility, higher systemic bone mineral density (BMD) was associated with increased risk of incident OA; however no association was seen with radiographic progression of existing knee OA [72]. Reasons for this association are not conclusive but this could be through mediation involving BMI, in that higher systemic BMD may represent higher BMI loading in the years preceding OA, itself a strong risk factor for OA [48]. Low BMD is associated cross-sectionally with reduced joint space width at the hip [73]. The lack of association between BMD and OA progression may be explained by the fact that once OA has been established and symptoms are developed, due to the reduction in physical activities in these individuals and therefore reduced loading of the joint: this may contribute to low BMD. There is evidence that although the apparent density of bone in OA may be increased, the bone itself is less mineralized, resulting in lower material density [74].

2.5.6 Nutrition

A number of studies have evaluated the role of dietary factors (particularly vitamins) in OA [75-77] and results have been inconclusive [75, 78]. Vitamin C has beneficial effects on collagen synthesis and, like Vitamin E, has antioxidant properties. Increased intake of these vitamins may protect against OA by reducing the oxidative damage to cartilage and joint tissues by oxygen-free radicals produced by chondrocytes in damaged cartilage [79]. Low vitamin D intake and reduced circulating serum vitamin D

may result in increased risk of hip OA although protection from incident knee radiographic OA was not associated with vitamin D status [80]. Selenium and iodine deficiency have been associated with Kashin-Beck osteoarthropathy [81]. A review however found inconclusive or contradictory findings on the efficacy of vitamin C or selenium for treatment of arthritis [82] which therefore warrants further investigation into the role played by these nutrients overall.

2.5.7 Obesity

Obesity is one of the most significant, risk factors for the development of OA, and numerous studies have shown a strong association between BMI and OA of the hip, knee, foot and hand [30, 83-85]. However, the mechanism(s) by which obesity contributes to the onset and progression of OA are not fully understood. The strong association between BMI, altered limb alignment and OA of the knee and the protective effects of weight loss support the classic hypothesis that the effects of obesity on the joint are due to increased biomechanical loading and associated alterations in gait causing breakdown of cartilage and damage to ligaments and other support structures [86-89]. Due to the association between OA and non-weight bearing joints such as the hands there may also be a role played by systemic factors in explaining the relationship between OA and obesity. Data suggests that metabolic factors associated with obesity may alter systemic levels of pro-inflammatory cytokines that are also associated with OA [90-92]. Total body fat was associated with decreased cartilage thickness, while lean mass was associated with increased cartilage thickness [93]. Adipose tissue is known to be metabolically active, secreting adipokines such as adiponectin, leptin, and resistin and tumour necrosis factor, which may provide a metabolic link between OA and obesity [94-96].

Based on results from a systematic review of prospective studies, obesity is a robust risk factor of knee OA with significantly higher odds of knee OA in obese individuals (Relative Risk = 4.55), and recommendations are that possible weight reduction should be taken into account for knee OA whenever a patient is significantly overweight [97].

2.5.8 Biomechanical factors

2.5.8.1 Occupation

Repetitive joint use may predispose to OA as seen in particular jobs requiring heavy manual labour for example repetitive hand movements have been found to lead to hand OA in females in the clothing industry [98], while jobs requiring kneeling and squatting were associated with increased risk of knee OA in a Chinese study [99].

Results from a meta-analysis, although hampered by publication bias for cross sectional studies and case-control studies suggested that occupational activities increase the risk of OA by about 1.6 times [100].

2.5.8.2 Physical activity

There is conflicting evidence on the impact of physical activity. While there is great benefit to the joint via muscle strengthening achieved through exercise (which in turn improves joint stability), physical activity may also increase vulnerability of the joint to damage [48]. There is some evidence that elite athletes are at higher risk of knee and hip OA [101] and compared to non-soccer players elite players have a higher risk of knee OA [102]. The mechanism by which elite-level physical activity results in increased risk of OA could be due to increased load bearing, higher rates of knee injury among soccer player and increased BMI as well as frequent squatting for example in weight lifters [102]. Moderate or light physical activity has not been shown to be a risk for OA, and in the absence of injuries moderate running and jogging did not appear to increase OA risk [103].

2.5.8.3 Alignment

Knee malalignment is a strong predictor for progressive knee OA, with varus alignment of the knee being associated with 3.6 times the risk of radiographic progression of medial knee OA [104], while valgus alignment is associated with lateral knee OA [105]. Alignment is a key determinant of load distribution and therefore any shifts from the neutral alignment results in changes to load distribution and therefore could be the reason for increased risk of OA in malaligned knees. This is a confounded issue: joint space narrowing and bony alterations occurring in OA may themselves result in joint malalignment, while malalignment itself may result in further alteration in joint loading and therefore accelerates diseases progression [48].The impact of malalignment was found to be greater in knees that have more severe baseline radiographic disease than knees with mild disease [106]. Knee malalignment has also been shown to be associated with the size and progression of bone marrow lesions and JSN [107].

2.5.8.4 Knee injury

A history of joint injury is a significant risk factor for the later development of OA. Knee injury whether specific such as meniscal tear, or unspecified such as a sprain or swelling was shown to be a major risk factor for OA development (results from a meta-analysis) with overall pooled odds ratios of 4.20 [108]. Other studies also support this finding and these report that acute injuries such as meniscal and cruciate ligament

tears, fractures and dislocations result in a significant risk of later development of OA [109, 110]. Apart from the direct effects of injury on joint tissues, injury may disrupt normal joint mechanics which contributes to later occurrence of OA [29].

2.5.8.5 Joint morphology/joint deformity

Due to the role played by load distribution and knee joint biomechanics in the development of OA, the geometric shape over which load is transmitted is important in the onset of OA. Using MRI data, three-dimensional bone shape has been shown to predict the onset of OA [111].The importance of bone shape in OA has been shown in abnormalities such as congenital hip dysplasia, femoracetabular impingement and subtle femoral differences that are shown to be associated with hip OA [112]. Joint shape alterations on radiography have been associated cross-sectionally with prevalent OA [113] however some of these changes may be related to problems during image acquisition for example positioning. Another factor that may alter joint biomechanics is limb length inequality (LLI). Participants with LLI greater than 2cm were twice as likely to have prevalent radiographic OA in the Johnston County OA Project [114], while the MOST study found that LLI of at least 1cm was associated with increased risk of symptomatic OA and was also associated with prevalent radiographic knee OA [115].

2.5.8.6 Muscle strength

The relationship between OA and muscle strength is not entirely understood. In a recent prospective study in which 96% participants had a clinical OA diagnosis in at least one joint, lower and upper limb weakness was very common (77% and 90% respectively) [116].

Muscle weakness and disuse atrophy may occur as a consequence of OA due to pain avoidance and reluctance to use the affected limbs [48]. However, there is evidence that muscle weakness may precede the onset of OA. Quadriceps muscle weakness was shown to be associated with increased risk of structural knee OA [117]. The role of quadriceps muscles in the distribution of load during gait function suggests that greater muscle strength may confer protection against development of OA [118]. Quadriceps weakness has been recorded in the absence of pain or atrophy, suggesting that this might be a risk factor for knee pain and progression of OA as a result of muscle dysfunction [119]. In a compromised joint (as characterised by joint laxity and malalignment) stronger quadriceps were associated with greater progression of existing radiographic OA [120]. The evidence as to whether increased muscle strength

is protective against knee OA is inconclusive and may therefore depend on other factors that affect load distribution.

Results show that the relationship between muscle weakness and OA varies by joint site. In the hand, greater grip strength was associated with increased risk of developing hand OA [121]. Those in the highest tertile of grip strength had a threefold increase in risk of OA in the proximal interphalangeal and metacarpophalangeal joints compared to those in the lowest tertile of grip strength. The mechanism could be that greater contraction forces increase joint loads during activity, and as these hand joints are subjected to the greatest forces during grip they are vulnerable to degenerative changes.

2.6 OA clinical findings

2.6.1 Symptoms

For symptomatic OA, the most common complaint is pain and it is the primary reason a person with OA will seek medical help from a physician. While joint pain is common, not all is attributable to OA and pain can rise from structures around the joint, such as inflamed tendons or bursae; this may contribute to the heterogeneity in OA definition. In a recent cohort study, (mean age 63) of patients with multi-joint pain, where the median number of painful joints was 6, only 51% of painful joints were diagnosed with OA [116].

There are many aspects of OA pain. Severity is usually enquired about, but description of pain varies with individuals and may be described as sharp, burning or a dull ache. Pain may also be episodic; a quarter of people over 55 years old will have an episode of persistent knee pain due to OA [36], defined as ever having knee pain on most days for at least a month plus knee pain in past year. Pain can also be described through its relationship with activity (worse on knee bending, or after a walk) and activity-related pain. In terms of pain trajectories, work from Peat and colleagues suggests that pain is largely stable over time [122].

OA patients can also suddenly experience a ‘flare’ up of symptoms known as an “OA flare” which is sudden and dramatic causing significant pain, disability and impacting on daily life. OA flare has long being considered just an exacerbation of pain occurring by episode but current research suggests that it may be much more complex phenomenon involving psychological aspects and impact of symptom. The OMERACT working group has proposed five domains as part of the definition of an OA flare to include: pain, swelling, stiffness, psychological aspects and impact of symptoms [123].

All of this makes characterisation of pain an important part in the OA diagnostic pathway.

Stiffness is another major OA complaint and in people afflicted by this condition, pain and stiffness are hard to distinguish. In people with OA there is usually stiffening of the joint due to inactivity which is usually reported as locking of the joint by most patients. This differs from inflammatory conditions because in OA this is usually short-lived and this is useful for diagnosis [124].

2.6.2 Examination findings

OA affects individuals in many complex ways such as observed or reported symptoms, loss of function, limitations of activities and impaired quality of life. The aim of clinical assessment is to characterise, establish cause and assess the direct impact of the problem to the patient. Woolf *and* Akesson [125] provide recommendations for taking medical history and performing clinical examination relevant to a musculoskeletal problem which could be adapted for knee OA. They also suggest that a rapid screening assessment should be included in all standard clinical examinations on any patient [125] and the advantage of such a standardised approach helps to improve competency of all doctors in assessment of musculoskeletal problems and also leads to improvement in the management of these conditions.

Clinical examination involves the assessment for bony swellings, which may be hard (when caused by osteophytes) or soft (involving synovial thickening and joint effusion). Crepitus is another notable sign (felt on palpation or heard as a crunching or popping sound of the joint when moved) and normally represents a late presentation of OA. Due to surrounding inflammation or effusion, there may be soft tissue swelling that may be tender or painful to touch. Other findings include abnormal bone formations which cause deformities and seen clinically as malalignment (e.g. varus deformity), muscle wasting due to disuse atrophy and restricted joint movement as a result of capsular thickening and painful muscle spasms.

2.7 Impact of OA

2.7.1 Impact on individual

The overall impacts of OA are substantial. Loss of function is compounded when more than one joint is affected and individuals are more likely to report some difficulty with occupational as well as certain leisure activities depending on the site and severity of OA. Compared to people without knee pain, individuals with a combination of knee, back, feet and hip involvement have a 60-fold increase in the likelihood of reporting

difficulties with tasks such as walking. Persons with just knee involvement have a 3-fold increase in risk compared to those free of knee pain [126].

Other adverse effects seen in people with OA include loss of independence and poor self-esteem which may also result in mental health problems such as depression which ultimately impairs quality of life. It is well known that depression is prevalent among people with OA [127]. The progressive nature of OA also leads to psychosocial decline which influences pain experience, clinical presentation and therefore affects diagnostic accuracy [128]. While mental health problems modify the association between self-reported limitations in activities and impairments due to OA, their effects are not entirely clear as selective under and over reporting if often seen in these studies ,however but it is worth noting that there is an alteration in symptom reporting as a result [129]. This uncertainty is also highlighted by a systematic review which revealed that although 20% of people with OA experience symptoms of depression and anxiety, it is unclear if this was significantly more than in people without OA (relative risk 1.17, 95% CI: 0.69-2.00) [130].

Having OA increases the chances of having other comorbidities; people with OA are twice more likely to have comorbid conditions than the general population [131]. There are numerous conditions frequently associated with OA namely diabetes, stroke, obesity, anxiety, cardiovascular disease and cognitive problems and although they generally don't influence the physician`s choice in terms of OA treatment, some like peptic ulcers may result in reduction in the prescription of anti-inflammatory drugs [132].

2.7.2 Impact on society

The impact of so many comorbidities with OA is that of increased health care expenditure overall through direct medical costs. It is also clear that OA and comorbidities significantly reduce the patients' quality of life and magnifies the impact of OA [133]; in addition the likelihood of functional decline increases when there is the combined effect of comorbidities and OA [134] with an estimated increase in risk of between 40–50% increased risk of disability pension in comparison with the general population [135].

OA has a huge economic impact that will continue to grow as a result of its growing prevalence. The socioeconomic burden ranges from one to two-and-a-half per cent of gross domestic productivity in developed countries, through a combination of direct and indirect costs [136]. Individuals with OA have twice the rate of absenteeism compared to non-OA individuals and their estimated direct and indirect costs are about two to

three times those of individuals without OA [137]. Individuals with OA were also reported to have significantly higher health care resource utilisation accompanied by lower work productivity compared to those without OA [138].

In the UK, according to the National Joint Registry over 81000 knees underwent knee replacement surgery in 2011, this represented a 6% increase from the previous year about 6% requiring revision surgery. Although multiple factors result in TKR, knee OA accounted for 97% of these replacements in 2010 which was worth £426 million to the health budget. OA represented more than 93% of all joint replacement indications with a total annual cost estimated at £852 million in the UK in 2010 [14].

2.8 Pathogenesis

As already noted, the OA process affects all joint tissues, and it is often conceptualised as failure of normal repair processes in the joint [139]. The risk factors have already been discussed previously. This section describes the individual tissue pathologies observed in OA joints.

2.8.1 Cartilage

As noted, cartilage has previously been given most attention in the pathogenesis of OA largely because of the striking changes seen in advanced OA. Joint surfaces are covered by a thin layer of cartilage resting on subchondral bone and while cartilage is neither innervated nor vascularized, subchondral bone is rich in both. Healthy cartilage is important in transmission of joint loads which it does through a cartilage matrix rich in proteoglycans and collagen. The normal joint needs biochemical remodelling of the cartilage matrix and biochemical forces to maintain its integrity. Changes in the joint cartilage include gradual proteolytic degradation of the matrix which results in increased synthesis of matrix components by chondrocytes [140]. At the molecular level, this results in early morphological changes such as cartilage fibrillation, cleft formation and loss of cartilage volume [140].

2.8.2 Subchondral bone

Events in the subchondral bone may be important in the initiation of OA; these include the development of osteophytes at the joint margin, decreased bone mineral content, changes in the vasculature and increased turnover of subchondral bone resulting in increased thickness of bone plate [141]. Changes in the bone lead to a reduced ability to maintain normal mechanical stresses, which may result in increased stress on the overlying cartilage [142]. Subchondral bone changes are recognized as a hallmark of OA, but normally occur later when degeneration is well established. Earlier bone

changes such as bone marrow lesions (BMLs) can be visualised from MRI, and are defined as regions of hyperintense marrow signal in fluid-sensitive, fat-suppressed MR image sequences. Histologically they are characterised by fibrosis, lymphocytic infiltrates and increased vascularisation, the latter thought to be responsible for the water signal seen on MRI [143]. BMLs have been linked to development of knee pain [144-146] and cartilage degeneration [147].It is unclear whether events in cartilage precede bone, or are concomitant with them or if bone changes cause cartilage damage. What it probably clear is that there seems to be some “bone-cartilage crosstalk”.

2.8.3 Synovium

Increased permeability at the junction between the articular cartilage and bone exposes the cartilage to an abnormal biochemical environment. Cytokines and other signalling molecules released from the cartilage, synovium, and bone affect chondrocyte function. Synovial inflammation is precipitated by cartilage debris and catabolic mediators entering the synovial cavity, and these chemokines and metalloproteinases further degrade cartilage and amplify inflammation leading a vicious circle [148].

In contrast to the diffuse synovitis seen in RA, the synovitis seen in OA tends to be more focal, with a study suggesting that abnormalities in the medial perimeniscal synovium were associated with medial chondropathy [149]. Although synovial changes are usually seen in advanced OA, they may be noted in early disease when there is little damage which suggests that synovial inflammation may actually be present early in the disease. As alluded to earlier, inflammation within the joint is thought to be the result of interaction between damaged cartilage and inflamed synovium, cartilage breakdown products are released into synovial fluid and phagocytosed by synovial cells which increase synovial inflammation. The synovium then produces pro-inflammatory products such as IL-1, IL-6 and TNF-alpha causing more inflammation and further release of proteolytic enzymes [148].

The molecular pathogenesis of OA is shown Figure 2. Because of joint trauma or overuse, altered biomechanics and instability contribute to tissue damage results in production of damage-associated molecular patterns (DAMPs), including cartilage extracellular matrix breakdown products that signal through receptors on synovial macrophages, fibroblast-like synoviocytes, or chondrocytes to induce the local production of inflammatory mediators. This inflammation-induced angiogenesis and increased vascular permeability causes an influx of plasma proteins capable of functioning as DAMPs. Acute and chronic production of inflammatory mediators

promote further cartilage degradation either directly or indirectly through their induction of proteolytic enzymes, amplifying a vicious cycle of innate immune activation in OA.

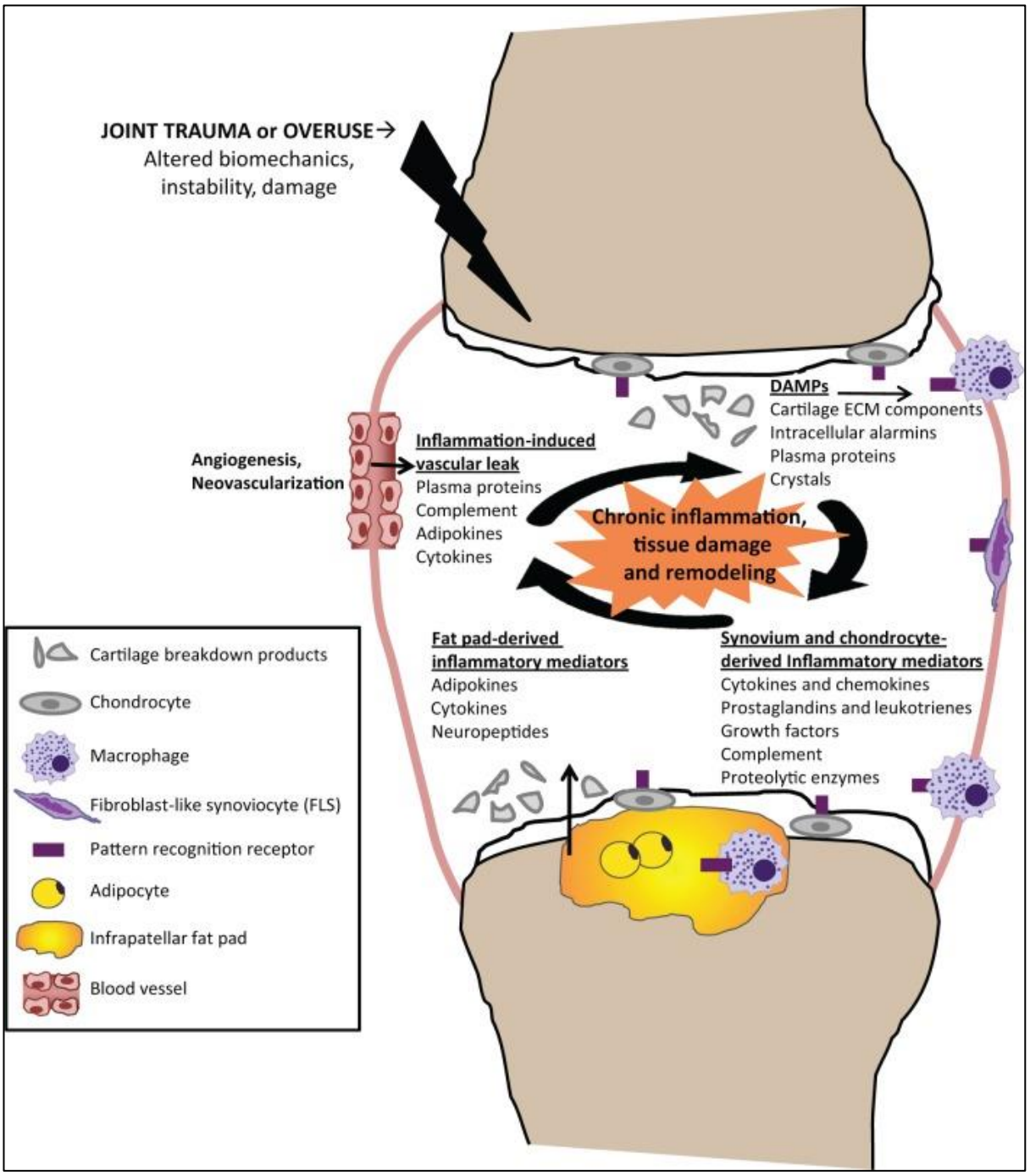


Figure 2.Schematic representation of chronic inflammation as a mediator of osteoarthritis

Adapted from [150].

2.8.4 Meniscus and ligaments

The menisci are important for their shock-absorbing and load distribution properties. Damage to menisci, as evidenced by tears and meniscal extrusion predisposes individuals to OA due to altered biomechanics within the joint [151, 152]. Meniscal damage will in general cause abnormal loading through the adjacent cartilage which has been demonstrated on MRI as early cartilage abnormalities which initiate a catabolic effect [153]. Ligaments are also an important feature in the OA pathogenesis. They function to provide stability and support for the joint, with suggestions that changes in the joints' supporting ligaments may be important in the OA pathogenesis [154]. The association between anterior cruciate ligament (ACL) damage and knee OA has been demonstrated [155] and MRI studies on OA subjects have shown the presence of a complete ACL tear within the OA joint in up to 22% compared with 3% in controls [156]. In addition periarticular muscular and neuronal problems that result in joint instability have been linked with OA development but there is lack of long-term follow-up studies on this.

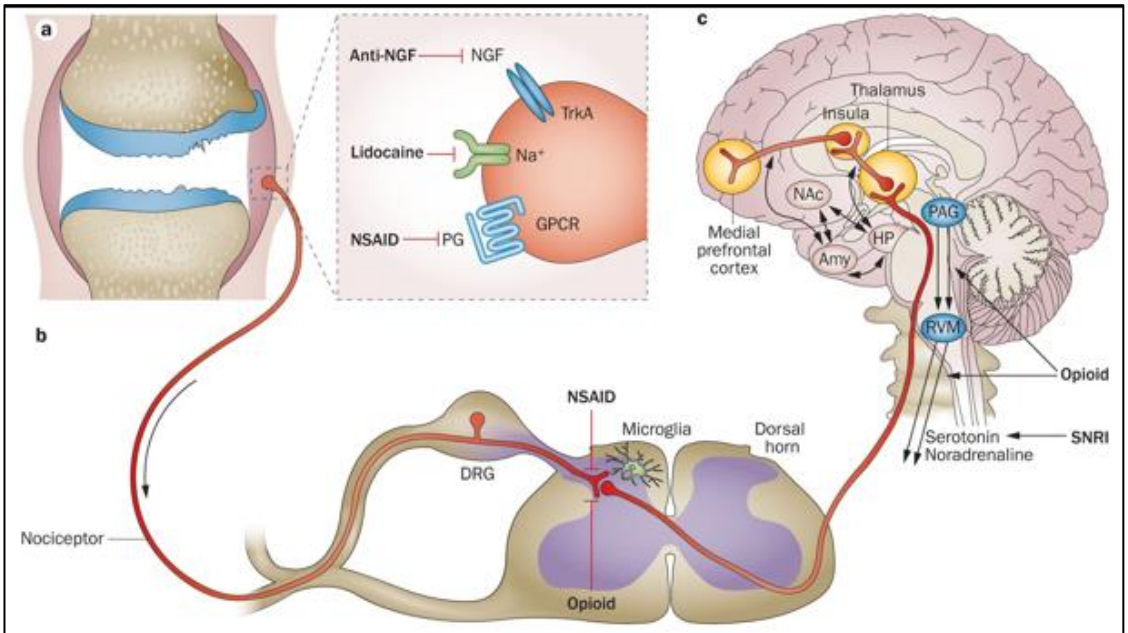
2.8.5 Pathogenesis of pain in OA

Pain in OA involves peripheral and central nociceptive mechanisms. The sources of OA pain are varied and characterising these sources is important for accurate diagnosis. This section describes the origin of pain in OA and the structural associations.

2.8.5.1 Origins of pain

The joint is a densely innervated organ, and its sensory innervation is predominantly towards proprioception and nociception. Nociceptors vary by joint tissue and by stage of OA. They are abundant in the joint capsule, ligaments, periosteum, menisci, subchondral bone and synovium [157, 158] but the cartilage is aneural and avascular in normal joints. Thus, pain can originate from many articular tissues, and cartilage unlikely to be the source of pain in early OA [159]. However as OA progresses there is neurovascular invasion which provides cartilage with the potential for nociception [160].

Physiologically, pain equals nociception, occurring when afferent nociceptive neurons innervating tissues are activated by stimulus of a mechanical, thermal or chemical nature [161]. The pain pathway is shown in Figure 3.



Abbreviations: Amy, amygdala; DRG, dorsal root ganglion; GPCR, G-protein-coupled receptor; HP, hippocampus; NAc, nucleus accumbens; NGF, nerve growth factor; PAG, peri-aqueductal grey; PG, prostaglandin; RVM, rostral ventromedial medulla; SNRI, serotonin–noradrenaline reuptake inhibitor

A Pain signals are detected by nociceptors in the periphery and carried to the dorsal horn of the spinal cord.

B The central terminals of the afferent nociceptors synapse with second-order neurones in the dorsal horn. Second-order neurons are either interneurones or projection neurons that cross to the contralateral side and carry the signal up the spinal cord. Central sensitization can occur through the strengthening of synapses and through the loss of inhibitory mechanisms. In addition, the activation of microglia contributes to enhanced pain sensitivity. Prostaglandins can also have a sensitizing effect in the dorsal horn, and NSAIDs can thus exert central analgesic actions, in addition to their peripheral actions. Opioids can inhibit incoming pain signals in the dorsal horn.

C Projection neurons relay pain signals to the thalamus, and to the brainstem. From there, the signals can be propagated to different areas of the brain, including the cortex. Descending pathways (black arrows), both facilitating and inhibitory, modulate pain transmission; descending inhibitory pathways release noradrenaline and serotonin onto the spinal circuits. SNRIs engage these descending inhibitory pathways. RVM neurones are opioid sensitive, and morphine has an analgesic effect through engaging descending inhibition.

Figure 3. Neuroanatomy of the pain pathway and analgesic targets in OA
Adapted from[161].

Central pain neurological mechanisms are important in the perception of pain. Central termini of afferent neurons enter the dorsal horn of the spinal cord and make their first synapse with interneurons as shown in Figure 3. Continued nociceptive transmission from the spine to the sensory cortex can be increased due to inhibition of the inhibitory central descending pathways and this is called central sensitisation [162]. Central sensitisation has many determinants and can be influenced by comorbidities including mood disorders, loneliness and sleeping problems in OA. Overall, central sensitisation

is due plasticity of the CNS, leading to increased uninhibited neuronal activity, reduced activation thresholds, and expansion of the receptive field, and is manifested as hyperalgesia and allodynia, even in areas outside the initial stimulated zone [162].Central sensitisation contributes to the lack of direct correlation between nociceptor activation and the pain experience which is characteristic of chronic pain [161].The enhanced pain experienced in response to a given stimulus reflects neural plasticity in the chronic pain of OA which involves both peripheral and central sensitisation. This phenomenon has been demonstrated by injection of saline into the tibialis anterior muscle of OA patients resulting in more intense pain than individuals without OA [163].

In acute pain, stimulus is transmitted from a peripheral nociceptor via afferent sensory neurones to spinal neurones for onward transmission to the thalamus and thereafter the sensory cortex. There are serotonergic and noradrenergic pathways within the CNS that provide descending inhibition to modulate and reduce the signal that is conveying the acute pain. In chronic pain, greater sensitivity in the peripheral nociceptors can increase this signal resulting in peripheral sensitisation. Examples of mediators of peripheral sensitisation include substance P, neuronal growth factor (NGF), calcitonin gene-related peptide, neuropeptide Y and vasoactive intestinal peptide.

Activity-related pain of OA is typically present in early OA before evidence of radiographic changes [164]. This pain in early OA is typically mechanical and may reflect the increased loading on subchondral bone [165]. Molecular mechanisms of mechano-sensation are poorly understood but in animal models sodium channel Nav1.8 has been shown to play a role. Nav1.8 is restricted in its expression to small primary afferent neurons and has been implicated in noxious mechano-sensation in mice with expression levels shown to increase in afferent neurons that innervate inflamed rat knees [166, 167].

2.8.5.1.1 Inflammatory mediators

Nociceptors express a broad range of receptors for ligands that can change the biochemical properties of neurones, such that they require lower thresholds to generate impulses or fire spontaneously when the receptors are engaged. These ligands include cytokines, chemokines, neuropeptides and prostaglandins [168].As a result of this peripheral sensitization, joint movement within the normal range becomes painful.

It is known that injecting proinflammatory agents into the joint cavity sensitizes afferent neurones [169]. The proinflammatory cytokines IL-1β and TNF can directly affect sensory neurons and can also trigger hyperalgesia. In addition to increasing cytokine

levels, inflammation enhances local levels of NGF, a major contributor to peripheral hypersensitivity [170].

NGF can be produced by articular cartilage, meniscus and synovium, and increased levels have been reported in synovial fluid from patients with inflammatory arthritis [171, 172]. Inhibition of NGF has been shown to be efficacious in many preclinical models of pain, including joint pain in rat autoimmune arthritis [173].

2.8.6 Structure-pain relationships

2.8.6.1 Structure and symptoms (limited to radiographs)

As noted earlier there is discordance between clinical and radiographic OA [16].In a survey of general practitioners the presence of radiographic OA was shown to influence the decisions for their management of OA, particularly resulting in increased referrals to secondary care due to the link with treatment decisions or current management strategies. The presence of symptoms of OA had little effect on this decision [174].It is important that the relationship between what is important to a patient i.e. reduction of pain and disability and the decisions based on radiographs are understood.

Symptoms are thought to precede the appearance of any radiographic changes seen on plain radiographs, namely osteophytes and joint space narrowing [175, 176] which suggest that a transition period exists between pre-radiographic and radiographic stages of OA. This period of premonitory symptoms known as a “prodromal phase” [164] is associated with MRI changes such as articular cartilage lesions, bone marrow defects and meniscal damage that appear to be associated with OA symptoms [21, 177].

Results from a nested case control study indicates that incidence of radiographic OA is associated with a prior increase in knee pain, stiffness and difficulties with activities of daily living. Symptoms that appeared earliest were pain on twisting or weight bearing [164]. Importantly, the appearance of joint space narrowing and the onset of a definite osteophyte were shown to be neither the beginning of OA pathogenesis nor the first indication of morphological changes on radiography as participants with KL Grade 1 were strongly associated with symptoms [164]. However after adjusting for KL Grade 1 there was persistence of an association between the symptoms and incident radiographic OA which poses a dilemma as the “prodromal phase” cannot fully explain existence of symptoms. Inherently these difficulties arise from factors alluded to earlier which are that OA has no universally accepted discrete point of onset and due to the

nature of the disease (repair and damage processes constantly occurring) there are challenges to applying the concept of a “prodromal phase” for OA but its importance is clear. However, with the development of new imaging techniques, definitions of early disease continue to evolve [178] posing a challenge to the concept of a “prodromal phase”. Given these novel imaging advances, the true existence of such a phase could be debateable.

The definition of symptoms does pose a challenge in evaluation symptoms and structure for example the concept of pain. Pain is an individual experience and comes from various sources therefore it is a challenge for researchers to standardise a definition of pain to encompass all levels and definitions. A systematic review highlighted this challenge as the definition of pain had an influence on the association with radiographic changes. Using definitions that explored “current” pain this symptom was found in 59-81% of participants that had radiographic changes, where the definition was “pain ever” lower estimates ranging from 20-59% were found [19]. Within study variation was also noted based on pain definition for example prevalence for “pain in or around the knee for one month” was estimated at 53% but this went up to 64% in the same group if this was limited to “pain in the last year” [179]. A further study showed that osteophytes predicted pain better in the knee during the last year than pain in the last month or “ever” which all provides little evidence that the recalled pain may be linked to radiographic structural changes [180]. Although structural symptoms may show concordance with symptoms in some cases, the problems with definition mean that results need to be critically assessed.

While there may be issues with symptom definition, the relationship between symptoms and structure may be confounded by radiographic factors themselves such as the radiographic view taken and also the grading system, which is more of an issue to do with imaging. Which views are used in the varying studies appears to have some impact upon the relationship of pain to radiographic knee OA. Using only the A/P weight bearing view, Claessens and colleagues found that the proportion of patients with knee pain that had radiographic knee OA was 36% [181]. Another study using both the A/P weight bearing and lateral views identified 53% participants [182] while Cittucini and colleagues using just the skyline view observed 53%. In terms of grading higher grades of OA (KL Grade 3 and above) were shown to be stronger predictors of pain than lower grades [183].

Concordance between symptoms and structure has been shown in some studies [179, 184, 185]. Using data from the MOST study, KL Grade 4 was strongly associated with knee pain (Odds Ratio 150) compared to KL Grade 0 and similar results seen in the

Framingham study (Odds Ratio 73). This suggests that a factor may be strongly associated with an outcome on the causal pathway but not a strong predictor of that outcome because of the multiple confounders that may exist in a multifactorial disease process and also due to the subjective nature of symptom assessment (mainly pain) [186].

In conclusion, assessment of symptoms and radiographic structure is affected by x-ray numbers and views taken, in studies where fewer views are taken there seems to be reduced prevalence of radiographic OA in a population with knee pain. Additionally, discordance exists between pain and x-ray findings in populations with radiographic knee OA. In addition to pain definition, the grading of radiographs also influences the estimates obtained when associations between pain and radiographic OA are evaluated. The wide variations in strength of association may exist purely because the variation in the radiograph grading which suggests that using a uniform definition of pain may not necessarily improve concordance. Pain could be the result of numerous sources that are not visible on the knee x-ray for example oedema or ligament injuries or other extra-articular sources [187] and because these can co-exist with other causes there are thus so many confounders to assessing this true relationship. The nature of study populations may have a bearing on the associations; younger age groups with knee pain are less likely to have structural OA changes seen on radiography[179] while ethnicity may also contribute to this variation [188]. It could also be hypothesised that in the presence of less radiographic severity there is more discordance due to milder forms of the disease being common. Future research could look into applying more uniform definitions or standardisation of these to allow comparability and also reduce variability.

2.8.6.2 Structure and symptoms (MRI)

OA structural pathology and pain relationships remain incompletely understood, but advances in the use of MRI has begun to improve our understanding about the specific pathologic features that likely play an important role in the aetiology of pain in OA. One barrier to relating the origin of pain in OA is the so-called “structure- symptom discordance”, which reflects the observation that some individuals have radiographic changes with minimal symptoms, while others have more significant pain with only minimal (if any) structural pathology noted on radiograph. Furthermore, OA is typically defined in epidemiologic studies into radiographic OA, based on radiographic findings such as osteophytes and joint-space narrowing and symptomatic OA (i.e., pain, aching, and stiffness). Such nomenclature highlights the recognised discordance between structure and symptoms noted in OA. Evaluation of symptoms (pain) is challenging

because of the multiple risk factors responsible for pain occurrence and severity as well as pain being a subjective and usually assed via patient self-report. Apart from imaging, improved study design is one approach to understanding the structure-symptom relationships for example when confounding inter-individual differences are accounted for (e.g. genetic and psychosocial factors) a string association between structure (using radiography) and pain was shown [186].

MRI studies have highlighted the importance of subchondral bone, particularly bone marrow lesions (BMLs) and synovitis/effusion as contributors to pain in OA [189-191]. Periarticular bone changes associated with OA are classified into distinct patterns based on the anatomic location and pathogenic mechanisms. These include subchondral plate thickness, alterations in the subchondral architecture, osteophytosis and development of subchondral cysts [189]. Bone changes that probably provide the most evidence base for their role in symptom genesis are BMLs. They reflect the histologic changes of fibrosis, trabecular microfractures, and other manifestations of bone remodelling; and have been shown to play a pivotal role in the symptoms that emanate from knee OA and its structural progression [146, 192]. They have also been related to pain severity [193] and incident pain[194]. While there may be some conflicting data, albeit from smaller studies with different methods, that suggest no relation of BMLs to pain [195, 196] on balance the data support a strong relation of bone marrow lesions to pain [190, 191]. Other features that may reflect as BMLs on imaging include periostitis (associated with osteophytosis), subchondral microfractures, and bone attrition are thought to be associated with pain.

Although more data is available for BMLs and they are arguably the most studied lesion in terms of structure-pain relationships, these associations with pain actually vary substantially in their magnitude of association. A recent systematic literature review by Barr *et. al* [197] concluded that subchondral bone features have independent associations with structural progression, pain and joint replacement. Considering the top 5 cohort studies (based on the quality score used in the Barr et.al. SLR) [67] the association or magnitude (beta coefficients and odds ratios) for BMLs was at best modest given the SQ nature of BML scores, and by also noting the lower limits of the confidence intervals (and wide 95% confidence intervals) reported in those studies. The results were as follows: [Foong et.al [198], BML size change (beta=1.53, 95% CI 0.37,2.70), Driban et.al.[199], BML volume change (beta=0.21, SE 0.07), Dore et.al.[200], BML size change (beta=1.13, 95% CI 0.28, 1.98); Kornaat et.al.[201], BML size increase (beta=2.0, 95% -8.0, 11.0 therefore not significant) and Moisio et.al.[202], BML scores reported odds ratios of 1.70 (95% CI 1.07, 2.69) for the lateral tibia and

femur but non-significant findings for the medial tibia and femur.] When the highest scoring cross-sectional studies were evaluated (12 scored highly); these too revealed modest associations and they also had wide confidence intervals: [Zhai et.al.[203] reported odds ratios (OR)=1.44 95% CI 1.04, 3.11 using an outcome of WOMAC pain ≥1; Sharma[204],(OR=1.96,95% CI 1.38,2.77 for BMLs with outcome of prevalent knee symptoms); Kornaat[196],(OR= 1.31, 95% CI 0.62,2.79 therefore not statistically significant); Lo[205] reported relative risks ratios of 1.3,2.1 and 2.3 for weight bearing pain in those with BML scores of 1,2 and 3 respectively vs the referent group BML=0) and Stefanik[206] found that having BMLs in either the medial or lateral patella femoral joint had 1.5 (95% CI ,1.1, 2.0) times the odds of knee pain compared with knees without BMLs. In case-control studies with the outcome (case status) of incident frequent knee pain Felson[194] reported OR 3.2 (95% CI 1.5, 6.8) in BML score increase of at least 2 units in cases and Javaid [177], (OR 2.8 95% CI 1.2,6.5).

For other subchondral lesions such as bone shape, fewer studies have been published: Everhart [207] found that a larger subchondral surface ratio was protective against OA symptoms (OR= 0.48, 95% CI 0.30, 0.75) while Ochiai [208], reported a positive correlation between pain on VAS and irregularity of femoral condyle contours (correlation coefficient, *r*=0.47). Recently Hunter et. al.[209], demonstrated that changes in bone shape were greater in participants with both pain and radiographic progression. In that nested case-control study, Hunter et.al. also found modest associations for change in bone shape at 24 months with pain progression (OR ranging from 1.16 to 1.23) between pain and non-pain progressors at 48 months. More recently Barr et. al. suggested that bone shape was associated with prevalent frequent knee symptoms, modest odds ratios of 1.25 on average per normalised unit of 3D bone shape vector (separately for femur and tibia) were found for their pain outcome which was defined as having knee symptoms (pain aching or stiffness) or medication use for knee symptoms on most days of one month in the past 12 months [210].

Synovitis is increasingly recognized as an important feature of the pathophysiology of OA [211]. Studies have shown that inflammation, assessed as effusion-synovitis and Hoffa-synovitis on non-contrast enhanced MRI [212], is an independent risk factor for incident radiographic knee OA [213, 214] and for radiographic and symptomatic progression [215] and plays a role in the development of centralized pain [216].Synovitis and effusion is frequently present in OA and correlates with pain and other clinical outcomes [22, 205, 217]. Studies have shown that changes in synovitis are associated with change in pain severity [190, 191, 218] and pain fluctuation [144]. The synovial reactions in OA includes synovial hyperplasia, fibrosis, capsular

thickening, activated synoviocytes and in some cases lymphocytic infiltrate (B- and T-cells as well as plasma cells) [219]. Causes of pain are thought to originate from include irritation of sensory nerve endings within the synovium from osteophytes and synovial inflammation that is due, at least in part, to the release of prostaglandins, leukotrienes, proteinases, neuropeptides and cytokines [157, 219].

Of note, BMLs and synovitis/effusion have been associated not only with pain presence or severity, but also with pain fluctuation [144]. Two clinical trials, one using zoledronic acid and the other using a patellofemoral knee brace, demonstrated reduction in pain along with reduction in BML volume, with the patellofemoral knee brace trial showing an effect on patellofemoral BMLs without affecting tibiofemoral BMLs [145, 220]. Synovitis is another attractive treatment target, with intra-articular corticosteroid injection commonly used, providing relatively short-term symptom relief [221]. Through modern imaging it has become evident that the synovitis is common and associated with pain, and not only offers a target for symptom modification but also a structural one [222].

Articular cartilage is both aneural and avascular and thus does not directly contribute to the typical OA symptoms of pain, stiffness and inflammation[159] particularly early on in the disease before neurovascular invasion that occurs in late disease has happened [160]. However there are studies that have demonstrated the importance of cartilage in symptom genesis [195, 223, 224]. These studies have shown that the association between cartilage and pain is generally weak which suggests that although cartilage is not a major determinant of symptoms in knee OA, it does relate to symptoms. OA is a whole joint disease and concurrently affects other tissues that do contain nociceptors. As cartilage degrades, cartilage degradation products are capable of inducing inflammation. However, evidence from cartilage studies that have demonstrated a relation of cartilage damage to pain have traditionally investigated the role of cartilage in isolation from other tissues and as such cannot provide insight into the independent contribution of cartilage pathology to pain. One cross sectional study suggested a relationship between patella cartilage reduction and knee pain severity [225]. Another cartilage study suggested that areas of denuded cartilage were related to symptoms [226]. The most likely pathogenesis for this is through secondary mechanisms such as exposing the underlying subchondral bone resulting in pain following the exposure of nociceptors, vascular congestion of subchondral bone leading to increased intraosseous pressure and synovitis secondary to articular cartilage damage. Therefore, knees exhibiting denuded areas of cartilage are more likely to concurrently have painful tissue pathology.

Another important knee joint structure, the meniscus is important in preserving joint integrity and prevention of further joint damage. There is data that suggests incident meniscal tears may play a limited role in symptom genesis through angiogenesis and associated sensory nerve growth [227]. In clinical practice a consequence of the routine use of MRI in clinical practice is the frequent detection of meniscal tears. Degenerative lesions, meniscal maceration or destruction are associated with older age and are almost universal in persons with OA [228]. In asymptomatic older participants meniscal tears were found in 67%, while in patients with symptomatic knee OA, a meniscal tear was found in 91% [229]. However, OA knees with meniscal tears were not more painful than those without tears, and the meniscal tears did not affect functional status [229]. Meniscal tears are nearly universal in persons with knee OA and are unlikely to be a cause of increased symptoms [230]. In middle-aged and older adults, there seems to be no causal association between meniscal damage and frequent knee pain but this relationship is likely confounded by the fact that pain and meniscal damage are related to OA.

Although very few in number, imaging studies have also investigated the importance of other periarticular lesions detected on MRI, and their association with knee pain. Hill et al. suggested that peripatellar lesions are equally prevalent among participants with knee pain and those without, but other periarticular lesions such as bursitis are significantly more common among subjects with knee pain and may contribute to pain in these individuals [231]. Another study found no association between all other imaging findings (including Baker cysts) with symptoms, but that the only important MRI determinants of pain were only effusion and osteophytes [196].

There have been challenges in assessing structure-symptom relationships. One major challenge is that the natural history of OA structural lesions and pain is not well defined. As a consequence studying the independent effects of each lesion is challenging [189]. As MRI technology becomes more widely used to identify various pathologic changes in the joint, more structures are now accessible but this brings a problem of choosing which lesions to include in the different statistical models. The result is that in some studies all structural lesions are then modelled in the same statistical model to obtain “independent” associations of various structural lesions with pain. It is then a challenge how to compare the magnitudes of effect of each structural lesion on the outcome of interest as their effects may not be totally independent. As the casual pathways are not fully understood, some lesions may represent the initiator or first exposure of cause while others may be promoters also on the causal pathway to increase expression of the disease. Therefore, standard statistical approaches risk

obtaining biased estimates due to different types of biases including selection bias, and the estimates for each lesion are not directly comparable amongst each other [232, 233]. Understanding the disease pathology is as important as image development. Thus, if adjustments are to be made to models to reduce confounding bias, the cause-effect relations between these factors need also to be defined in order to avoid adjustment bias (or over-adjustment and collider stratification bias) [234]. In the assessment of meniscal extrusion and bone marrow lesions on their risk to knee pain as an example, and assuming meniscal extrusion often occurs before BML, any attempt to compare the effects of each lesion from the same model yields difficulty. The effect estimate for BML reflects its total effect on risk of knee pain, while the effect estimate for meniscal extrusion represents its direct effect on pain through pathways other than through BMLs. Therefore these two effect estimates are not directly comparable [189]. Also, as alluded to earlier there maybe issues such as collider-stratification bias [234, 235] which in this example may bias the regression estimate for meniscal extrusion owing and requires appropriate statistical methods to overcome [236, 237].

Another difficulty is that knowing how much of the “true” variance in pain is actually attributable to structural changes is presently unclear. And, with so many lesions added into statistical models this is even more difficult to ascertain. One reason for this could be that structural OA features are assessed late into the disease when numerous pathologies are already present [189]. Also, as shown in one study some radiographically normal painless knees showed at least one type of abnormality on MRI, and MRI-detected abnormalities were equally highly prevalent in those with pain [21]. Improved study design is also key to better understand causality and also reduce potential confounding and the assessment of structure-pain has also been hampered by use of observational studies that do not adequately address this. Therefore the true magnitude of structure on pain has to date not been adequately quantified. Furthermore, as highlighted before the most studied lesions are BMLs however these analyses assessed lesions scored using ordinal SQ scores (usually ranging from 0-3) that are known to be less responsive over time. Definition of change in SQ measures is challenging due to various BML score combinations and there are drawbacks with SQ measures, such as comparing a summed score for BMLs when only one of six sub-regions scores the maximum and the other five score zero for example. Also pain outcome measurements in most cases were categorical, dichotomised to severe (WOMAC pain scores of 2-4) vs not severe pain further impacting on whether any associations found are “clinically meaningful”.

2.9 Symptom measurement in OA

2.9.1 Pain pathophysiology

Amongst knees with ROA, little change is observed in knee pain over six years which suggests that knee OA pain is characterised by persistent rather than worsening symptoms [238].However when large increases in radiographic severity are observed there seems to be worsening pain reported [239]. In summary pain in OA involves complex mechanisms intertwined with those of OA pathophysiology, yet also distinct. OA pain comes from tissue damage, peripheral sensitisation, central sensitisation [159] and neuropathic pain [240] therefore attempts to measure progression need to be cognisant of this. While OA pain has important structural associations with MRI-determined pathology, these other mechanisms must be acknowledged as important determinants of OA pain. More detailed discussion on symptom measurement is discussed below.

2.9.2 Measures of symptom progression

Overall, none of the options available to measure structural damage or cartilage changes over time have been completely satisfactory in measuring OA disease progression. Although most measurement methods can detect changes in relation to the impact of OA on cartilage, these changes have however not been clearly correlated with beneficial outcomes for patients [241]. There have been tools such as proposed for clinical evaluation of symptom-modifying OA drugs however they are not appropriate for evaluation of DMOAD therapy [242] and as such the decision to perform a TKR seems the most relevant endpoint for phase 3 clinical studies as it characterises treatment failure [241].

While TKR could be a reasonable end-point based on its clinical relevance there are issues such as standardisation of the need for TKR that may hamper its use. The decision leading to TKR is not always driven by OA-specific disease severity but could also be influenced by surgeon choice and the patient`s willingness for surgery [243]. Also due to differences in resource allocation across different countries for instance the USA where private health care and public health policies differ from the UK the rates of TKR also differ [244]. Another factor is that patient expectations, which have been shown to vary geographically, may have an impact on the rate of TKR uptake [245] and ultimately may bias use of TKR as a measure of progression. On a practical point of view, the use of TKR raises feasibility challenges because OA is a slow progressing disease, and use of TKR as an outcome measure would require several years of follow up and large patient samples which renders its use impractical.

OA symptom progression is not well defined and could refer to progression of pain severity or incidence of pain in individuals as that usually seen in cohort studies. The selection of specific measures depends on the measurement objective and could be at the joint level, or be generalised for OA, health status and health-related quality of life. As such a combination of measures is often employed to provide robust evaluation, and a core set of measures to be collected in OA clinical research are now recommended [246, 247]. Recommendations are that an assessment of the individual`s quality of life should be included in the measures but because this may capture other comorbidities not related to OA, OA-specific quality of life measures are also recommended [248]. Measurement of symptoms is possible through the use of standardised patient reported outcome measures (PROMs) [246, 249, 250].

The key complaints in people with OA are pain, activity limitations, changes in mood (anxiety and depression) and sleep disturbances. The consensus is therefore that for OA clinical studies, assessments should capture patient-reported pain, physical function, quality of life measure and also performance-based measures of function [246, 247] and consolidating these tools is currently a change in OA [251]. Table 2, not exhaustive provides a summary of selected generic and OA-specific measures commonly used.

2.9.2.1 Measurement of pain

Evaluation of pain can done using visual analogue scales, numeric rating or using standardised questionnaires that ask patients to quantify pain severity. Causes of pain in knee OA however remain poorly understood, and the discordance in structure and symptoms suggests that structural pathology alone cannot account for variation in pain severity [252]. Pain measurement is made difficult because it can be characterised by intensity or severity, the quality of pain (burning or aching for instance) and the pain related to activity or function, sleep disturbance and psychological impact [165].

Numeric rating scales and VAS are used as generic pain measures and these usually measure pain severity over a defined period. The most commonly used knee OA-specific pain scales include Western Ontario And McMasters Universities Osteoarthritis Index (WOMAC) [253] pain scale and the knee Injury And Osteoarthritis Outcome Score (KOOS) [254] however there is evidence that peripheral and central sensitisation contribute to pain in some people with OA thus measures such as the McGill pain questionnaire [255], painDETECT [256] and self-reported Leeds Assessment and Neuropathic Symptoms and Signs [257] pain scales have been developed to identify such subgroups of individuals. These additional outcome measures have been designed with cut-offs to indicate pain sensitisation warranting further investigation for

pressure-pain threshold or heat-pain thresholds. Another measure to assess two OA pain domains (intermittent and constant pain) is the Measure of Intermittent and Constant OA pain (ICOAP) which assesses the frequency, severity and impact on sleep, mood and quality of life [255].

2.9.2.2 Assessment of activity limitation

Activity limitation measurement tools have also been developed but these are largely self-reported as opposed to clinician-led questionnaires. While there may be different measures specific to hand OA and other regions by design there seems to be on overlap between hip-knee measures. For the knee and hip, WOMAC is common to both while and KOOS is specific to the knee and the Hip Disability and Osteoarthritis Outcome score (HOOS) [258] for the hip. The WOMAC physical function subscale uses the same 17 items for hip and knee and this was later shown not to capture items of sufficient difficulty [259] resulting in KOOS and HOOS being developed to address their limitation. KOOS and HOOS have the 17 items but in addition a subscale of higher demand activities, sports and recreational activities.

Many of these measures were developed to reflect functioning of a specific joint but more often the items represent the entire functional unit and accessory structures for example in walking. Questionnaires are usually self-reported and as such they include various sources of measurement error. Activity questionnaires also fail to capture scenarios where individuals no longer partake of certain activities by choice regardless of their physical functioning, therefore performance-based measurements are useful to assess physical functioning in OA.

Performance based measures complement the self-reported questionnaires mentioned previously. They are usually based on standardised activities such as walking or rising from a chair. To evaluate lower extremities the 30-second chair stand testy, 40-m fast paced walk test, stair climb test, 6-minute walk test and timed up-and go (TUG) test are commonly used in trials and measure lower body functional status [260, 261]. Results from these assessment have been shown to correlate moderately with assessments from self-report [262]. In hand OA similar measure include the grip strength and pinch grip measurements [263].

2.9.2.3 Assessment of sleep and mood

A number of patient reported measures are used to evaluate sleep quality in OA. These include Pittsburgh Sleep Quality Index [264], Sleep Disorders Questionnaire [265], and Chronic Pain Sleep Inventory [266] and these measures have now been validated for hip and knee OA with cut-offs defined for poor sleep quality.

Depression and anxiety are a common finding in OA [130] and evaluation of these is important as these conditions are known to affect the interpretation of pain and fatigue which can be confounded in the presence of mood disorders [267]. Most clinical studies have assessed mood using the Hospital Anxiety and Depression Scale (HADS) [268] and Center for Epidemiologic Studies Depression Scale (CES-D) [269] and one study found a link between depressive symptoms and structural OA progression [270].

2.9.2.4 Assessment of fatigue

Currently no OA-specific fatigue measures exist but generic fatigue measures available may be used in OA clinical research to assess overall fatigue, and fatigue as a multi-dimensional disorder (physical, emotional and physical fatigue) [271]. Fatigue is commonly reported in OA and affects the quality of life of an individual (affects pain and sleep) [271, 272]. VAS, numeric rating scales, the Profile of Moods State Fatigue Scale (POMS-Fatigue) [272] and the Functional Assessment of Chronic Illness Therapy Fatigue Scale (FACIT-Fatigue) [273] all measure overall fatigue while the Multidimensional Fatigue Symptom Inventory (MFSI) [274] asses both mental and physical fatigue, thought to be important in people with OA [271] and probably a more important measure compared to the overall measures.

Table 2. Selected outcome measures for use in OA clinical research

Construct	Generic measure	OA specific (Hip or knee)
Pain	VAS	WOMAC pain
	Numeric rating scale	KOOS pain
	McGill Pain Questionnaire	HOOS
	PainDETECT	ICOAP
	S-LANSS	
Activity limitation	Timed-chair stand	WOMAC function
	40-m facts paced walk test	HOOS
	Timed up-and-go test	KOOS
	Stair climb test	
	6-min walk test	
	Grip strength	
	Pinch strength	
Sleep	VAS	
	Numeric rating scale	
	Sleep Disorders Questionnaire	
	Chronic Pain Sleep Inventory	
Mood	CES-D	
	HADS	
	Personal Health Questionnaire Depression Scale (PH-8)	
Fatigue	VAS	
	Numeric rating scale	
	POMS-Fatigue	
	FACIT-Fatigue	
	MFSI	

2.10 Structural assessment in OA

A number of MRI studies have now demonstrated that multiple structural abnormalities are seen frequently in the knee joint supporting the view that OA involves the whole organ (Figure 4) [195, 223, 275-277].

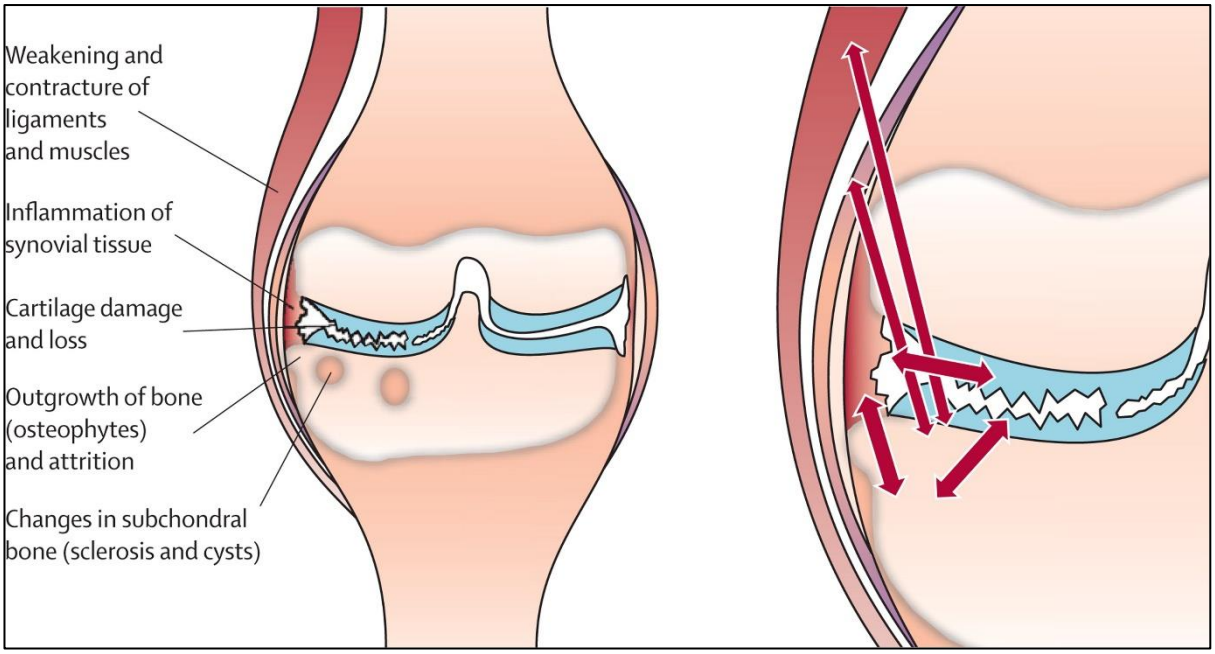


Figure 4. Schematic drawing of an osteoarthritic joint

The different tissues involved in clinical and structural changes of OA are shown on the left. On the right the bidirectional interplay between cartilage, bone, and synovial tissue involved in OA is depicted.

Adapted from [278].

Knowing the interrelationships between different tissues is important in determining the chronology of events leading to OA which is valuable in order to tackle individual lesions early, thus avoiding progression to more advanced stages of the disease [279].

2.10.1 Whole organ assessment

A variety of different MRI sequences have been developed for whole organ assessment of OA owing to the different study needs e.g. morphologic or quantitative evaluation and the fact that different tissues are involved. Selecting the appropriate MRI sequences to study specific OA features is important; in general, fluid-sensitive fat suppressed sequences (e.g. T2-weighted, proton-density-weighted, or intermediate weighted fast spin-echo sequences) are useful for evaluation of cartilage, bone marrow, ligaments, menisci and tendons [280]. These sequences are important when assessing focal cartilage defects and BMLs while gradient-recalled echo (GRE)-type sequences such as 3D-spoiled gradient echo at steady state (SPGR), double echo steady state (DESS) and fast low angle shot (FLASH) are not as they are prone to artefacts which hinders accurate image interpretation [281].

Although MRI has remained largely a research tool, new scoring systems have been developed for quantifying pathological changes occurring in the different tissues in the

knee and hand, an important step in refocusing our attention from just evaluating cartilage. These scoring systems have initially been semi quantitative scoring systems such as the Whole-ORgan Magnetic resonance imaging Score (WORMS) [282], the Knee Osteoarthritis Scoring System (KOSS) [283], the Boston Leeds Osteoarthritis Knee Score (BLOKS) [284] and the MRI OsteoArthritis Knee Score (MOAKS)[212].

The following text describes the SQ assessment of the 4 well-described systems as introduced previously, for whole organ assessment based on MRI. The figures thereafter describe the anatomical scoring positions of each one and the table compares the systems.

WORMS has been extensively used in MRI studies worldwide including the OAI, MOST and Framingham study. Articular surface features (cartilage and subchondral bone) are scored in 5 sub regions and 4 patellofemoral sub regions are scored. By dividing the knees into sub regions and then aggregating each lesion type in each compartment, a whole knee score is calculated. Scoring in this way allows multiple lesions occurring within a sub region to be scored together which simplifies data analysis which is advantageous over a lesion based approach as there is no risk of misclassification if lesions coalesce or split during longitudinal evaluation. The WORMS is currently the only whole joint system that includes subchondral attrition.

KOSS was introduced in 2005, and its description of OA features is very similar to WORMS but has different sub regions and does not score attrition, bursitides, loose bodies and ligaments. For the medial and lateral side, the anatomical regions scored by KOSS are the patella facets (divided by the patella crest), femoral trochlear facets, femoral condyles and tibial plateau. The KOSS scoring system also includes effusion, menisci, synovitis, and Bakers cysts but unlike WORMS each scoring grade is differentiated by the size of the lesion. The KOSS scores BMLs, subchondral cysts and cartilage individually unlike WORMS that scores these cumulatively. The 2 systems both score meniscal morphology however KOSS additionally scores meniscal subluxation.

BLOKS, introduced in 2008 applies a lesion-oriented system unlike WORMS and this offers advantages for longitudinal analysis of individual lesions. The sub regions are based on weight-bearing regions of the knee rather than the patellofemoral joint. In each tibio-femoral compartments, 2 sub regions are scored for cartilage and subchondral bone and 4 patellofemoral sub regions exist in the BLOKS. A single sub region for the tibia is defined and a “weight bearing” portion of the femoral condyle. For lesions extending across multiple sub regions, the lesion is assigned to its most involved region.

MOAKS was designed based on the relative advantages and disadvantages of the existing scoring systems at the time. The MOAKS incorporates sub regional assessment, a refinement of the BML and meniscal scoring. WORMS and MOAKS both divide the femur and tibia into anterior, central and posterior regions while BLOKS divides the femoral condyles into trochlear and weight-bearing regions as described previously. The WORMS, BLOKS and MOAKS scoring systems share a similar delineation of medial and lateral regions of the femur.

While these systems have greatly improved whole organ assessment and have contributed to the description of the natural history of OA, these systems are SQ and some of the ordinal measures used in these systems are not interval scale and therefore progression from one grade to another is not exactly uniform. There remains some controversy regarding which method is a better validated outcome in knee OA, as shown by the different longitudinal findings between BLOKS and WORMS. Using the same knees, WORMS BML scores found a stronger association with cartilage loss than the BLOKS scoring however BLOKS meniscal tear scoring showed a stronger association with cartilage loss than the WORMS's meniscal scoring [22]. Another study found that unlike BLOKS, the WORMS (BML score) was not associated with pain and also less strongly associated with cartilage loss [228].

Some advances in imaging including the use of quantitative 3D measures of bone, meniscus and cartilage are now being realised in the hope that they afford more sensitivity to change, better reproducibility and higher responsiveness. There has also been improvements in SQ measures by including within-grade changes to improve responsiveness in longitudinal assessments, in recognition of the superior sensitivity that quantitative measures have [285]. Sub regions used in WORMS, MOAKS and BLOKS scoring systems for knee OA are shown in Figure 5.

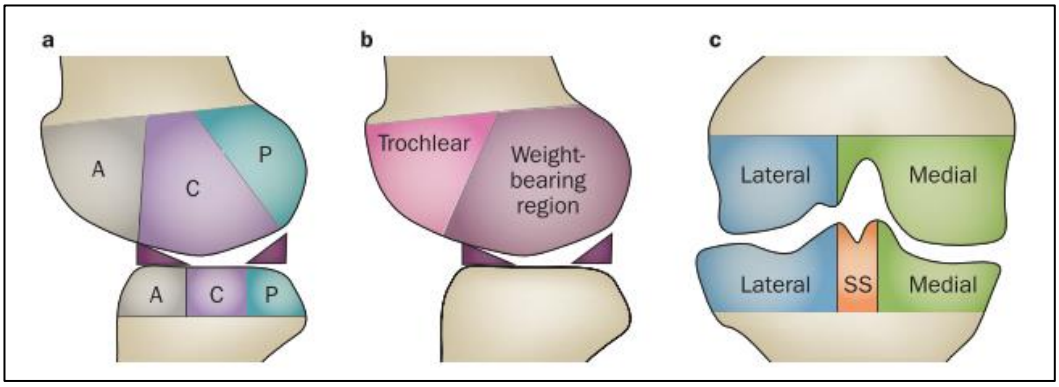


Figure 5. Regions used to score different scoring systems

a) WORMS and MOAKS divide the femur into anterior (A), central (C) and posterior (P) regions
b) BLOKS separates the femoral condyle into trochlear and weight bearing sub region
c) The medio-lateral division of the femur common to all 3 scoring systems

2.10.2 Reviewing traditional measures of disease progression

Despite a huge unmet need for therapies that improve both symptoms and limit structural damage associated with OA, the value of joint imaging is still dependent upon the clinical or research question and the characteristics of the imaging modality. Radiography is not usually required for routine clinical diagnosis unless differential diagnoses or decision to support a TKR is sought. Also, the fact that treatment is mainly symptomatic, and with poor correlations seen between radiographic findings and symptoms further justifies the lack of such imaging routinely.

As mentioned before, OA involves all tissues within the joint including cartilage, subchondral bone, synovium and periarticular tissues therefore in research different imaging modalities such as MRI and ultrasonography are increasingly used. While MRI permits assessment of multiple tissues, radiography remains the most studied in knee OA. This section reviews radiography in more detail and other imaging modalities thereafter.

2.10.2.1 Radiography

The ability to quantitatively assess structural and symptomatic progression is an important step in understanding how OA progresses. Structural changes on radiography are considered the primary outcome variables for assessing progression of OA. Depending on the joint studied, several measures are currently used to assess radiographic progression including individual (e.g. JSN) and composite measures of radiographic pathology (e.g. KL grade). The majority of studies focus on the measurement of structural OA 'progression'. However structural severity is poorly correlated with patient symptoms. The incidence of joint replacement (reflecting symptom severity and structural progression) and incidence or progression of clinical symptoms represents more patient-centred outcomes.

It is widely accepted that radiography is the method of choice for evaluation of the efficacy of DMOADs [241, 286, 287]. Radiography is a reliable, easily reproducible, relatively cheap, and does not require specialist facilities when used to measure long term outcomes in OA. There are however limitations with the use of radiography due to the need for standardization of methods (such as joint positioning) which may lead to variations. Given that only slight changes are seen over time, the source of variation needs to be reduced greatly to improve the ability to detect changes. There are a number of factors contributing to variation and measurement error in radiography [288,

289], such as the measurement process, reader experience as well scoring method used.

2.10.2.1.1 Pathologic features assessed by radiographs

The radiographic hallmarks of OA include joint space loss, osteophyte formation and subchondral bone changes such as sclerosis and cyst formation[290]. Initially radiographs may not show all findings, with only minimal JSN being present and these spaces are normally asymmetric and with disease progression subluxations may occur with formation of osteophytes [291]. Subchondral sclerosis is a characteristic feature of OA and this occurs as cartilage loss increases, appearing as an area of increased radio-density while joint collapse occurs in advanced stages [292, 293].

Osteophytes have the appearance of outgrowths at articular margins, capsular insertions and central articular regions, formed through endochondral ossification and associated with increased bone turnover [294, 295]. Earlier changes in the OA pathogenies occur in the subchondral bone, beneath articular cartilage and involves subchondral cortical plate and bone trabeculae thickening [71]. The thickening or sclerosis represents a failed bone homeostatic process following bone micro fracturing resulting in increased bone turnover which produces excess hypo mineralised osteoid bone, incapable of load dissipation and ultimately results in joint destruction [296].

JSN which represents the distance between bones has been previously assumed to represent cartilage thickness due to radiographic imaging in weight bearing joints but recently through advances in MRI has been shown to be a composite measure of meniscal pathologies such as extrusion and degeneration as well as cartilage pathology [297].Radiographic JSN is however important is in distinguishing OA from other arthritic diseases e.g. RA as marked JSN is more typical of later stages of OA.

In established OA, subchondral bone consists of excess hypo mineralised osteoid unable to distribute load effectively. This is associated with bone attrition and on radiographs is seen as flattening of the bone surfaces which impairs the joint congruity, resulting in further stress on the articular cartilage. A vicious cycle then develops as cartilage loss in weight bearing compartments coupled with denudation of bone is associated with further attrition and alteration in joint alignment. Incongruent bone shape within diarthrodial joints is recognised as a predisposing factor for adverse biomechanics and failure to dissipate load evenly.

Through the use fractal signature analysis, radiographic images can be further characterised for subchondral bone morphometry, which provides a description of subchondral bone microarchitecture and properties such as trabecular size, number,

and spacing and cross linking [71, 298]. Studies have shown that increased radiographic severity was associated with thicker trabeculae and less space between them [71] while others have demonstrated that alterations in knee trabecular morphometry was associated with JSN, cartilage thinning and TKR [299-301]. A summary of pathologic changes in OA and corresponding radiographic changes and linkages are sown in Figure 6.

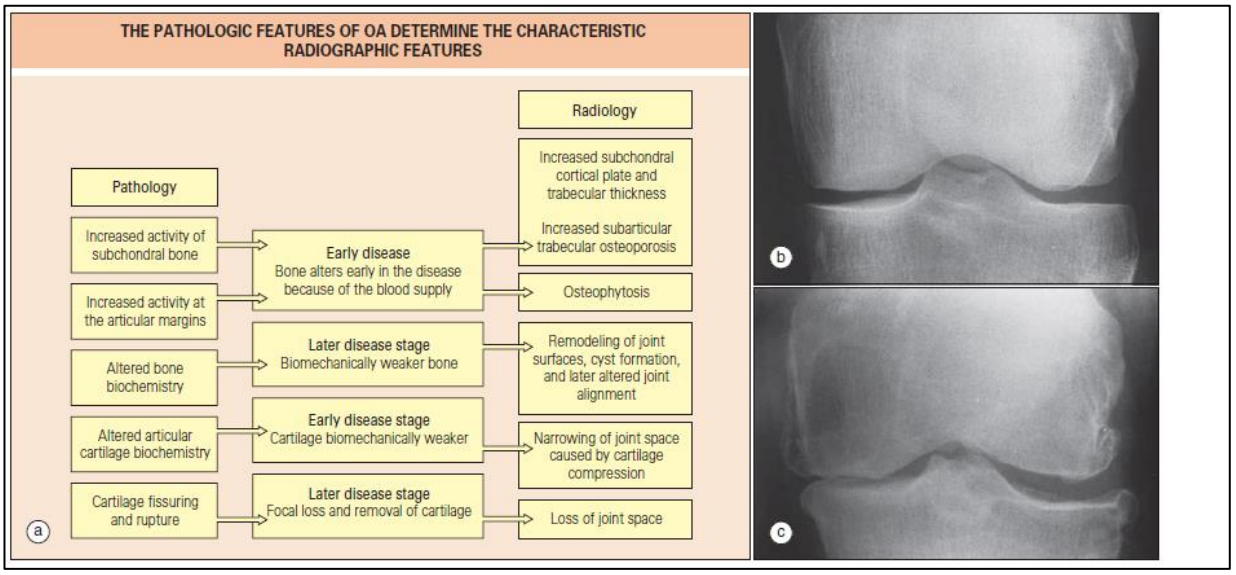


Figure 6.Summary of pathologic changes and corresponding radiographic changes.

(a) Pathologic changes in osteoarthritis and the corresponding characteristic radiographic features visualized in plain radiographs of joints with early- and late-stage disease.
(b) Radiograph of a healthy knee.
(c) Radiograph of a knee with late disease showing most of the features listed in **(a)**

2.10.2.1.2 Quantification of pathology on radiographs

2.10.2.1.3 The Osteoarthritis Research Society International (OARSI) atlas classification

Various radiographic measure are currently employed, such as direct measurement of interosseous distance for JSN and other studies requiring classification or grading just use visual comparison with a standard [290]. JSN correlates with erosion and is a sensitive predictor of knee OA when it affects the medial tibio-femoral (TFJ) or lateral patella-femoral joints [182]. The OARSI atlas provides distinct scoring systems for each joint and describes JSN, osteophytes with a semi-quantitative score (0-3), where 0=normal; 1 =mild or 1-33%abnormal; 2=moderate or 34-66% abnormal and 3 severe or 67-100% abnormal. The presence or absence of subchondral sclerosis, cysts, bone attrition and also joint malalignment is also evaluated [18].

JSN is considered the current standard radiographic atlas for OA with good intra-observer and inter-observer agreement in clinical studies [291]. Other atlas–based radiographic measure such as that used by Nagaosa have been trialled and showed comparable reproducibility to OARSI atlas, but showed discordance in some aspects of the grading [302].

Evidence of predictive and construct validity of OARSI JSN has been provided. In the MOST study, OARSI knee JSN was significantly associated with cartilage and meniscal damage, and extrusion after adjusting for age, gender and BMI [303]. Knees with higher OARSI JSN grades showed thinner cartilage in the medial TFJ than those without JSN and also showed larger subchondral bone areas in the OAI study [304]. In a study evaluating the association between grades of three ROA measures with pain severity, all measures showed an association with pain but OARSI JSN grade and KL grade had greater association with OA pain than OARSI osteophyte grade [186]. As OARSI JSN grade has construct validity, it appears to offer advantages over KL grade in terms of its reliability and responsiveness.

Use of JSN as either an absolute measure or grade although now commonly used, does however have methodological problems [305, 306]. This measure has also been shown to be poorly responsive in knee OA [307-309]. In a study comparing agreement and responsiveness between categorised metric JSW, KL grade and OARSI JSN grade, categorical metric JSW had comparatively better reliability and responsiveness and OARSI JSN had the lowest inter-observer reliability [310].

2.10.2.1.4 Kellgren Lawrence grade

Some semi-quantitative methods such as KL grade have been developed based on a global assessment combining several features: grade 0 = absence of any sign of radiological OA; grade 1 = possible osteophytes only; grade 2 = definite osteophytes and possible JSN; grade 3 = moderate osteophytes, and /or definite narrowing of joint space; grade 4 = large osteophytes, marked narrowing of joint space, severe sclerosis and definite bone deformity [17].

Despite being widely used the KL grade-system has been criticised for the inconsistencies seen when interpreting the scores [305], and the prominence given to osteophytes at all sites is problematic because minor changes in rotation of the knee can affect osteophyte appearance. There are already existing challenges with standardisation of positioning which also compound these problems [311]. Another criticism is the moderate reliability of these scores [306, 307] and lack of sensitivity for example only increases of two more KL grades in knees over 4 to 5 years were

associated with increasing pain and dysfunction in one study [239]. Insensitivity to structural progression has been demonstrated in the MOST study also where structural progression was defined using JSN grades and had KL grade been used instead, less than half of those deemed to have progressed would have done so using this measure [312].

The KL grade does not represent a truly interval scale where individual categories are equidistant from each other in severity which is important to recognise, as progression from one grade to the next is not comparable for all starting points in the scale. The OA threshold of KL 2 has been inconsistently defined in longitudinal studies which have led to some heterogeneity in definitions of OA. Almost all definitions of KL 2 require a definite osteophyte, while some do not mention JSN, and some require JSN [313]. A study using both MRI and X-ray data found that 4 % of knees that were KL grade 2 but (OARSI JSN=0) had cartilage loss while 44% that had KL grade 2 but had (OARSI JSN=1) showed cartilage loss [312]. As a result KL grades seem to encompass a broad spectrum of OA pathology which results in inaccuracies in definition of progression.

Despite its limitations KL has advantages in that the score acknowledges the “whole joint” knee OA concept and incorporates this to reflect multi tissues in the scoring system. Baseline knee KL score has also been shown to predict knee replacement in the OAI. Semi-quantitative measures such as KL are, despite the limitations more convenient to perform, rapidly scored and require little complex imaging technology but are not sensitive to clinically meaningful change in symptoms.

2.10.2.1.5 Progression on radiography

Radiography has been used extensively to provide a classification system and thus been useful in defining OA in epidemiological cohorts, especially systems such as KL grading. Radiographic JSW in persons with knee OA is moderately responsive to change (SRM of 0.33, increasing to 0.57 in studies longer than 2 years duration) [314] however, these estimates are much lower than MRI cartilage measures [315, 316]. Radiographic grade (assed by KL grade) has demonstrated predictive validity for joint replacement [317]. Recently, Wirth and colleagues found much improved responsiveness when applying location-specific measures of JSW instead of standardized minimum JSW (mJSW) [318]. Eckstein and colleagues using the same cohort (OAI) then showed that a change in location-specific JSW was a stronger predictor for knee replacement than mJSW [279]. Both studies suggest that regulatory agencies need to revise mJSW as a structural endpoint for DMOAD intervention trials, and instead use location-specific measurement of JSW in future studies.

Trabecular bone structure reflects the structural progression of OA and can be measured by fractal analysis of radiographs. Longitudinal studies have shown that alterations in the trabecular bone can predict incident radiographic JSN, MRI-measured cartilage thinning and TKR [299-301]. However, as highlighted previously the precision of radiography remains an issue, with large numbers of participants needed to demonstrate change over time in clinical trials.

2.10.2.2 Magnetic Resonance Imaging

MRI provides good soft tissue contrast and covers the entire joint providing a 3D view of the knee, allowing for the visualisation of different sections in any plane. Through the use of MRI we can visualise structural pathology of the hyaline cartilage, the menisci, synovium, bone shape, bone marrow, and ligaments regardless of their location [319]. MRI abnormalities are seen more frequently with increasing radiographic OA severity but as shown before, MRI detects OA structural pathologies in the pre-radiographic [320] and also radiographic phases of OA [8].

MRI works through application of a magnetic field and then use of different pulse sequences to visualise the tissues within joints. After a radiowave pulse, the protons within different tissues are aligned and then resume a state of equilibrium by which they can be measured as T1 and T2 relaxation signals. The protons in each tissue (e.g. fat, fluid, and muscle) have different T1 and T2 relaxation times, and this provides soft tissue contrast based on the different signal sizes coming from each tissue and therefore different tissues will have optimal sequences [321] to best characterise them.

2.10.2.2.1 Pathologic features assessed by MRI

Cartilage abnormalities are now more readily detectable by MRI, with improving accuracy and reliability. From epidemiologic studies of knees at risk of knee OA, MRI-assessed cartilage damage has been associated with incident radiographic OA at follow-up, and worsening cartilage damage associated with incidence of OA in a dose-response relationship. The same study also showed cartilage damage was associated with incident and persistent knee pain [320]. More advanced techniques have now been developed to assess cartilage changes even before macroscopic degeneration has occurred through the evaluation of glycosaminoglycan concentration and collagen properties. Techniques such as delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) [322], diffusion-weighted imaging, and T2 and T1 rho mapping procedures can provide information on the composition and structure of the cartilage matrix [323] but are however limited to research and not well validated yet owing to challenges in standardisation.

MRI is more precise at detecting synovial effusion and hypertrophy and contrast enhanced MRI remains the gold standard for its assessment. Joint effusion is best detected on fat-suppressed proton density–weighted or T2-weighted fast-spin echo sequences. The use of MRI has shown that the prevalence of inflammation in OA is much greater than previously thought with some estimates as high as 80% and 90% in large knee cohorts [22, 23]. Quantitative MRI markers of synovitis include synovial membrane thickness, synovial volume, and the rate of synovial enhancement after intravenous injection of contrast agent. Although contrast enhancement improves the precision of detecting an inflamed synovium [324] accurate quantification of synovitis can be achieved without using contrast agents [325] and with recent concerns over the potential toxicity of gadolinium-based contrast agents, such methods may need further study. The importance of imaging synovitis and effusion is that their presence has been associated with pain, while synovitis in knee OA is associated with structural progression longitudinally.

Other tissue pathologies such as meniscal tears, degeneration and extrusion are associated with knee OA and are detected by MRI, with prevalence of meniscal tears seen with increasing radiographic OA [297]. Also, ligament abnormalities seen on MRI may be an early feature of OA as described previously[156]. The importance of MRI in assessing the whole joint is clearly invaluable; meniscal damage has been associated with incidence and worsening BMLs [326], which are the most commonly studied bone pathology on MRI, and been associated with longitudinal structural progression, pain severity and TKR [107, 200, 327]. Bone attrition typically seen in advanced OA is also commonly seen with BMLs [328] and has been shown to associate independently with cartilage loss [329].In addition meniscal pathology and malalignment have been shown to co-exist with bone attrition [330, 331]. With the recognition of the importance of muscle strengthening (in particular quadriceps strengthening exercises) as a treatment for OA pain relief [51, 332, 333], newer studies have now attempted to assess the independent muscle imaging risk factors for OA structural progression [334] and this is an area of ongoing research.

As seen for other subchondral bone pathologies, osteophytes develop in association with increased mechanical load for example as a consequence of meniscal degeneration [335], ACL tears [336], and cartilage damage [337].Although marginal osteophytes are seen with radiography, central osteophytes are more easily seen with MRI. MRI studies have also demonstrated that osteophytes are common in people with no knee symptoms even in the absence of radiographic changes [21]. MRI-determined

osteophytes are independently associated with knee OA structural progression but not with pain [67].

There is growing recognition for the role of bone in both early- and late-stage knee OA. Three dimensional bone shape, as measured using MRI provides new perspectives into structural progression in OA. OA represents a failure to effectively dissipate load within the joint tissues, and Wolff's law describes how bone shape readily changes in response to mechanical forces acting on it [296]. Changes in bone shape are evident over 10 years before radiographic measurements pick up structural change and bone shape continues to demonstrate progressive change throughout the course of the disease. In the OAI, change in bone shape longitudinally, formation of osteophytes, and widening of articular surfaces, was shown to be very consistent in participants with knee OA [338]. Also from the OAI, in a sub study comparing OA to non-OA knees over a period of four years, the area of the medial femoral condyle of OA knees was observed to change 3 to 4 times more rapidly than non-OA knees showing that MRI-defined bone shape was a responsive measure of progression [339]. Furthermore, bone shape is independently associated with TKR [340], onset of radiographic knee OA [111] and pain and structural progression [209].

2.10.2.2.2 Quantification of pathology on MRI

MRI measures can be classified into SQ and quantitative outcomes, with tissues being assessed by morphological features or composition. The most common semi – quantitative MRI measures that assess multiple tissues in the knee joint typically using an ordinal scale (e.g., 0, 1, 2 and 3) were reviewed in section 1.9. Each of these measurement tools divides the knee into various anatomical sub regions and uses ordinal scales to describe the structural severity of multiple tissues within these sub regions. However, some SQ systems focus a single tissue for example cartilage defects [341], BMLs [146], ligaments [342], meniscus [343], and synovitis [344].

Quantitative imaging involves the segmentation of tissues, which is usually performed manually or using semi-automated methods. The segmentation of tissues permits the quantification of specific tissue characteristics ranging from cartilage thickness, volume, BML volume, 3D bone shape, synovial volume, and muscle area. Longitudinal measurements of cartilage volume or thickness provide quantifiable data to assess structural progression of OA [345, 346] but because the rate of change in osteoarthritic joint tissues is very slow therefore detecting structural progression using cartilage loss great reliability and accuracy in the segmentation process. Compared with quantitative measures, SQ systems use a simpler and quicker process for assessing structural pathology than segmentation of tissues, but the ordinal scales are less sensitive to

change. Also, the imaging sequences required for quantitative measures differ from SQ grading because the precise distinction of the osteochondral interface and cartilage surface is essential, and spatial resolution is of greater importance [345, 346].

Core measures that comprehensively describe cartilage morphology, and its longitudinal change have been described [347] and the responsiveness and reliability of SQ and quantitative measures of BMLs and cartilage in knee OA have shown to be good [348]. As a consequence, the measurement of longitudinal structural change with SQ measures has evolved to incorporate “within-grade” scoring changes to improve longitudinal sensitivity to change.

The implications for knee research is that MRI SQ and quantitative measures have started to be used as clinical outcomes in structure-modification DMOAD trials. This reflects the opinion of the OARSI working group that recommend MRI cartilage morphology assessment as primary structural endpoints in clinical trials, and acknowledges the rapid evolution of quantitative MRI assessments of subchondral bone and synovium [315, 348]. While radiography is favourable in terms of cost and accessibility compared to MRI for endpoint definition, its insensitivity dictates that large numbers of participants are still required and longer follow up duration necessary to demonstrate change. Reduction in trial size and duration, if MRI were used would therefore offset the costs associated with MRI imaging analysis and could make OA trials favourable for the pharmaceutical industry with a view to develop DMOADs.

2.10.2.2.3 Progression on MRI

Although not required for diagnosis in clinical practice, MRI has shown great promise due to its capacity to visualise the whole knee joint. As previously stated, synovitis is important in OA and is a possible target for structure modifying OA drugs. MRI-detected effusion was shown to predict increased cartilage loss and pain [349].

BMLs, the most studied of the potential bone biomarkers cannot be detected on MRI as areas of high signal intensity. Besides their association with pain, BMLs have been associated with other OA risk factors such as increased loading due to obesity, joint malalignment, and meniscal pathology, as well as pain and structural progression in knee OA [191, 337, 350] . A recent study found that large BMLs were associated with structural damage and progression after 48 months, whereas small baseline BMLs were of less clinical relevance while a decrease of BMLs over time, although related to decreased pain, did not predict improvement of structural aspects of OA [199]. Importantly in the regions that showed decreases in BML size, a trend toward increased cartilage defects and increased JSN was seen. This could have clinical trial

relevance in that baseline BML size rather than change in BML size longitudinally could be more important than longitudinal in predicting OA progression and may provide a better trial endpoint.

MRI has been useful in demonstrating changes in early OA. Roemer et al [214] found that presence of Hoffa synovitis, effusion synovitis, medial BMLs and medial meniscal damage increased the risk of OA prior to incident radiographic OA, and that having more MRI features increased the risk compared to just having one. In cases of existing radiographic OA, MRI-defined cartilage thinning (presence and worsening over time) is a robust predictor for radiographic knee OA progression [351].Meniscal extrusion detected by MRI has also been reported as a separate risk factor for OA progression in the tibiofemoral joint [352].

Modern imaging analysis using MRI-derived 3D measures has now resulted in more accurate quantification of structure. This has provided new insights into the importance of bone shape in OA. In the OAI, bone shape showed greater change, in a linear manner annually in participants with higher KL grades than those selected for persistent KL grades of 0. Bone shape has also been shown to be a more responsive measure of OA progression than radiographic JSN or MRI cartilage thickness [339]. Furthermore, bone shape has been shown to predict TKR[340] while 3D image analysis of BMLs recently showed that these features were closely aligned with adjacent areas of cartilage loss, thereby supporting a biomechanical origin for these features [353]. 3D bone shape has also been shown to predict incident radiographic OA, as well as radiographic and pain progression in large analyses from OAI [111, 209].

2.10.2.2.4 Measurement error in MRI

While MRI is an important advance to imaging studies it is also subject to measurement errors. Measurement error is inherent in MRI with many potential sources of variability and introduction of bias during image acquisition, processing and analysis. The degree of bias is dependent on the quality of the image, the application of a measurement protocol, patient factors and interactions among these.

Measurement errors due to the technology may include gradient-induced distortions leading to image warping because of magnetic field inhomogeneities. Other factors such as manipulation of the resolution are a result of a trade-off between acquisition time and the desired resolution which may affect the level of detail obtained, and because these are built-in to the protocol will vary across radiologists and institutes causing significant source variation [354]. Another technical limitation inherent to MRI is

presence of partial volumes which affect measurement reliability by decreasing clarity in the target region being measured. Partial volumes occur when the MRI slice thickness exceeds image resolution or the structure being measured. The effect of these sources of biases is that of increased effect sizes, the effect sizes also vary according to the tissue being measured. In one study, the effect sizes for partial volume inclusion compared with partial volume exclusion exceeded the effect sizes attributable to measurement reliability. This suggests that if unaccounted for, partial volumes in measurement protocols could produce a spurious, statistically significant results [355].

Imaging processing also presents some source of bias. Segmentation is user-dependant with different reliability indices observed. Also, depending on whether it is manual or automated this could result in variations and potential bias. Manual segmentation is often time-consuming and prone to errors due to various inter- or intra-operator variability studies. Registration and smoothing algorithms also vary and could potentially introduce measurement bias [356].

At the patient level, sources of measurement error can arise from mal-positioning during imaging resulting in non-standard measurements taken across studies. In knee OA, one potential limitation and source of bias is that patients are imaged in a supine position which could be produce measurement error for example in assessing the meniscus as changes could be more responsive under load [357]. Another potential source of variation could be the activities performed by participants prior to the scan. High impact physical activity shortly before an MRI scan can interfere with the results of the diagnosis, resulting in false positives. When contrasts agents are used as part of the MRI protocol, exercised muscle will take up more glucose leading to misdiagnoses [358].

2.10.2.3 Ultrasonography

Ultrasonography is a real-time imaging modality that permits dynamic assessment of joints, providing a 3D multiplanar aspect. US is non-invasive, cost-effective, and widely accessible which makes the evaluation of multiple joints in a single sitting and the repeated evaluation of peripheral joints feasible. US machines consist of a transducer array with generates US pulses and receives returning echoes from tissues thereby generating an image. Known as grey scale US, images can be anechoic, hyper- or hypo-echoic and used to demonstrate joint effusion, synovitis and bone erosions. Doppler US makes use of the Doppler Effect to visualise and measure blood flow. US waves used for generating US images are elastic mechanical waves requiring elastic or viscoelastic medium. As a result, US cannot visualise structures beneath or within tissues that are not viscoelastic (e.g. pathologic changes deep to the cortex such as

bone marrow lesions) or those very deep to the skin for the echo to be detected (e.g. cartilage in the centre of the knee). Most work assessing the validity of US has been based on inflammatory arthritis, in which it was shown to be more reliable than radiography. However, US imaging is operator dependent and is therefore also limited by the proficiency of the ultrasonographer [359].

2.10.2.3.1 Pathologic features assessed by ultrasonography

Although US can assess a wide spectrum of structures, including the bony cortex, tendons, ligaments, bursae, aspects of the menisci its main use in OA has been to detect synovial pathology, where it performs better than radiography. In OA, US readily demonstrates joint effusion and synovial thickening with one study showing a prevalence of just under 50% for synovitis or effusion on US examination of a symptomatic knee [24]. US-detected knee synovitis was also shown to be associated with WOMAC pain in OA cross-sectional analysis in a study with very high inter-observer agreement [360], while in longitudinal studies effusion was associated with TKR[361] and furthermore US could detect short-term synovial response in knee OA In particular, power Doppler score following corticosteroid therapy [362].

As US cannot penetrate bone structures, it is not feasible to visualise the central load-bearing regions of the joint which typically degenerate in OA and thus its use in assessing cartilage pathology is limited. However in cross sectional studies, cartilage pathology on US has been associated with radiographic JSN [363] and while US can be useful in detecting already established osteophytes in OA joints, other subchondral bone changes such as BMLs cannot visualised making US less useful for early OA changes [364]. While analyses of US-determined meniscal degeneration in the knee indicate good reliability [360] MRI has advanced meniscal imaging and established the importance of meniscal degeneration in the pathogenesis of OA, therefore US-determined meniscal assessment requires further validation before considering its use in OA clinical trials. One advantage of US over MRI is in the imaging of extruded meniscus which is more feasible with US for weight bearing studies [365].

The US pathologies described (synovitis, cartilage damage, medial meniscal damage, and osteophytes) in knee OA have been scored using SQ scoring methods which although show fair to good reliability are hampered by poor responsiveness as mentioned before for SQ measures. In a slowly progressing disease like OA US measures are unlikely to be recommended as primary structural outcomes in knee OA trials, however in hand OA the OARSI guidelines suggest that for hand OA US may be used to assess structural damage [366].

2.10.2.4 Other modalities to assess structure

Other less commonly used, but useful imaging methods in the knee include bone scintigraphy, computed tomography (CT) and positron emission tomography (PET).

Bone scintigraphy reflects alterations in bone metabolic activity through use of radiopharmaceutical agents that accumulate rapidly in bone by adsorption to the mineral phase of bone. Imaging detects areas of isotope accumulation, which represent areas of high bone turnover. In hand OA, studies using scintigraphy showed increased bone activity before any radiographic changes were seen and in patients with chronic knee pain scintigraphy had good correlation with MRI-detected subchondral bone lesions [367]. Scintigraphy (likely through the detection of BMLs) has predicted disease progression in OA in a 5-year follow up study [368]. Scintigraphy is cheap and readily available but despite its high sensitivity; shows low specificity and is associated with a significant radiation dose which severely limits its use in OA clinical practice.

CT has been well used in spinal imaging. It is a cross-sectional digital imaging method that uses advanced radiographic technology, and is particularly effective at visualising cortical bone and also useful when validation of MRI imaging is required for subchondral bone. When used with contrast medium, CT is comparable to MRI for qualitative assessment of knee cartilage and could be the image of choice when MRI is unavailable or contraindicated. However, CT currently has no established role in OA trials due to the high doses of radiation associated with it and because of its low soft tissue contrast, gives no detail on the intrinsic structure of cartilage. Recently, 3D CT scanning has been used in hip OA to assess osteophytes and JSN [369] and further work could make this modality relevant to accurately measure structural disease.

The role of PET in assessing OA is not well established due its poor anatomical resolution, cost, and exposure to ionising radiation. PET is a type of nuclear medicine imaging that allows comprehensive imaging of the whole joint, including soft tissues and bone [370, 371] and works by demonstrating metabolic changes in target tissues, reflecting glucose metabolism in OA [372]. However in a study of individuals with or at risk of radiographic hip OA, PET signal was associated with increasing radiographic severity although this was cross-sectional [373]. Using fluorescence–PET, synovitis and BMLs have also been shown in OA but again this was cross-sectional [374].

2.10.2.5 Progression on other imaging modalities

With superior images of the bony cortex and soft tissue, computed tomography (CT) may serve as a reasonable gold standard in OA research when validating MRI bone morphology such as osteophytes however its main disadvantages are its low soft-tissue contrast and higher radiation doses than that of other modalities. Quantitative 3D analysis of the proximal femur using CT was shown to identify changes in cortical bone thickness that correlated to structural hip OA [369]. Nuclear medicine imaging for example positron emission tomography (PET) have not been applied as much in OA clinical research possibly due to higher radiation levels. However, Kobayashi and colleagues have demonstrated that an increased uptake of radioactive fluoride as a bone-imaging tracer represents early abnormalities in the subchondral bone, found prior to radiographic JSN [375] however it is unclear if these findings demonstrate very early OA without replication in other cohorts. A longitudinal study demonstrated that baseline PET signals predicted incidence and progression of OA as well as worsening of pain in the hip [376].

In summary, with modern imaging, short-term changes of novel outcome measures may better reflect long-term changes in structural outcomes and thus make randomized trials more feasible. Although there are currently no licensed DMOADs, it is likely that novel imaging techniques will be used to quantify structural changes and more efficiently establish new therapies. Imaging biomarkers of joint structure are essential in understanding the natural progression of OA at a time when a new pharmacological agents are being tested for their potential structure modification. We now understand that the best responsiveness in progression in clinical trials can be achieved using either quantitative cartilage thickness or 3D bone shape. Radiography will continue to be important for diagnostic purposes and in clinical studies MRI seems mandatory.

2.10.3 New developments in structural assessment

The use of MRI techniques to investigate tissue pathology has become increasingly widespread in OA research. MRI-determined quantitative and SQ measurements are now being utilised as clinical outcome measures in structure-modification DMOAD trials reflecting the current opinion of the OARSI working group that recommended MRI cartilage as a primary structural clinical trial end point, and also acknowledged the emergence of quantitative MRI assessments of subchondral bone and synovium. Use of MRI imaging biomarkers of the knee brings a significant improvement in sensitivity,

responsiveness in assessing structural progression compared to radiography-derived biomarkers. Several reliable and validated SQ scoring systems (e.g. BLOKS, WORMS and MOAKS) now exist. These have since been utilised across cross-sectional and longitudinal observational epidemiological studies. However, these approaches are time-consuming and require expert musculoskeletal radiologists to improve inter-operator reliability. Quantitative assessment of the knee joint from MRI data has increased our understanding of the natural history of the complex OA disease process. Quantitative measures are more sensitive to change than SQ measures [377]. New developments in OA have included the careful measurement of articular cartilage, bone, and menisci. Quantitative MRI measures are described here.

2.10.3.1 Cartilage

For cartilage, thickness and volume measures derived from MRI have emerged in recent years as important quantitative measurements of a joint's OA status. The increase in number of studies has been aided by careful validation studies showing that thickness and volume of cartilage can be estimated accurately by careful quantitative measurement of images from high resolution MRIs [345, 378, 379].These two measures have excellent reliability [315] while the responsiveness to change has been demonstrated in a number of studies [380-382] showing that MRI cartilage loss was more sensitive to change that radiographic JSW [314].Head to head comparisons between MRI-derived cartilage measures and radiographic JSW also showed superior responsiveness for MRI [382, 383].

Work on validity of cartilage measures continues to grow. Cross-sectional and longitudinal studies have demonstrated that cartilage loss was associated with both prevalent and incident knee pain [202, 384]. Higher rates of cartilage loss were shown to predict knee replacements [385-387] although the relationships between MRI and lesions and knee replacements are generally closer to the time of surgery reflecting concurrent pathology. The rates of cartilage thickness and volume loss was associated with other typical OA risk factors such as obesity [382] ,malalignment [151] and meniscal pathology [388].

Other advances in MRI cartilage quantification include cartilage compositional measures such as dGEMRIC which permits the visualisation and characterisation of the ultra-structure and biochemical structure of cartilage [322]. dGEMRIC has demonstrated construct validity [323, 389, 390] and also demonstrated good reproducibility and reliability [391]. Other dGEMRIC studies have demonstrated an association between lower baseline dGEMRIC values and incident JSN in patients with partial meniscectomy followed up for 11 years [392] while longitudinally a decrease in

dGEMRIC predicted an increase in cartilage thickness, this inverse relationship representing swelling of cartilage in the early stages of degeneration [285].

2.10.3.2 Bone

With advances in technology, precise measurements of established tissue change such as the loss of articular cartilage, and in addition, demonstration that bone itself is a highly responsive tissue, usually ignored in clinical trials is beginning to shape the development of new imaging modalities. Currently the best studied MRI bone lesions are BMLs (as shown in Section 2.9.2), which contribute to both development and progression of radiographic disease, and to pain. Clinical trials have provided proof-of-concept support for BMLs as a relevant treatment target and imaging endpoint for knee OA [145, 220]. Such trials provide a rationale for pursuing treatment interventions that target bone pathology, and for developing imaging biomarkers that take advantage of the ability of bone to change more rapidly than cartilage.

One of the novel bone MRI measures, 3D knee bone shape has great promise as an imaging biomarker and has shown reliability, improved responsiveness and demonstrated predictive validity (see Section 2.13.3). Through these advances in bone MRI assessment, emerging evidence also suggest that knees generally progresses linearly over time and the determinants of the rate of change is how one moves along this 3D bone shape [338].

Bone shape has also provided new insights into OA pathogenies revealing a “pie-crust” osteophytic ridge of bone around the articulating surface of the femur of diseased knees, accompanied by widening and flattening of the condyles (Figure 7). This deposition of bone could affect other joint structures and their measurements including apparent JSN, alignment, and meniscal and ligamentous insertion sites. Also, these alterations in joint geometry, combined with the changes in the material properties of bone in OA, can lead to changes in joint congruity, affecting the distribution of biomechanical loads, all of which may contribute to OA pathology.

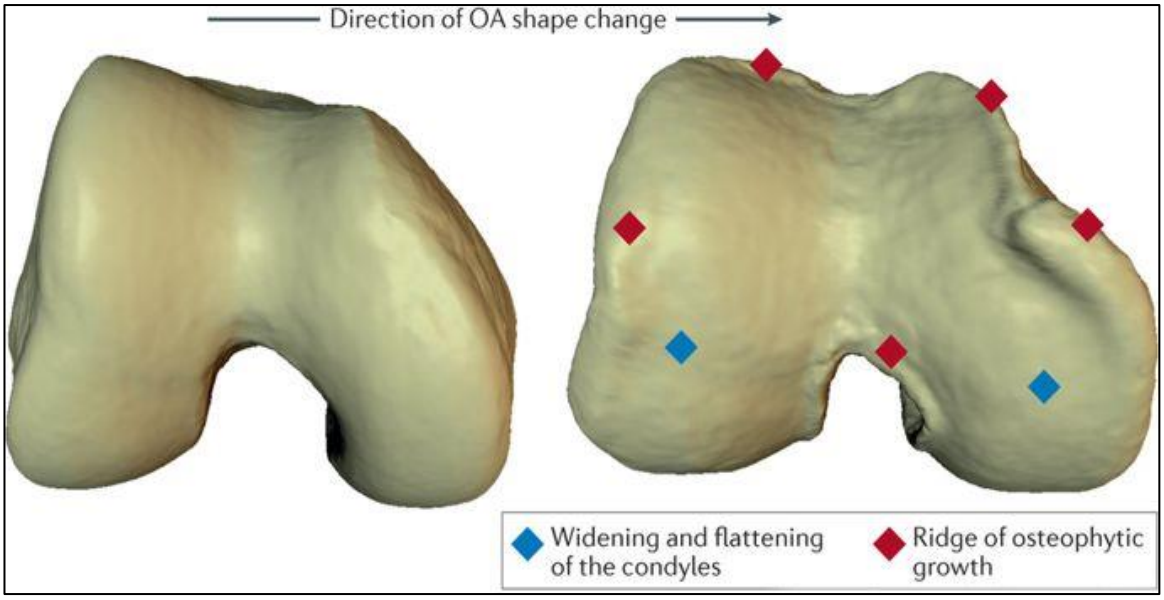


Figure 7.Three-dimensional femur bone shape changes in osteoarthritis
Reproduced with permission from [393].

2.10.3.3 Meniscus

Improvements in the field of imaging now allow for both meniscus and effusions to be segmented using similar technology to cartilage and bone segmentation. Compared to the improvements seen for bone and cartilage, meniscus measures are still in their infancy, but recent advances on meniscus are introduced in section 2.13.3 and discussed in more detail in Chapter 4.

In conclusion, all the imaging advances outlined here have the potential to deliver OA trials of the future with smaller patient cohorts and shorter time scales, providing new insights into disease progression and making trials viable that might otherwise have been uneconomic. However we have not yet harnessed the full potential of MRI biomarkers and utilised all of three dimensional tomographic information that MRI has to offer.

2.10.4 Machine learning

Supervised machine learning removes the need for prior human input into data interpretation, enabling reproducible evaluation of very large datasets. Statistical shape modelling (SSM), an example of supervised machine learning, enables accurate and fully automated derivation of 3-dimensional bone, cartilage, synovium and meniscal shape from MR images [10, 11]. SSMs employ principal components analysis to efficiently represent complex geometric shapes and, in so doing, facilitates the

quantification of different structures enabling more accurate analysis of structural progression.

Previously, defining the shape of similar tissues has been achieved by the process of manual segmentation which is a labour-intensive and time-consuming process requiring that the edge of the structure of interest be drawn around to define its shape. With machine learning, computers can be trained to recognise shapes (and their patterns) consistently and thus build a training set of such images. If the geometric properties of the target tissue are manually segmented in a substantial population of training set images, the segmented regions of interest can be analysed for their geometric shape properties using statistical analysis to create a SSM. This 'trained' SSM is a shape-recognition model that has been "taught" to seek certain shape patterns in a subsequent target image. This process involves the analysis of geometric shape, identifying landmarks on the edge of the target tissue that are consistent at a population level. Identification of these consistent landmarks and their automated application to subsequent target images can automate the segmentation and substantially reduce the labour-intensive approach of manual segmentation. Details on the technical aspects of SSMs and their development for knee OA are covered in Chapter 3 (Methods). The summary below reviews the specific tissues that have quantified using SSMs for knee OA and further developments in the field.

2.10.4.1 Cartilage

In MRI studies of knee OA, cartilage thickness and volume are commonly investigated as morphological parameters. There have been cross-sectional studies showing that cartilage of normal knees (controls) is thicker than cartilage of patients with knee OA, while longitudinal studies have shown cartilage thinning with the progression of the pathology [394]. The challenge however, is that the changes reported are minute and often difficult to detect, especially across the large regions that are commonly analysed and could be masked by measurement errors. New techniques have now been developed that enable cartilage partition into smaller and more specific regions that are anatomically corresponded. Consistent measurements are then taken using an anatomically segmented bone surface to produce detailed maps of cartilage thickness for the entire femur, tibia and patella [395]. Using careful manual segmentation of cartilage it has been shown that the central medial femur area shows the most rapid cartilage loss, a finding that has in the past been overlooked using other global measures of cartilage as these are averaged out by cartilage gains in other regions [396]. Recent advances have also resulted in automated methods being developed

because manual careful segmentation is labour intensive and would not be feasible to perform in the segmentation of large datasets with long follow up.

2.10.4.2 Bone

Change in bone shape such as flattening of subchondral bone (attrition) as seen on radiography has long been known to be associated with OA symptom and structural progression [397]. OA is thought to be a largely mechanically driven process [111] and Wolff's law dictates that bone can readily change its shape in response to stresses acting upon it [398, 399]. Such alterations may be feasibly assessed in a practical period, making it a useful target for therapy or clinical trial endpoint. It is also possible that subtle differences in bone shape or geometry could lead to abnormal joint loading and thus predispose to OA.

Using 2D bone shape derived from SSMs, Haverkamp showed that 2D knee bone shape was associated with structural OA severity as measured by presence or absence of cartilage defects [113]. In the hip, Agricola and colleagues used SSMs to describe the 2D shape of hip subchondral bone and showed that the shape of the hip could predict total hip replacement, suggesting its use as a radiographic biomarker for progression. However, no shape variations were related to clinical symptoms [400]. Lane and colleagues examined the association of incident hip OA with variations in 2D proximal morphology, assessed by SSM and found that in elderly Caucasian women variations in the relative sizes of femoral heads and neck were determinants of radiographic hip OA [401].More recently Wise and colleagues have also shown using SSMs that the shapes of the distal femur and proximal tibia differed by gender. They suggested that gender differences in risk of OA could therefore be due to bone shape differences [402]. They further tested mediation effects and found that bone shape was actually a mediator of the relationship between gender and incident knee OA [403].Longitudinal work using 2D bone from SSMs also suggests that trajectory groups based on the tibia and femur shapes exist, with these being associated with gender and OA [404].

3D SSM technology has now been extended and successfully applied to subchondral bone pathology which includes bone area and shape, bone marrow lesions, osteophytes, cysts and attrition. Imorphics Ltd, Manchester, UK are a commercial company that have developed an automated segmentation method for defining the three dimensional shape of knee bones using the next generation of SSMs called active appearance models. Imorphics delivers advanced image analysis technology for the analysis and interpretation of 3D medical images to support medical device and pharmaceutical companies. Through these advances is it now possible to quantify the

undulating 3D surfaces of the tibia, femur and patella from MRI. The use of SSMs in this way allows for the full parameterisation of the shape of each subject knee in terms of the population mean and shape variation learnt during the model training phase. This parameterisation has further been adapted and used to construct a bone shape vector of OA versus non-OA shape that has proven useful to identify knees at risk of developing OA [111].

3D bone shape measures have been shown to predict total knee replacement, radiographic and pain progression, and onset of radiographic OA [111, 209, 340], with the limitation that these were nested case-control studies. These measures have also shown promise as clinical trial measures by demonstrating superior responsiveness indices compared to JSW [339] and also to other well studied bone biomarkers such as BMLs [405]. Most studies have assessed BMLs using SQ methods that use ordinal scores that are not informative of spatial information in BMLs and also cartilage. Recently Bowes and colleagues using SSMs have shown a strong spatial relationship between BMLs and cartilage loss [406].

2.10.4.3 Meniscus

Quantitative measurement of damage to the meniscus could serve as a useful biomarker of OA progression [407]. While the meniscus appears as a simple shape, dysfunction of this structure may take many forms and appear as damage/alteration to meniscal volume, extrusion of the meniscus, or a general failure of meniscal competence, resulting in the spreading and maceration of the surface. More detailed analysis of the meniscus is covered in Chapter 4.Briefly, SSMs have been employed to study a number of potential measures of meniscal deterioration longitudinally with mixed results [407, 408]. These models have also been useful in assessment of where spatial changes occur in the meniscus and in doing so providing construct validity for the application of SSMs to study meniscal pathology.

In conclusion, the use of MRI and SSMs have become increasingly widespread which bodes well for future of OA trials. Compared to SQ scoring, the use of automated quantitative measurements will increase the reproducibility of the measurements made, thereby increasing their sensitivity to change. As these automated measurements are usually rapid to compute in comparison to manual methods, large datasets can be run simultaneously.

2.11 Limitations of previous analytic techniques

Correlation and regression techniques are some of the most commonly used statistical methods in medical epidemiology. Correlation analyses test the linear relationship or dependency between two variables (usually continuous) while linear regression examines the association between one outcome variable and one or more explanatory variables or covariates. The last few years have seen advances in computer systems and statistical software packages which have also resulted in complex multivariable regression methods being frequently used in medical research and extending our knowledge beyond what correlations and simple linear regressions could reveal.

Despite these positive changes, some researchers may not have the technical expertise to follow instruction manuals that accompany statistical packages and some may not have the necessary statistical expertise to choose appropriate statistical methods and also interpret results correctly. This has led to misuse of statistical methods in clinical research and it is clear that statistical analysis and reporting in medical journals need to be improved [409]. The OA field has not been spared and while a formal review of the methods used in OA research is not available, the summaries below are a narrative review of current pitfalls and issues related to study design, recruitment (inclusion criteria), outcome measures, and analysis and reporting. However, some aspects are not unique to or limited to OA research.

2.11.1 Analyses of single tissues

While a “whole organ” imaging approach is advocated in OA, a big challenge still remains on how to appropriately assess these interrelationships statistically as most of the available measures are designed to reflect single tissues. The commonly used scoring systems like MOAKS, WORMS and BLOKS discussed previously each provide scores for a single pathology for example a cartilage score or meniscus scores separately, and even then this could be a regional score (medial or lateral side for instance). These scoring systems are also SQ in nature and in many cases ordinal, as are the many PROs used to evaluate pain progression for example, but are then incorrectly treated as interval scales in numerous analyses with some studies even reporting means from these scores. Other attempts to use the ordinal data has included discretising the data into (yes or no) for presence or absence of abnormalities or collapsing categories and then applying logistic regression techniques to these “new” binary outcomes, but it is well known that discretising data into categories reduces the amount of information available from the data and this results in loss of statistical power and thus less robust analyses [410]. Also, having different cut points

for different MRI features might yield somewhat different results across studies and makes generalisability more difficult. In the 4 studies one each for bone [146], cartilage(JSN) [8], synovium [213] and meniscus [151] all these used an ordinal score inefficiently by treating these as interval or dichotomised the data to a yes or no response.

A relevant, but separate issue is that measuring interrelationships is further compounded by the fact that trying to add up these scores to produce a composite whole organ score is rendered meaningless as these scoring systems (based on ordinal and not interval scales) do not have any known normative values to represent the whole joint, thus deciding what a meaningful severity score that represents overall whole joint pathology is difficult.

Another methodological weakness in OA research is that many studies that have tried to assess structure and pain relationships for example have thus far investigated associations between one tissue and a chosen outcome, for example meniscal lesions with pain, or cartilage with pain independently yet pathologies affecting these structures are known to occur concurrently. Therefore, attempting to extrapolate findings or effect sizes from one tissue to another, or combining effect sizes obtained from these individual tissue studies in an attempt to understand the overall or synergistic effect is a difficult exercise, especially because depending on how they interact in vivo, their effects may not necessarily be additive on the chosen outcome.

2.11.2 Analyses of changes in multiple tissues

In the studies that have incorporated multiple tissues in their modelling, this has mostly been done without careful consideration of how the ordinal scoring systems works and what the impact of having two or more ordinal predictors in the same model means. Therefore from a linear modelling perspective in terms of modelling assumptions, these analyses (using ordinal predictors) are most likely to have violated these and the inferences from such models need to be made with caution. As linear and logistic regression models represent the effect of the covariates in an additive and multiplicative scale respectively, it is challenging to see how the effect of one level of an ordinal variable influences another level of an ordinal predictor as these do not have a pre-specified quantitative hierarchy. The pitfalls of treating ordinal measures as interval measures have been described before [411, 412] including violation of the assumptions for parametric statistics and furthermore it has been acknowledged that the extremes of the ordinal responses tend to be less used than the central choices.

The effect of this is that intervals near the extremes may be further apart than those near the centre.

Secondly, the issues around bias and confounding in these studies lacks important detail because while multiple tissues may have been considered (e.g. adjusting for covariates or including other tissues in the model) in most cases the causal relationships between different tissues are not been well-defined. A few examples of this includes studies for example by Sharma and colleagues who did well to try and model all sources of structural pathology using an outcome of pain however how these pathologies interact was not considered [204] while Roemer and colleagues [214] in a case-control study similarly did well to include a number of pathologies to represent the multi tissue nature of OA and performed conditional logistic regression in a model including Hoffa synovitis, BMLs, cartilage and meniscal damage. The study found that the concomitant presence of multiple pathologies increased the risk of OA. While this is an important finding, the causal pathways which the study aimed to investigate are still not fully understood. It is possible that some lesions may represent the initiator or first exposure of cause while others may be promoters also on the causal pathway to increase expression of the outcome (pain for example) thus confounding any associations that may be present or absent. In some cases this may lead to biased estimates or spurious associations appearing just by chance because of this inappropriate covariate adjustment, for example in the case of adjustment for mediators [413].

A review on the appropriate use of statistical analyses in OA highlighted that study design and statistical analysis are interconnected. A clear description of the study design is the basis for the correct choice and interpretation of the statistical analysis. However, logical relations between research question(s), study design, and statistical analysis are currently not clear in most OA manuscripts at present including those that have assessed multiple tissues [414]. What most studies have termed longitudinal analyses may have involved recording an outcome at say “X” number of years and then investigating the baseline effects of a covariate on this endpoint. This is often then referred to as longitudinal associations in the results and conclusions, but demonstrating cause and effect in these settings is difficult and these studies do not use the data efficiently for example where covariates change over time and by so doing have different effects over time, this effect cannot be captured. Inaccuracies in the analysis of longitudinal research are common, usually when repeated hypothesis testing is applied to the data, as it would for cross-sectional studies. This leads to an underutilisation of available data, underestimation of variability, and increased

likelihood of type II errors (false negatives) [415]. Commonly used methods to analyse longitudinal data include change score analysis (difference between outcomes measured at two time points), repeated measure ANOVA (uses two main factors and an interaction term to assess group differences over time) and multivariate ANOVA (repeated responses over time treated as multivariate observations) [416].

Another problem with current measurement involves the use of radiography-derived measures that are thought to be unreliable and unresponsive compared to novel 3D measures of structure. Due to the problems relating to 2D projection in radiographs and the object`s pose (the combined relative position and rotation of the bones), and the SQ nature of currently used MRI measures, what is perceived to be a lack of association or lack of responsiveness in many studies could just be small changes masked by large measurement errors despite how well designed or robust the analyses are. The responsiveness indices (SRMs) reported for radiography generally range between 0.3-0.4 and while MRI measures have resulted in improvements in both reliability and responsiveness [315, 417] these still fall short of novel 3D measures [339]. From a feasibility/practical perspective, these assessment methods are also operator-dependent and time-consuming, making it difficult to analyse features for large datasets.

2.11.3 Could new analytical methods be appropriate?

Improved outcome measures that are accurate, reliable and responsive are essential for OA research and these issues have been highlighted previously in Section 2.8.6.2. In an attempt to improve methodology, OMERACT working groups have been established. A systematic review by the OMERACT Rasch group demonstrated how the lack of interval scaling as commonly is the case for ordinal measures compromises the validity of clinical effects reported in numerous rheumatology studies [418]. They recommended that it was crucial to promote the use of the Rasch interval scale in measurement. Currently an international collaboration has been established to provide data for the establishment of Rasch transformed scales for commonly used PROMs in rheumatology.

Appropriate study design is essential not only for OA research but also medical research in general. Using previously suggested versions for the “hierarchy of evidence” [419] (Figure 8), it is important for careful consideration as to what the typical research question should be before designing a study. Appraisal of these study designs has been published before [420] and in OA research there have been far fewer RCTs performed with the majority of studies being case-control or cohort studies. This

has resulted in many studies that show an association but all have the limitation of being unable to prove causation and normally conclude by suggesting that further studies are needed.

Newer methods to address confounding and bias involve the use of directed acyclic graphs (DAGs) [236]. DAGs are useful in establishing which covariates might operate as potential confounders, mediators, or competing exposures in the multivariable regression analyses, achieved through construction of a causal paths drawn from established and hypothesized functional relationships between the exposure and outcome, and each covariate. Such models are invaluable for the specification and verification of the statistical analyses and results in appropriate adjustment and the most parsimonious model being chosen without the risk of over adjustment and thus reduction of statistical power which would otherwise occur [421].

Other related methods to address mediation or test mediation effects include formal statistical testing of mediation [413] for example using marginal regression models [422, 423]. These models are designed to control for the effect of confounding variables that change over time in longitudinal models and could be useful in disentangling the confounding that exists especially when trying to assess multiple tissues in OA.

The study design and analysis should be described with enough detail to enable replication in future studies by including clear research questions and how the study design and statistical analyses are a suitable choice for the study design, as this will vary greatly for example matched studies require different designs from unmatched ones while longitudinal studies require special types of analyses compared to cross-sectional ones.

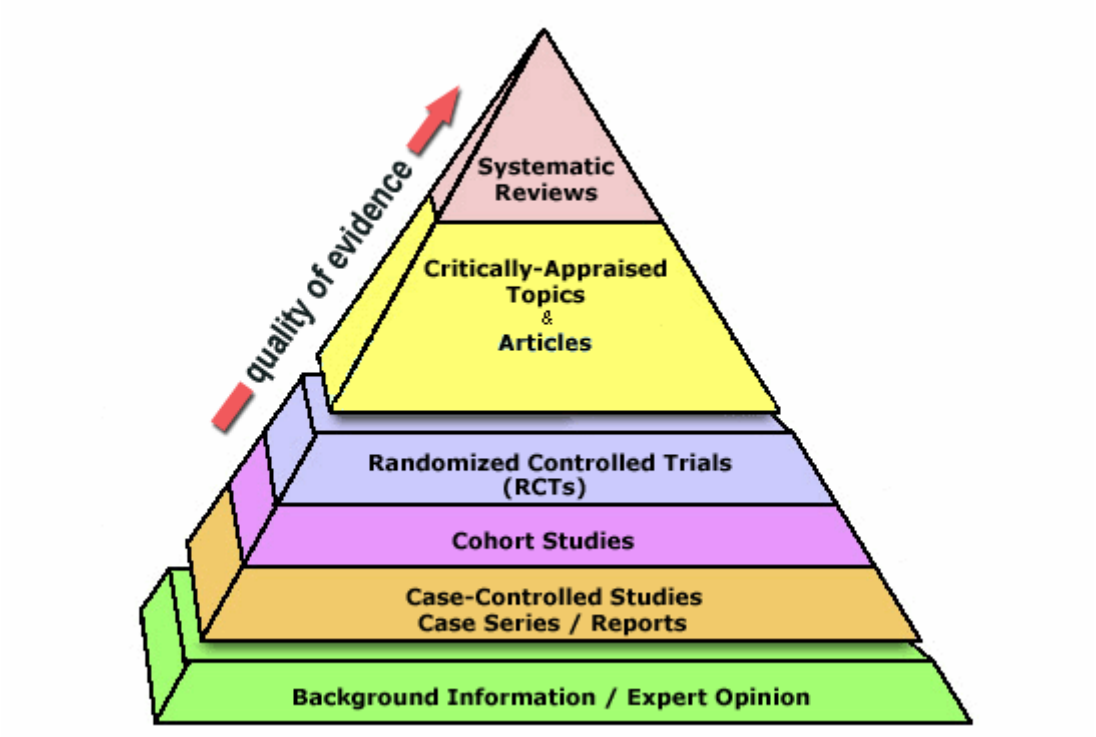


Figure 8.Hierachy of evidence

Understanding the overall progression of OA is critical to the timing of therapeutic interventions and design of effective clinical trials. Disease progression can be evaluated via longitudinal studies that measure outcomes repeatedly over time in relation to risk factors. Longitudinal data allows assessment of multiple disease aspects: changes of outcome(s) over time in relation to associated risk factors, timing of disease onset, and individual and group patterns over time. Assessing longitudinal temporal changes is important to studying specific time patterns of clinical impairments that could be missed otherwise [424]. Moreover, compared to cross-sectional studies, longitudinal studies often have less variability and increased statistical power. Longitudinal data analysis is however, complicated, by practical and theoretical issues including correlated data between adjacent timepoints, irregularly spaced data collection visits and missing data. Advanced statistical methods and improvement in software can handle these complexities, but knowing when to use these methods, checking model assumptions, and interpreting their output correctly still remain a challenge and not adequately dealing with these can lead to inappropriate and inaccurate analysis [416].

Among the different statistical methods reviewed for analysing longitudinal data, the mixed effects modelling is the most flexible and designed to handle multiples challenges of longitudinal data. As such, it is recommended by the FDA in analysis of

observational studies and clinical trials. Another useful method for longitudinal data is use of generalised estimating equations (GEEs) which are traditionally intended for hierarchical data (such as knees nested within patients in the OAI). Both GEEs and mixed effects models allow time-invariant and time-variant predictors and handle irregularly collected data and missing data without the need for imputation. GEEs are robust to misspecification of the repeated measures' correlation structure and are not computationally intensive to run. However, GEEs are not useful for analysing the correlation structure of the repeated responses and their utility is mainly for assessing the regression relationship between covariates and repeated measures. Secondly, GEEs assume missing data are missing completely at random (MCAR), which may not always hold true in OA studies. Another limitation of GEEs is that the usual model fit comparison indices like Akaike information criterion (AIC), Bayesian information criterion (BIC) and likelihood ratio tests cannot be used to compare different GEE models, unlike mixed effects models.

Mixed effects models assess both the regression relationship between covariates and repeated responses, and also the correlation structure of the repeated response. These models capture correlations of repeated measures using "random effects" that serve to describe cluster-specific trends over time, which is useful in understanding inter-individual variability in longitudinal responses and cluster-specific predictions. Mixed effects models multi-level hierarchical modelling that allow predictions for each data hierarchy level and an advantage over GEEs is that one can perform hypothesis testing on correlation parameters since these are directly estimated. Another advantage of mixed effects modelling is that they are more robust to missing data and assume missingness is missing at random, which is more general than the MCAR assumption of GEEs. However, mixed effects models also have limitations for example their computational complexity over GEEs particularly with nonlinear mixed effects models as these involve time-consuming numerical integration over the random effects. Another limitation is that the models rely on the correct specification of the mean and correlation structure of the repeated responses for valid hypothesis testing and conclusions or else results may be biased [425], unlike GEEs which are more robust to such departures.

Recently, other more advanced statistical techniques have been used to model longitudinal data. Such techniques like structural equation modelling (SEM) see in which sub types such as latent curve growth modelling (LCGM) and latent class growth analysis exist are discussed in more detail in Chapter 3 (Methods) and then applied in Chapters 6 and 7. The LCGM approach essentially follows the same premise as mixed

effects model, except that growth is formulated in a general SEM framework rather than as an extension of the regression framework in mixed effects models. LCGM models are confirmatory factor analysis (CFA) models with an imposed factor mean structure and particular constraints to yield estimates of growth. The basic idea of the LCGMs is that there is some overall mean trajectory for the entire sample, but each individual receives random-effect estimate(s) to capture how their particular growth curve differs from the overall trajectory. The main conceptual difference is that the random effects are specified as latent variables in a CFA, rather than as randomly varying regression coefficients; however, these two notions can be shown to be mathematically equivalent, although not identical depending on the data structure [426].Compared to ANOVA even mixed effects models, LCGMs offer several advantages including incorporating latent variables, simultaneously analysing parallel process growth models, ease of testing different trajectories of change and comparison of change across sub-samples.

Ultimately it is improvements in OA imaging i.e. use of 3D measures that provide quantifiable measurement of structure that could improve imaging studies. Using 3D quantitative analysis provides a solution to these recognized imaging shortfalls. As mentioned before, 3D image analysis uses the statistics of shape and image information, derived from a training set of images. This automated segmentation is capable of accurate identification of the shape and appearance of bone, cartilage and meniscus providing an accurate, rapid and highly reliable solution for analysing large imaging datasets [10, 339]. A major benefit is that the 3D imaging measures are not influenced by the pose of the object [11]. With superior reliability and responsiveness these measures can detect change where present and for a disease that has a long “incubation period” such measures may be useful in reducing the follow-up times or clinical trial lengths therefore their utility in OA studies will be invaluable in future.

2.12 Current management of OA

2.12.1 Current guidelines and treatment aims

Prior to commencement of therapy, the diagnosis of OA should be confirmed by a combination of history, physical examination and appropriate clinical symptoms as advised by NICE and EULAR guidelines [26, 427, 428], and knowledge of risk factors as suggested before. Radiographs and laboratory analysis are not necessary for the clinical diagnosis of OA but in the presence of atypical features that suggest the presence of diagnoses other than OA, such tests may be used for differential diagnosis of inflammatory arthritis, septic arthritis or malignant bone pain [427]. Management of

patients with OA is focused on symptom relief (including pain), limiting further physical dysfunction through maintenance of joint mobility and stability and improving health related quality of life (sleep quality, improved mood), education about OA, and attempts to slow the progression of structural damage in multiple tissues (cartilage, bone, ligaments and, muscles).

Current management strategies for knee and hip osteoarthritis are in broad agreement across the guidelines of the various stakeholder organisations that include professional societies, research bodies and regulatory (government agencies) [429, 430]. Recently, a large number of evidence-based guidelines formulated from important musculoskeletal organisations. These are derived from expert opinion, published literature, and patient opinion (all three sources are valid for comprehensive guidelines). There is broadly good agreement across these guidelines in which therapies they recommend as shown in the Table 3 as per the recommendations up to 2014 [431].

Table 3. Evidence based guidelines for OA treatments

Guideline	NICE 2014	OARSI 2014		EULAR 2013	ACR 2012		
	All sites	Knee	Multi	Knee & Hip	Hand	Knee	Hip
Education or self-management	+	+	+	+	(+)	(+)	(+)
Exercise and/or physiotherapy (water-based and land-based)	+	+	+	+	NE	+	+
Weight loss in obesity	+	+	+	+	NE	+	+
Aids, adaptations, braces and footwear (site-specific)	+	(+)	(+)	+	(+)	(+)	(+)
Transcutaneous electrical nerve stimulation	+	NR	–	NE	NE	(+)	NE
Acupuncture	–	NR	NR	NE	NE	(+)	NE
Thermotherapy (for example, hot packs or spas)	+	NR	(+)	NE	(+)	(+)	(+)
Topical NSAIDs	+	+	NR	NE	(+)	(+)	NR
Oral NSAIDs (lowest possible dose)	+	(+)	(+)	NE	(+)	(+)	(+)
Paracetamol	+	(+)	+	NE	NE	(+)	(+)
Cyclooxygenase 2 inhibitors	+	(+)	(+)	NE	(+)	(+)	NR
Topical capsaicin	+[#]	(+)	NR	NE	(+)	–	NE
Opioids (for refractory pain)	(+)	NR	NR	NE	–	(+)	NR
Serotonin–noradrenaline reuptake inhibitor	NE	(+)	+	NE	NR	(+)	NR
Nutraceuticals (for example, glucosamine and chondroitin sulfate)	–	NR	NR	NE	NE	–	–
Intra-articular corticosteroids	+	(+)	+	NE	–	(+)	(+)
Intra-articular hyaluronic acid	–	NR	–	NE	–	(+)	NR
Duloxetine	NE	NR	+	NE	NE	(+)	NR
Risedronate	NE	–	–	NE	NE	NE	NE
Strontium ranelate	–	NE	NE	NE	NE	NE	NE
Surgery (lavage or debridement)	–[∞]	NE	NE	NE	NE	NE	NE
Surgery (total joint replacement or arthroplasty) (site-specific)	(+)	+	NE	NE	NE	NE	NE

+ = treatment is unconditionally recommended; (+) = treatment is conditionally recommended; – = treatment is not recommended; NE = treatment is not evaluated; NR = no recommendation for treatment despite reviewing the evidence; # =excluding hip OA; ∞ =unless there is a clear history of mechanical knee locking.

Adapted from [431]

The OARSI recommendations for management of knee OA were produced using a Delphi process for four different patient groups, depending on the number of joints affected (knee only and multijoint) and comorbidity profiles [332].The ACR guidelines

were developed in a process that rates existing scientific evidence with an expert panel to develop evidence-based recommendations. These guidelines emphasize different pharmacological and non-pharmacological therapies that are joint specific and extend previously published EULAR recommendations [26, 428, 432, 433] .Both non-pharmacological and pharmacological interventions are used, separately but more commonly in combination, specifically for the patient`s preferences, OA type and comorbidities.

2.12.1.1 Latest OARSI guidelines for knee OA

The most recent OARSI guidelines [434] were developed to update and expand on the prior guidelines by developing patient-focused recommendations derived from expert consensus and based on objective review of high quality meta-analyses. The guidelines recommend that core treatments for knee OA should include education on arthritis and structured land-based exercise programs, dietary weight management in combination with exercise and mind-body exercise such as Tai Chi and Yoga, regardless of comorbidities.

2.12.1.1.1 Level 1A (≥75% in favour & >50% strong recommendation)

Topical NSAIDs were strongly recommended for individuals with knee OA that had no comorbidities Topical NSAIDs were also recommended for patients with GIT or cardiovascular comorbidities and for patients with frailty as these had minimal adverse events. In individuals having knee OA and concomitant widespread pain, no interventions were strongly recommended.

2.12.1.1.2 Level 1B (≥75% in favour & >50% conditional recommendation) & level 2 (60-74% in favour)

Aquatic exercise, gait aids, cognitive behavioural therapy with an exercise component, and self-management were the recommended non-pharmacologic options for individuals with no comorbidities, and for individuals with GIT or cardiovascular comorbidities or with widespread pain and/or depression.

Oral NSAIDs were conditionally recommended for individuals with no comorbidities and the panel recommended use of a proton-pump inhibitor (PPI). For individuals with GIT comorbidities, COX-2 inhibitors were Level 1B and NSAIDs with PPI, Level 2. For individuals with cardiovascular comorbidities or frailty, use of any oral NSAID was not recommended. However, a GCP statement specifies that when NSAIDs may be used at the lowest possible dose, for the shortest possible treatment duration in those with more favourable safety profiles

Intra-articular corticosteroids (IACS), IA hyaluronic acid, and aquatic exercise were Level 1B/Level 2 treatments dependent upon comorbidity status. GCP statement applying to intra-articular (IA) treatments for all comorbidity subgroups is that IACS may provide short-term pain relief, whereas Intra-articular hyaluronic acid (IAHA) may have beneficial effects on pain at and beyond 12 weeks of treatment with a more favourable safety profile than repeated IACS.

Conditionally recommended for patients with widespread pain and/or depression included any oral NSAIDs, duloxetine, IACS, hyaluronic acid and topical NSAIDs. The use of acetaminophen/paracetamol was conditionally not recommended (Level 4A and 4B), and the use of oral and transdermal opioids was strongly not recommended (Level 5). A treatment algorithm was constructed in order to guide clinical decision-making for a variety of patient profiles, using recommended treatments as input for each decision node.

2.12.1.2 Latest ACR guidelines for knee OA

A collaboration between ACR and the Arthritis Foundation recently updated the 2012 ACR recommendations for the management of hand, hip, and knee OA. Based on the available evidence strong recommendations were made for exercise, weight loss in overweight or obese patients, self-efficacy and self-management programs, tai chi, cane use, tibiofemoral bracing for tibiofemoral knee OA, topical NSAIDs, oral NSAIDs, and IACS for knee OA. Conditional recommendations were made for balance exercises, yoga, cognitive behavioural therapy, patellofemoral bracing for patellofemoral knee OA, acupuncture, thermal modalities, radiofrequency ablation, topical NSAIDs, topical capsaicin, acetaminophen, duloxetine, and tramadol [435].

2.12.2 Non pharmacological and surgery

There is consensus that patient education should be included as part of the core OA treatment to encourage self-management. With relatively low cost and no appreciable side effects this makes this a core treatment option. Patients with OA should be made aware, and this information reinforced at subsequent consultations, what the OA disease process entails and how OA reflects a failed repair process, what their personal risk factors are (e.g. obesity) and their prognosis. Each educational delivery should be tailor-made to the individual based on their illness perception and educational capability [436].

A physical exercise plan should be provided for patients. Systematic reviews have shown that exercise therapy decreased pain and improved function in patients with OA [437, 438] and demonstrated moderate effect sizes (ranging from 0.34 to 0.63 for pain

and 0.25 to 0.41 for function) [332]. The types of exercise need to be adapted to presence or absence of painful episodes to ensure they are tolerable and thus improve adherence as efficacy is usually better in complaint patients. The types of exercise should be varied and could include manual therapy such as stretching to improve range of motion, and also activities that build on muscle strength [439] starting with low-impact exercise and then gradually increasing the intensity according to the individual's capacity. Broad aims of physical therapy are usually to reduce pain, optimise range of motion and physical activity, improve muscle strength and reduce functional limitations.

It is recommended that all overweight or obese symptomatic OA patients be given weight loss advice [88, 440, 441]. Current evidence suggests that weight loss is associated with significant improvement in both pain and function, despite the small effect sizes reported (0.20 for pain) these were found to be significant [332]. Due to the fact that most OA patients that are overweight patients tend to have commodities such as cardiovascular and metabolic diseases, weight loss not only improves their OA-related pain related but is beneficial against these too. This group should be offered a dietician's review or dietary advice as well as individualised weight loss programs that include education and self-help.

Assistive devices and aids (for example, canes or crutches) are useful for everyday activities and should be held on the contralateral side to the affected knee to help reduce load. Other measures such as corrective footwear for individuals with OA (appropriately fitting shoes with no heel elevation, thick shock-absorbing soles and adequate plantar arch support) could be seen as useful adjunct therapies. The use of orthotics and braces may be considered mainly to improve symptoms and correct abnormal biochemical loading at the joint, however ACR only conditionally recommends their use while OARSI guidelines suggest their use as directed by an appropriate specialist as evidence for their use so is not compelling [442-444]. Other non-pharmacologic interventions recommended by ACR are acupuncture and transcutaneous nerve stimulation (TENS) [445, 446]. OARSI classified the benefit of acupuncture as uncertain and for TENS their recommendations were that is was inappropriate for multijoint OA and uncertain for exclusive knee OA [332].

2.12.3 Pharmacological therapy

Table 4 classifies the various pharmaceutical treatments based on risk scores, benefit scores and composite risk and benefit scores for OA treatment.

Table 4. Risk scores, benefit scores and composite benefit risk scores for OA pharmaceutical treatments in the knee

	Risk scores		Benefit scores		Benefit and risk scores	
	No comorbidities	Comorbidities	No comorbidities	Comorbidities	No comorbidities	Comorbidities
	Mean (1-10)	Mean (1-10)	Mean (1-10)	Mean (1-10)	Mean (1-100)	Mean (1-100)
Paracetamol	3.4	4.5	4.5	4.4	34.0	28.3
Avocado soybean	1.6	1.8	3.5	3.5	33.2	32.6
Capsaicin	2.6	2.8	5.1	5.1	42.6	41.8
Intra articular corticosteroids	2.8	3.6	6.5	6.4	53.8	47.1
Chondroitin symptom relief	1.1	1.3	3.8	3.9	37.8	38.0
Chondroitin disease modification	1.1	1.3	2.7	2.7	27.0	26.5
Diacerein	3.8	4.0	3.7	3.7	26.6	25.7
Duloxetine	4.0	4.7	5.3	5.4	37.2	34.0
Glucosamine symptom relief	1.4	1.7	3.9	3.9	37.4	36.3
Glucosamine disease modification	1.4	1.7	2.7	2.7	26.3	25.3
Hyaluronic acid	3.1	3.8	4.1	4.2	32.4	30.5
NSAIDs (topical)	2.7	3.5	6.0	5.9	49.8	44.7
Opioids (transdermal)	4.8	6.1	5.2	4.9	31.7	24.2
Opioids(oral)	5.5	6.5	5.6	5.4	30.7	24.0
Risedronate	3.2	3.3	2.7	2.7	20.9	20.4
Rosehip	1.8	1.9	3.3	3.4	30.3	30.7

Adapted from [332].

2.12.3.1 Topical agents

Topical agents should be considered a first line therapy in patients with symptomatic knee OA. The most widely used topical agents contain capsaicin, lidocaine and NSAIDs. Topical capsaicin, a chilli pepper extract that depletes neurotransmitters in sensory terminals thus attenuating the central pain transmission from the joint has demonstrated efficacy in double-blinded trials. It is generally recommended as supplementary analgesic for hand and knee OA. Topical NSAIDs have also demonstrated efficacy in knee RCTs, some showing the same efficacy as oral NSADs and accompanied by lower incidence of GIT side effects favours their use [447-449].

2.12.3.2 Oral analgesia

Previous evidence from high quality RCTs suggested oral paracetamol to be effective for pain relief in OA, with effect sizes of 0.21 reported [450]. However, new evidence suggests that the effect sizes are much smaller (0.10) if only high quality trials are considered [449] and that it may have greater toxicity than was previously reported [451]. Despite this, current guidelines suggest that paracetamol should be the first line treatment.

The next line of analgesia to consider if more is required include oral NSAIDS, selective COX-2 inhibitors and then opiates, but with consideration of the greater risk of toxicity particularly with increasing age and co-morbidities. Head to head comparisons show that NSAIDs are more effective that paracetamol for pain relief [452], however GIT side effects are more common with NSAIDs use alone than with paracetamol alone, with higher risk of serious GIT effects if both are used in combination [453].

Nutraceuticals, including glucosamine sulphate and chondroitin sulphate products, are natural compounds consisting of glycosaminoglycan components. Their mode of action remains controversial and none are licenced by the FDA and are therefore marketed as health food supplements. Nutraceuticals are not recommended due to the lack of evidence of efficacy and uncertainty regarding clinically important analgesic or structural benefits, however small benefits in pain-relief has been reported in low quality trials [437, 454].

Hyaluronic acid (HA) is a high molecular-weight polysaccharide that is naturally occurring and a major component of normal synovial fluid and cartilage, preparations of which are injected to provide pain relief. Due to its visco-elasticity it may provide lubricating and shock absorbing properties. Conclusions from studies are that HA injection gives a small but significant symptomatic effect lasting up to 26 weeks

[455],however intra-articular HA is not recommended for OA by the NICE guidelines [427] but ACR guidelines conditionally recommend its use in individuals older than 74 years with knee OA pain that is refractory to conventional pharmacological therapies [445]. The intra-articular injection of corticosteroids is a useful short-term adjunct in the treatment of moderate to severe OA pain, which may facilitate muscle strengthening and exercise.

Other therapies such as diacerein (anthraquinone derivative) shown to inhibit interleukin 1 production and demonstrating efficacy with a small effect size of 0.24 for pain [456], are still considered as uncertain under OARSI recommendations. Duloxetine was shown to be effective for chronic knee pain and can be used alone or combined with NSAIDs in patients with persistent pain [457]. Other opioid analgesics although shown to be efficacious for pain relief are however rarely indicated because of their safety profile [458].

2.12.4 Surgical interventions

Surgical interventions for OA are classified into 3 groups depending on the main objective of treatment: joint debridement or lavage to improve current symptom state, osteotomy to limit the risk of structural progression and joint replacement in advanced disease to improve symptoms. Arthroscopic debridement and lavage are not recommended as treatment for OA [459], except for patients with signs of internal derangement such as meniscal tears or when there is a clear history of true mechanical locking. Osteotomy to correct malalignment could be beneficial as this is one of the important OA progression risk factors [460] however the safety information is lacking hence the variability in uptake of this treatment. Joint surgery is considered the “gold standard” treatment and should be considered if a patient suffers persistent symptoms despite adequate use of the non-pharmacological described and pharmacological interventions described earlier. Indication for surgery should be made based on the severity of the patient’s symptoms and functional limitation rather than just severe radiographic features [461].

2.12.5 Massive unmet need

The ultimate goals for OA treatment are disease modification, analgesic and anti-inflammatory efficacy. However, the current pharmacological therapies manage only pain and inflammation as there are no licensed structure-modifying therapies and even so, the effect sizes are relatively small [462, 463]. Furthermore, commonly used therapies like NSAIDs and COX-2 inhibitors have been associated with serious GIT, renal and cardiovascular adverse effects. While IA steroids may be offered to patients

to substantially relieve pain, their benefits are short-lived and there is widespread concern that frequent injections could possibly lead to cartilage damage. Ultimately therefore, many patients with knee or hip OA will undergo surgery to replace the diseased joint. Knee replacement reduces pain but joint prostheses have a rigid life expectancy and revision surgery offers less favourable outcomes. In the future when we better understand the structure-pain relationships there could be a potential to perform minor tissue-specific procedures instead of total knee replacements. It is therefore apparent that there are unmet needs in OA management, in terms of both the safety of treatment options and the ability of therapies to modify disease progression. Novel drug trials are therefore warranted and to make them attractive to the pharmaceutical industry it is important that the measurements of disease progression are well characterised thus ensuring optimum numbers of people are recruited and the lengths of follow up reduced. This thesis aims to characterise some potential imaging biomarkers as a step towards their potential utility in future OA trials.

2.13 Thesis Aims

The hypotheses underlying this thesis is that the use of novel machine-learning derived 3D quantitative multi-tissue imaging biomarkers of bone and meniscus, and advanced statistical techniques will provide valid measures for use in knee OA clinical trials.

As part of their validation the hypothesis was that 3D meniscal measures would demonstrate good reliability, responsiveness, feasibility, demonstrate construct validity, be associated with known OA risk factors and be useful additions to existing structural OA trials measures. The 3D bone shape measures would be compared to existing bone imaging biomarkers for construct validity, and responsiveness and it was hypothesised that these would provide superior responsiveness indices.

Using the largest knee OA cohort to date, the Osteoarthritis Initiative with at least 8 years of follow up date the longitudinal change in bone would be investigated with the hypothesis that 3D bone shape changed linearly over time and that the determinants of this change were the same for the three knee bones (femur, tibia and patella). Lastly the thesis aimed to investigate the hypothesis that change in 3D bone measures of structure was not homogenous over time, and that variation existed over time yielding distinct trajectories of change. These longitudinal analyses would employ advanced statistical techniques.

Chapter 3 Methods

This chapter details the methodologies that were used in the studies included in this thesis. To characterise different imaging biomarkers extensively, a mixed statistical approach using quantitative methods was adopted. All data used in this thesis was obtained from the Osteoarthritis Initiative (OAI) an ongoing multicentre observational study set up in the USA and designed to focus primarily on knee OA as is described in detail in Section 3.1. The MR images characterised in this thesis were quantified into 3D measures obtained via AAMs as described previously from annual MR measures from the OAI. Further details on AAMs and SSMs are detailed in Section 3.2 of this Chapter.

In this thesis, the term imaging biomarkers is often employed, as biomarkers are not only “wet”/laboratory measures. Further details on the definition and validation of biomarkers are outlined in section 3.3.The specific statistical analyses applied in the thesis are described in section 3.4 in the different sub headings relevant to that chapter. Methods relevant to Chapter 4 on responsiveness and reliability are described in section 3.4.1. This responsiveness study analysed 86 participants selected from the progression cohort of the OAI. Chapter 5 assessed the relationship between a more established bone biomarker (BMLs) and an emerging (potential biomarker) bone shape, in an OAI sub-cohort of 600 participants identified for development of potential biomarkers using a combination of descriptive statistics, responsiveness and multilevel modelling. An overview of advanced statistical techniques of structural equation modelling (including latent growth curve modelling to model longitudinal change, and growth mixture models to assess trajectory classes of change) relevant to Chapters 6 and 7 are described in section 3.4.2.These analyses utilised the full OAI dataset over 8-year follow-up. After establishing the number of trajectories, variables predictive of trajectory class membership were tested using multinomial logistic regression.

3.1 Introduction to the Osteoarthritis Initiative (OAI)

The OAI is a public-private partnership (PPP) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Four clinical centres were awarded contracts following NIH peer-review: University of Maryland (subcontract site: Johns Hopkins University), Memorial Hospital of Rhode Island / Brown University ,Ohio State University and the University of Pittsburgh and a data coordinating centre at the University of California

San Francisco. These academic centres carried out the research and provided access to public databases and bio-specimens. Private funding partners include pharmaceutical companies, Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. Private sector industry partners provided study planning, scientific input, financial support ,data and bio specimens. Public funder provided contract funding, financial support, bio-specimen repository and also housing the public database. Governance is through the OAI steering committee which provides scientific oversight and there also exists an Observational Study Monitoring Group. The NIH and private sector funding have contributed US $ 60 million towards this PPP, with US $ 22 million coming from the private sector (equal contributions from each company).

The OAI cohort was set up to focus primarily on knee OA. At inception its main aim was to develop a public domain research resource to facilitate the scientific evaluation of biomarkers for knee OA as potential surrogate endpoints for disease onset and progression. The OAI is to date the largest publicly-available database exploring OA progression. The success of the OAI as measured through various outputs (data, publications for example) have been through international collaborations (see Figure 9) involving clinical sites, the NIH sponsor, other public and private sponsors, data coordinating centre, consultants, contractors and sub-contractors and vendors across the world.

While the broad aims of the OAI were to develop a cohort suitable for studying the natural history of OA and the risk factors for onset and progression of knee OA, it also aimed to determine the validity of radiographic, magnetic resonance imaging, biochemical and genetic measurements as potential biomarkers and surrogate endpoints for knee OA (data collected described in more detail in Section 3.4). The study had an initial recruitment target of about 5000 individuals (4,000 who did not have definite knee OA on enrolment, but considered at high risk of developing new OA into the *Incidence* sub cohort, 800 who had symptomatic and radiographic knee OA into the *Progression* sub cohort and 200 in the *Reference* (Non exposed) control group). Based on evaluation of progress midway through recruitment, enrolment targets into *Progression* cohort was increased to 1200 in response to the relatively large number of interested individuals who were eligible for this cohort in combination with below goal recruitment to the *Incidence* cohort.

In total 4796 individuals (aged 45-79) were recruited and assessed annually. These comprised 1389 participants in the progression group,3285 in the incidence group, and 122 in the control group. Enrolment was achieved between February 2004 and May 2006.

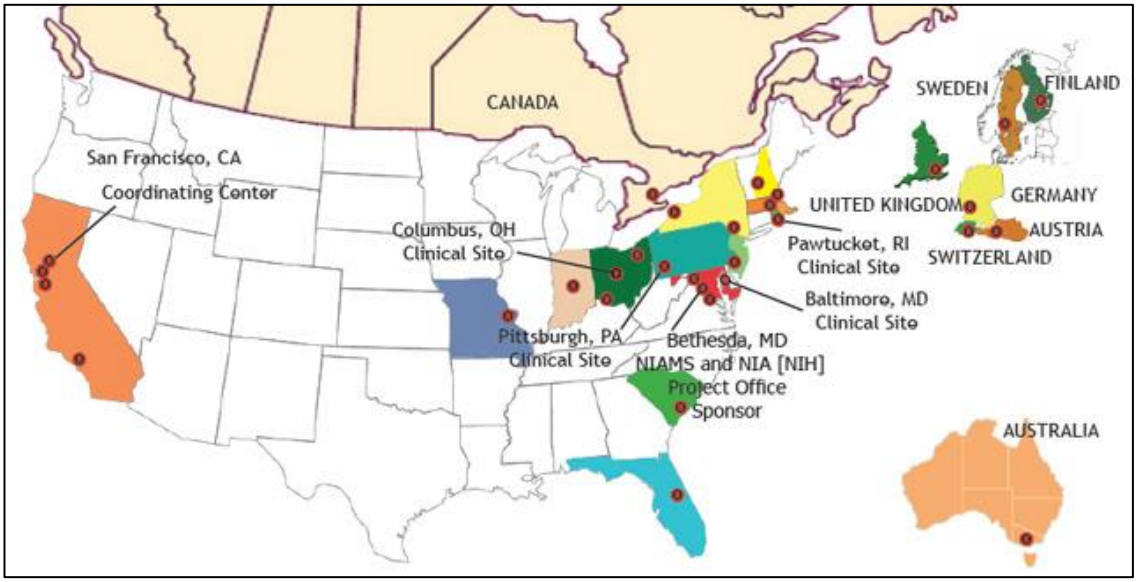


Figure 9. OAI worldwide collaboration

These maps pinpoint the locations of the clinical sites, the NIH sponsor, other public and private co-sponsors, the data coordinating centre, consultants, subcontractors, and vendors –world-renowned experts working together.

3.1.1 Inclusion criteria

The overall recruitment goal was to obtain approximately equal numbers of males and females, to be aged between 45-79 and at least 23% from ethnic minorities. Prevalent symptomatic OA definition for OAI encompasses both presence of frequent knee symptoms (FKS) and radiographic features and is similar to the ACR criteria for clinical knee OA [464]. As symptoms of knee OA are often intermittent and many years may elapse before they become monotonic/chronic there is no clearly defined point of onset. Degenerative changes precede and predict the incidence of radiographic knee OA [214] therefore there are limitations to the existing definition of the OA progression group.

To be in the *Progression sub cohort* participants were required to have in at least one knee both the following:

Frequent knee symptoms (defined as having had had knee pain in the last 12 months for at least one month) and radiographic knee OA defined as a definite tibio-femoral

osteophytes (OARSI grade 1-3 equivalent to K-L grade ≥ 2 on a fixed flexion radiograph.

Incident sub cohort classification was defined as having no symptomatic OA in either knee but at risk of developing symptomatic OA. Incidence was defined as the first occurrence in either knee of both FKS and radiographic OA in the same knee. For feasibility of recruitment and to enrich each stratum so that reasonable number of incidents would be recorded age-eligible persons would be classified as high risk depending on age band as follows:

Age 45-79: participants had to have FKS or frequent use of medication for treatment of knee symptoms (use of all types of medication on most days of a month in the past 12 months), or infrequent knee symptoms (pain, aching or stiffness in or around the knee at any time in the past 12 months but not on moist days for at least one month) and should have at least one eligibility risk factor.

Age 50-69: any of FKS or frequent use of medication as above or be overweight or have 2 or more eligibility risk factors.

Age 70-79: any of FKS, frequent use of medications or at least one risk factor

List of risk factors

- Knee symptoms which can be any of FKS or infrequent and frequent use of medication
- Overweight (greater than 93kg in males and 77kg in females aged 45-69 and greater than 97kg in males and 81kg in females between 70-79 years old)
- Knee injury defined as history of knee injuring causing walking difficulties for at least one week.
- Knee surgery defined as any history of knee surgery
- Family history of total knee replacement in a biological parent or sibling
- Heberden`s nodes
- Repetitive knee bending
- Age 70-79

The control sub cohort was defined as those having no pain, aching or stiffness in either knee in the past year, no radiographic OA (OARSI osteophyte grade = 0 and JSN grade = 0) and no eligibility risk factors.

3.1.2 Data collection

3.1.2.1 OAI data access

The database is publicly available and upon agreeing to the privacy terms and conditions of the Data Use Agreement for Limited Data Set(s) users may download datasets for use in line with the end user agreement. Access to bio-specimens and images is upon receipt of an application and approval by the OAI committee responsible. Data is publicly available at https://data-archive.nimh.nih.gov/oai/.

3.1.2.2 Recruitment and enrolment

The original design was for each recruiting centre to contribute 25% towards recruitment but actual recruitment numbers differed by about 22% (from the highest to the lowest). Informed consent was obtained from each participant in line with federal government guidelines and obtained prior to any screening or enrolment procedures. The recruitment process involved 4 stages:

- Initial contact designed to reach the target population through focused mailings, including to identified populations with OA, local newspaper adverts, presentations at community or civic groups, at church and a dedicated website about knee pain and knee OA
- Initial eligibility interview (IEI) by telephone to assess if individuals qualified for the study. During this process pre-screening was done such that if age and gender sub-cohorts were already full then participants would not undergo the IEI
- Those qualifying following telephone screening had a screening clinic visits scheduled where additional assessments were performed.
- Following success at the screening visit participants would then have an enrolment clinic visit at which the majority of baseline data was collected and MRI exams performed. Enrolment visits could span more than one day to complete all baseline imaging.

3.1.2.3 Primary outcome assessments

To understand the natural history of the disease the OAI collected at baseline and at each follow-up visit core knee OA status and knee OA outcome measures (clinical and imaging). Selection of primary measures was guided by recommendations of the OMERACT III task force on OA research on core measures for OA clinical trials. Of the core set of recommended outcomes, pain, physical function, patient global assessment and joint imaging comprise four domains. Table 5 summarises what data was collected at each visit during the recruitment process and Tables 6 and 7 summarise the frequency of data collection.

Table 5. Recruitment process and data collected at each point

Stage	Data collected
Initial eligibility interview	Contact information, demographics (age, gender, ethnicity) Frequent knee symptoms and frequent medications for knee symptoms status Additional screening risk factors (weight, history of knee injury and surgery, knobby fingers, frequent knee bending, TKR in parent or sibling) Assessment of exclusions (having a TKR or planning one, RA and inflammatory arthritis, MRI contraindications, serious comorbid conditions likely to interfere with participation, plans to relocate, clinical trial participation & non ambulatory status
Screening visit	History of arthritis diagnoses, family history of knee or hip TKR Knee symptoms in the past 12 months and past 30 days & knee pain severity in the past 30 days Activity limitations due to knee symptoms in the last 30 days Detailed history of knee injury and surgery, hip symptoms in the past 12 months Back, shoulder, elbow, wrist, hand, ankle and foot symptoms in last 30 days temporomandibular symptoms Menopausal symptoms and pregnancy Standing height, weight, body size and knee size for MRI eligibility Bony enlargement of DIP joints Standing PA fixed flexion radiograph of both knees
Enrolment visit	Marital status and household occupancy ,education, health care access and health insurance, income Comorbidity index, fracture history, smoking history, weight history, current alcohol consumption Medical Outcomes Study Short-Form 12 (SF-12), CES-D for depressive symptoms Block Brief 2000 Food Frequency Questionnaire WOMAC and KOOS for each knee, participant global assessment of knee symptoms impact Current knee bending activities, Physical activity Scale for the Elderly (PASE) Inventory of all prescription medications used in the past 30 days. Current use of all medications (prescription, over the counter, supplements & nutraceuticals for joint symptoms) Past use of bisphosphonates, knee injections (HA acid, steroids) and complementary/alternative medicine for joint pain MRI of both knees and thighs and radiographs of each knee for subjects in progression cohort (fluoroscopic guidance) PA radiographs of the right hand, standing bilateral radiograph of the pelvis Biological specimens (urine and fasting blood specimen),abdominal circumference, blood pressure and heart rate Knee examinations (anserine bursitis, patellar tenderness, crepitus, effusion/swelling, knee alignment, joint line tenderness Bilateral isometric quadriceps and hamstring strength Physical performance measures (20metre walk, 400 meter walk, rapid chair stands)

3.1.2.4 Clinical variables assessed - overview

Frequent knee symptoms defined as “pain, aching or stiffness in or around the knee on most days “for at least one month during the last 12 months” is the definition of symptomatic OA used in the OAI, along with radiographic findings of OA. Use of the WOMAC, KOOS and other questions are used to complement the investigation of symptoms and are administered regardless of frequent knee symptom status.

Global knee pain severity was assessed and recorded using an 11-point (0-10) scale which assessed the participants’ pain severity during the past 30 days and also past 7 days. A participant global assessment focusing on overall impact of knee problems on their sense of well-being during the past 30 days was collected and that is also an 11 point (0-10) scale.

Knee pain, stiffness and knee related physical function was assessed using the WOMAC, to characterise subjects` knee symptoms the OAI used the WOMAC pain with activity and stiffness scales and to evaluate knee-related disability the WOMAC disability scale. The 5-point Likert scale version of the WOMAC questions were used and modified to ask about the right and left knees separately during the past 7 days.

The non-WOMAC components of the KOOS were included in order to evaluate knee symptoms and function under different activity conditions than evaluated by the WOMAC. The 5-point Likert scale version was used and assessed right and left knees separately. The rationale of using the KOOS was to extend the target population of the WOMAC to younger and middle age participants with knee injuries and post-injury arthritis. In addition questions about activity limitation due to knee pain in the last 30 were adapted for use in OAI and have been validated as measure of disability and been found responsive to a variety of medical conditions and injury.

General health and function status was measured using the Medical Outcomes Study Short Form 12 (SF-12) and abbreviated version of the SF-36 which is a generic health-related quality of life instrument consisting of 12 questions and covering 8 health domains (physical functioning, social functioning, role-physical, role-emotional, mental health, energy/vitality, pain and general health perception.

Measures of walking ability and endurance included the 20-metre and 400-metre walks which are self-paced endurance tests and are modified to increase tolerability in elders and those with physical impairment. Leg strength was another clinical measure

assessed and this was measured using the Good Strength isometric chair which measures the maximal force during isometric contractions of the right and left quadriceps and hamstring muscles.

3.1.2.5 Image acquisition and assessment- overview

An important goal of the OAI was to support development of imaging markers that indicate the presence of OA, or increased risk of OA even when radiographic changes are minimal or absent thereby predicting subsequent disease course. Central assessments available are semi-quantitative (SQ) and qualitative of OA- related pathologies defined as either present or absent or measured on ordinal scales ranging from normal to abnormalities or severity e.g. K-L grade, JSN grade, MOAKS,WORMS,BLOKS. The OAI also collected quantitative measures on continuous scales for example widths and volumes and also 3-D measurement of shape of knee structures which are provided as dimensionless scores related to a mean shape.

Participants that had at least one follow up visit have SQ scores available from baseline to 48 months for K-L grade and JSN grades. Furthermore for those that had a KL≥ 2 at any time point other radiographic features such as osteophytes, sclerosis and cysts were also scored. Individuals that only had a baseline radiograph taken only have a quasi-KL grade score as the only SQ score and this is obtained at the screening visit. Quantitative longitudinal measures are available from baseline to 48 months and these include minimum medial compartment JSW, and fixed location JSW measurements at various positions in both medial and lateral compartments.

3.1.2.6 Biological specimens and other measures

Part of the OAI aims was to develop an archive of biological specimens available to investigators for testing and validating OA biomarkers. Blood and urine specimens were thus collected at baseline and at follow-up visits. Blood is available as serum and plasma in order to assess different biomarker assays. DNA is also available for the entire cohort. Other measures available in the OAI database at baseline include physical activity, medication use, comorbid conditions and food frequency questionnaires (Table 5).

3.1.2.7 Frequency of data collection

Data was collected when participants visited one of the four designated OAI clinical research centres (each equipped with a 3.0 Tesla MRI scanner, for imaging the knee and nearby radiology facilities). MRI, radiography, biochemical and genetic markers were collected at baseline and additional specimens at collected at each of the four

annual follow-up visits. Clinical data and joint status including the risk factors for progression and development of knee OA were obtained by questionnaire and examination at baseline and selectively updated at the yearly follow-up visits. The frequency of data collection (restricted to measures relevant to this thesis) are shown in Table 6 and 7 while questionnaires and operation manuals for examinations are can be accessed online at: https://nda.nih.gov/oai/ upon agreeing to the privacy terms and conditions of the Data Use Agreement for Limited Data Set(s).

3.1.3 Limitations of the OAI database

The OAI database is a cohort of individuals from North America who volunteered to participate in the research project. This cohort may be considered a self-selecting sample of healthier and more educated or affluent individuals compared to the full spectrum of clinical practice, and may therefore limit generalisability to the wider population from different economic, social and cultural backgrounds worldwide. For example, physical activity and weight status may have different and measurable effects on OA in those in the general population not at high risk of OA, and considering BMI was part of the inclusion criteria into OAI, this may introduce bias and confounding to studies using the OAI database. Bias and confounding need to be addressed in any analyses utilising the OAI database, and care taken in terms of generalising any findings from the OAI. More on this is discussed in Chapter 8 (Discussion Chapter).

Table 6. Examination measures relevant to thesis and their scheduled frequency

Measurement	Screening visit	Enrolment visit	Follow-up Visit				
			12 - months	Interim 6- months	24- months	36 - months	48 - months
Height, standing	X				X	X	X
Weight	X		X		X	X	X
Knee examination							
Alignment (by goniometer)		X	X		X	X	
Knee pain location					X		
Performance Measures							
20-metre timed walk		X	X		X	X	X
400-metre timed walk		X			X		X
Chair stands timed		X	X		X	X	X
Isometric quadriceps and hamstring strength		X	X[1]		X		X
MRI							
Right and left knee		X	X	X[2]	X	X	X
Right and left thigh		X	X[1]		X		X
X-ray-knee							
Bilateral PA fixed flexion view	X		X		X	X	X
Unilateral fluoroscopic guided view (one or both knees)		X[3]	X[3]		X[3]		
Unilateral lateral view (both knees)		X[4]	X[1,4]			X[4]	
Bilateral full limb for mechanical alignment			X[5]		X[1]		

[1] Obtained in those participants eligible for this measurement at the previous visit but for whom a valid measurement was not obtained.
[2] Obtained in the knee that had the extended set of sequences at baseline, usually the right knee.
[3] Obtained in a subset of Progression sub cohort participants at 2 clinical centres.
[4] Obtained in Reference (Non-exposed) controls.
[5] Obtained in Progression sub cohort participants.

Table 7.Questionnaire measures relevant to thesis and their scheduled frequency

Questionnaire / Interview Measures	Initial eligibility	Screening visit	Enrolment visit	Follow-up Visit				
				12 - months	Interim 6- months	24- months	36 - months	48 - months
Demographics (age, gender, ethnicity, education, marital status, residency, income)	X		X				X	
Employment, current and past			X	X		X	X	X
Health care and health insurance			X	X		X	X	X
Knee Symptoms, function & QOL								
Frequency of knee symptoms & medication use for knee symptoms, past 12 months, 30 days	X	X		X	X	X	X	X
Knee pain 0-10 rating scale, past 7, 30 days		X	X	X	X	X	X	X
WOMAC, past 7 days			X	X	X	X	X	X
KOOS, past 7 days			X	X	X	X	X	X
Participant global assessment of knee impact			X	X	X	X	X	X
Limitation of activity due to knee Sx, past 30 days		X		X		X	X	X
Work disability due to health problems			X	X		X	X	X
History of inflammatory arthritis/other arthritis	X	X		X		X	X	X
SF12			X	X		X	X	X
CES-D (depressive symptoms)			X	X		X	X	X
Comorbidity Index			X			X		X
Health behaviours and OA risk factors								
History of knee surgery (incl TKR)	X	X		X		X	X	X
Family history of total knee and hip replacement	X	X				X		X
Fracture history			X	X		X	X	X
Tobacco and alcohol use			X					X
Physical activity (PASE), past 7 days			X	X		X	X	X
Dietary nutrient intake (Block Brief 2000), 12 months			X					
Medication								
Prescription medication inventory, past 30 days			X	X	X	X	X	X
Current medications/supplements for joint symptoms			X	X	X	X	X	X
Knee injections for arthritis			X	X	X	X	X	X
Past use of selected medications			X	X	X	X	X	X
CAM treatments for joint Sx, past 12 months			X			X		X

3.2 Quantification of MRI measures in 3D

The novel quantification technology used in this thesis was provided by Imorphics Ltd (Manchester, UK). While clinical images capture anatomical data in 3D, a human reader is constrained to recognise and appreciate this data as in a 2D projection as a series of slices. The understanding of 3D shape is not easily recognised in SQ scoring, and only relatively simple 3D geometric measures may be made by the human observer. Describing the shape of naturally occurring organs like joints that adapt to their environment is much more challenging but represents an important source of in vivo information. This is because the shapes are often complex, asymmetrical structures and there is significant variation between individuals at a population level.

Statistical shape modelling (SSM), a form of active appearance modelling (AAM) is part of a broad group of techniques that fall under the branch of supervised machine learning. SSMs employ principal component analysis to reduce complex 3D geometric shapes to a single metric value. Using machine learning methods, the appearance of a tissue can be learnt, and that learning applied to automatically find and delineate the same tissue in new, unseen images. With respect to imaging developments in OA, if the geometric properties of the target tissue are manually segmented in a substantial number of training set images, the segmented regions of interest can be analysed for their geometric (shape) properties using statistical analysis to create an SSM. This 'trained' SSM is a shape-recognition model that has learnt what shape to look for in a subsequent target image. This process involves the analysis of geometric shape identifying landmarks on the edge of the target tissue that are consistent at a population level. These consistent landmarks and the automated application of AAMs to subsequent target images can automate the segmentation and substantially reduce the labour-intensive approach of manual segmentation. The advantages of this are that measurements obtained in this way are highly accurate and repeatable, and enables quantification of any systematic change. This learning phase can be independent of prior expert opinion on what may be considered important covariates with clinical outcome and avoids measurement bias.

3.2.1 Derivation of bone measures

MRI images were obtained from the OAI. These were high-resolution sagittal 3D dual-echo at steady-state water-excitation (DESS-we) knee MRI images acquired using a 3T MRI system (MAGNETOM Trio, Siemens Healthcare, Erlangen, Germany) at the four OAI clinical sites described previously. Additional parameters of the full OAI sequence protocol and sequence parameters have been published before [465]. Images were acquired at baseline, on recruitment into the OAI and at 1, 2, 4, 6, and 8 years of follow-up. Due to changes to the MRI hardware between the year 4 and 6 follow ups, and to avoid potential problems of systematic errors between measurements, the baseline to 4-year follow-up images were used in the main analyses in Chapters 5, 6 and 7, while full OAI data (up to 8 year follow up) was used as part of sensitivity analyses in Chapter 6 having applied adjustments for these known MRI changes.

3.2.1.1 Statistical Shape Modelling

Femur, tibia and patella bone surfaces were automatically segmented from 3T DESS-we images using active appearance models (AAMs) by Imorphics (Manchester, UK) in a multistage process. First, a training set of 96 knee MRIs, using the DESS-we sequence, was used to build AAMs. This training set was selected to contain examples of each stage of OA with knees fulfilling each KL (43 KL0 and KL1, 7 KL2, 28 KL3, 18 KL 4), giving a broad coverage of Kellgren–Lawrence Grade (KLG) from OAI subjects. Anatomical regions were identified on the mean bone shape (Figure 10).

As discussed previously, AAMs have proven to be a successful supervised machine learning method that can produce a segmented knee bone surface with sub-millimetre accuracy. During the model building process, 69 principal components for the femur shape, 66 for the tibia, and 59 for the patella were built. Once trained, AAMs automatically segment bones in MR images by matching principal components of shape and appearance using the least squares sum of residuals. As a consequence of this process, each time a new image is searched, the distance along each of the principal components for the object is recorded. This has the effect of reducing shape dimensionality; in the case of the femur, this reduces a triangulated mesh of over 100,000 points to 69 floating point values, one for each principal component in the model. All models within this study were generated in order to account for 98% of the variance in the shape data from the training set.

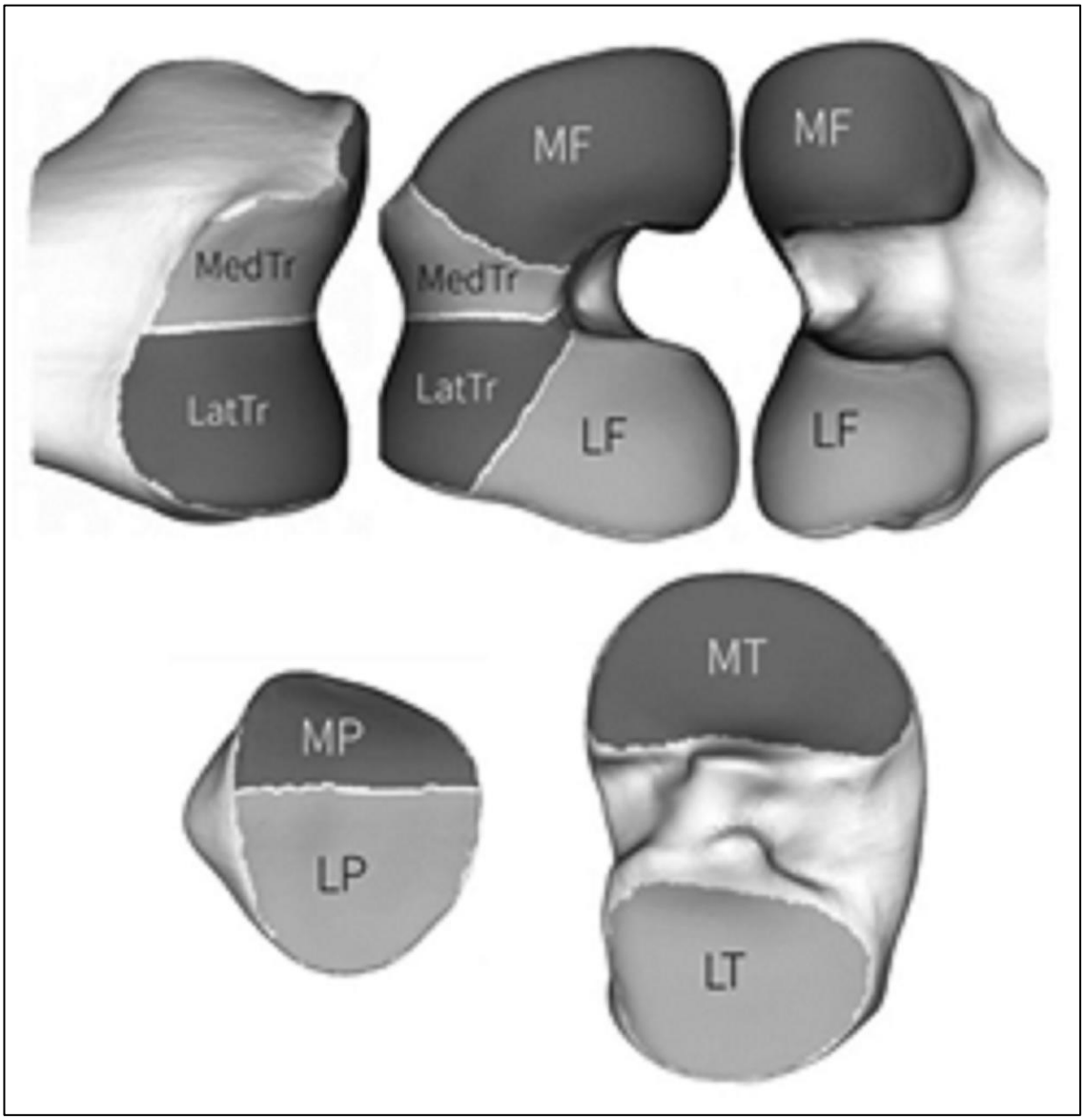


Figure 10. Anatomical regions used in this study, displayed on the mean shape for each bone.

The medial femur (MF)/MedPF and the lateral femur (LF)/LatPF boundaries were defined as a line on the bone corresponding to the anterior edge of the medial or lateral meniscus in the mean model. The MedPF/LatPF boundary was defined as the centre of the trochlear groove in the mean model.

The construction of an AAM produces a “shape space,” spanned by the set of principal components used to describe the training set of examples. Within this shape space, an “OA vector” was created using a second independent training set. This was defined by passing a line through the mean shape of the population of points with OA (OA Group; all knees with KLG ≥2 at each of 0, 1, 2, and 4 years), and the mean shape for a population without OA (Non-OA Group; defined as those with KLG of 0 at each of the same time points) using linear discriminant analysis (LDA). LDA has been used

successfully in pattern recognition and machine learning to find features that can differentiate objects into separate groups such as in facial recognition. This process further reduces the shape dimensionality to a single scalar value, which is the distance along the LDA vector for each bone or combination of bones. The point sets of the femur and tibia were combined, and a shape model was built for the combined femur and tibia shape. A combined model of femur, tibia, and patella was also constructed. LDA was performed with the principal components for each bone as inputs, with each example labelled as OA or non-OA. The distribution of the femur shapes in the training set is shown in Figure 11.

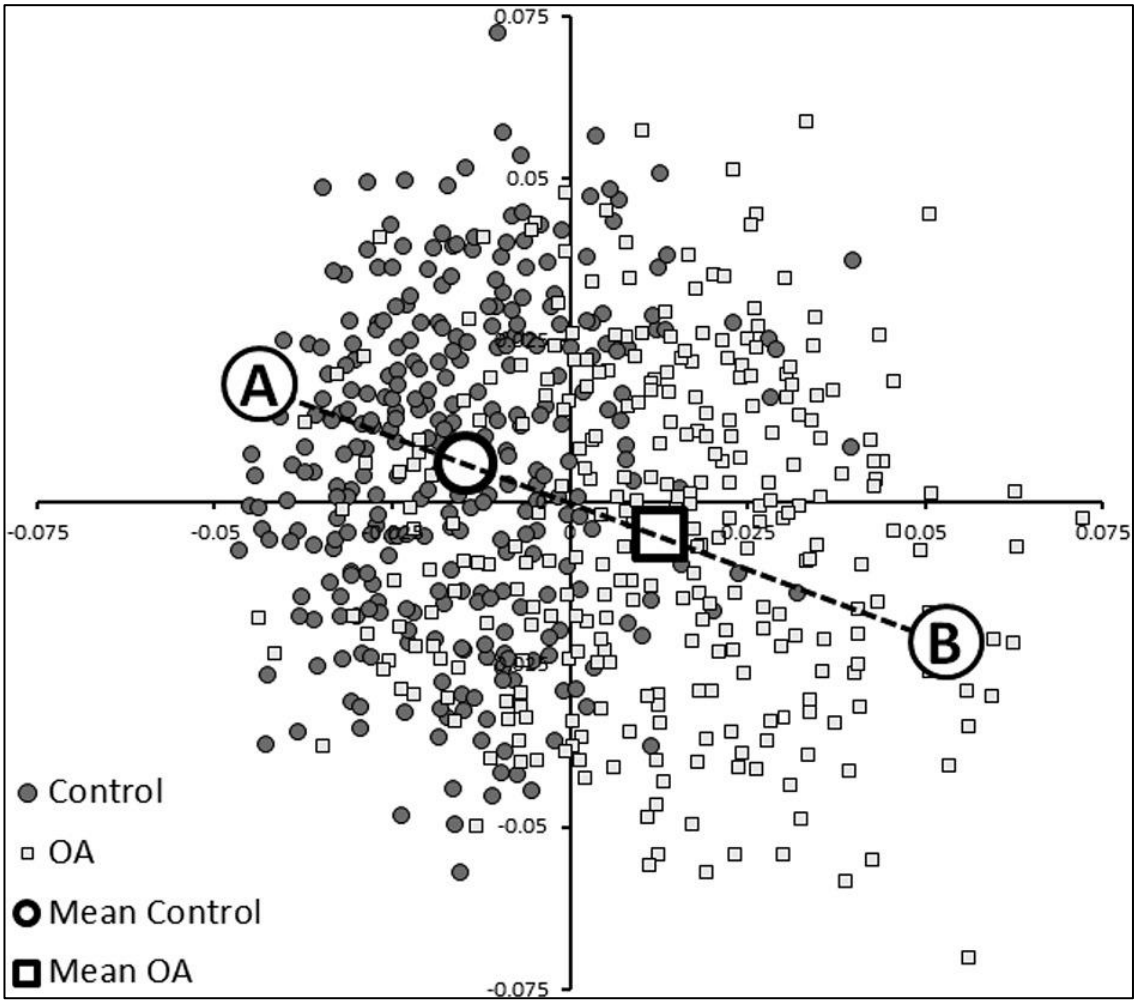


Figure 11.Sammon plots illustrating the shape distributions of the femurs used in the training set. Reproduced with permission from[111].

Linear discriminant analysis was used to determine the best single vector that discriminated the 2 groups (e.g., non-OA; [A] versus OA [B]). The results for each individual femur are encoded as 70 principal components, creating a 70-dimensional value. The Sammon plot reduces these 70 dimensions into 2 dimensions while preserving the distances between shapes as far as possible. Individual bone shape is represented using the same principal components and are projected orthogonally onto the vector. The labels “A” and “B” represent shapes at the 95% confidence boundary of a line drawn between the mean non-OA (control) and mean OA shapes.

3.2.1.1.1 Bone shape

Distances along the OA vector are subsequently normalised to a z-score (here termed “bone shape unit”), with the mean shape of the Non-OA Group represented as the OA vector origin, 0; 1 unit represents 1 standard deviation of the Non-OA Group along the OA vector (positive values toward the OA Group) representing worsening structure. Each parameterized bone shape was projected orthogonally onto the OA vector to specify the corresponding bone shape unit. Representative examples of the changes in femur bone shape, are shown in Figure 12.

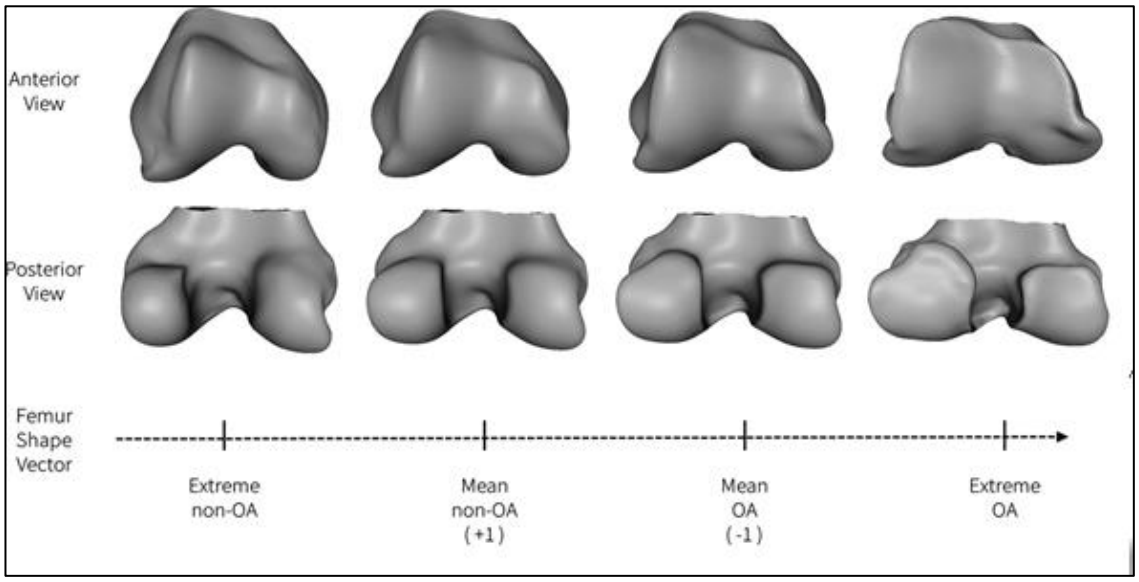


Figure 12. Shape examples along the femoral shape vector

The shape vector is calculated by taking the principal components of the mean non-OA shape, and of the mean OA shape, and drawing a straight line through them. Individual bone shape is represented using the same principal components and are projected onto the vector. Distances along the vector are normalised so that +1 represents the mean non-OA shape and −1 the mean OA shape

3.2.1.1.2 Bone shape examples

Figure 13 depicts examples of the OA and non-OA shapes for the population within the LDA training set represented by “A” and “B” in Figure 11 above.

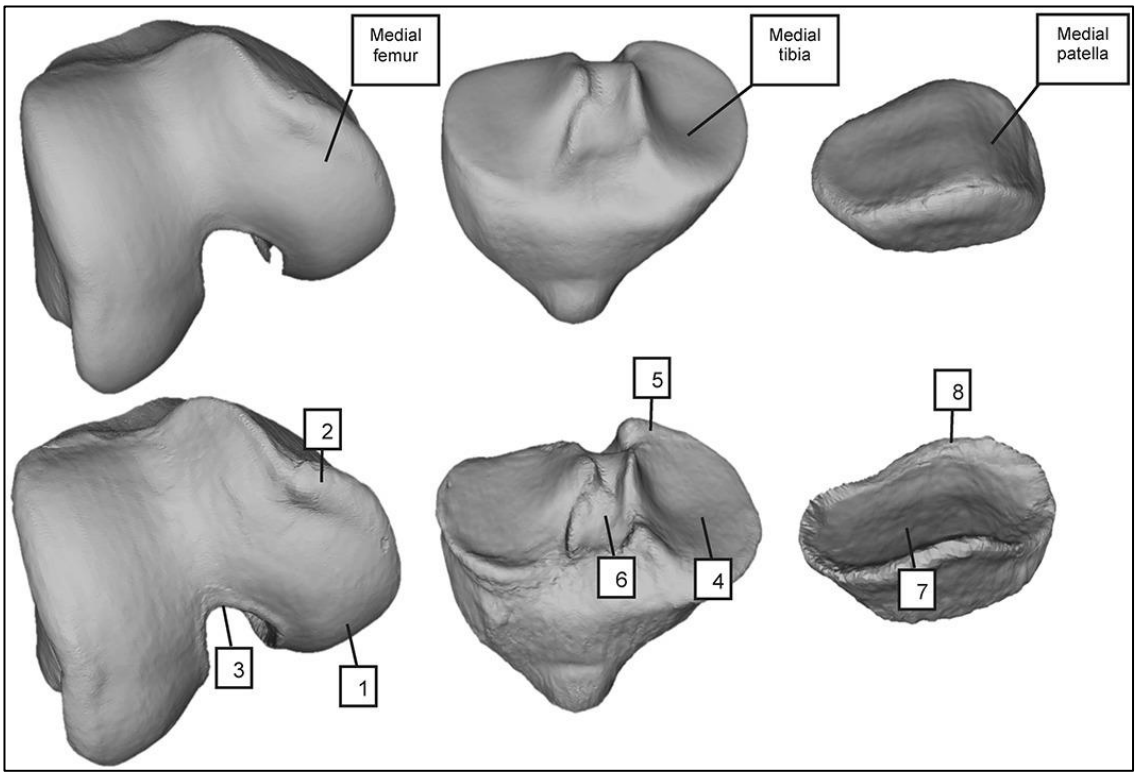


Figure 13. 3D representation of OA and control shapes

Top rows shows the control shapes for the femur, tibia, and patella and the bottom row corresponding OA shapes. With OA, the femur shape changes include widening and flattening of the condyles (**1**), an increased ridge of osteophytic growth around the cartilage plate (**2**), and narrowing of the notch (**3**). Tibia shape changes include widening and flattening of the condyles (**4**), an increased ridge of osteophytic growth around the cartilage plate (**5**), and tibial spines drawing closer together (**6**). The patella demonstrates similar patterns of increased cartilage plate size (**7**) and osteophytic ridge (**8**)

3.3 Biomarkers

The National Institutes of Health Biomarkers Definitions Working Group defined a biomarker as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” [466]. Biomarkers generally refer to a broad subcategory of objective indicators of medical state observed from outside the patient and can be assessed accurately and reproducibly.

3.3.1 Biomarkers, clinical endpoints and surrogate endpoints

While by definition biomarkers are objective, quantifiable characteristics of biological processes, they may not always correlate with a patient's experience and sense of

wellbeing. Also, due to biological variations that may exist in the population it may be difficult to predict presence or absence of diseases. In contrast, clinical endpoints are variables that characterise a participant`s health and wellbeing from their perspective [466]. Consensus has generally been that clinical endpoints are the primary, and to an extent the only relevant endpoints of all clinical research since the goal of clinical practice is to improve symptoms.

When used as outcomes in clinical trials, biomarkers are considered surrogate endpoints, acting as substitutes for clinically meaningful endpoints having been validated for their clinical relevance [466]. There are a number of advantages to using biomarkers as surrogate endpoints in trials. Primary clinical endpoints, such as survival or in the case of OA progression TKR, can occur so infrequently that their use in clinical trials can be highly impractical, or even unethical. Use of established biomarkers as surrogate endpoints also reduces the risk of harm to subjects and may allow for the design of smaller, more efficient studies.

3.3.2 Validation of biomarkers

To be considered a surrogate endpoint, there must be solid scientific evidence to determine its relevance and validity, often through epidemiological, therapeutic, and/or pathophysiological assessments. Validation requires certain assessment of certain domains and once a threshold is exceeded then the biomarker may be used as a surrogate measure. The biomarker needs to be both valid and reliable. OMERACT filter encapsulates the concepts of validity by requiring truth, discrimination and feasibility [467]. Truth requires that the biomarker measures what it is intended to measure in an unbiased and relevant way. Many components are required to validate a biomarker and the validation process is not black and white but a continuous process requiring appraisal of evidence.

3.3.2.1 Validity

Truth summarises the concepts of face, content, construct, and criterion validity. A biomarker should be valid, in that it measures what is it is sets out to measure. Face and content validity refers to the ability of a measure to include representative range of content of what is being studied (on the face of it reflect the OA outcome of interest). For imaging biomarkers of OA they should capture the intended pathophysiologic features of OA such as pain. Construct validity pertains to whether a biomarker behaves in a way that is expected from theoretical and practical perspective. This is generally assessed by comparing the prospective OA biomarker measure with other measures of the same OA domain. In imaging, for a structural imaging biomarker of OA

to have construct validity it should be associated with another structural measure of OA such as the KL grade or JSN or provide similar results. Criterion validity is a special case of construct validity where a measure is compared with the “gold standard”. In OA this is a difficult phenomenon to fully validate as a gold standard measure currently does not exist. In OA, TKR would be the closest measure to such a “gold standard”, and an OA imaging biomarker could demonstrate criterion validity by being directly predictive of TKR which becomes the ‘criterion’. Criterion validity can be classed into concurrent and predictive validity. Concurrent validity describes cross-sectional associations between biomarker and endpoint while predictive reflects longitudinal associations.

3.3.2.2 Discrimination

This requires that the biomarker be able to discriminate between situations or different groups at one time point, or multiple time points by measuring change. Related to discrimination is the demonstration of reliability, sensitivity to change or responsiveness and precision.

3.3.2.3 Feasibility

Feasibility in the OMERACT Filter encompasses the practical considerations including its ease of use, time to complete, monetary costs, and interpretability. For imaging studies such considerations include equipment and infrastructural costs, training for personnel, burden/difficulty for the patient, and ease of retrieval of information.

3.3.2.4 Responsiveness

The ability to detect small, but meaningful changes is an important attribute of a potential OA imaging biomarker. To show responsiveness the biomarker should show changes in ways that are consistent with OA measures of the same process and thereafter objective estimates of the standardised change in the biomarker over time can be computed. Changes should also be interpretable and appropriate for the population they are being applied. Moreover application of such biomarkers be feasible, with no additional burden on researcher or participants.

3.3.2.5 Reliability

In the context of biomarkers, reliability refers to the degree to which the results obtained by a measurement and procedure can be replicated. One way of testing this is to use a test-retest reliability of the biomarker either as an intra-observer reliability using one tester of if by several observers then reliability is expressed as an inter-

observer reliability or inter-rater agreement. Different reliability measures of assessment exist, including intra-class correlation coefficient (ICC), and smallest detectable differences for continuous measures, Kappa statistics for categorical measures or expressing reliability using the coefficient of variation (defined as the ratio of the standard deviation to the mean).

3.3.3 Development of imaging biomarkers for OA

The main biomarkers in current development for OA are biochemical and imaging markers. The many challenges related to biomarker research and development have been clearly articulated by The Biomarkers Definitions Working Group [466]. Current guidance for measuring clinical efficacy in disease modifying therapy development in OA is JSN, recommended by both the FDA and European Agency for the Evaluation of Medicinal Products (EMEA) as the imaging endpoint for DMOADs.

At present therefore structural progression is determined by plain radiography, but it is possible that newer technologies may be approved including biochemical markers, or MRI, once appropriately validated leading to an accelerated pathway for new OA therapies. One challenge is that the current approval of potential OA therapies requires that structural progression be linked to clinical benefit either at the time when the structure was measured or be predictive of it. Therefore it is important that improvements in OA structural features are ascertained that are more likely linked to the clinical symptoms or serve as a surrogate for a clinically meaningful outcomes. Currently there is little consensus on what constitutes a meaningful clinical endpoint for OA e.g. development of symptomatic radiographic OA or a virtual total joint replacement [468].

In 2008, the OARSI-FDA OA Assessment of Structural Change (ASC) Working Group was launched in response to a 2007 FDA notice seeking a critical appraisal on issues pertaining to treatment and prevention of OA. The group reviewed and synthesized published data and their findings suggested that there was insufficient data to make any conclusion on the predictive validity of JSW and change in JSW for clinical outcomes beyond a specific trial duration (typically 1–2 years). Responsiveness of JSW measures pooled over multiple studies was generally low for all knee radiography techniques, with higher responsiveness for fluoroscopic semi-flexed views and longer studies (~2 years+) [9, 191, 315].The Working Group recommended MRI measures of cartilage morphology on the basis of their preferable validity and responsiveness [9]. Cognisant of the need to assess the whole knee joint and also due to the growing literature on MRI quantification of non-cartilage features, the most promising MRI

measures identified in systematic reviews with respect to reliability, responsiveness and validity, were quantitative cartilage morphometry, cartilage defects and bone marrow lesions on semi-quantitative analysis, bone shape/ attrition and subchondral bone area [4]. These particular parameters were subsequently selected for inclusion in the FNIH OA Biomarkers Consortium study and research recommendations developed through a consensus process by the ASC Working Group.

3.3.3.1 Recent update on OA biomarker validation

In 2010, a working group was formed that included biomarker experts from the NIH, FDA, academia, and industry that could bring new solutions to OA biomarker development and accelerate implementation for OA therapeutics. The immediate focus of this group was to use standardized methods for biomarker validation [469]and qualification [470] in OA, using readily available well developed observational and clinical trial datasets chiefly, the OAI. Their overarching objective was to establish the predictive validity of disease progression biomarkers and assess the responsiveness of several imaging and biochemical markers pertinent to knee OA (detailed in Chapter 4).

3.4 Thesis statistical analysis

Various statistical techniques are employed in this thesis and this section provides a brief background on applications relevant to this thesis Chapter 4 ,6 and 7. Chapter 4, investigates the responsiveness and reliability of meniscus measures. Chapter 5 employs linear regression, and multilevel modelling techniques which have been extensively used in medical research as well responsiveness techniques used in Chapter 4 and detailed later. Methods relevant to Chapter 6 and 7 include longitudinal analyses (structural equation modelling, latent growth curve analysis and growth mixture models) detailed in this section.

3.4.1 Methods relevant to Chapter 4

Once domains to be measured have been established for a particular tool, it is important to assess whether the tool is truthful (validity), whether it has discriminatory capability (reliability and responsiveness) and whether it is feasible. A broad overview of the methods relating to psychometric properties of meniscus measures are discussed below.

3.4.1.1 Responsiveness

Responsiveness is the ability of a measure to detect change over time. This can be classified as internal or external responsiveness. In the latter this relates to change over a pre-specified time-frame and in the former reflects change compared to an

external reference measurement [471]. For the meniscus two measures of group level internal responsiveness, effect size (ES) and standardised response mean (SRM) were considered. The effect size is the difference between the mean baseline scores and follow-up scores on the measure, divided by the standard deviation of baseline scores. Thus, a measure that has a high level of variability at baseline in relation to mean change scores will have a small effect size. Responsiveness is a ratio of observed change and the standard deviation reflecting the variability of the change scores. This is calculated as the average difference divided by the standard deviation of the differences between the paired measurements. Thus, a measure with a high level of variability in change scores in relation to mean change will also have a small SRM value. Validated benchmarks for effect size exist: 0.20 or less represents a negligible effect, a value between 0.20 and 0.50 small effect, a value between 0.50 and 0.80 represents moderate effect and a value greater than 0.80 represented large effect.

Other measures of responsiveness exist such as the paired t-test and one thought to be the most superior, Guyatt's responsiveness index [472]. The paired t-statistic however focuses on the statistical significance of the observed change in the measure. Statistical significance depends on the magnitude of the observed change, but also depends on sample size and the variability of the measure. As a result paired t-tests are not routinely used when assessing responsiveness as sample size is not related to responsiveness.

The Guyatt`s index is derived from the formula : Guyatt Index= $\Delta x\ 2 * MSE_x$

where Δx = minimally clinically important change on the measure and MSEx is the mean squared error of X obtained from an analysis of variance (ANOVA) model that examines repeated observations of the measure in clinically stable subjects. Alternatively, where there are only two observations of the measure (e.g., before and after an intervention) MSEx is the standard deviation of the individual change scores in clinically stable patients [472]. Despite its perceived superiority, the index is not widely used because the minimally clinically important change is not yet known for a number of measures and also because that minimal clinically important change for a measure may vary across different patient populations [473].

External responsiveness can be measured by assessing meniscal measures and comparing the outcomes (e.g. change over two time-points) with a patient`s health status (remained the same, worsened or improved) or to corresponding change in a reference measure. Other methods of external responsiveness include the receiver operating characteristic (ROC) curves, which describes responsiveness in terms of sensitivity (probability of the measure correctly classifying patients who demonstrate change on an external criterion of clinical change) and specificity (probability of the

measure correctly classifying patients who do not demonstrate change on the external criterion). It represents the probability that a measure correctly classifies patients as improved (worse) or unimproved (not worse). The ROC curve provides a very useful overview of the relationship between a measure and an external indicator of change but the disadvantage is that the external clinical change be dichotomized (e.g., improved worsened). This sacrifices information on the magnitude of change in the external criterion.

3.4.1.2 Reliability

Reliability is defined as the overall consistency of a measure over time in a stable population and is calculated using test-retest methods. Where the outcome measure is measured by one individual then this reliability is expressed as intra-observer reliability or intra-rater agreement and where multiple raters are used, inter-observer agreement or reliability is measured. As the meniscus measurements studied were on a continuous scale the intra-class correlation co-efficients (ICC), were used to evaluate the intra-rater reliability for each meniscal measure and evaluated according to the following standard: poor ≤ 0.40, fair = 0.40 - 0.70, good = 0.70 - 0.90, excellent ≥ 0.90. To assess absolute reliability the standard error of measurement (S.E.M.), the smallest detectable difference (SDD) and the SDD as a percentage of the baseline value were also considered.

The SDD is a measure of the variation in a scale due to measurement error. Thus, a change score can only be considered to represent a real change if it is larger than the SDD [474]. While S.E.M, a measure of absolute reliability, provides estimates for the error size of each measured score and is an indicator of the reliability of indices, SDD is defined as a reliability level of 95% of the S.E.M between measured scores [475]. It measures the sensitivity of changes in measured values and, together with the S.E.M, is a change index reflecting the reliability of indicators. The S.E.M was calculated from the square root of the error variance of the ICC (√VarError) and SDD calculated as 1.96 X √2 X S.E.M. The ICC was calculated with a two-way mixed effects model for absolute agreement [475].

3.4.2 Methods relevant to Chapter 5

3.4.2.1 Data sources

Chapter 5 utilised a sub-study of 600 OAI participants identified for development of potential biomarkers from the Foundation for the NIH Biomarkers Consortium (FNIH). The FNIH was designed to establish the predictive and concurrent validity and

responsiveness of biomarkers relevant to knee OA progression. Its primary aims were to:

- Examine the relationship between putative efficacy of intervention biomarkers (biochemical markers, imaging features on x-ray and MRI and their progression over 1 and 2 years) and clinically relevant outcomes over a 4-year follow-up period)
- To identify the most responsive biomarker(s) of OA progression
- To develop a risk score based on baseline values of selected biomarkers that predict clinically relevant outcomes

More detail on the FNIH project are provided in Chapter 5.

3.4.3 Methods relevant to Chapter 6 & 7

3.4.3.1 Overview of structural equation modelling

Chapters 6 and 7 employ structural equation modelling (SEM) techniques, which encompass a collection of statistical techniques that allow examination of relationships between several independent (IV) and dependent variables (DV). It can best be defined as "a class of methodologies that seek to represent hypotheses about means, variances and covariances of observed data in terms of a smaller number of structural parameters defined by a hypothesised model" [476]. The IVs and DVs can be factors or measured variables. SEM can be viewed as an extension of the general linear model that enables a researcher to test a set of regression equations simultaneously. With advancements in software development, new SEM software can test traditional models and additionally examine much more complex relationships and models such as confirmatory factor analysis and time series analyses that wold be too complex otherwise.

SEM developed largely from the social sciences where because of the complexity of social reality including the latent character of many social phenomena, sophisticated methods and techniques of statistical analysis are required to determine the cause-effect relationships among many variables of interest hence the growth of SEM there [477]. The need to improve on traditional regression analysis is due to its less flexible nature: only one DV can be assessed at a time, the direction of the relationship (usually causal) can only be tested in one direction and the handling of error associated with outcomes is not well accounted for with traditional regression models [478].

SEM combines path analysis (multiple regressions concerned with the association between measured or observed variables) and factor analysis (which analyses the extent to which measured items capture latent variables). The SEM model is divided

into a measurement model part and a structural model part and is driven by a conceptual theory about a set of variables and how they relate. The measurement model allows for testing how well the observed variables combine to identify underlying hypotheses usually through CFA, while the structural model has equations that specify how variables are related i.e. the direction which can be casual or for example covariances. As SEM is an iterative technique where the direction of the relationships can be changed to fit the model, it is fundamental that the model is hypothesis driven , by evidence from literature or a conceptual understanding of the variables [479].

There are two major types of variables in SEM, observed and latent variables. Observed variables are measured directly such as WOMAC pain, history of knee surgery while latent variables are not directly measured but inferred constructs based on the selected variables for example the mean starting points for individuals` health status (intercept) or rate of change. Adequacy of SEM models is evaluated using a set of model fit indices (see 3.5.5 for more details).

While SEM has a number of advantages over traditional multivariate methods such as:

- Explicit assessment of measurement error
- Estimation of latent variables via observed variables
- Model testing where a hypothesized structure can be imposed and assessed as fit to the data

It is important to understand the assumptions underpinning SEM such that data need to be interval and normally distributed and the model to be tested needs to be specified *a priori* and then tested. Some SEM techniques cannot deal with missing data and therefore needs to be dealt with before modelling [480].

3.4.3.2 Latent growth curve analysis

Latent growth curve analysis (LCGA) was developed from SEM and offers a flexible solution to the problem of analysing change over time. Analyses of change over time have traditionally been performed using repeated measures ANOVA or ANCOVA type analyses. More recently longitudinal multilevel models, where observations are nested within subjects have been used to address these questions. Compared to ANOVA/ANCOVA and even multilevel approaches LCGMs offer several advantages including incorporating latent variables, simultaneously analysing parallel process growth models, the ease of testing of different trajectories of change and comparison of change across sub-samples. LCGA methods can also be extended to growth class analysis which allows for the identification of homogenous subpopulations within the larger heterogeneous population.

Through LCGA important longitudinal questions can be answered, such as

- What is the shape of change over time?
- Does change vary between subjects?
- Which variables predict the rate of change?
- Does change in one variable predict change in another?

LCGA differs from MLM by analysing the data at the subject level, these repeated observations are treated as measures or indicators of latent variables representing the level of and change in the construct assessed over time. Time is modelled as a latent variable and the effect of time on the outcome variable is estimated.

The latent growth curve describes the trajectory of a variable, y in terms of its intercept (value at baseline) and slope (rate of change). Thus: $y = a + bx$; where a is the value of the dependent variable, y at baseline (when time, x =0) and b is the change in y per unit of time (x). In this example, a positive value of b indicates that y increases with time. The complexity of the model can be altered to include polynomial functions which describe non-linear change in a variable. For example, a quadratic growth curve is described as: $y = a + bx + cx^2$ .Here,c indicates the quadratic growth rate (also known as acceleration) and x^2 , the quadratic form of the time scores. A negative value of c suggests that the rate of change of y slows with time, whereas a positive value indicates that the rate "accelerates" with time. Addition of further polynomial growth factors is possible, given an adequate number recorded values of variable y, but interpretation of the resulting model becomes complex.

3.4.3.2.1 Model estimation

For the present study, three separate latent growth curve models were considered for each bone tissue (femur, tibia and patella) over 8 year follow-up (9 time points) and a parallel process model to observe how change in each tissue related to change in the other tissues over 8 years. Analysis was carried out using Mplus Version 8.1,Los Angeles, CA: Muthen & Muthen.

Linear models were initially fitted, the effect of time on the outcome was fixed at zero for baseline measurement (0% of the change has occurred) and 4 for the outcome at 4 years (100% of the change has occurred). To deal with the non-linearity of growth patterns other modelling options such as polynomials and piecewise growth models were considered. The polynomial function was modelled by adding quadratic terms and for the piecewise function, models were fitted by creating joints or break points of the mean curves at different time points (year 4-8).

The final models were selected based on indices of model fit (described in Section 3.5.4.2.1),assessing whether model behaviour was keeping in line with expectations from published evidence, and with preference for a parsimonious model. The results of the analyses were reported in terms of the intercepts, which represented the mean baseline shape and for the entire cohort, and the linear and quadratic slope parameters, which represented the changes in the outcome variable with time. The resulting model-implied growth curves were then plotted and compared to the observed mean values from the sample population. The variances of the latent variables were also noted, together with the indices of model fit.

Parallel process models were fitted next: the LGCMs of femur, tibia and patella were applied to the data simultaneously and covariances between the latent covariances (intercept and slope) were estimated. With parallel process models it was possible to determine how change in femur or tibia for example affected change in patella and vice versa. A schematic of the linear models are shown in Chapter 6.

Having determined the best model (linear, quadratic or piecewise), conditional latent growth models were fitted to test whether variation in growth parameters was related to covariates. As covariates were added to the unconditional model, the significance of the variance accounted for by the covariates was tested by fitting a reduced model in which the covariates` effect on the growth parameters were constrained to be zero and conducting the appropriate chi square test. A fully adjusted model was also compared similarly.

3.4.3.2.2 Measures of model fit

Multiple test statistics and fit indices are used to determine model accuracy to ensure the model accurately represents the relationships among the constructs and observed variables. While there is agreement that fit indices should be evaluated in terms of significance and strength of estimated parameters, variance accounted for in endogenous observed and latent variables and how well the overall model fits the observed data, as indicated by a variety of fit indices. There is however still disagreement amongst researchers and/or statisticians as to what constitutes good fit and what indices need to be considered [478].

The most basic and universally reported fit statistics for any SEM model is the chi-squared statistic which is applied in a wide range of test scenarios to test whether the observed data departs form the expected or proposed model. A significant chi-square indicates the relationships between the variables in the model are significantly different from what would be expected if the model was a true representation of the data. A

small chi-square value indicates superior fit of the model, and together with the number of degrees of freedom of the model can be used to obtain the probability, p, that the specified model is appropriate and could be applied to another sample of the same population and achieve a chi-square statistic that is the same, or greater. Thus, a higher value of p indicates superior model fit and significance level of p>0.05 often chosen [481]. However when the sample is large, differences between the observed and expected covariance matrices that are trivial can cause significant chi-square statistics even in the presence of good model fit [480]. This has therefore led to the development of fit indices to allow more objective model testing. These fit indices can be classified as absolute fit indices and incremental fit indices.

Commonly used absolute fit indices in SEM include the standardised root mean square residual (SRMR) and root mean square error of association (RMSEA). These are simply derived from the fit of the observed and expected covariance matrices and the maximum likelihood minimization function. RMSEA corrects for a model's complexity, when two models explain the observed data equally well, the simpler model will have the more favourable RMSEA value. A RMSEA value of 0 indicates that the model exactly fits the data [482]. The SRMR index is based on covariance residuals, with smaller values indicating better fit. The SRMR is a summary of how much difference exists between the observed data and the model. The SRMR is the absolute mean of all differences between the observed and the model implied correlations. A mean of zero indicates no difference between the observed data and the correlations implied in the model thus a perfect fit [483].

Incremental fit indices compare a chi-square for the model tested with the chi-square from the "null" or "independence" model. The null model generally specifies that all observed variables are uncorrelated (there are no latent variables).Most incremental fit indices are computed using ratios of the model chi-square and null model chi-square and degrees of freedom of the models and all have values that range from 0 and 1.Examples include comparative fit index (CFI), Normed Fit Index (NFI) and Tucker Lewis Index (TLI). The CFI compares the improvement of the fit of the proposed model over a more restricted or null model, which specifies no relationships among variables. CFI ranges from 0 to 1.0, with values closer to 1.0 indicating better fit [483].

There has been much research regarding which combination of fit indices to best use and what the suitable cut-offs for each one should be to signify adequate fit .*Hu and Bentler* [484] examined various cut-offs for many of these measures under various conditions (sample size and model complexity) and suggest that in order to minimise Type I and Type II errors a combination of one incremental fit index, typically CFI is used (values> 0.95 indicating a good model) alongside the SRMR (good models < 0.08) or the RMSEA (good model <0.08). *Kline* [480] suggest that a minimum four

indices should be reported: the model chi-square, RMSEA, CFI and SRMR. The four measures discussed here (chi square, RMSEA, SRMR and CFI) were chosen to describe model fit of the latent growth curve analyses of bone shape. Additionally, standardised residuals covariances, which are a measure of the residual difference between the model implied covariances and those derived from the sample population, were also reported. Assumptions of normality were checked for the models.

3.4.3.2.3 LGCM measurement issues

When measuring any construct over time one would expect, and aim for certain properties in the measurements.

- When observations are collected with short intervals between them that there will be a pattern of within-subject autocorrelation between adjacent and nearly adjacent time points . This can be modelled by adding correlation paths between adjacent and nearly adjacent observed variables.
- Another issue is the equality of variance (homoscedasticity) over time. It is expected that the same observations being measured at different time points will exhibit the same residual variance. In Mplus software this is modelled by fixing observed variances to be equal across time.
- Another measurement issue occurs when using latent variables as measures of a construct at multiple time points. On such occasions there is a measurement model specified for each time point. Measurement should be operating in a consistent way across time. An initial step in testing LGCMs with repeated factors is to test the measurement model`s invariance across time. This is done by fixing equivalent factor loadings to be equal across time, and then testing for improvement in model fit when they are freed. If the model fit does not show a significant improvement it can be argued that the measurement model is consistent.
- Sample size: for an LGCM there are effectively two sample size issues, the number of time points for which data is collected and the number of participants. The number of time points determines the complexity of the growth curve that can be fitted (the technical number of time points required for being able to fit a model with m growth factors is m+1). Sample size was found to influence the convergence rates of models, with larger samples resulting in fewer improper estimates and failures and recommendations of at least n=100 thought to be necessary to obtain model convergence and stable parameter estimates. Smaller samples may be used when additional time points are present and when the variances of the slope and intercept factors are expected to be low [485]. However as the number of observation increases so does the

probability of rejecting the model based on the chi-square statistic, hence the use of fit indices to assess model fit. MacCallum *et al* offer a method of computing the minimum number of subjects required level of power that involves model acceptance and rejection via RMSEA [486].

3.4.3.3 Growth mixture models

LGCMs describe change over time and assumes that everyone in that population changes the same way through the estimation of one overall mean intercept and slope. In some cases there may unobserved heterogeneity within the sample and the overall mean while relevant in giving a picture of overall change does not capture this variability fully as there may exist a subset of individuals whose growth trajectories are significantly different from the overall estimate. For example some participants may have more rapid change in their bone shape and may have more advanced OA than others. Mixture modelling broadly aims to detect unobserved heterogeneity by allowing parameters to vary across a multinomial latent class variable. When the underlying model is a LGCM, the addition of a latent class results in a growth mixture model (GMM). By allowing the LGCM parameters to vary across classes, GMMS can identify mutually exclusive subgroups of individuals that share a similar growth curve.

The conventional growth model is a multilevel, random effects model where intercept and slope vary across individuals and this heterogeneity is captured by random effects (continuous latent variables) [487]. As shown this approach assumes that the growth trajectories of all individuals can be adequately described using a single estimate of growth parameters. GMM, relaxes this assumption and allows for there to be differences in growth parameters by use of using latent trajectory classes which allow for different groups of individual growth trajectories to vary around different means. The results are separate growth models for each latent class, each with its unique estimates of variances and covariate influences [487]. This modelling flexibility is the basis of the GMM framework.

Latent class growth analysis (LCGA) is a specific form of GMM, in which participants with a specific growth trajectory are identified, but the variation within groups is assumed to be zero (mean intercept, linear and quadratic slopes are estimated for each group, but the variances of these parameters are constrained to zero) [488]. Ideally researchers have an a priori hypothesis on the number of latent classes expected and more often these models are used exploratory with the objective of finding the optimum number of classes needed to best describe the data. One of the challenges of mixture modelling is that the ideal number of classes is not truly estimated through the modelling but the number of classes pre-specified in advance

and multiple models estimated ,each with a different number of classes and a “best fitting” model chosen.

There is some debate surrounding the most suitable indicator of the optimal number of growth trajectories, which model fit index to use and problems with model convergence [487]. The standard criteria are the degree to which the latent classes can be meaningfully interpreted, the fit of the data and the quality with which it classifies individuals into latent classes. Currently, this is determined by finding the model with the smallest Bayesian information criteria (BIC) value and a significant Lo, Mendell, and Rubin likelihood ratio test (LMR-LRT) statistic. More recently, simulation work has demonstrated that while the BIC performed the best among the information criteria based indices, the bootstrap likelihood ratio test (BLRT) proved to a better indicator of classes across all of the models considered [488]. The values of AIC and BIC are compared for models with successive numbers of trajectories, smaller values of BIC and AIC indicating improved fit. In addition entropy, LMR-LRT and the bootstrap likelihood ratio test (BLRT) are also considered. For the latter 2 tests, a significance level of $p<0.05$ indicates that the model fit is superior to the model with one with less trajectory. Entropy (ranging from 0 to 1) indicates how well subjects from the sample are classified into the trajectories described with values approaching 1 indicating superior classification [489].

An important issue often raised with use of GMMs is the problem of non-convergence and local solutions [490] .Mathematically modelling sample distributions that consist of a mixture of many different kinds of sub distributions is extremely difficult and results in convergence issues due to likelihood estimation problems (e.g. local minima and maxima and singularities). Like other methods such latent class analysis (LCA), GMMS are also susceptible to local solutions. The problem of local solutions is where during curve estimation a largest value (maximum) or smallest value (minimum) that a function takes is identified for only a given area on that curve, but that is not necessarily the largest or smallest value for the entire curve (i.e., the global minimum or maximum) and this has been well known for some time in LCA [491] .GMM parameters are estimated by the method of maximum likelihood and are iterative in nature. Ideally, the iteration results in successful convergence on the global maximum solution, that is, the parameter estimates associated with the largest log-likelihood. However issues arise when the algorithm cannot distinguish between a global maximum and a local maximum and as long as it reaches some maximum, the algorithm terminates. Fortunately, the Mplus software incorporates the use of random starting values, with sufficient user flexibility, to avoid local solutions in GMM [487].

For the longitudinal bone shape analysis LCGA results were reported in terms of:

- The number of trajectory classes identified
- The number of cases and proportion allocated to each class
- Average latent class probabilities for most likely latent class membership
- A description of each trajectory
- A graphical representation of each trajectory, compared to the mean values observed for cases grouped into each class
- Model fit indices

Once the number of trajectories was identified and each trajectory class described as above, variables predictive of class membership were tested using multinomial logistic regression. Predictor variables applied to this analysis were: age, body weight, ethnicity (white and non-white), WOMAC knee pain and history of knee surgery (yes or no).

3.4.3.3.1 Summary of GMM analysis

Having established the growth curve for each bone shape (femur, tibia and patella) separately using LGCM and determined the best model fit (linear, quadratic or piecewise), GMM was conducted. In the GMMs for each bone separately, estimation of parameters was performed in two steps. First, 1–6 class models were specified and ran to identify a model with the optimal number of latent classes. When selecting the best fitting model and optimal number of classes, the Log-likelihood, AIC, BIC, BLRT and the classification quality or entropy model fit statistics were used in combination. Following successful convergence and having determined the model with the optimal number of classes, the growth models were re-ran by including covariates and then estimated the multinomial logistic regression coefficients of the latent classes on the covariates.

Missing growth data were estimated using a Full Information Maximum Likelihood (FIML) method in which parameters are estimated using all available observations in the dataset, under missing at random (MAR) assumption. The assumption is that the probability that data are missing does not depend on the missing data but may depend on the observed ones.

Chapter 4

Where does meniscal damage progress most rapidly? An analysis using three-dimensional shape models on data from the Osteoarthritis Initiative

This chapter presents the results of the investigation into the spatial distribution of meniscal change and also assesses which meniscal pathologies change more rapidly during a 1-year follow up in a cohort typical to that included in OA clinical trials. Published in Osteoarthritis & Cartilage,2017.

4.1 Background

The role of meniscus pathology in knee OA is thought to be an important part of a complex OA pathogenesis process, and is somewhat understood from epidemiologic studies. However, quantitative meniscal measurement remains at a very early stage of development. Due to the heterogeneous and complex array of morphological changes that occur, accurately measuring meniscal pathologies is still a challenge. A number of meniscal constructs such as volume, extrusion, thickness (or height) and tibial coverage (area of the tibia covered by meniscus) have been studied previously [492-494] and nomenclature for these has been suggested [492].

The knee menisci are two crescent-shaped discs of fibrocartilage located between the surfaces of the femur and tibia in the medial and lateral compartments of the joint (Figure 14).The normal meniscus is wedge-shaped, with a flat surface facing the tibia and a concave surface facing the femur [495]. Healthy menisci protect the articular cartilage from concentrations of stress and are therefore important in load distribution [496-498]; consequently impairments in these structures results in damage to articular cartilage and may then lead to the development of OA [151].While the importance of the meniscus in OA initiation and progression is well appreciated and has been demonstrated in various studies [297, 326, 343, 495, 499-501], there is however a paucity of data on the detailed changes in meniscal pathology that occur during OA progression. Such information would be useful to determine whether the meniscus as a single structure could have properties to qualify it as a biomarker of OA progression. Also unknown is whether the meniscus could add or improve to existing imaging measures` responsiveness indices when combined with other tissue biomarkers. What

is clear is that studies on the meniscal pathologies are increasingly of relevance to the development of meniscal repair and replacement therapies.

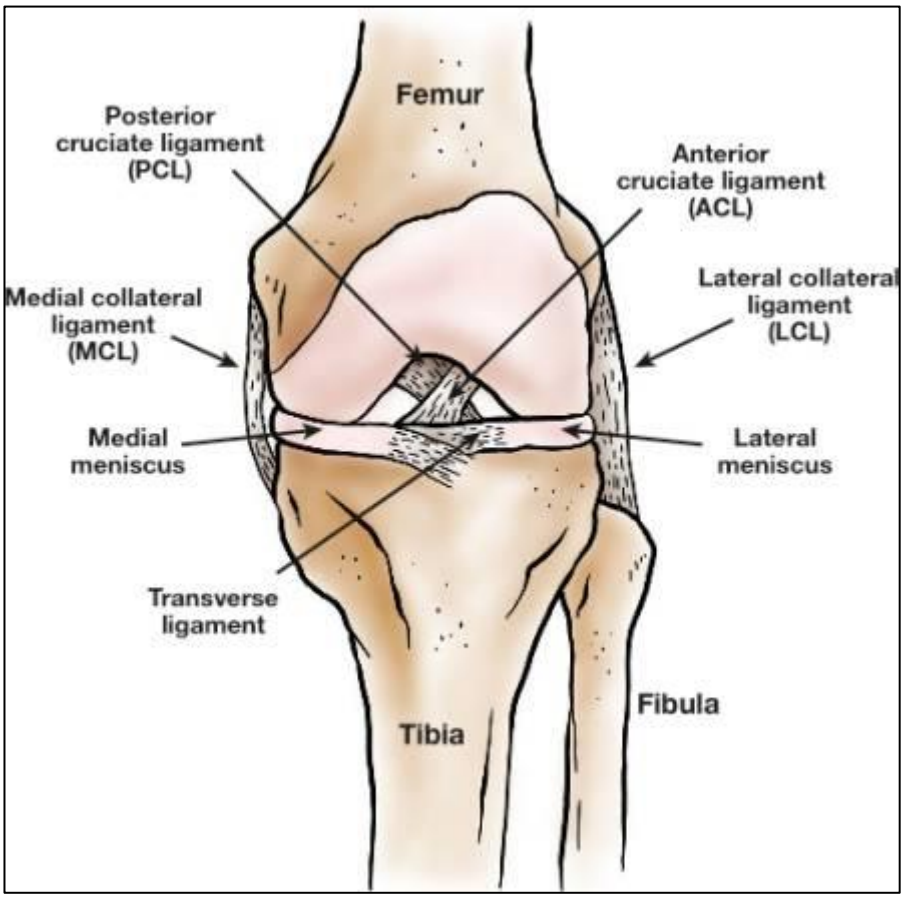


Figure 14 Anatomy of the knee joint: anterior view
Reproduced with permission from [498].

Current MRI semi-quantitative scoring [193, 212, 282] has been insightful in assessing the nature and location of meniscal pathology but may be insensitive to change as there is less scope for individuals to change by a full grade score over observation periods of 1-2 years, the feasible time for clinical studies [502] and as has been discussed previously for their use in clinical trials very large numbers would therefore be required. Another problem to do with the measures is that these SQ scores are derived from ordinal measures and their limitations were highlighted previously in Chapter 3.

Recently, through advances in imaging the quantification of meniscal volume has been achieved through segmentation of MRI images [493] and using 3D meniscal volume the effects of meniscal volume evaluated for OA and non-OA knees [503, 504].Also, manual and automated segmentation methods have been used to quantify meniscus extrusion in 3D [407, 503],and more recently automated segmentation of knee menisci

has been proposed and trialed with promising results [505].However the use of meniscal measures in clinical trials is still not well appreciated.

Through the application of statistical shape models of the meniscus, accurate information about where change occurs, and also how such change can be monitored is starting to emerge. Such information is important in the process of characterising whether the meniscus in its own right could be a biomarker of knee OA progression or if it would add to responsiveness when combined with other tissue biomarkers. What is clear is that such properties are increasingly relevant in the development of meniscal repair and replacement therapies. In view of the current difficulties in establishing where specifically OA damage first occurs [214],and in an effort to develop interventions that are responsive early in the disease, the meniscus may provide important insights to this.Adding the meniscus to other traditional measures will likely improve responsiveness of OA progression measures.

4.2 Aims

The aim of this study was to apply quantifiable novel 3D image analysis in a cohort typical to that included in clinical trials, to determine the spatial distribution of change, and the meniscal pathologies most associated with change during 1-year of OA progression.

4.3 Methods

4.3.1 Participants

This study used the first release (OAI public-use data sets 0.B.1 and 1.B.1, n=160) of the OAI progression cohort. These subjects had both frequent knee symptoms (defined as “pain, aching or stiffness”) in the past 12-months and radiographic tibiofemoral-OA (defined as definite tibiofemoral osteophytes or Kellgren-Lawrence (KL) grade ≥2) in one knee. This specific subsample was drawn from potential “fast progressors”, chosen as most likely to undergo cartilage loss, as described previously [327].

Specific inclusion criteria for this study were: evidence of medial JSN, medial JSN > lateral JSN, evidence of medial osteophytes, greater than 1º of varus mal-alignment, and availability of baseline and 12-month MRI images. In the current study, one knee per subject was selected at random and where both knees fulfilled the inclusion criteria, the knee with the greater medial joint space narrowing (JSN) was selected. Participants were excluded if they underwent arthroscopy, meniscal surgery or ligament repair

between baseline and at follow-up 12 months later. This resulted in 86 participants being selected (each with a pair of knee images at two timepoints).

4.3.2 Quantitative image analysis

MR images were acquired using Siemens-3T-Trio-Systems using the double-echo-in-steady-state-sequence (DESS) in the sagittal plane. The DESS sequence produced a 160-slice image with a high spatial-resolution and signal-to-noise ratio. This optimised morphological analysis of menisci and facilitated segmentation. Meniscus segmentation and morphometry from DESS was previously shown to yield acceptable reliability and agreement with measurements from a coronal intermediate-weighted turbo spin echo (IW TSE) sequence [494] as well as high intra-observer reliability [492, 506, 507]. DESS offers the advantage of providing better resolution with thinner slices (1.5 vs 3 mm) than the coronal IW TSE sequence, and a better delineation of the tibial plateau cartilage surface area [508].

Segmentation was performed by Imorphics (Manchester, UK). The medial and lateral menisci in the chosen knees were manually segmented by an expert segmenter who had passed a segmentation training protocol, requiring a coefficient of variation lower than 3% on paired test images. The segmenter was single-blinded to time point but not to subject. This careful manual segmentation was done using Endpoint software (Imorphics, UK). The segmentation method employed in this study was completely manual although it used SSM technology to accurately represent spatial change.

Segmented contours were converted to 3D surfaces using a marching quads algorithm, followed by quadratic smoothing techniques. Bone surfaces in the tibia were identified by automated segmentation using statistical shape modelling, a form of AAMs as described previously [339]. Fig 15a shows the mean shape of the menisci derived from the 86 individuals recruited for this study. Using AAMs a dense set of anatomically corresponded points was automatically identified on the tibia bone surfaces, which then enabled meniscal measurements to be taken in a consistent manner, and has the advantage of also correcting for patient shape and size (Figure 15b). Three dimensional images of the shape and position of the menisci relative to the tibia for each knee and time point were generated for visual review. Several measurements were taken (see Figure 15). Four meniscus measures for volume, thickness, extrusion, and tibial coverage were calculated each for the medial and the lateral sides.

Volume was calculated using Gauss' theorem for measuring volume, in which the volume is calculated by summing the vector product of the centroid, area and normal of each surface triangle [509]. Volume was derived as total volume (mm^3), after

systematically excluding the meniscal attachments from Figure 15b. Meniscal roots can be difficult to segment due to their visibility, and this measure excluded them by cutting the menisci at the boundary of the hyaline cartilage on the medial and lateral tibial plateaus.

Meniscal thickness was obtained using the corresponded points on the tibial bone (Figure 15c), obtained by subdividing the meniscus into three approximately equal segments (anterior, central, and posterior) (Figure 15d), and reported as a mean value for each region. Total thickness was the mean of all points in the combined 3 regions. Figure 15g shows how thickness measures were obtained using the underlying correspondence points. Next, sub-regional measures of thickness (anterior, central and posterior) were obtained, these would subsequently be assessed for their responsiveness as single regions of thickness and compared for responsiveness against total thickness. Tibial coverage refers to the area of cartilage-covered bone that the meniscus directly overlies; this was calculated as the area of tibia which returned a thickness measure of >0 (mm) which would represent the meniscus.

Using a novel 3D technique, extrusion measures of the medial meniscus were derived by first identifying the outermost points of the tibial plateau from the previously identified correspondence framework on the tibia, and then fitting a spline through those points. This line was extended into a plane in the sagittal direction, which is used to cut the meniscus (Figure 15e). Any volume extruded beyond this cutting plane was calculated as extruded volume (Figure 15f). Currently, the method for assessing extrusion involves drawing a vertical line at the tibial joint margin on a single coronal MRI slice and extrusion past this point is measured in millimetres [8].

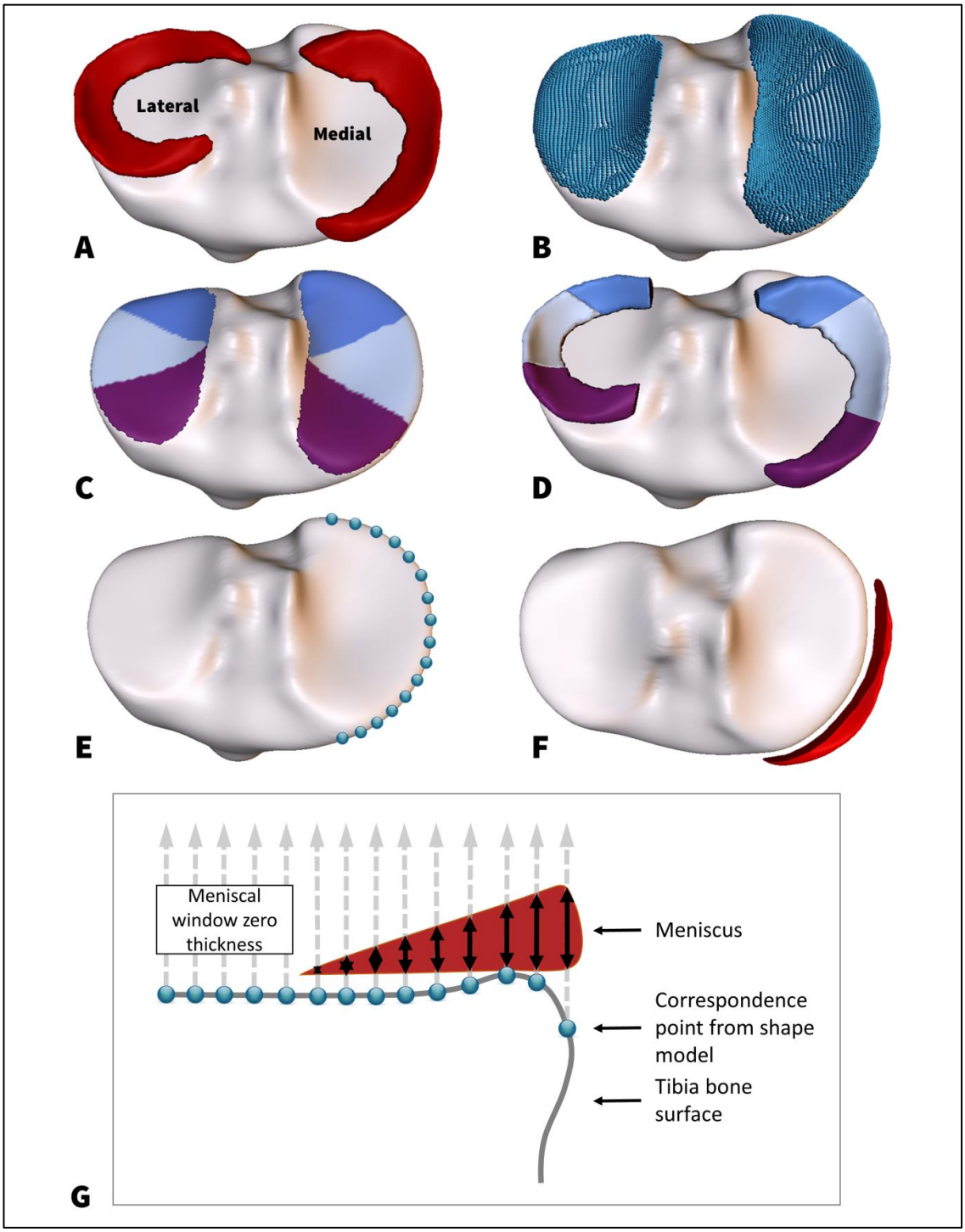


Figure 15 Identification of anatomical regions and measurement

Figure A shows the mean shape of the menisci for this group of 86 individuals. Figure B shows the anatomical correspondence points (blue spheres) from the tibia bone shape model which are used to subdivide the tibial plateaus, from which measurements are taken. Figure C shows the anterior (purple), central (light blue) and posterior (dark blue) regions on the lateral and medial tibial plateaus, selected using the correspondence points, and D shows the mean meniscus split into 3 regions for each meniscus. Figure E shows the correspondence points identified along the outer boundary of the medial tibia. These points are joined into a line, and extruded into a plane in the superior direction, which cuts the meniscus into an inner and outer section. F shows the extruded section. Figure G shows how thickness measures are taken using the underlying correspondence points on the tibia bone.

4.3.3 Reliability

A reliability study was performed on an independent sample of 20 participants (23% of the main study population) with no OA or mild OA. Careful manual segmentation as in the main study was using manual segmentation, with the repeat performed by the same individual blinded to subject. Intraclass correlation coefficients (ICCs) were used to evaluate the intrarater reliability for each meniscal measure, while the standard error of measurement (S.E.M) [475]; smallest detectable difference(SDD) [474] as well as SDD as a percentage of the baseline value were employed to assess absolute reliability. ICCs were evaluated according to the following standard: poor ≤ 0.40, fair = 0.40 - 0.70, good = 0.70 - 0.90, excellent ≥ 0.90. The SDD was calculated as 1.96 X $\sqrt{2}$ x S.E.M.

The SEM, a measure of absolute reliability, provides estimates for the error size of each measured score and is an indicator of the reliability of indices[475]. The SDD [474] is another measure of absolute reliability and is used with the small reference difference (SRD) [475].The SDD is defined as a reliability level of 95% of the S.E.M between measured scores. It measures the sensitivity of changes in measured values and, together with the S.E.M, is a change index reflecting the reliability of indicators [475]. Lower S.E.M and SDD values indicate higher reliability of the accuracy and precision of the measured values. When the SEM value is less than 10% of the average measured value or the highest measured value, the measurement error is small, and therefore, the measurement is reliable [475].

4.3.4 Statistical analysis

Statistical analysis was conducted using STATA software, Version 13 (College Station, TX, 2013) and MedCalc for Windows, Version 15.6 (MedCalc Software, Ostend, Belgium). Measures of dispersion, the mean and standard deviation (SD) of the difference at 1 year follow-up were determined for all meniscal measures. Group level internal responsiveness was assessed using two measures, effect size (ES) and standardised response mean (SRM), to ensure comparison of the magnitude of change in a standardised manner, for each measure [37]. The SRM was calculated as the mean change divided by the standard deviation of the change score. Confidence intervals for SRMs were estimated using the bias-corrected and accelerated bootstrap methods, because in small samples the estimate of the standard deviation may be biased [38].The following validated benchmarks were used: 0.20 or less represented a

negligible effect, a value between 0.20 and 0.50 represented a small effect, a value between 0.50 and 0.80 represented a moderate effect and a value greater than 0.80 represented a large effect [471].

A paired student's t-test compared baseline and 12-month means to evaluate whether any changes were significantly greater than zero. Graphical checks were performed to ensure statistical assumptions were met prior to performing t-tests and these were satisfactory. SRMs were evaluated on the 86 participants as these were assumed to be homogenous in terms of their expected change over 1-year and therefore satisfied the assumptions for the SRM method. Based on the final selected sample of 86 a retrospective power calculation showed that this study had 80% power to detect an effect size of 0.31.

Four measures as described previously were assessed on the medial and lateral sides, and thickness was further evaluated using sub-regions. To adjust for multiple comparisons (on the 14 tests performed), a Bonferroni correction adjusted for mean correlation of the meniscal measures was applied and the level of significance set at (α=0.008) [39].

Furthermore, to assess if responsiveness varied based on OA risk factors 3 strata were derived and exploratory analysis performed of these. These sub-groups were based on three demographic factors known to be important in OA: age, gender and body mass index (BMI) as previously described in Chapter 2. The strata were created based on median age of the sample (age<62 and age≥62), gender (males and females), and obesity status using WHO cut-offs (BMI≥30 and BMI<30). Responsiveness was also compared between the group of participants that self-reported previous arthroscopy or meniscectomy at baseline, and the rest of the group. Lastly, exploratory analyses compared 3D medial extrusion against the MOAKS medial extrusion scores.

4.4 Results

4.4.1 Demographics

Median (IQR) age was 61.5 (52-71), 49% were women and 78% identified their ethnicity as white. The mean BMI ± SD was 31.1 ± 4.60 kg/m² and median (IQR) pain score of 5.44 (2.4-6.3) as measured using the Western Ontario MacMaster Universities Osteoarthritis Index (WOMAC) scales. About 30% self-reported having had a meniscectomy/arthroscopy on the chosen study knee (Table 8). The 74 of the 160 participants excluded from the study had very similar characteristics to those selected (age 61.0 vs 61.5 and gender 53% vs 49% respectively) see Table 8. As expected,

visual appraisal of images confirmed the heterogeneity of meniscal pathologies, Figure 16 demonstrates these using examples from participants included in this study.

Table 8: Characteristics of 86 participants in meniscus study

	Included in the study N=86	Excluded from study N=74
Age, years, median (IQR)	61.5 (52-71)	61.0 (53-69)
Gender, female	42 (49)	39 (53)
Ethnicity, white	67 (78)	65 (88)
BMI, kg/m², mean (SD)	31.1 (4.64)	29.4 (4.57)
Height, m, mean (range)	1.7 (1.5-1.9)	1.7 (1.5-1.9)
High school education or less	21 (24)	7 (10)
Study knee, Right	43 (50)	35 (47)
Arthroscopy /meniscectomy on study knee	25 (29)	17 (23)
Health care insurance	84 (98)	71 (97)
WOMAC pain score, median(IQR)	4.1 (2.4-6.3)	3.5 (2.0-6.0)

Values are N (%) unless stated. m (metres). BMI (body mass index) IQR (interquartile range)

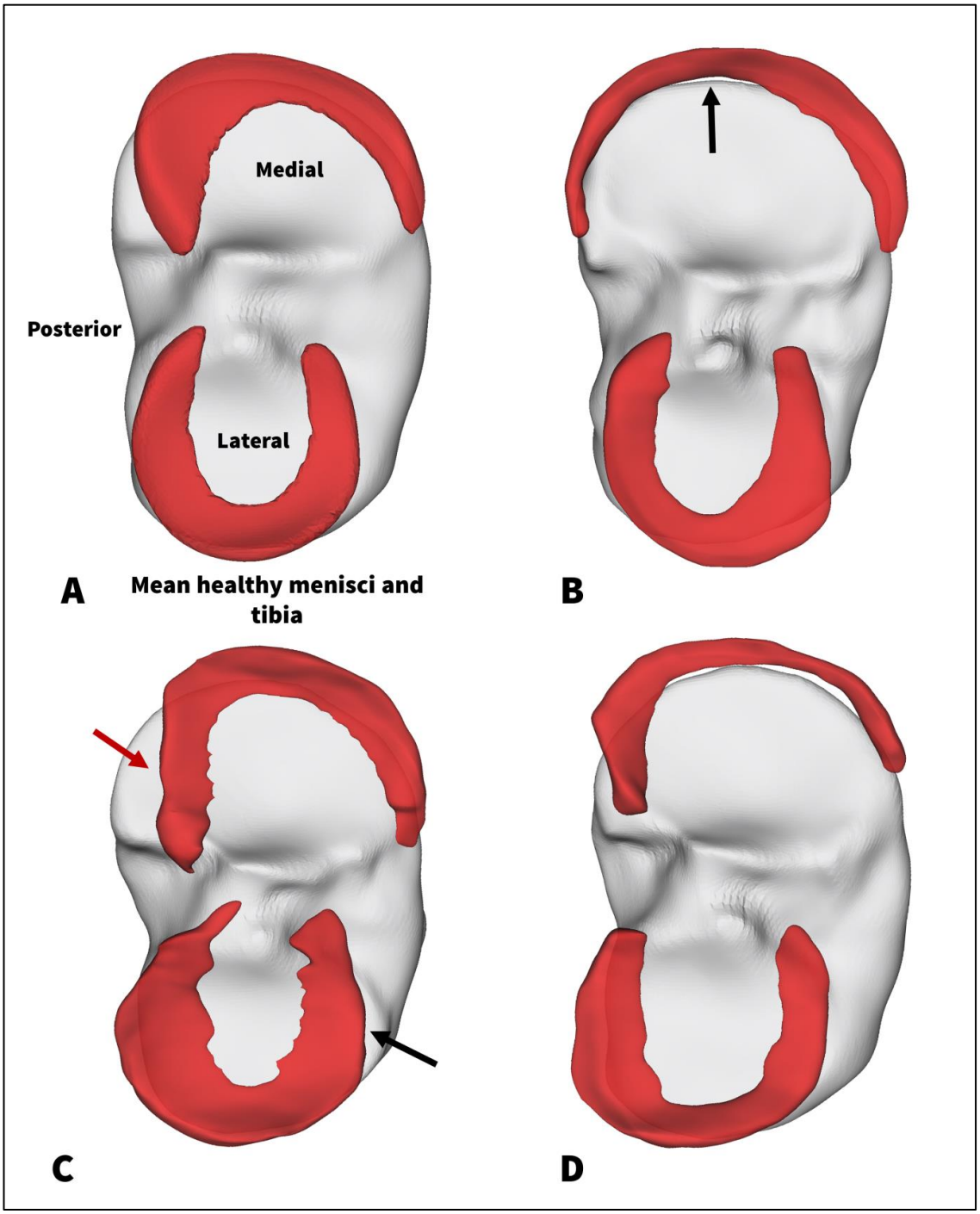


Figure 16 Variety of meniscus shapes; examples from the data set and healthy mean shape

Menisci are shown in red, with slight transparency to visualise extrusion beyond tibial bone. Figure A shows the mean medial and lateral meniscus shape from a group of healthy (KL0) knees from the OAI for comparison with cases. B shows a damaged medial meniscus, which is much thinner than the healthy meniscus, the central section is almost all extruded beyond the tibia. C shows both the medial and lateral menisci deformed by a tibial osteophyte (red arrow, posterior medial osteophyte pushing the meniscus anteriorly; black arrow anterior lateral osteophyte pushing the meniscus posteriorly). D shows both menisci are damaged.

4.4.2 Reliability

For both medial and lateral measures ICC values realised were very high (good reliability), with the lowest value seen for lateral extrusion (ICC 0.97, 95% CI 0.92, 0.99) and highest for medial tibial coverage (ICC 0.99, 95% CI 0.97, 0.99) (Table 9). Also, low SDD values were seen in the repeatability study also representing good repeatability of the segmentation methods. The SDDs (SDD as % of baseline) on the medial side for volume, extrusion, thickness and coverage were 32.2 mm^3 (1.9%); 15.7 mm^3 (9.2%); 0.03 mm (2.6%) and 9.2 mm^2 (2.3%) respectively, all very small values relative to the average for the sample. Similarly low SDD values were seen for the lateral measures: 55.5 mm^3 (3.6%) for total volume; 9.7 mm^3 (16.2 %) extrusion; 0.03 mm (2.3%) thickness and 6.1 mm^2 (1.6%) for lateral meniscal coverage.

Table 9 . Results from reliability study of 20 participants

Measure	Mean diff	ICC	S.E.M	SDD
Meniscal coverage				
Lateral	+6.82	0.99	2.2	6.1
Medial	-3.26	0.99	3.3	9.2
Trimmed volume				
Lateral	+47.27	0.99	20.0	55.5
Medial	+9.05	0.98	11.6	32.2
Medial extrusion	-4.53	0.97	5.7	15.7
Meniscal Thickness				
Lateral	+0.03	0.98	0.01	0.03
Medial	-0.01	0.99	0.01	0.03
Total volume				
Lateral	-86.3	0.98	30.9	85.7
Medial	+2.94	0.99	19.8	54.9

4.4.3 Longitudinal change at 1-year

Medial volume decreased by 1.1% from baseline while medial extrusion showed an increase of 4.1%, but neither changes were statistically significant ($p>0.05$) and assessed against the previously determined SDD, change in extrusion exceeded SDD (Table 10). Mean medial total thickness showed a decrease of 6.1% from baseline to follow-up ($p<0.001$) while mean tibial coverage decreased by 4.4% ($p<0.001$) with both changes exhibiting change greater than SDD. On the lateral side no changes were greater than SDD except for volume, however none were statistically significant. Analysis of the sub-regions however, showed a significant increase of 2.6% for mean central thickness ($p<0.001$), although the absolute mean change was very small (0.05 mm) (Table 10).

Visualisation of the spatial location of change in meniscal thickness is shown in Figure 17. Greatest change in thickness was consistently seen in the posterior region of the medial meniscus.

4.4.4 Responsiveness

The SRM and ES values are reported in Table 10 and 11. In the primary analyses on the medial side volume and extrusion measures showed no significant change (Table 10) while medial meniscal thickness (SRM - 0.35, 95% CI -0.55,-0.14) and tibial coverage (SRM of -0.36, 95% CI -0.58,-0.13) showed moderate responsiveness. Changes on the lateral side did not show any statistically significant changes, with the highest SRM seen being that of total thickness SRM = +0.29 (95% CI 0.12,0.50). However, the regional measure of central thickness on the lateral side showed a small positive response (SRM +0.33, 95% CI 0.13, 0.51) (Table 11).

Analysis of whether the thickness measures analysed as sub-regions on the medial side would improve sensitivity compared to total thickness measures revealed that responsiveness in the posterior thickness was similar to total thickness, central thickness was less responsive, and anterior thickness did not change. As highlighted before in the lateral side no pattern was seen and the results were not consistent as only the central region showed change.

Table 10 Changes in medial meniscus measures

Meniscal measure	Baseline	12 months	Change (95% CI)	% change (95% CI)	SRM (95% CI)	ES	p-value (t-test)
Volume (mm3)							
Total volume	2527.69	2498.97	-28.72 (-108.89,51.46)	-1.1 (-0.04,2.03)	-0.08 (-0.27,0.13)	-0.02	0.48
Extrusion (mm^3)							
Extruded volume	507.26	528.12	+20.86 (-2.56,44.27)	+4.1 (-0.50,8.72)	+0.19 (-0.03,0.40)	+0.08	0.08
Area (mm2)							
Tibial coverage	414.74	396.32	-18.42 (-29.33,-7.52)	-4.4 (-7.07,1.81)	-0.36 (-0.58,-0.13)	-0.12	<0.001*
Thickness (mm)							
Total thickness	1.14	1.07	-0.07 (-0.11,-0.03)	-6.1 (-9.64,-2.64)	-0.35 (-0.55,-0.14)	-0.16	<0.001*
Anterior thickness	0.40	0.41	+0.01 (-0.02,0.03)	+2.5 (-5.00,7.50)	+0.04 (-0.18,0.26)	+0.02	0.71
Central thickness	0.81	0.76	-0.05 (-0.10,-0.01)	-6.1 (-12.35,-1.23)	-0.27 (-0.47,0.04)	-0.11	0.02
Posterior thickness	2.16	2.00	-0.16 (-0.24,0.07)	-7.4 (-11.11,3.24)	-0.38 (-0.53,-0.21)	-0.20	<0.001*

ES: Effect Size. SRM: Standardised response mean. *: significant p-value when using paired student's t-test.

Table 11 Changes in lateral meniscus measures

Meniscal measure	Baseline	12 months	Change (95% CI)	% change (95% CI)	SRM (95% CI)	ES	p-value (t-test)
Volume (mm3)							
Total volume	2131.21	2177.11	+45.90 (11.75,80.05)	+2.2 (0.55,3.76)	+0.29 (0.01,0.50)	+0.05	0.009
Extrusion (mm^3)							
Extruded volume	25.77	25.02	-0.75 (-8.44,6.93)	-2.9 (-32.75,0.32)	-0.02 (-0.23,0.19)	-0.01	0.85
Area (mm2)							
Tibial coverage	507.24	513.11	+5.87 (0.69,11.06)	+1.1 (0.14,2.18)	+0.24 (0.03,0.44)	-0.06	0.03
Thickness (mm)							
Total thickness	1.88	1.92	+0.04 (0.01,0.06)	+2.1 (0.53,3.19)	+0.32 (0.12,0.50)	+0.09	0.04
Anterior thickness	1.90	1.92	+0.02 (-0.008,0.05)	+1.1 (-0.42,2.63)	+0.16 (-0.07,0.37)	+0.05	0.15
Central thickness	1.95	2.00	+0.05 (0.02,0.08)	+2.6 (1.03,4.10)	+0.33 (+0.13,0.51)	+0.09	0.002*
Posterior thickness	1.84	1.88	+0.04 (-0.004,0.09)	+2.2 (-0.22,4.89)	+0.19 (0.01,0.38)	+0.06	0.07

ES: Effect Size. SRM: Standardised response mean. *: significant p-value when using paired student's t-test.

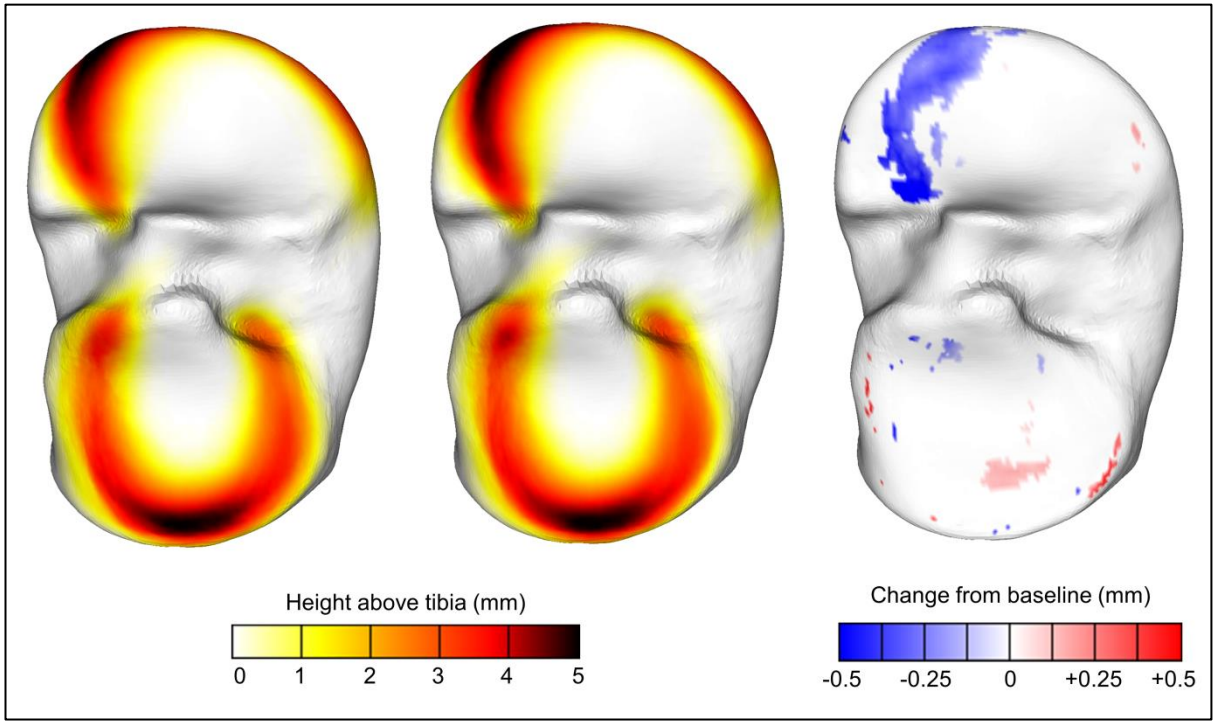


Figure 17 Mean thickness of baseline and 12 month menisci, and difference map

Left hand figures show mean thickness (height above the tibia) at baseline (furthest left) and 12 months (middle image). Measurements were taken as shown in Figure 15G. The figure on the right shows the areas which showed significant change, Blue represents thinning of the meniscus, and red is thickening.

4.4.5 Exploratory analyses of drivers of change

After stratifying for age, gender, BMI or previous arthroscopy/meniscectomy mean differences in meniscal measures at 1-year were not substantial nor statistically significant (Table 12). Responsiveness indices (SRMs) were similar within each stratum.

Analyses also investigated ceiling effects, as menisci could on average be already relatively extruded hence less likely to extrude further. This was achieved by dividing the dataset into quartiles based on volume extruded in the medial meniscus at baseline, and assessing the amount of change in extrusion over time, summarised by quartile (Table 13). Overall, positive change over time was seen in all quartiles, with greater change seen in those people falling in the quartiles having more baseline extrusion suggesting that ceiling effects were not important. While change was greatest in the 3rd quartile, those in the 4th quartile extruded a greater volume than those in quartiles 1 and 2, suggesting that even those with most extrusion at baseline continued to extrude further.

Other exploratory analyses included examining the relationship of the volume measured in this study against a validated extrusion measure in the central slices of MOAKS scores. However, as not all participants had MOAKS scoring in the OAI, only 27 participants from this study had MOAKS scores available. The relationship between the novel 3D extrusion volume measures and MOAKS extrusion scores were assessed descriptively using a box plot (Figure 18). As the MOAKS score increased so did the median 3D volume extruded, however knees with MOAKS scores of 2 or 3 contained a much wider range of extruded volumes than those with a score of 0 or 1. This suggests that the relationship between novel 3D quantified volume measures, and a SQ score based on extrusion at a central slice is about as close as might be expected. In these 27 participants only 3 of the 21 participants that had scope to progress (6 had a baseline MOAKS score of 3) showed progression from a MOAKS extrusion score of 1 (between 2mm to 2.99 mm) to 2 (3mm to 4.99 mm). Using a novel 3D extrusion score there was no statistical difference between baseline and follow-up in these individuals, however the trend was towards an increase in mean extrusion; and at the participant level 13/27 showed extrusion changes greater than measurement error.

Table 12 Longitudinal change in meniscus measures after stratification

	Total volume (mm^3)	Volume Extruded (mm^3)	Tibial coverage (mm^2)	Meniscal Thickness (mm)
Meniscectomy status				
Single meniscectomy	-111.38 (-273.30, 50.54)	-12.67 (-57.78, 32.44)	-30.95 (-25.46, -1.69)	-0.09 (-0.18,-0.01)
None	+3.28 (-90.26, 96.83)	+33.83 (6.36, 61.31)	-13.57 (-56.05, -5.86)	-0.06 (-0.11,-0.01)
Difference between groups (95% CI)	114.66 (-63.45, 292.77)	46.51 (-5.03, 98.05)	17.38 (-6.79, 41.56)	0.03 (-0.11,-0.03)
p-value	0.20	0.08	0.16	0.60
Age				
< median age	-31.66 (-132.05, 68.71)	10.90 (-18.40, 40.20)	-13.35 (-27.50, 0.80)	-0.03 (-0.08,0.01)
> median age	-25.77 (-155.10, 103.57)	30.81 (-6.72, 68.34)	23.50 (-40.56, -6.44)	-0.11 (-0.18,-0.04)
Difference between groups (95% CI)	-5.90 (-167.22, 155.43)	-19.91 (-66.83, 27.01)	10.15 (-11.69, 32.00)	0.08 (-0.006,0.16)
p-value	0.94	0.40	0.36	0.07
Weight status				
Obese	-28.37 (-117.21, 60.48)	21.48 (-11.90,54.87)	-19.30 (-35.77, -2.88)	-0.07 (-0.12,-0.02)
Non-obese	-28.97 (-153.93, 95.99)	19.99 (-13.24,53.21)	-17.22 (-30.66, -3.77)	-0.07 (-0.14,-0.03)
Difference between groups (95% CI)	0.60 (-162.90, 164.11)	-1.50 (-49.25,46.26)	2.08 (-20.17, 24.32)	0.00 (-0.08,0.09)
p-value	0.99	0.95	0.85	0.98
Gender				
Male	-72.83 (-222.99, 77.32)	15.57 (-24.71, 55.86)	-24.00 (-42.20, -5.81)	-0.08 (-0.15,-0.01)
Female	+17.50 (-36.31, 71.31)	26.39 (1.75, 51.03)	-12.58 (-24.81, -0.34)	-0.06 (-0.11,-0.01)
Difference between groups (95% CI)	-90.33 (-250.52, 69.84)	-10.82 (-57,89, 36.26)	-18.42 (-29.33, -7.51)	-0.02 (-0.11,0.07)
p-value	0.27	0.65	0.30	0.64

Values are paired mean differences (95%CI)

Table 13 Volume extruded longitudinally based on baseline quartiles

Quartiles of 3D baseline meniscus volume	Volume extruded in 12 months, mean (SD) (mm^3)
1	38 (31)
2	74 (96)
3	131 (71)
4	96 (87)

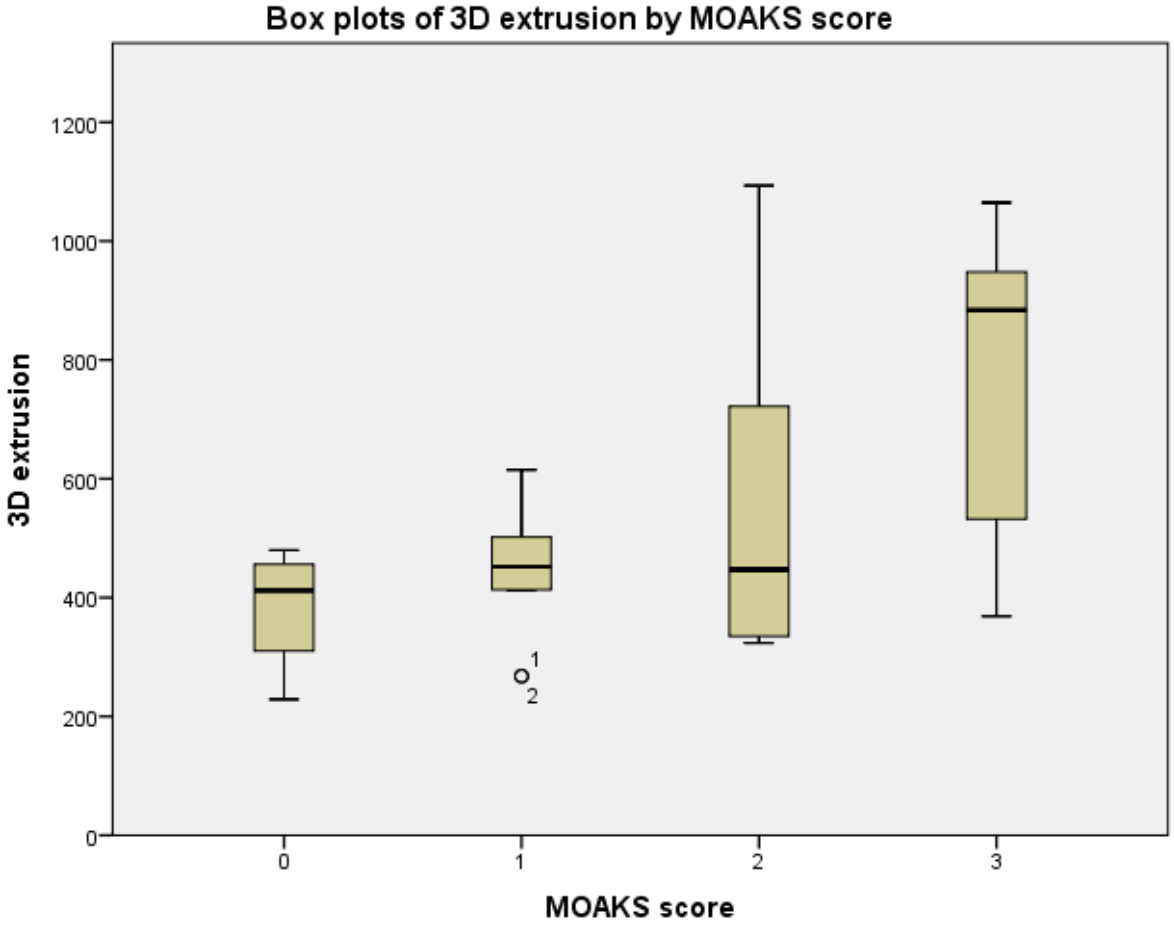


Figure 18 The relationship between 3D extrusion and MOAKS extrusion

4.5 Discussion

This is the first study using SSMs to assess the longitudinal change of a range of 3D meniscal pathologies, in an OA cohort typical of that seen in OA clinical trials. The spatial location of meniscal damage was found to occur predominantly in the posterior sub region of the medial meniscus. Longitudinally the most responsive meniscal measure was tibial coverage which changed (reduced) by 4.4% (SRM = -0.41) during follow up. Although most change was demonstrated in the medial posterior thickness measure (7.4% reduction in 1-year) (SRM -0.38), responsiveness in that region was similar to that of tibial coverage because the change in thickness was subject to more variation. The responsiveness of these meniscal measures was better than those seen for 12 month radiographic joint space width measures (SRM = -0.22) and MRI cartilage thickness measures (SRM = -0.32) found in one study[318]. Minimum radiographic joint space width still represents the FDA standard for demonstrating structural therapeutic benefits for knee OA. Results from a systematic review showed that studies with similar follow-up to this study (1-2 years) reported a pooled SRM of 0.25 for JSW [314]. A major benefit of SSM technology stems from the 3D registration capability that corrects for both size and shape of knees; this may be the reason for the strengths of this study including very good repeatability.

Of the 4 primary measures the meniscal pathology demonstrating the most responsiveness to change was medial tibial coverage (SRM = -0.36). This finding is similar to another smaller-sized study employing 3D meniscal measures that also found tibial coverage to be the most responsive meniscal measure, although they measured responsiveness at 2-year follow-up and reported a higher SRM of 0.82 [407]. The responsive decrease in coverage seen could be as a result of diminishing tibial coverage in OA-affected subjects due to meniscal destruction and radial displacement [297]. No significant changes were seen for lateral coverage which could possibly be due to the study inclusion criteria of medial OA progression. The importance of tibial coverage in OA has been demonstrated by Bloecker and colleagues previously [504], in their study knees with medial JSN showed substantially less tibial coverage by the meniscus and this was thought to be via the mechanism of reduced mechanical protection of the articulating surface. When combined with central femoral cartilage measures, tibial coverage was shown to explain 66% of the variance in mJSW[510]. Tibial coverage is also thought to be associated with meniscal extrusion (discussed later in this section). Wenger and colleagues [503] in another study using the OAI cohort demonstrated a cross-sectional association between OA knees and less tibial coverage and more extrusion. A methodological strength of the current study was the

application of a bootstrapping method to provide confidence in the SD estimates for SRMs, as estimating SD from small populations can become sensitive to outliers. Tibial coverage is very responsive and from a feasibility perspective should be easier to perform for research groups without access to specialist 3D measurement. Previous work from Leeds OA research group (unpublished data) used the area of the meniscal window (either as measured in mm^2 or as a percentage of tibial bone) to represent a similar construct and was found to be the most responsive meniscal measure. This measure intrinsically relates the size of the (shrinking) meniscus to that of the (expanding) tibia but does not correct for this tibial expansion [339] which could result in systematic over-estimation of changes as bone changes could be concurrently happening, and sometimes at a much faster rate than those happening in the meniscus. The measure of meniscal coverage used in this study is not affected by tibial size.

Overall, medial thickness measures decreased significantly at one-year follow-up and appeared moderately responsive compared to other measures. This study found a 6% reduction at 1-year follow up that was both statistically significant and in excess of measurement error. This result is consistent with findings from a 2-year pilot study that found a significant reduction of about 4% in meniscal height over the tibia (similar to the measure for total thickness used here) [407], however that study only measured thickness in one region. In addition changes in three sub-regions of the meniscus were evaluated, some of which appear to provide promising measures of change based on their responsiveness. Similarly for thickness, in a study with 257 participants Hunter *et al* found a reduction in thickness on the medial side which was associated with cartilage loss at 15 and 30 months follow up [297].Cross–sectional analyses however have not shown any differences in meniscal thickness measures between OA and non-OA knees [492, 503, 504], and future studies could evaluate if the longitudinal changes in this measure are associated with OA progression. While sub–regional analysis showed that most change occurred in the posterior region of the meniscus, measuring the whole meniscus thickness was more responsive (SRM= -0.35) than using three separate regions. Separating the regions into smaller sub-sections offered some advantages, higher responsiveness for posterior thickness (SRM = -0.38) but this measure may be noisier as it was only accompanied by a 7.4% overall change. Surprisingly the lateral thickness measures showed increases during follow up but these were not statistically significant except for the central thickness sub-region; and even this showed less overall change i.e. 2.6% vs 6.1% on the medial side. This difference in direction of change could be attributed to an inclusion criteria for medial progression and the possibly higher measurement error for the lateral measures.

In terms of magnitude a decrease in medial volume was found, but surprisingly an increase on the lateral side although both changes were not statistically significant. Studies assessing meniscal volume previously have yielded conflicting results: one study reported greater lateral volume in OA knees compared to non-osteoarthritic knees with no differences in medial volume [503], while another study from the OAI showed no differences in either compartment over time [504]. A pilot study evaluating 2-year longitudinal data [407] like this study, found a longitudinal decrease on the medial side. Using quantifiable measurements of meniscus volume is important as these could provide a more robust means of tracking substance loss longitudinally in patients with knee OA. The volume measurements used in this study however came with many technical challenges. Firstly, manual segmentation of volume proved difficult as damaged menisci and meniscal roots can assume various complex shapes which results in a laborious process as these structures need to be separated before derivation of the volume measure. Secondly, correctly determining where the roots begin is itself a technically challenging process. Therefore, variation in volume results could potentially be a result of measurement error due to the varying techniques employed by different studies in deriving, and also measuring meniscal volume. Some of these studies did not report how their change scores varied with measurement error, therefore what might be perceived as a lack of sensitivity in volume could actually be small longitudinal changes masked by large measurement noise. Because segmentation of meniscal volume is laborious, these varying findings for volume highlight the need for further investigation. The lack of responsiveness observed in this study and the difficulty in segmentation could undermine its use as a potential tool for clinical trials from a technical and feasibility perspective.

This study employed a novel way of measuring extrusion on a 3D plane which facilitated the calculation of extruded volume. Notably this study found poor responsiveness for meniscal extrusion which was surprising because extrusion has previously been linked to various OA features in longitudinal and cross-sectional studies [343, 503, 511] and is arguably one of the most studied meniscal measures. Meniscal extrusion measured using SQ methods has also been specifically associated with cartilage volume loss longitudinally [382, 388] and is thought to contribute to subchondral bone changes [512].However, this finding suggests it may be a less responsive measure in a cohort selected for clinical trial characteristics. A possible reason for this is that the quantitative measures of meniscal extrusion used were such that they assessed the entire 3D meniscus and are not just confined to single slices, as done in previous studies [343], and it may be that these measures were measuring a somewhat different meniscal construct to that assessed by current SQ measures. The

3D methodology used in this study may also explain why no substantial relationship was found between decreased tibial coverage and increased meniscal extrusion, as has been reported previously in other studies although a trend towards this was observed. Bruns and colleagues in their study using controls from the OAI reported increased meniscal extrusion that was not associated with meniscal coverage, which they postulated could be due to increased bulging of the peripheral meniscal margin and less radial displacement [513]. As previously established, meniscal extrusion is a combined construct of radial displacement and change in meniscal width [503, 514]. Few studies have directly evaluated the internal responsiveness of meniscal pathologies and specifically for extrusion using 3D technology, only one other study reported such indices in their longitudinal analysis. In that study using 3D, similarly Bloecker *et al* also found poor responsiveness for meniscal extrusion (SRM = +0.22) in the central five slices and longitudinal change seen was not statistically significant. However, their measure for extrusion distance across the entire meniscus (including anterior and posterior horns) was significantly different over a 2–year period although responsiveness still poor to moderate (SRM 0.32) [407].

Meniscal extrusion is associated with cartilaginous changes in the knee such as JSN, cartilage loss, chondral lesions and meniscal tears. Extrusion is important in the development of other knee pathologies as it impairs load transmission [515]. Following extrusion there is alteration of meniscal function and load distribution leading to compartmental instability. As a result, forces at the femoral and tibial bone surface increase the susceptibility of subchondral bone to trauma during dynamic movements of the knee. It has been shown that one compensatory response to increased load through the medial compartment, is expansion of the tibial plateau leading to increased bone area, ensuring mechanical load redistribution to enhance the mechanical competence of the bone [197]. In this 12 month cohort, little change in meniscus extrusion was noted. The inclusion criteria for medial progression meant that more extrusion was expected on the medial side than the lateral side; in fact 65% of participants in this study did not demonstrate any extrusion the lateral side. The methodology used to identify the outer limit of the tibia differs from other methods, in that it uses all of the 3D information from the tibia to generate a plane, outside of which is considered extrusion. The plane is constructed using points in the shape model which may fall in areas which become osteophytic, and these may be handled differently in other measurement systems. This is potentially an important difference between this method and other methods which score maximum thickness at one slice in the coronal or sagittal plane, and may explain some of the differences between findings in this study and others. It is certainly possible that the differences found here

are related to the novel measurement approach. 3D extrusion from this study may not be the same construct as the extrusion found in individual slices as per MOAKS scoring.

Exploratory analyses aimed at evaluating if any drivers of change existed based on specific factors did not yield any important results, with suggestions that responsiveness varied by weight status (obese vs non-obese using WHO cut-offs) for total thickness and that of tibial coverage varied by meniscectomy status, although both findings should be interpreted with caution in view of the sample size. Patient size has an effect on the size of the medial plateau, a point highlighted by Stone *et al* [516]. Although hampered by a smaller sample size the head–to-head comparisons between 3D extrusion and an established (validated) score for extrusion on MOAKS does provide some validity for the 3D extrusion measure. The recurring themes such lack of sensitivity of SQ measures were also seen in this study: based on the MOAKS scores only 3 participants showed increased extrusion after 12 months compared to approximately 50% in the same sample that showed extrusion changes greater than measurement error.

The development of disease modifying osteoarthritis (OA) drugs has been a frustrating process, in part due to lack of valid and responsive biomarkers to change[4], creating a vicious cycle where large numbers of people are required for trials resulting in higher costs to pharmaceutical companies who have thus become reluctant to pursue this area [5, 6].To date OA biomarker development has focused mainly on cartilage measures, with cartilage relatively well validated as an OA imaging biomarker[351, 517] while measures reflecting subchondral bone changes have also demonstrated their potential as imaging biomarkers [209, 339, 340]. This study provides preliminary evidence that meniscal measures have the potential for use in clinical trials and should now be investigated for their ability to add discriminatory power in OA progression assessment.

4.6 Limitations

In terms of limitations, it should be noted this work was focussed on a cohort typical of that in clinical trials and does not necessarily reflect the meniscus natural history in a general population. Knees were selected for medial progression only. Like most of the reported MRI meniscal studies, this study used non-weight-bearing images; changes in the meniscus might be more responsive under load. Segmentations were performed on DESS images, which offer the best compromise for identification of multiple OA tissues (here meniscus and bone) but may not be the optimal sequence for detecting particular

meniscal pathologies. OA is a long-term disease, and 12 months is insufficient to study the long-term pathogenesis of menisci in the OA knee, and it would be useful to follow OA knees for a much longer period, especially using shape modelling to quantify any spatial change which occurs, while removing confounding by the pose of the knee.

The repeatability of the method is likely to provide an optimistic assessment of measurement precision, as only healthy menisci were used for the test-retest manual segmentation method due to resource constraints. As this was the first step in developing a new imaging biomarker, and given that manual segmentation is very time-consuming, the study aimed to use healthy menisci that would be reasonably consistent in shape and size for characterising the repeatability of the overall measurement pipeline. This may have overestimated the repeatability of the methods. However, based on this preliminary work it seems likely that in the future meniscal segmentation may be fully automated using statistical shape models. One of the key segmentation aims was to ensure that meniscal shape was identified as accurately as possible to avoid averaging effects.

Another limitation of the study was the inclusion of participants that self-reported having a meniscectomy at baseline. This inclusion may have resulted in a slight bias in the changes seen. However for pragmatic reasons, and to ensure the sample size was not further reduced the participants were kept in the study. Results from a sensitivity analyses however showed that there were no differences between the group that reported a meniscectomy at baseline and those that did not. Also, as the original inclusion criteria was that of participants thought to be “fast progressors” keeping these participants was justifiable. Care was however taken to ensure no participant reporting meniscectomy at 1-year follow-up was included.

As this study was designed as an exploratory study to determine if indeed there were any changes in a 12 month period, a feasible period for a putative clinical trial there was no normative data to initially power the study on and 86 knees were recruited mainly on the basis of convenience. However, a post hoc power calculation suggested the study has 80% power to detect an effect size of 0.31 which represents a small-to-moderate effect. Comparatively, the sample size in this study was similar to previous work using 3D meniscal measures. Also, the study design was primarily to understand where most change occurred, and used responsiveness as a tool to understand the answer to this spatial question.

4.7 Conclusion

Using modern image analysis has provided evidence that the spatial location of meniscal damage in patients at risk of medial progression was predominantly in the posterior sub region of the medial meniscus. Musculoskeletal radiologists have consistently reported changes in the same regions in the meniscus, thereby providing face validity for these novel measures. In this 12 month OA knee cohort, medial tibial coverage and thickness were the most responsive measures of change, with change comparable to other MRI outcomes and better than radiographic JSN. However, as clearly demonstrated in Figure 17, the type of morphological pathology may vary across cohorts. The proposed 3D measures are not suggested to replace SQ meniscus scores as these possibly measure meniscus pathology only indirectly. They do however provide potentially more robust and responsive measurements and in the case of tibial coverage, a measure that current SQ measures do not assess.

Chapter 5

The relationship between two different measures of osteoarthritis bone pathology, bone marrow lesions and 3D bone shape

This chapter presents the results of the investigation into the relationship between 2 potential imaging biomarkers, bone marrow lesions (BMLs) and 3D bone shape. MRI-detected bone pathologies in OA provide promising opportunities as treatment targets and imaging endpoints. While BMLs have been frequently studied, 3D bone shape obtained from supervised machine learning techniques (with statistical shape models) provides a highly accurate, novel measure that is specific for OA and highly responsive over time in OA progression. Given the growing literature on bone shape, it is important to understand the relationship between these 2 bone pathologies, and their relative usefulness (or responsiveness) in clinical trials. *Published in Osteoarthritis & Cartilage,2018*

5.1 Background

Magnetic resonance imaging (MRI) has provided insights into the development of osteoarthritis (OA) and helped demonstrate the importance of subchondral bone pathology [67].Bone is important in OA pathogenesis and biomarker development, and bone marrow lesions (BMLs) are one of the most studied of these bone pathologies [518]. BMLs are high signal MRI lesions that have been associated with other pathologies and symptoms [199, 353, 519], and their predictive validity for OA progression has also been reported [199].

Research on another OA pathological manifestation, change in 3-dimensional bone shape, has emerged. Bone shape which incorporates both spreading of bone and osteophytic changes [111, 209] has shown to be more responsive than current radiographic and standard MRI measures of cartilage for assessing OA progression [339], predictive of incident radiographic OA [111], and associated with joint replacement [340].

While it is appreciated that subchondral bone changes play an important role in OA pathogenesis [518] the relationship between these two measures (BMLs and 3D bone shape) remains poorly studied. It is important to understand if they represent a single

construct or different parts of the OA process, and to further explore their use as imaging biomarkers.

5.2 Aims

The aims of this study were to assess the relationship between BMLs and 3D bone shape in cross-section and over time. Another aim was to compare the responsiveness of both pathologies over 24 months follow-up.

5.3 Methods

5.3.1 Participants

This project utilised a sub-set of 600 OAI participants identified for development of potential biomarkers from the Foundation for the NIH Biomarkers Consortium (FNIH) OA. Clinical and imaging data was available for baseline, 12 months and 24 month time-points in this sub-group. Inclusion and exclusion criteria are described below, (this was broadly participants having KL grade of 1, 2 or 3 at baseline, availability of knee radiographs and MRIs of appropriate quality at baseline and 24 months). Exclusion was failure to meet the radiographic or pain progression due to ceiling effects (minimum JSW less than 1.00 mm and /or the Western Ontario and McMaster Universities Arthritis Index (WOMAC) pain >91 on a 0-100 scale). All 600 participants were used in this study.

5.3.1.1 FNIH nested case-control study

As part of the FNIH, a nested case-control study of knee OA progression was performed,(see Figures 1-3 for the design overview). All 600 participants, one index knee per subject, were selected. Eligible subjects were those with at least one knee with a KL grade of 1, 2 or 3 at baseline and availability at baseline and 24 months of knee radiographs, knee MRI without artefacts that would interfere with image analysis, stored serum and urine specimens and clinical data. Participants were excluded if they had a total knee or hip replacement or metal implants in bone from baseline through 24 months due to potential effects on biochemical markers which would confound analyses.

Eligible knees were classified for radiographic and pain progression from baseline to 24, 36 and 48 months. Knees that were unable to meet criteria for radiographic or pain progression due to ceiling effects at baseline (minimum medial joint space width <1.0mm and / or WOMAC pain >91 on 0-100 scale) were also excluded.

5.3.1.1.1 FNIH definition of radiographic progression

Radiographs were assessed for KL scores and semi-quantitative (SQ) joint space narrowing (JSN) based on the OARSI atlas. The minimum joint space width (minJSW) in the medial femorotibial compartment (MFTC) was measured using automated software as described before. Radiographic progression was defined as a decrease in minJSW of ≥0.7 mm from baseline to 24, 36 or 48 months. This cut-off was determined based on the mean and SD of one year changes in medial minJSW in 90 OAI reference control group knees with a KLG of 0 and WOMAC pain scores of 0 at both time points. Such a decrease of ≥0.7 mm was found to have a 10% probability of being due to measurement error and is consistent with values for the minimum detectable differences in medial minJSW using other methods [520].The ICC for test-retest reliability of change in medial minJSW in OAI knees over 36 months is 0.96.

Knees were excluded if they had poor and/or inconsistent positioning (defined in terms of MFTC tibial plateau rim distance) on knee radiographs at one or more visits that would make measurement of MFTC JSW unreliable. Additionally knees with predominantly lateral compartment JSN at baseline or during follow-up, were also excluded to avoid misclassification on radiographic progression when based only on minJSW changes in the medial compartment.

5.3.1.1.2 FNIH definition of pain progression

Pain assessment was performed using the WOMAC pain subscale. Progression was defined as a persistent increase in pain from baseline to 24, 36 or 48 months of ≥9 points on a 0-100 normalized score, based on the MCID for pain worsening [521, 522]. Pain persistence required a pain increase of ≥9 points at two or more time points from the 24 to 60 month pain assessment, therefore knees were excluded if participants did not have enough follow-up time points after the first increase in WOMAC pain data above the MCID, to determine if this increase was persistent.

5.3.1.1.3 Selection of participants based on case-status

Four knee categories were defined in the FNIH based on the outcome in an index knee (one index knee identified per participant):

Group 1 - knees with both radiographic and pain progression

Group 2 - knees with radiographic but not pain progression

Group 3 - knees with pain but not radiographic progression

Group 4 - knees with neither radiographic nor pain progression

Participants with a knee that already met the criteria for radiographic and pain progression at 12 months were excluded. This was because the primary predictors were biomarker changes during the first 24 months of follow–up and outcomes were defined based on changes at 24 months and after. One knee was selected at random if a participant had 2 knees with the same outcome. On the basis that molecular biomarkers are person-level variables measured in serum or urine, participants were excluded if outcomes were inconsistent between the knees. Furthermore requirements for consistency between outcomes in an index knee and the contralateral knee were as follows:

- If index knee fell into Group 1 the contralateral knee could fall into any of the 4 groups
- If index knee was in group 2 then the contralateral knee could not have pain progression but could have radiographic progression or no progression
- If index knee was in group 3 then contralateral knee could not have radiographic progression but could be in group 3 or no progression group
- For an index knee in group 4 its contralateral knee could not be any group except group 4.

Additionally for outcome groups 3 and 4 when looking at the contralateral knee, radiographic progression included lateral compartment SQ JSN progression, and if a knee did not have the medial compartment JSW data then MFTC SQ JSN data was used. In all progressor groups ,both radiographic and pain progression in the contralateral knee included having a knee replacement at 36 or 48 months.

The initial sample size target for the 4 groups was initially set at 200,100,100, and 200 respectively. However for better covariate balance among the groups, knees selected for the 4 groups were frequency matched for 15 strata of KL grade (1,2 or 3) by BMI category (<25; 25 to <27.5; 27.5 to <30; 30 to <35; ≥35). In the end the achieved sample sizes in the groups were 194,103,103 and 200 respectively. Figures 19-21 adapted from the OAI database available at https://data-archive.nimh.nih.gov/oai/ outline the FNIH study design and participant flow chart.

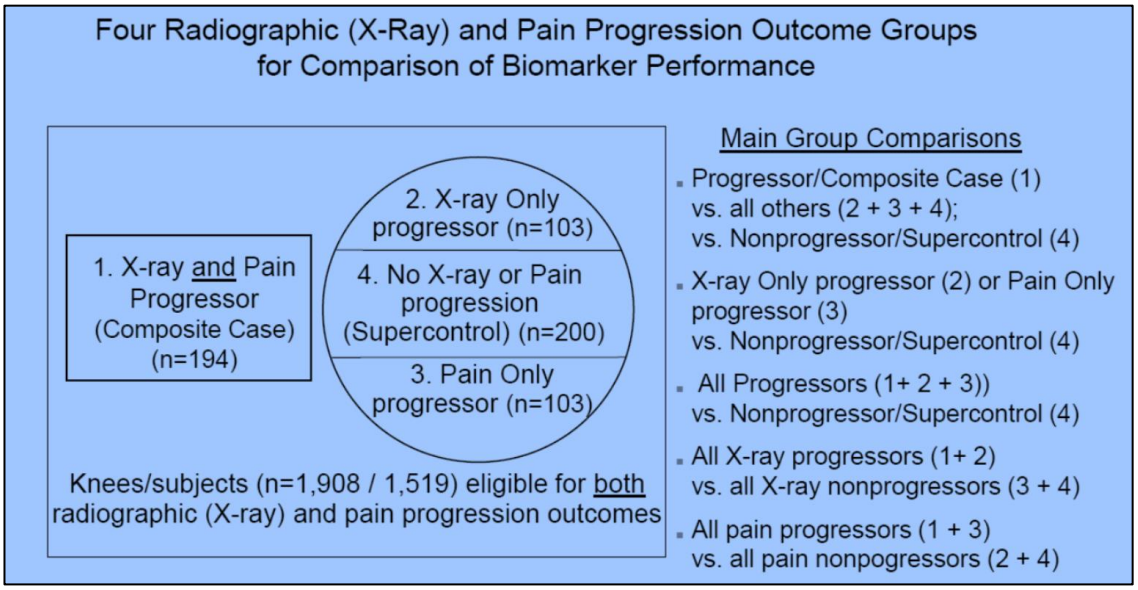


Figure 19 Study analysis plan

Adapted from https://data-archive.nimh.nih.gov/oai/.

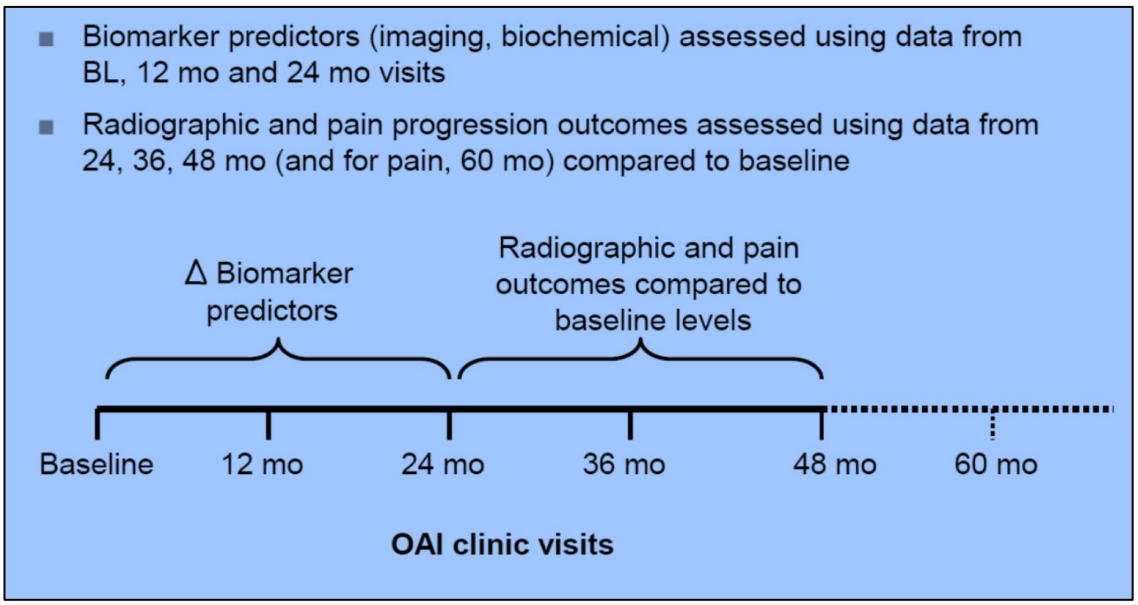


Figure 20 Measurement of key variables in FNIH study

Adapted from https://data-archive.nimh.nih.gov/oai/.

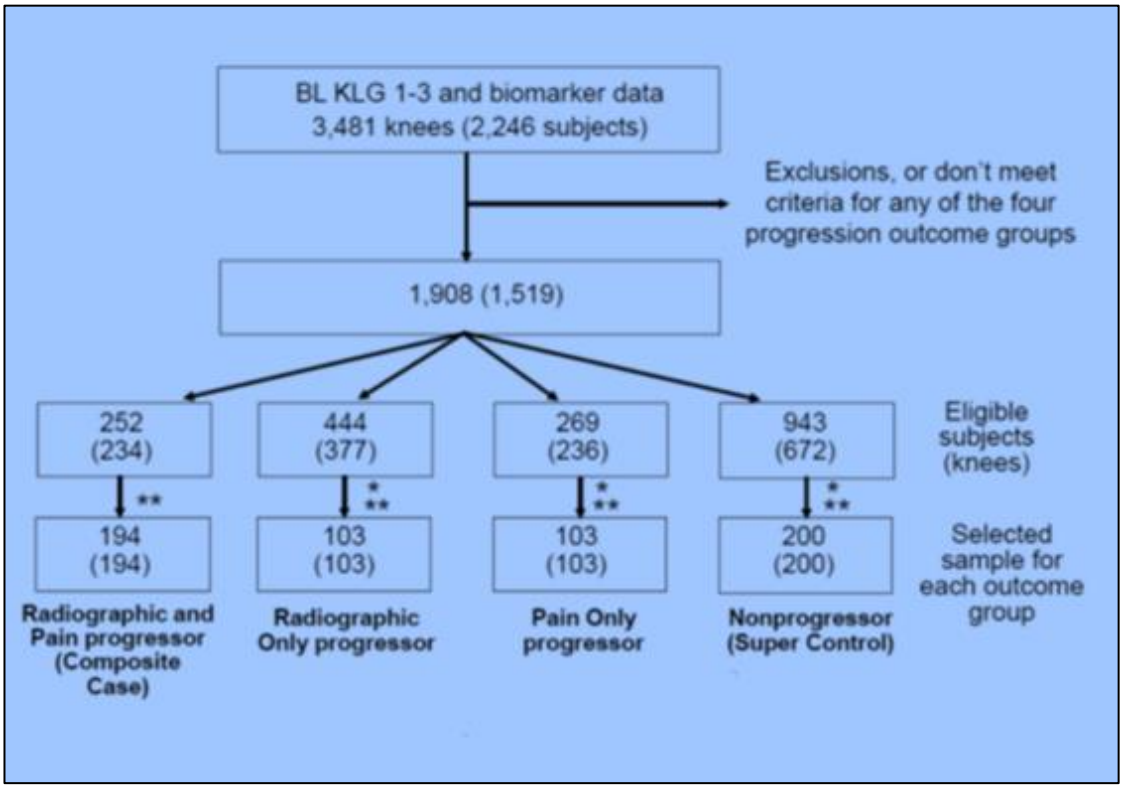


Figure 21 FNIH Participant flow diagram

Adapted from https://data-archive.nimh.nih.gov/oai/.

5.3.2 FNIH data relevant to study

Clinical data for all participants is available under the *AllClinical00* for baseline, *AllClinical01* for year one and *AllClinical03* for year two data sets of the OAI. Imaging data obtained from the FNIH for this current study included longitudinal semi-quantitative BML scores obtained from the Boston Core Imaging Lab (BICL) scored using the MOAKS system and 3D bone shape obtained from Imorphics Ltd. Details of the derivation of bone shape measurements are discussed in Chapters 2 and 3.

5.3.2.1 MRI variables and reading methods

The variables measured in the FNIH include:

•Scores for cartilage morphology (lesion size and depth) in 14 anatomical locations in the knee.

• Scores for the size and number of bone marrow lesions (BMLs) in 15 anatomical locations.

• Scores for osteophyte size in 12 anatomical locations.

• Scores for meniscal damage for anterior horn, body and posterior horn of both medial and lateral menisci, as well as scores for meniscal signal abnormalities, root tears, meniscal hypertrophy and meniscal extrusion.

• A score for synovitis at infra-patellar fat pad and one for synovitis/effusion in the whole knee.

• Scores for cruciate ligament teats (ACL and PCL) and extra articular features (e.g. cysts, bursitis).

5.3.2.2 Image acquisition and scoring

For this study the sagittal and coronal IW TSE sequences, the sagittal 3D DESS WE and the axial and coronal multiplanar reformats of the DESS were used. Data is available for baseline, 12-month and 24-month visit data. The OAI Coordinating Centre blinded the MR images to the OAI Release ID before transferring images to BICL were they were assessed paired and with known chronological order. Cartilage morphology, BMLs, osteophytes, meniscal damage, ACL/PCL tear, synovitis and effusion and extra articular features such as cysts and bursitis were scored.

5.3.3 Anatomical locations for MRI scoring

The scoring for cartilage and BMLs is performed in a number of anatomical locations, for the lateral and medial tibio-femoral compartments five sub regions are employed: in the tibial plateau (anterior; central and posterior) and two for the femoral condyle (central and posterior). Similar locations are used for the medial side.

5.3.3.1 Cartilage and BML locations

Figure 23 shows the anterior (A), central (C) and posterior (P) sub-regions on the lateral side used in scoring cartilage and BMLs as used in the WORMS scoring system.

For MOAKS, the anterior portion of the lateral femoral condyle is considered part of the patella-femoral compartment as too the anterior of the medial femoral condyle therefore the P-F compartment comprises 4 sub regions (2 from the femur and 2 from the patella). For BMLs there is an additional sub-spinous region associated with the insertion of the cruciate ligaments and this region is not associated with either lateral or medial side as highlighted in Figure 23.

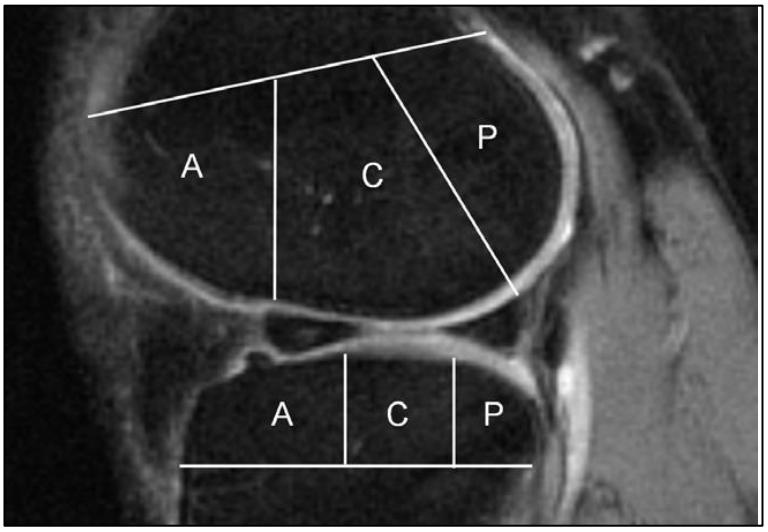


Figure 22 Cartilage and BML scoring regions used in MOAKS

Figure 4 shows the anterior (A), central (C) and posterior (P) sub-regions of the lateral femoral condyle and lateral tibial plateau used in WORMS. There are similar regions defined for the medial side of the knee.

Adapted from [212].

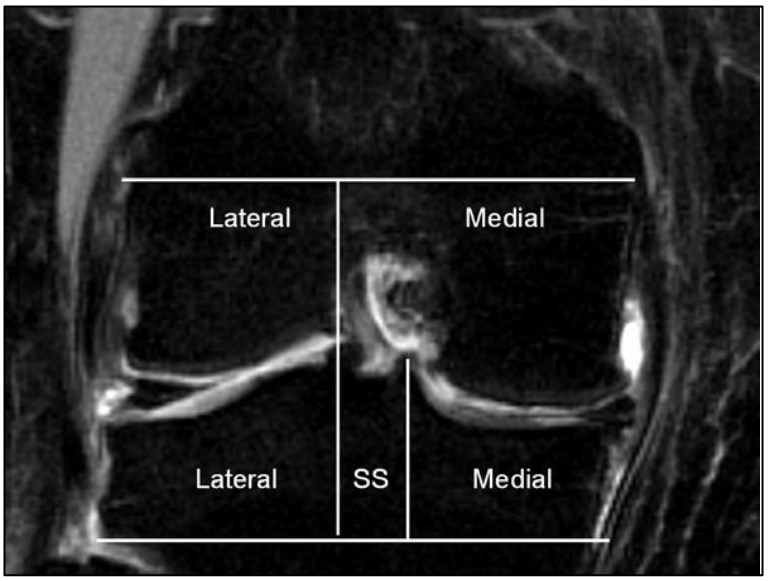


Figure 23 Delineation of medial and lateral sides and defining the sub-spinous region.

Figure 5 Shows the lines delineating medial and lateral sides of the femur and tibia, along with the definition of the sub-spinous region (SS) used only for scoring bone marrow lesions in MOAKS.

Adapted from [212].

5.3.4 Cartilage scoring

The size of any cartilage lesion is scored on a 4 point scale (0-3) based on

(i) the percentage that the lesion affects (size of any cartilage loss (partial or full thickness) as a % of the surface area of the sub-region)

(ii) a separate score for the percentage of the region affected by full thickness cartilage loss ranging from 0-3

where 0 represents none; 1 represents <10% of the surface area or region; 2 represents 10-75% of the surface area of the region and 3 represents > 75% of the surface or the region.

In the FNIH dataset these 2 scores are combined into a single number where the portion before the decimal represents the score for the size of the lesion and these score after the decimal point represents the score for the amount of full thickness cartilage loss. For example, a value of 3.1 lesion represents a large lesion that covers more than 75% of the surface area of the sub-region, but has only a small amount of full thickness cartilage loss covering less than 10% of the surface area of the sub-region.

At follow-up visits only, a special value of 0.5 is assigned to reflect that although the score is the same as at the previous visit, a definite worsening has occurred (within-grade worsening). A special value of -0.5 is used to record when a within-grade improvement has occurred.

5.3.5 BML scoring

For each sub-region there are 3 scores, one for the percentage of the volume of the sub-region affected by BML, one for the number of BMLs within the sub-region and a 3rd score for the percentage of the lesion that is a BML as opposed to a cyst (Table 14).

Table 14 BML scoring

Size of BML (0-3) (including any associated cysts)	**Number of BMLs counted within the sub-region (0-2)**	**% of lesion that is BML (vs cyst) (0-3)**
0: none	0: no BMLs in subregion	0: none
1: < 33% of sub-regional volume	1: a single BML in the subregion	1: < 33%
2: 33-66% of sub-regional volume	2: a pair of BMLs in the subregion	2: 33-66%
3: >66% of sub-regional volume		3: > 66%

5.3.6 Osteophyte scoring

In the P-F joint, osteophytes were scored at 4 locations on the patella (superior, inferior, medial and lateral) and also 2 locations on the anterior portion of the femur (medial and lateral). For the medial T-F joint osteophyte size around the medial tibial plateau was scored as well as the size at 2 locations (central and posterior). For the lateral T-F joint the same 3 locations were scored. Osteophytes are scored on a 4 point scale: Grade 0 = none, Grade 1 = small, Grade 2 = medium, Grade 3 = large.

5.3.7 Scoring meniscal damage

Each meniscus, lateral and medial was split into 3 sub-regions namely anterior horn, meniscal body and posterior horn. The presence and type of tear was scored separately for each of those regions. Meniscal extrusion of each meniscus was scored (in the medial-lateral direction) and anterior extrusion of the lateral horn was also scored.

0: normal meniscus

1: signal abnormality that is not severe enough to be considered a meniscal tear

2: radial tear

3: horizontal tear

4: vertical tear

5: complex tear

6: partial maceration

7: progressive partial maceration (only used for follow-up visit scores)

8: complete maceration

The presence of meniscal hypertrophy, meniscal extrusion or meniscal cysts is also recorded.

5.3.8 Scoring of synovitis and effusion

Synovitis was scored in the infra-patellar pad based on signal abnormalities in Hoffa`s fat pad. The presence and size of the synovial effusion was scored. This study used non-enhanced MRI sequences therefore the effusion score can include both effusion and synovitis as these cannot be differentiated using these methods.

5.3.9 Computation of BML scores for this study

The aim of the study was to compare BMLs and bone shape but for consistency these had to be for corresponding regions as closely as possible. The BML scores were thus scored separately for the femur and tibia.

Four BML scores were computed

1. “BML total size” score (computed separately for the femur and tibia by summing the BML size scores in those regions, which combined the 6 sub-regions in the femur for a total possible score =18 and similarly for the tibia). These regions in the femur were the femoral condyle central region; femoral condyle posterior region and femur anterior region on both the lateral and medial sides. For the tibia these were tibia anterior region, tibia central region and tibia posterior region and similar to the femur, both lateral and medial sides.
2. “BML total number” score (by summing the number of BMLs in each sub-region) as defined above.
3. “BML maximum size” score by taking the highest grade across the femur regions described before (ranging 0-3) and similarly for the tibia
4. “BML total sub-regions” which was calculated by summing the total number of sub-regions within the femur/tibia affected by any BML (ranging from 0-6) and like before treating each region separately.

5.3.9.1 Defining bone shape healthy limits

To provide a meaningful interpretation for bone shape values, bone shape "healthy limits", were defined as the upper 95th percentile of normal knees (bone shape ≥0.96 on vector scale). The smallest detectable difference (SDD) for bone shape was determined from a reliability study using an independent sample of 885 OAI participants from the OAI that had KL zero for 4 years consecutively in both knees, and this was set at 0.24 bone shape units for the femur shape vector and 0.59 bone shape units for the tibia.

5.3.10 Statistical analysis

Mixed statistical methods were performed.

i. The baseline correlation between BML measures and bone shape was assessed using Spearman's correlation coefficients.

ii. The proportions of participants with shape vector scores outside "healthy limits", were compared between participants having BMLs and those without descriptively.

iii. Linear regression analysis was performed to evaluate associations between bone shape and presence of BMLs at baseline (binary outcome), adjusting for age, sex, body mass index (BMI), physical activity score (PASE) and Kellgren Lawrence (KL) score. Presence of a BML was defined as having any BML total size score of at least 1 in the corresponding region.

iv. The incidence of BMLs at 24-month follow-up was reported descriptively and compared to changes in bone shape (for changes greater than the smallest detectable difference (SDD) as defined previously. Incident cases were defined as any knee that was previously scored zero for total BML size and then scored one or above at subsequent visits, separately for femur and tibia.

v. The longitudinal relationship between change in femur shape and change in total BML size was assessed using multilevel linear models, incorporating the effect of time (years 0, 1, and 2) while adjusting for covariates as before. Initially unconditional growth models were assessed for the 3D bone and BML measures, with the intercepts and the effects of time specified as random effects thus allowing them to vary across individuals. The effect of baseline variables (total baseline BML size at baseline) in predicting change in bone shape was modelled by fitting models with baseline total BML size and an interaction term (the product of baseline total BML size x time) to the unconditional growth model for 3D shape. The interaction term tested whether BMLs predicted change in 3D shape over time. Lastly, BMLs were modelled as

time-varying predictors by fitting a multilevel model that included both time and BMLs as independent variables and adjusted for covariates as before. The level of significance was set at $p<0.05$.

vi. Group-level internal responsiveness was assessed using standardised response means (SRMs), and to aid comparison with previous FNIH studies the “maximum BML size” score and “total BML sub-regions” were included in these analyses. SRMs were analysed within each outcome group, since expected changes were assumed homogenous within these groups. Responsiveness was also explored by KL grade.

5.3.11 Results

5.3.11.1 Baseline findings

The mean (SD) age was 61.5 (8.88) years, 59% female with mean (SD) BMI of 30.7 (4.78) and median (IQR) PASE score of 154.5 (102-214). Based on these characteristics this cohort was similar to that of the full OAI cohort. Prevalence of BMLs in the femur was 71% and 41% for the tibia at baseline (Table 15). The distribution of bone shape measures based on visual inspection of the histograms, seemed reasonably normal with no evidence of departure from normality (Figure 24 and Figure 25).

At baseline, 26% participants had femur bone shape scores outside healthy limits with similar proportions seen in the tibia (24%).The distribution of bone shape vector scores was associated with the OA progression groups from the pre-defined FNIH case-control study. For instance the radiographic and pain progression group (Group 1) had on average a more positive femur shape, representative of a more “OA like” shape while controls had on average relatively more negative bone shape scores representative of less OA severity. Bone shape vector scores outside healthy limits were more likely in participants that had BMLs compared to those with no BMLs (31% vs 14% respectively, chi-square (1DF) = 17.50, $p<0.001$).

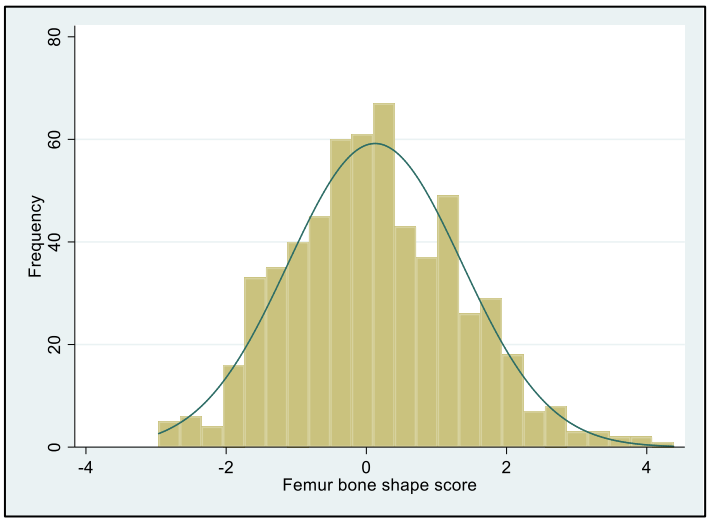


Figure 24 Femur bone score distribution overlaid with normal distribution curve

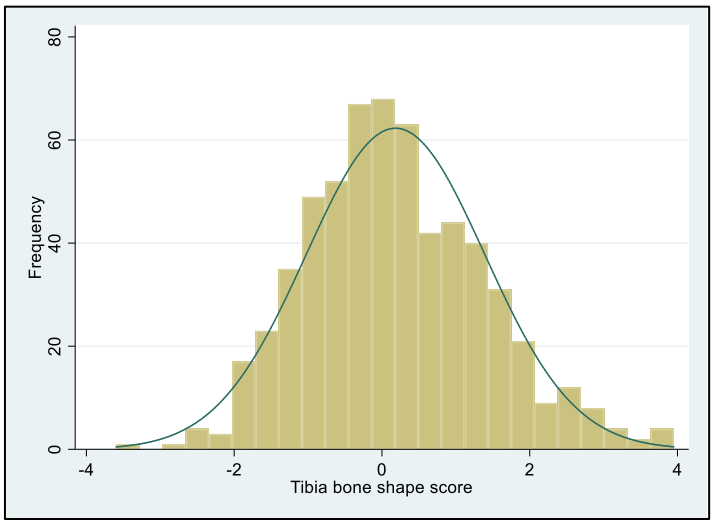


Figure 25 Tibia bone score distribution overlaid with normal distribution curve

Table 15 Clinical and radiographic features at baseline

	Radiographic and pain progression N=194	Radiographic progression only N=103	Pain progression only N=103	No progression N=200
Baseline findings				
Age, mean (SD)	62.0 (8.8)	63.1 (8.3)	59.2 (9.1)	61.5 (9.1)
Sex, female n, (%)	110 (57)	46 (45)	67 (65)	130 (65)
BMI, mean(SD)	30.7 (4.8)	30.7 (4.7)	31.1 (5.0)	30.5 (4.8)
Physical activity Scale for the Elderly, median (IQR)	148.5 (102-202)	176.5 (114-246)	156.0 (115-235)	150.0 (89-208)
KL grade n, (%)				
1	24 (12.4)	14 (13.6)	13 (12.6)	24 (12)
2	84 (43.4)	47 (45.6)	61 (59.2)	114 (57)
3	86 (44.3)	42 (40.8)	29 (28.2)	62 (31)
Femur shape vector	+0.35 (1.29)	+0.31 (1.21)	-0.04 (1.08)	- 0.11 (1.23)
Tibia shape vector	+0.34 (1.25)	+0.35 (1.26)	+0.03 (1.11)	+0.02 (1.17)
Femur BML, present, n (%)	155 (80)	79 (77)	65 (63)	124 (62)
Tibia BML, present, n (%)	103 (53)	53 (51)	32 (31)	60 (30)
Patella BML, present, n (%)	143 (74)	65 (63)	68 (66)	141 (71)

Moderate positive correlation was seen between femur shape and femur BML total size, r (598) = 0.31, p<0.001 while a small positive correlation was seen for the tibia, r (598) =0.16, p<0.001. Analyses with total BML numbers revealed similar associations (r = 0.30 for femur and r = 0.14 for tibia, all p<0.001).These relationships were also represented using correlation plots (scatterplots and dot-plots) as shown by Figure 26 and Figure 27, also highlighting modest associations for each comparison.

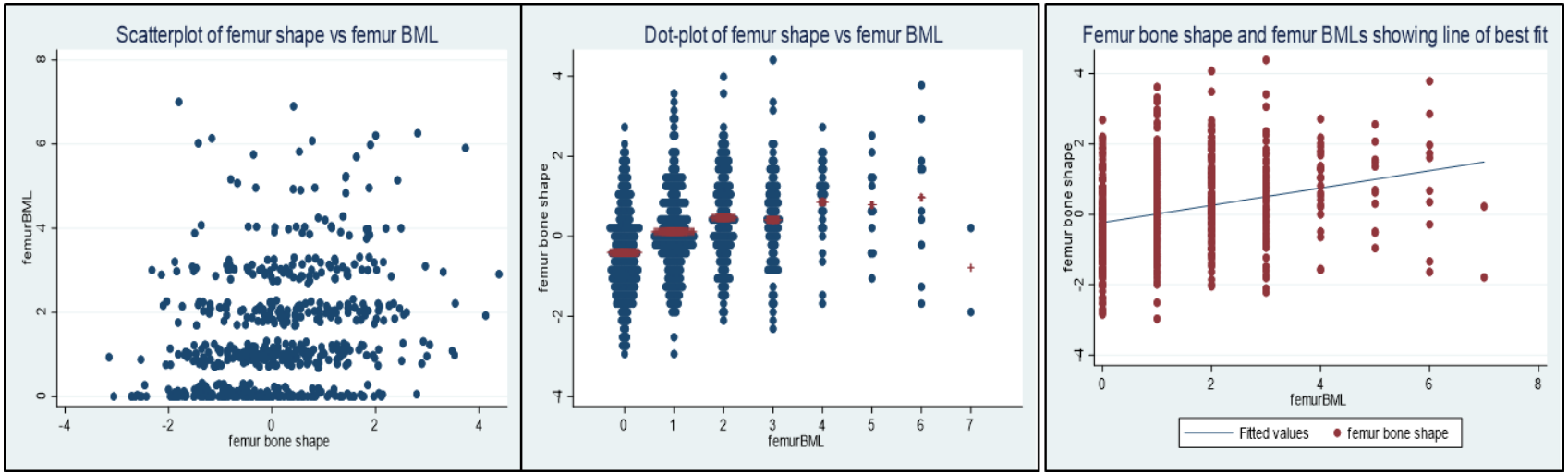


Figure 26 Plots comparing femur bone shape and femur BMLs

*red horizontal bars represent means

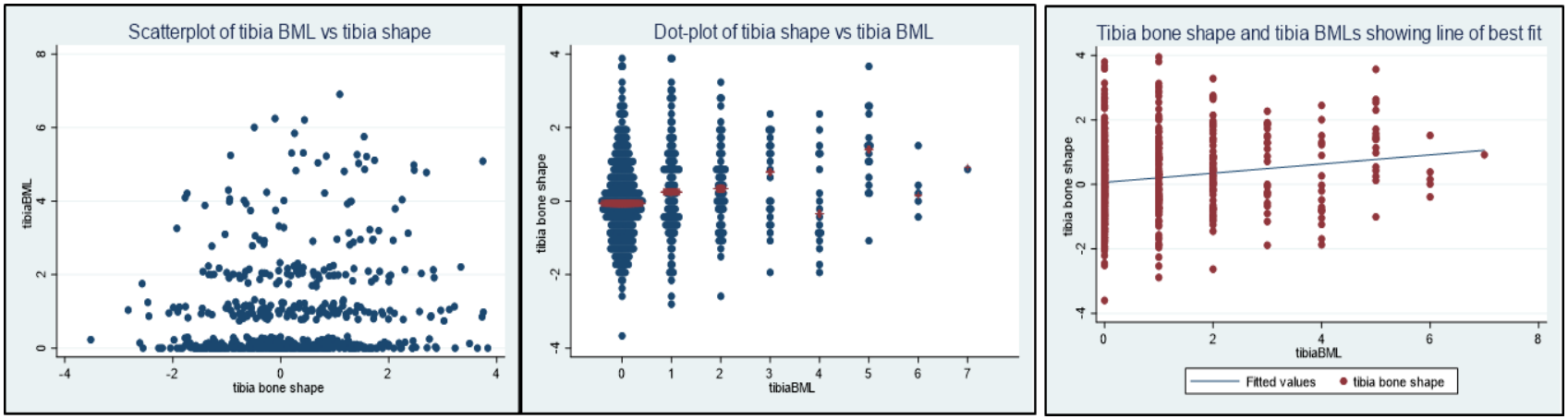


Figure 27 Plots comparing tibia bone shape and tibia BMLs

*red horizontal bars represent means

5.3.11.2 Linear regression

Univariable models showed statistically significant associations for both the femur and tibia bone shape modelled independently with their corresponding BML size scores (tibia beta coefficient = 0.57, 95% CI 0.38, 0.77). The association between bone shape scores and KL grade were also in the expected direction and statistically significant in both bones, compared to a reference group for KL grade 1, higher KL grades had more positive bone shape scores (representing worsening) (Table 16).

After adjustment only the femur showed a statistically significant association between presence of a femur BML at baseline and 3D femur shape (adjusted beta coefficient 0.49, 95% CI 0.30, 0.68) indicating a more positive femur vector (indicative of "increased OA") in individuals with BMLs at baseline, with a difference equivalent to 0.5 x SD of non-OA knees. The effects of adjusted covariates were marginal and similar in both bones (Table 16).

Model diagnostics were performed graphically and the residuals (see Figure 28 below for femur) seemed reasonably normally distributed, hence it could be assumed there were no departures from normality.

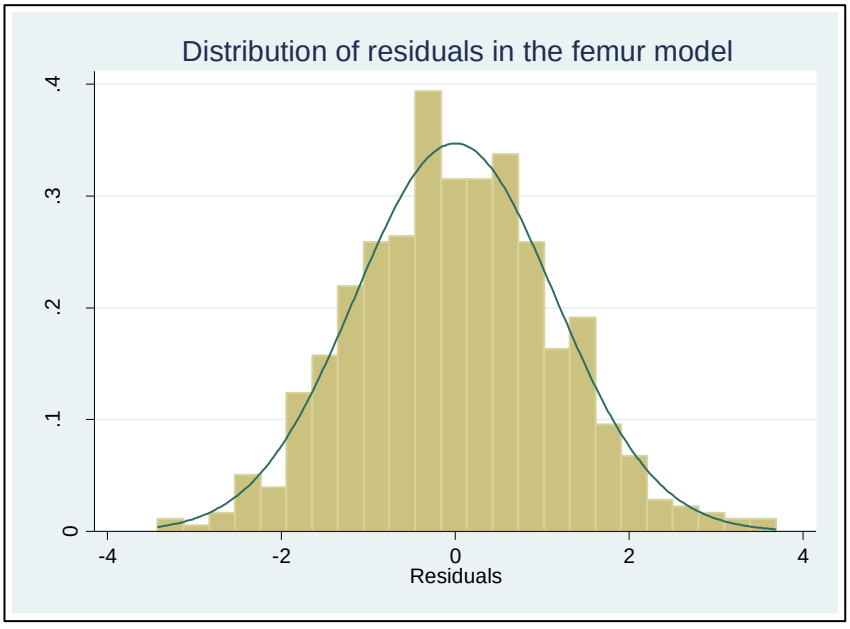


Figure 28 Residual plot from femur linear model

Table 16 Association between bone shape and BMLs at baseline

Univariable models	Coefficient (95% CI)	p-value	Multivariable models	Coefficient (95% CI)	p-value
Cross-sectional models					
Femur BML (present)	0.75 (0.54,0.96)	<0.001*	Femur BML (present)	0.49 (0.30,0.68)	0.03*
			PASE (square root)	-0.02 (-0.05,0.01)	0.10
KL grade (ref=KL1)			Age	-0.01 (-0.01,0.01)	0.83
Grade 2	0.50 (0.21,0.79)	0.001*	BMI	0.03 (0.01,0.05)	<0.001*
Grade 3	1.30 (0.99,1.60)	<0.001*	Gender (ref= female)	-0.97 (-1.15,-0.80)	<0.001*
			KL grade (ref= KL1)		
			KL grade 2	0.35 (0.08,0.61)	0.01*
			KL grade 3	0.94 (0.66,1.22)	<0.001*
Tibia BML (present)	0.57 (0.38,0.77)	<0.001*	Tibia vector	0.07 (-0.13,0.27)	0.50
			PASE (square root)	-0.01 (-0.04,0.02)	0.38
KL grade (ref=KL1)			Age	-0.01 (-0.01,0.01)	0.86
Grade 2	0.67 (0.39,0.95)	<0.001*	BMI	0.02 (-0.01,0.04)	0.08
Grade 3	1.36 (1.07,1.66)	<0.001*	Gender (ref=female)	0.20 (0.01,0.39)	0.04*
			KL grade (ref= KL1)		
			KL grade 2	0.62 (0.33,0.90)	<0.001*
			KL grade 3	1.33 (1.02,1.65)	<0.001*

*statistically significant

5.3.11.3 Incident BML findings

Over the 2-year follow up period a total of 53 “incident” cases of femur BMLs were seen and 70 in the tibia (baseline and year one 29 “incident” BML cases were recorded and 24 between year one and year 2). In the tibia BMLs incident cases were N=30 between baseline and year one and N = 40 between year one and year two.

At 24 month follow-up,124 (21%) participants had not developed any femur BMLs at all 3 time points while for the tibia this was 282 (47%) participants. When both bones were considered, 85 (14%) participants did not develop either a tibia or femur BML at any time point. There was evidence of BML size fluctuation over time. While in most participants (40%) femur BML size remained the, 38% showed improvement and 22% worsening. In the tibia these findings were similar proportionally more remained the same: 60% remained the same and those that improved or worsened were very similar (19% and 21%) respectively (Table 17).

Of the incident femur BMLs, 21(40%) showed bone shape changes greater than SDD compared to 211/547 (39%) in participants with no incident BMLs. In the tibia 49% of those with incident tibia BMLs showed changes greater than SDD for the tibia bone while 29% with no incident BMLs showing such bone changes.

Table 17 BML size status between baseline and year 2

Measure at baseline	Worsened	Remained the same	Improved
Femur BML size			
0 (n=177)	44 (25)	133 (75)	-
1 (n=180)	46 (26)	66 (36)	68 (38)
2 (n=119)	29 (25)	24 (20)	66 (55)
3 (n=75)	11 (14)	13 (18)	51 (68)
4 (n=26)	3 (12)	0 (0)	23 (88)
5+(n=23)	1 (4)	2 (9)	20 (87)
Overall BML size	134/600 (22)	238/600 (40)	228/600 (38)
Tibia BML size			
0 (n=352)	57 (16)	295 (84)	-
1 (n=120)	34 (30)	43 (35)	43 (35)
2 (n=66)	15 (23)	11 (17)	40 (60)
3 (n=22)	4 (18)	6 (27)	12 (55)
4 (n=18)	3 (17)	3 (17)	12 (66)
5+(n=22)	3 (14)	4 (18)	15 (68)
Overall tibia size	116/600 (19)	362/600 (60)	122/600 (21)

5.3.11.4 Longitudinal associations

Univariable multilevel models (Table 18) revealed that bone shape vectors tended to start off with a “mild OA state” at baseline (Intercept (S.E) for femur = 0.12 (0.05) and for tibia 0.18 (0.05). Over time these vectors became more positive (indicating worsening), slope = 0.11 for the femur and slope = 0.12 for tibia (both statistically significant, p=0.02 and p<0.001 respectively).

Figures 29 and 30 show a randomly selected 20 participants and how their femur bone shape values changed over 3 time points. The change was fairly linear and a better fit for bone shape than BML size, higher order models such as quadratic models could not be tested as there was only 3 measurement occasions in the FNIH.

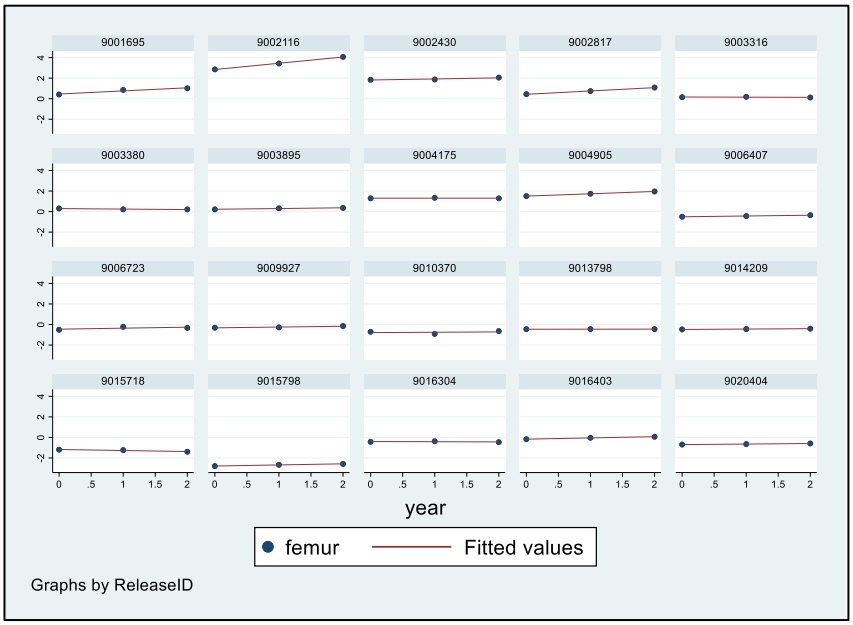


Figure 29 Femur bone shape change over time

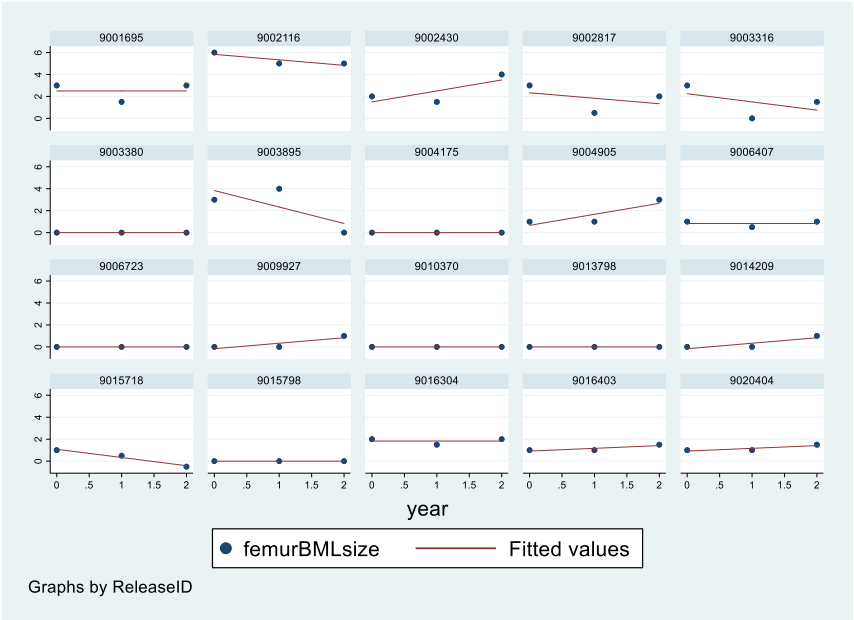


Figure 30 Femur BML size change over time

In BMLs however, there were differences noted: although both intercepts were significant ($p<0.001$) only the femur BML slope was significantly different from zero (slope = -0.11) and this represented a decrease over time signifying a general trend towards improvement, while the tibia BML slope was not different from zero (Table 18).

When modelled simultaneously to include the effect of an interaction with time, an increase in total BML size over time was associated with increase or worsening of the shape vector over time [beta coefficient =0.011, $p<0.001$]. Similar baseline effects were seen in the tibia, increased baseline tibia total BML size was related to more positive (more OA-like) tibia shape [beta coefficient = 0.15, $p<0.001$]. However there was no statistically significant associations between tibia BML and tibia bone over time although the trend suggested increased tibia BML size was associated with a more positive tibia shape vector (indicative of “worsening OA” state).

Table 18 Multilevel modelling of bone shape and BMLs

Univariable models	Estimate (std. error)	p-value	Multivariable models	Estimate (std. error)	p-value
Unconditional models			**Multivariate models**		
Femur vector intercept	0.12 (0.05)	<0.001*	Femur baseline BML	0.24 (0.03)	<0.001*
Femur slope	0.11 (0.01)	0.02*	Femur BML slope	0.01 (0.002)	0.007*
Femur BML intercept	1.37 (0.06)	<0.001*			
Femur BML slope	-0.11 (0.03)	<0.001*			
Tibia vector intercept	0.18 (0.05)	<0.001*	Tibia baseline BML	0.15 (0.04)	<0.001*
Tibia slope	0.12 (0.01)	<0.001*	Tibia BML slope	0.01 (0.01)	0.43
Tibia BML intercept	0.77 (0.05)	<0.001*			
Tibia BML slope	0.04 (0.03)	0.13			

5.3.11.5 Responsiveness

Bone shape was more responsive than both SQ total BML size and total BML number scores in all regions over 2 years [femur shape (SRM = 0.89, 95% CI 0.72,1.02) vs femur total BML size (SRM -0.13, 95% CI -0.26,0.02)]. Similar results were found when responsiveness was compared by KL grade with bone shape being more superior (Table 19). Bone shape responsiveness was consistently highest in the groups that showed combined radiographic and pain progression and radiographic only progression groups.

When compared to previously published BML summed scores (maximum BML size score and the total BML sub-regions), bone shape was also shown to be much more responsive.

Table 19 Responsiveness of bone shape and BML measures over 24 months

	Radiographic and pain progression N=194	Radiographic progression only N=103	Pain progression only N=103	No progression N=200
2 year responsiveness, SRM (95%CI)				
Bone shape				
Femur	0.89 (0.72,1.02)	1.02 (0.85,1.20)	0.46 (0.31,0.61)	0.61 (0.49,0.72)
Tibia	0.84 (0.70,0.97)	0.76 (0.56,0.96)	0.26 (0.07,0.43)	0.47 (0.33,0.69)
BMLs				
Femur BML total size	-0.13 (-0.26,0.02)	-0.15 (-0.35,0.07)	-0.31 (-0.51,-0.13)	-0.24 (-0.37,0.13)
Femur BML total number	0.38 (0.26,0.50)	0.20 (-0.02,0.41)	0.17 (-0.04,0.35)	0.27 (0.14,0.38)
Tibia BML total size	0.11 (-0.02,0.26)	0.14 (-0.04,0.31)	-0.04 (-0.23,0.15)	-0.01(-0.16,0.11)
Tibia BML total number	0.37 (0.24,0.51)	0.31 (0.12,0.51)	0.19 (0.00,0.33)	0.16 (0.02,0.29)
Femur BML maximum size	-0.05 (-0.19,-0.09)	-0.01 (-0.19,0.21)	-0.20 (-0.38,0.01)	-0.11 (-0.25,0.02)
BML total sub regions	-0.02 (-0.15,0.13)	-0.03 (-0.23,0.15)	-0.13 (-0.32,0.07)	-0.06 (-0.20,0.08)

5.4 Discussion

This is the first study to examine the relationship between 3D bone shape and a relatively well studied bone pathology, BMLs. The study investigated their inter-relationship and relative responsiveness as imaging biomarkers. A moderate positive correlation was shown between bone shape and total BML scores in cross-sectional analyses and also that femoral and tibial bone shape were associated with prevalence of BMLs. Longitudinal analyses also showed that change in bone shape was associated with change in BMLs. These findings therefore provided construct validity for 3D bone shape as a potential biomarker. This is plausible since worsening OA status (as measured using bone vector) has been shown to relate to incident radiographic OA and its progression [67, 339] and BMLs have been associated with OA prevalence and progression [199, 519]. 3D femoral bone vector is also independently associated with incident radiographic OA and total knee replacement [67, 340]. The relative importance of the femur (over the tibia) may be explained by its larger articulating surface area compared to the tibia.

A recent systematic review [67] concluded that subchondral bone features (mainly BMLs), and also bone shape were independently associated with clinical features such as pain and joint replacement [111, 340]. While it is appreciated that BMLs are associated with symptoms in most studies, these studies have nonetheless reported effect sizes that are bound by wide 95% CIs with the lower limits of CIs close to 1.0 for ORs and 0 for beta regression coefficients [67] which represents weak effect sizes. These findings were detailed in Chapter 2. Another SLR also found that knee pain was associated with BMLs but suggested that these associations may need further exploration as the level of evidence was judged to be moderate and the lower bound of the effect size (Odds ratios in this case) was not reported and also that one quality study found no association with pain [190]. There have been far fewer studies exploring bone shape as it is a relatively new measure and consequently more evidence exists for BMLs. Recently Hunter and colleagues in a longitudinal study demonstrated modest associations between changes in bone shape and pain progression [209].

Bone shape change and BMLs could possibly represent different pathological processes. 3D bone shape, which is not easily appreciated by a reader, incorporates both spreading of bone and osteophytic changes [111, 209]. BMLs are high signal MRI lesions that represent areas of trabecular remodelling, fibrosis and necrosis and are

strongly associated with adjacent cartilage loss; they may represent a larger 'field of change' in bone [523]. BMLs are part of a dynamic process and not a constant finding which probably represents just a snapshot in time, coupled with their variability such as regressing over time (discussed later) make it a challenge for clinical trials that may want to use them as part of a clinical inclusion or even outcome measure. In terms of their pathogenic relationship within this 2 year study, bone vector changes (beyond that of non-OA knees) were seen in individuals with no incident BMLs at follow-up, suggesting bone shape change may precede BML formation [518].

Clinicians have long appreciated that distinctive changes in bone accompany OA, but only recently have the tools been available to physically quantify the changes. Over the last few years the emergence of automated SSM technology has enabled accurate and reliable quantification of tissue structures and the availability of large datasets to test these has also helped. Subchondral bone structure changes markedly during OA, and it has long been assumed that this occurs secondary to cartilage degeneration. However, for various conditions that are associated with OA, it is also known that bone structural changes occur in the absence of cartilage degeneration. Specific pathologic changes, such as osteophytes, a definitive sign of radiographic OA, are a clear indication that bone changes occur in early OA. Clinicians are starting to appreciate that radiographs do not meaningfully convey the 3D-structure of bone and its associated changes and acknowledge bone changes in OA.

Longitudinal analyses from this study confirmed that modest associations exist between bone shape and BMLs, adding to the earlier cross-sectional findings. In terms of their relative use as imaging biomarkers in OA clinical trials, this study demonstrated that bone shape is a much more responsive measure than SQ assessment of BMLs. There is limited literature comparing OA imaging biomarkers to date. Using the same FNIH cohort, Hunter *et al.* showed that bone shape was associated with radiographic and pain progression longitudinally [209], while imaging biomarkers of bone (including bone shape and BML measures used in this study) were only weakly associated with OA biochemical biomarkers; however bone shape and BMLs were not directly compared [524]. In the same cohort, Collins *et al.* used SQ imaging biomarkers of OA progression to explore the effect of a combination of joint structures on OA progression and reported that changes in BMLs were not significant predictors of progression in models that already included cartilage, meniscus, and effusion markers [215].In this study even after adjusting for KL grade there was a still an association between BMLs and bone shape. Bone has previously been shown to be more responsive than radiographic measures of progression such as JSW and also MRI-derived cartilage

thickness measures [339] and other studies with similar follow-up duration reported low responsiveness for BMLs [525].

While this study found that bone shape in general follows a linear trajectory of change over time, the evolution of BMLs in OA can be very variable and this study arguably showed a better linear change in bone shape than BMLs; some of which may be related to the differences in measurement. Different studies have shown that BMLs can fluctuate in size or regress, one study found that subchondral lesions including BMLs regressed completely at 30-month follow-up [526] while another suggested BMLs are unlikely to resolve over time and in fact often get larger [527]. The findings for tibia BMLs were consistent with those from a Dutch study that showed 20% of BMLs decreased or resolved after 2 years [201] without treatment although a higher proportion (38%) reduced in the femur.

Apart from the studies highlighted above there have been few studies assessing both bone features in a head-to-head comparison, or including both in the same analysis model. Using the full OAI cohort, the Leeds OA research group found that bone shape was predictive of both prevalent and incident knee symptoms in adjusted models that did not include BMLs, and were stratified by gender. However, after adjusting for BMLs these associations were attenuated and were not statistically significant. It should be noted that in the analyses adjusting for BMLs, the sample size in the male model for example was 111 individuals compared to 1692 individuals in the previously fully adjusted pain models without BMLs. The reason for this sharp decrease in sample size was due to limited availability of MOAKS scores in the OAI. When the same analyses were repeated in a sub-sample of individuals thought to have pre-radiographic OA features defined as KL grade of zero in both knees, no association between bone shape and knee pain was found [210]. Unsurprisingly, BMLs were associated with pain in these analyses as has already been highlighted previously. Sharma and colleagues using the same sample found associations between BMLs and pain [204].

It would have been interesting to know how the patella vector relates to patella BMLs. However, in this study these comparisons were not performed due to the way patella BMLs are scored using the MOAKS scoring system. In the MOAKS the anterior (or trochlear portion) of the lateral femoral condyle is considered part of the patello-femoral compartment since it articulates with the lateral facet of the patella and similarly the anterior of the medial femoral condyle, which articulates with the medial facet of the patella is part of the patellofemoral compartment. Therefore the patella-femoral compartment comprises 4 anatomical sub-regions, 2 from the femur and 2 from the

patella. As a result it would not be possible to perform direct comparisons between patella vector and patella BMLs as these were not derived from comparable regions.

5.5 Limitations

There are limitations to this study. Firstly in trying to understand the temporal nature of different bone pathological changes, participants were only followed over a 2 year period. Secondly the study used selected participant data, chosen for the presence or absence of structural/pain progression and may not represent a broader population sample. Definition of change in SQ measures was challenging due to the use of various BML score combinations and there are drawbacks with use of SQ measures, such as comparing a summed score for BMLs when only one of six sub-regions scores the maximum and the other five score zero. Also, BMLs fluctuate in size over time which is likely to reduce their responsiveness. The weak relationship seen between BMLs and 3D bone shape could be due to the fact that the precise location of BMLs is unknown. BMLs on MRI are typically subchondral in location. However, a proportion may occur in the central region of the knee and are related to knee ligament attachments. Some of these ligamentous BMLs may also include a cyst-like component. It is possible that both these BML types have no relationship with subchondral bone shape. Lastly, with only three time points only linear models could be tested and higher order polynomial terms such as quadratic or cubic models could not be incorporated. However model fit in this study was satisfactory.

5.6 Conclusions

This study has provided construct validity for bone shape, provided preliminary evidence for the temporal order of MRI-detected OA bone pathologies and demonstrated the better responsiveness of 3D bone shape over semi-quantitatively assessed BMLs over time periods typical of a clinical trial.

Bone shape has shown its potential as a biomarker but its natural history over time is still unknown. It would be useful to establish how this changes longitudinally and also assess what clinical factors influence this. Thus far, the most studied bone shape measure has been femur bone and this study included the tibia, providing preliminary evidence that these pathologies are related in terms of “structural disease” starting point and rates of change, and both had more superior responsiveness compared to BMLs. However, with only three time points the natural history of bone shape was not comprehensively tested and lacked data on patella bone shape.

Chapters 6 and 7 using the entire OAI participants and 8-year follow-up will be useful to assess longitudinal change. This will also use advanced statistical models that adequately account for the correlation of bone shape measures over time and also capable of assessing all three bone shapes simultaneously in the same model.

Chapter 6

Determinants of osteoarthritis 3D bone shape and its change in the three knee bones: a latent growth modelling approach on 37,583 MR images from the Osteoarthritis Initiative

This chapter presents the results of the investigation into the longitudinal changes seen for the three knee bones, femur, tibia and patella using latent growth modelling analysis, a form of structural equation modelling. Published in Osteoarthritis & Cartilage, 2019.

6.1 Introduction

Despite the growing literature on bone shape particularly for the femur, its natural history over time is still unknown. Femur 3D bone shape provides a responsive biomarker of knee OA, but it is unclear whether this is only femur-specific. Chapter 5 provided further evidence that femur bone shape was more responsive than tibia in a group selected for biomarker development in a 2-year study. It is important to understand the longitudinal relationship between the three knee bones, understand what factors determine this, and to understand which bone might provide the most responsive measure of change in OA clinical trials. It is already well appreciated from MRI studies that OA is a whole-joint disease and a whole joint approach is important in understanding the complex OA pathogenesis [8, 21], hence the need to assess all three bones forming the knee joint.

Evidence for the utility of bone as an knee imaging biomarker continues to grow [67, 111, 209, 339, 340, 393, 405], with the femur being most studied. Femur bone shape is thought to progress linearly towards an “OA” like shape over the disease course but this has not been formally assessed and neither the longitudinal relationships between individual knee bones been established. Limited data so far suggested the tibia and patella bones showed limited predictive association with knee replacement (where there were small numbers of replacements) [340], and while tibia bone shape has been reported as less responsive than the femur, this was in a very small cohort [405].

The Osteoarthritis Initiative (OAI) provides a sufficiently large dataset with long duration of follow-up to assess if there is a single disease process in the OA knee, and to

understand which bone might provide the most responsive measure of change in a longitudinal setting.

6.2 Aims

This study aimed to examine the association between the 3 knee bones (femur, tibia and patella) at baseline, to describe their latent growth patterns, and to assess the factors associated with baseline and longitudinal change. This is important in providing normative data on the natural history of these novel 3D bone measures.

6.3 Methods

6.3.1 Participants

This study included all 4796 participants from the OAI, using both knees per participant from baseline to 8 year follow-up (9 time points).

6.3.2 Bone shape measures

Quantitative 3D bone shape data was assessed by Imorphics (Manchester, UK) from 3T DESS-we images using SSMs as described previously [10], separately for each bone in methods similar to the ones highlighted in Chapter 3 and 5. The anatomical regions for derivation of bone shape measures were the whole distal femur, proximal tibia bones [111, 209, 339] and the patella. Briefly, an OA vector is constructed for each bone, defined as the line passing through the mean shapes of the bone with and without OA, parameterised as shape components. Individual bone shapes are projected orthogonally onto this vector. Zero was defined as the mean position along the vector for those with a KL score of 0 for 4 consecutive years (“Non-OA group”) and each +1 unit increase represented a change of 1 SD unit away from this Non-OA group. The training set for the segmentation model, and a separate training set for determining the mean shape of the OA and non-OA bones were independent of the test set. Reproducibility of the shape models has been reported previously [111, 528].

6.3.2.1 Classification of shape vectors outside healthy limits

Participants were dichotomised per bone into those with shape vector scores outside “healthy limits” and herein described as “OA bone” for that particular bone shape and those without. The definition of OA bone was those bones which fall above the 95th percentile of the Non-OA group derived from a sample of 885 healthy individuals with KL0 in both knees over 4 years. This dichotomisation was necessary for the descriptive analysis detailed below.

6.3.3 Statistical analysis

Descriptive analyses were performed in STATA Corp, V.13.1 and latent growth curve modelling (LGCM) performed in Mplus software V.8.1. Analyses were performed independently for each bone and stratified by gender, as there are known gender differences in tibia and femur shapes [402, 529]. At baseline Pearson's correlation coefficients and scatterplots were analysed for: femur vs tibia; tibia vs patella and femur vs patella as continuous 3D measures. Using the dichotomisation of OA bone, proportions of bones classified as OA bone were compared for all different possible bone combinations and these in turn reported descriptively for their associations with baseline knee pain, radiographic OA status and knee replacement status at follow-up.

For the LGCM using continuous 3D bone data, missing growth data was estimated using a Full Information Maximum Likelihood method. First, an overall and a multigroup LGCM was fitted to estimate overall mean curves under the assumption of homogeneity of growth patterns in each group and then gender-specific (i.e. males vs females) analyses performed thereafter. Linear models were initially fitted, and to deal with the non-linearity of growth patterns other modelling options such as polynomials and piecewise growth models were considered. For the polynomial function quadratic terms were included and for the piecewise function, models were fitted by creating joints or break points of the mean curves at different time points (year 6-8).Three separate LGCMs were considered for each bone (femur, tibia and patella) over 4 year follow-up (5 time points) and a parallel process model was then fitted by modelling the growth of all 3 tissues and allowing their intercepts and slopes to covary amongst each other, which allowed for the observation of how change in each bone related to change in the other bones longitudinally. Sensitivity analyses were repeated using 8 years of follow-up (9 time points). When selecting the best-fitting model several fit indices were considered including, the chi-square test, root-mean square error of approximation (RMSEA), comparative fit index (CFI), and the standardized root mean square residual (SRMR).

After establishing the best-fitting models further tests were performed using covariates (age, body weight, ethnicity, knee pain and history of knee surgery) chosen as OA risk factors [30, 530-532], to establish if they were predictive of the intercept and slope; and how much of the variance in growth factors these explained. As covariates were added to the unconditional model, the significance of the variance accounted for by the covariates was tested by fitting a reduced model in which the covariates` effect on the growth parameters were constrained to be zero and conducting the appropriate chi-square test between the 2 models.

Each participant contributed two knees to the analyses however the LGCM methodology is unable to account for the within-person correlations. Therefore, the right knee from each participant was selected and as sensitivity check, repeated all analyses on the left, as well as parallel process models for left and right knees which all yielded similar estimates for the mean intercept and slope.

6.4 Results

6.4.1 Descriptive analysis

Table 20 gives the mean values of each bone separately by gender for each year. Mean baseline values for all bones were within 1 SD of their corresponding mean Non-OA shape. As expected bone shape scores increased over follow-up time indicating worsening OA structural status. The full OAI sample of 4796 participants was included (but 9580 knees at baseline due to some missing data) was analysed.

There was a positive correlation between femur and tibia bone shape (r = 0.68) with slightly lower correlation between femur and patella (r = 0.55) and tibia vs patella (r = 0.45) (Figure 31). For each additional unit of femur score, tibia and patella scores were about 0.6 and 0.4 units lower.

In the 3775 knees (39% of knees) with bones classified as OA, the most frequent pattern 936 (25%) involved having all three bones classified as OA, 721 (19%) were exclusively OA femur, 15% OA tibia, 11% OA patella. A combination of femur and tibia OA was seen in 20% of the knees while just 95 (2.5%) had an exclusive combination of tibia and patella OA bones. Knee combinations involving OA femur had the highest frequency of knees undergoing knee replacements during follow-up, however all combinations had similar pain scores (Table 21).

Table 20. Means and standard deviations of three bone shape measures in the OAI

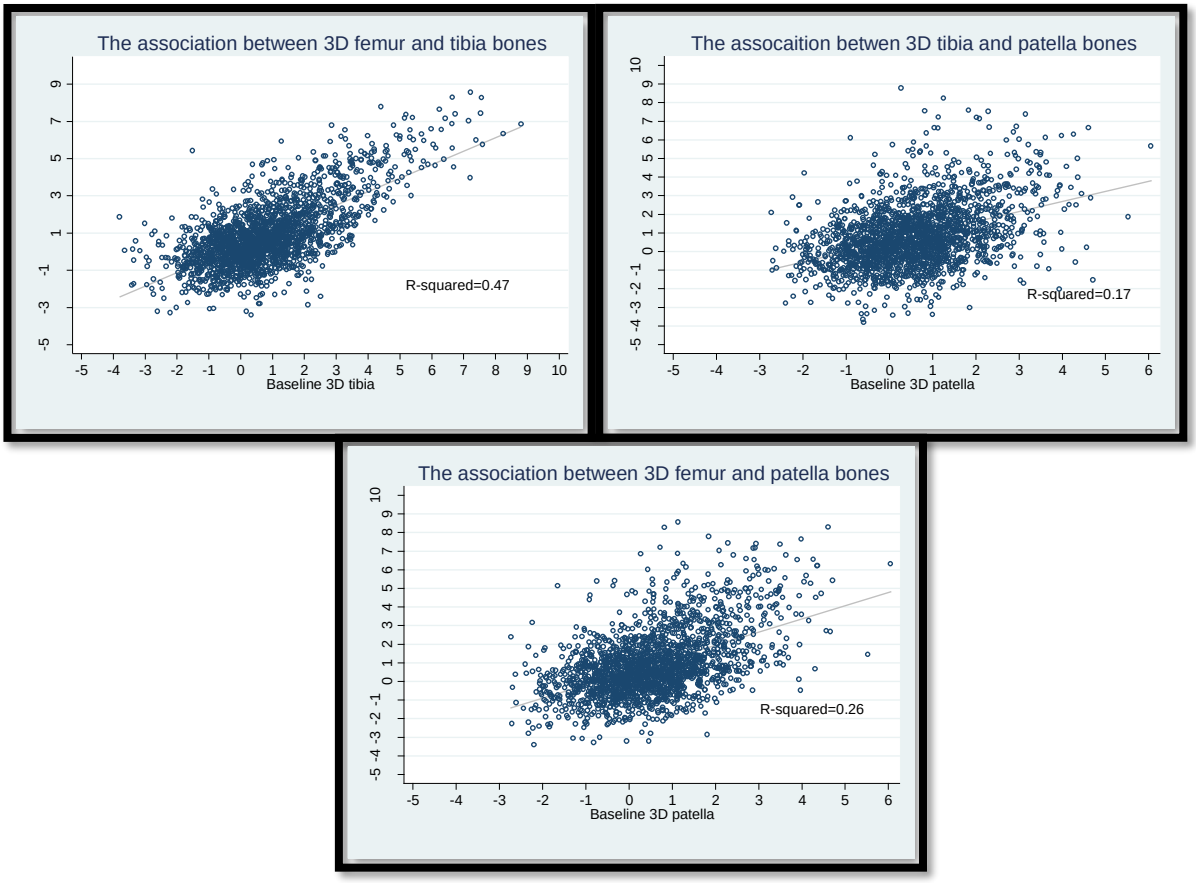
	Males						Females					
	Femur left	**Femur right**	**Tibia left**	**Tibia right**	**Patella left**	**Patella right**	**Femur left**	**Femur right**	**Tibia left**	**Tibia right**	**Patella left**	**Patella right**
Baseline	0.88 (1.79)	0.92 (1.75)	0.65 (1.64)	0.83 (1.67)	0.47 (1.25)	0.56 (1.26)	0.99 (1.75)	1.10 (1.81)	0.76 (1.47)	0.91 (1.51)	0.59 (1.29)	0.77 (1.31)
Year 1	0.92 (1.81)	0.97 (1.80)	0.67 (1.68)	0.84 (1.70)	0.47 (1.24)	0.57 (1.26)	1.07 (1.81)	1.18 (1.85)	0.79 (1.49)	0.96 (1.55)	0.63 (1.30)	0.83 (1.32)
Year 2	1.01 (1.86)	1.05 (1.81)	0.66 (1.70)	0.88 (1.71)	0.53 (1.26)	0.63 (1.27)	1.17 (1.85)	1.27 (1.87)	0.80 (1.52)	0.97 (1.57)	0.69 (1.31)	0.88 (1.35)
Year 3	1.00 (1.92)	1.07 (1.83)	0.66 (1.76)	0.88 (1.69)	0.50 (1.29)	0.61 (1.27)	1.21 (1.85)	1.37 (1.91)	0.81 (1.54)	1.03 (1.60)	0.70 (1.33)	0.91 (1.36)
Year 4	1.05 (1.90)	1.10 (1.84)	0.68 (1.71)	0.92 (1.68)	0.56 (1.27)	0.64 (1.26)	1.30 (1.90)	1.41 (1.91)	0.88 (1.53)	1.08 (1.59)	0.76 (1.33)	0.97 (1.35)
Year 6	1.00 (1.92)	1.02 (1.85)	0.66 (1.70)	0.86 (1.69)	0.54 (1.24)	0.68 (1.25)	1.35 (1.94)	1.47 (2.00)	0.98 (1.56)	1.17 (1.61)	0.80 (1.30)	1.06 (1.34)
Year 8	1.26 (1.94)	1.29 (1.87)	0.52 (1.70)	0.71 (1.70)	0.65 (1.28)	0.73 (1.27)	1.63 (1.98)	1.71 (1.99)	0.93 (1.62)	1.06 (1.62)	0.90 (1.39)	1.15 (1.38)

Values are mean (SD)

Table 21. Knee combinations and association with structure and pain

Exclusive knee combination	Affected bones	KL grade (>2)	WOMAC >=8	TKR during study
	N (%)	N (%)	N (%)	N (%)
All 3 bones	936/3775 (25)	444/936 (47)	60 (6)	136 (16)
Femur and tibia	728 (19)	337/728 (46)	43 (6)	78 (11)
Femur only	721 (19)	188/721 (26)	33 (5)	46 (6)
Femur and patella	285 (8)	38/285 (13)	10 (4)	38 (13)
Tibia and patella	95 (3)	24/95 (25)	4 (4)	6 (6)
Patella only	432 (11)	17/432 (4)	15 (3)	17 (4)
Tibia only	578 (15)	74/578 (13)	17 (3)	12 (2)
All healthy	5662 (60)	303 (5)	256 (5)	72 (1)

Figure 31. Scatter plots with linear fit for pairwise comparisons between bones



6.4.2 Linear and quadratic growth curve models for each bone

6.4.2.1 Overall (one group) models

The linear growth models showed excellent fit to the data in all tissues (Table 22), thus the growth models described hereafter contained 2 parameters, the intercept and slope (and 95% CI), with these parameters allowed to covary. Although the quadratic models showed marginally better fit than that seen for linear models (Table 22), for reasons of parsimony and interpretability and also due to the fact that the fit seen in the linear models was very good and the estimates very closely matched, there was no advantage in choosing quadratic rather linear models. Additionally estimates for the quadratic terms were very small (~ 0.00).

Inspection of the growth curves (shown in Figures 32-34) also suggested a good fit for linear models, with no significant differences between the fitted curve and the sample means; and the 2 curves almost perfectly overlying each other. Graphically the femur growth curves showed better fit than tibia and patella curves. As highlighted earlier the quadratic model was only marginally better than the linear model (but model fit was excellent for linear model, and now supported by graphical evidence of good linear fit. Therefore, the linear function was optimum and allowed the choice of model herein to be deemed linear.

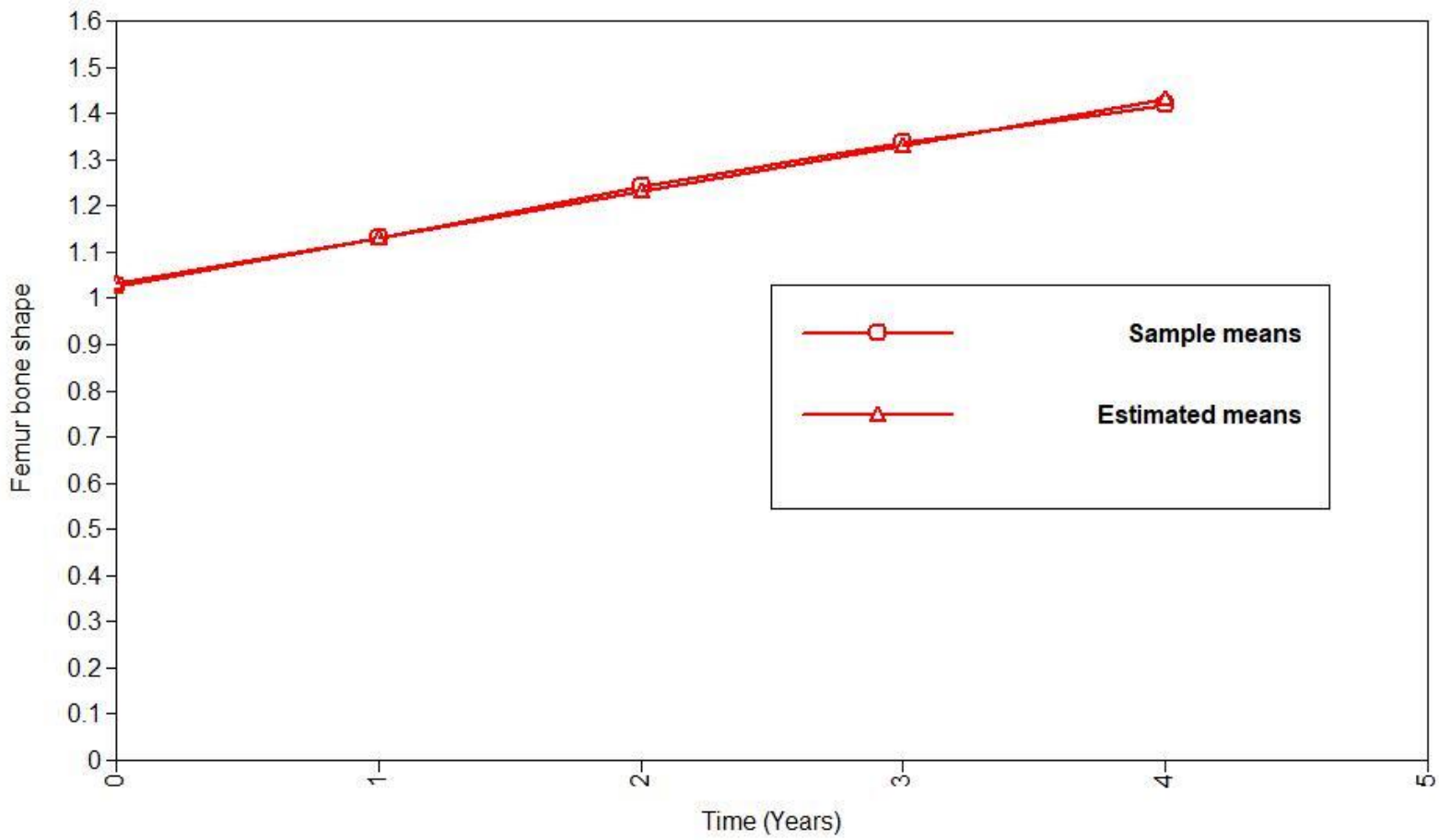


Figure 32. Graph shows how the mean femur bone shape changed over time, compared to the means estimated by the linear growth curve model

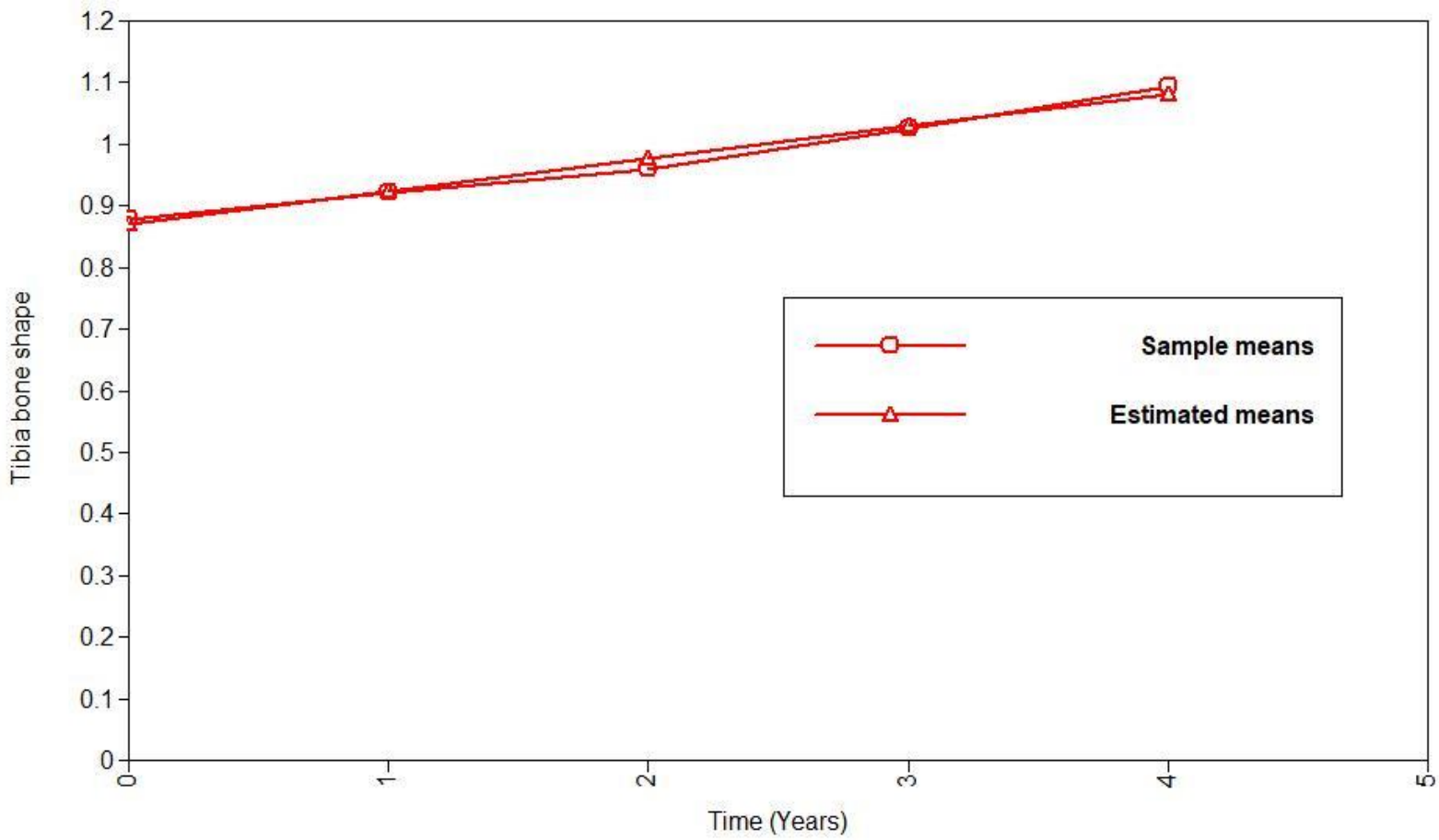


Figure 33. Graph shows how the mean tibia bone shape changed over time, compared to the means estimated by the linear growth curve model

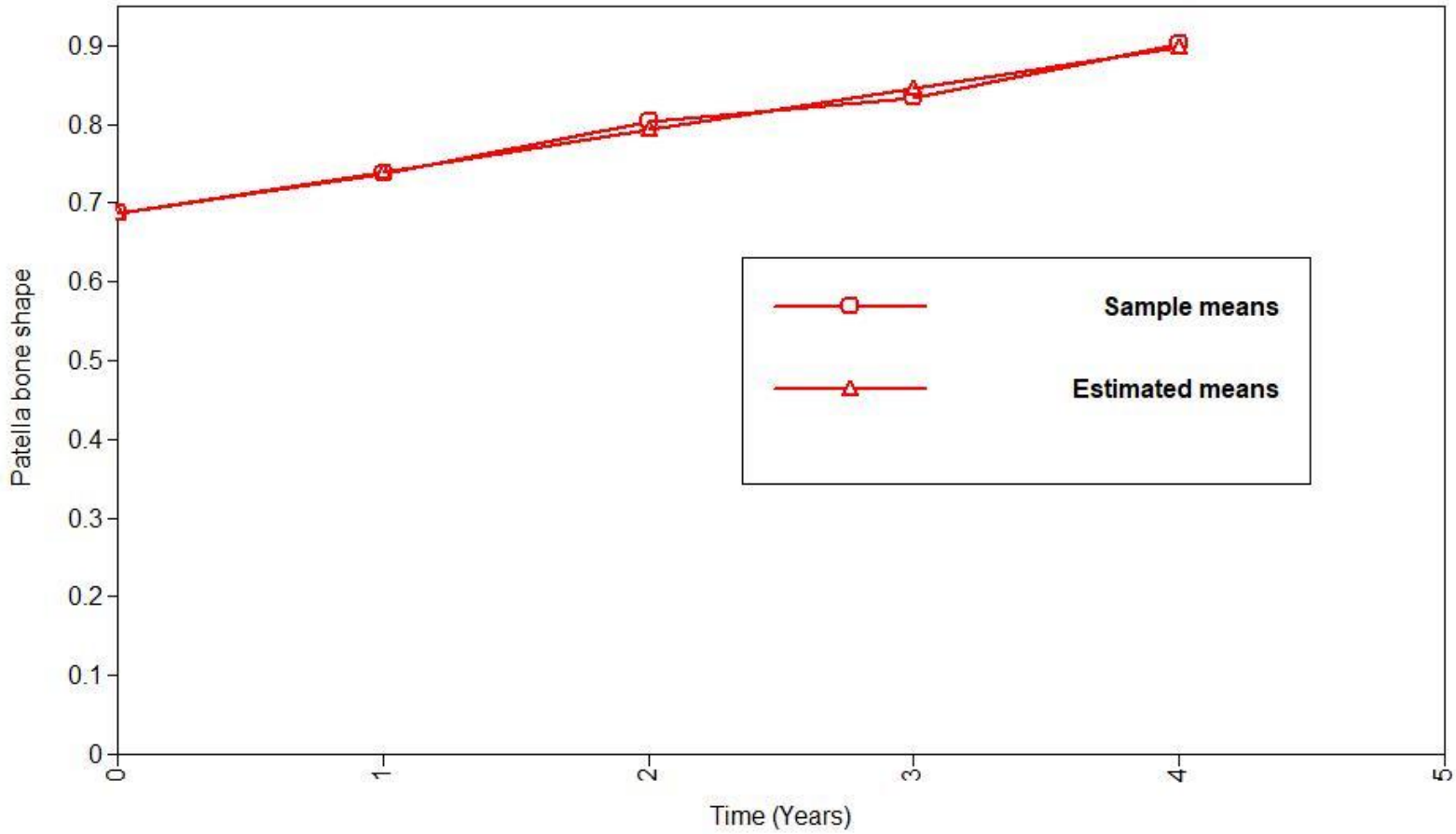


Figure 34. Graph shows how the mean patella bone shape changed over time, compared to the means estimated by the linear growth curve model

6.4.2.1.1 Femur

Data from 4748 participants were included in this analysis. Model fit in the one-group femur model was good: (chi-square value = 213, DF = 13, CFI = 0.99, RMSEA = 0.057 and SRMR =0.004). The linear model fit was expected, as the mean values of femur increased at each time point (as shown in Table 20). The average femur intercept (1.03, p <0.001) and average slope (+0.10, p <0.001) were significantly different from zero, indicating that femur shape significantly increased over time (representing worsening over time). The increase in femur was 0.06 units greater per unit of baseline femur (intercept/slope covariance = 0.06,95% CI 0.06,0.07) (Table 22).

6.4.2.1.2 Tibia

Model fit in the tibia was equally good : (chi-square value = 150, DF = 13, CFI = 0.99, RMSEA = 0.047 and SRMR =0.006). The average femur intercept (0.87, p <0.001) and average slope (+0.05, p <0.001) were significantly different from zero, indicating that like seen for the femur, tibia shape significantly increased over time (representing worsening over time). The increase in tibia was 0.03 units greater per unit of baseline femur (intercept/slope covariance = 0.03,95% CI 0.02,0.03) (Table 22).

6.4.2.1.3 Patella

As seen for the femur and tibia , the one-group patella models also showed good fit to the data: (chi-square value = 86, DF = 13, CFI = 0.99, RMSEA = 0.043 and SRMR =0.008). Statistically significant intercept = 0.69 and slope = +0.05 (p<0.001) also indicated that patella bone shape increased over time (representing structural worsening). The increase in patella was 0.01 units greater per unit of baseline patella (intercept/slope covariance = 0.01,95% CI 0.00,0.01) (Table 22).

6.4.2.2 Multi-group growth models by gender

Similar to one-group models, linear growth models separated by gender showed good fit in both genders and for all bones, and as shown in Table 23 there was slightly better fit seen for the femur than tibia and patella. Across all bones, females had marginally higher intercepts compared to males (although in both sexes this was approximately within 1 SD of the mean Non-OA group, (femur intercept ~1.10 in females vs 0.93 in males) & (tibia intercept ~ 0.91 vs 0.82 in males). The greatest rate of change was seen in the femur followed by tibia and patella. Slopes in females were also higher for all bones, approximately twice that of males (tibia slope 0.07 vs 0.04 in males). This

implies that females on average started with slightly worse structural disease than males and deteriorated at a faster rate than males, in the case of femur in females bone shape increased at about 0.12 units per year. The model-estimated intercepts, or mean baseline femur, was 1.10 (95% CI 1.04,1.16) in females and 0.93 (95% CI 0.87,1.00) in males, and as this was close to the sample mean baseline femur of 1.10 and 0.92 units respectively in males and females (Table 20), this further supported the choice of a linear growth model.

However there was considerable inter-individual variation as indicated by the estimates of variance (Table 24), and as shown in Figure 35 and 36 for the femur in a random sample of 200 males and 200 females (approximately 10%). These plots showed that the trend in femur varied considerably between individuals. The variation was similar for the tibia and patella. Further investigation into the variation in all bone shape growth curves between participants was therefore warranted and is provided by analysis of the effects of covariates on the growth factors later in this Chapter and also latent growth class analysis (Chapter 7).

Table 22.One-group models in the 3 bones

Bones	Intercept (95% CI)	Slope (95% CI)	Quadratic	Intercept & slope Covariance (95% CI)	Adjusted BIC	Chi square (DF)	RMSEA (0.06)	CFI/TLI (>0.95)	SRMR (0.08)
Femur									
Linear model	1.03 (0.99,1.07)	0.10 (0.09,0.10)	-	0.06 (0.06,0.07)	16 464	213 (13)	0.057	0.997	0.004
Quadratic model	1.025	0.114	-0.004	0.086	16 387	115 (9)	0.050	0.998	0.001
Tibia									
Linear model	0.87 (0.83,0.91)	0.05 (0.05,0.06)	-	0.03 (0.02,0.03)	20 036	150 (13)	0.047	0.997	0.006
Quadratic model	0.879	0.031	0.006	0.050	19 985	79 (9)	0.040	0.999	0.002
Patella									
Linear model	0.69 (0.66,0.72)	0.05 (0.05,0.06)	-	0.01 (0.00,0.01)	29 704	86 (13)	0.043	0.996	0.008
Quadratic model	0.688	0.052	0.000	0.03	28 078	48 (9)	0.030	0.999	0.005

95% CI only quoted for linear models.

Table 23. Multi-group growth models in the 3 bones

Model	Intercept (95% CI)	Slope (95% CI)	Intercept/slope covariance (95%CI)	RMSEA (<0.08)	CFI (>0.95)	SRMR (<0.08)
Multi group models						
Femur				0.067	0.995	0.004
Male model	0.93 (0.87,1.00)	0.07 (0.07,0.08)	0.05 (0.04,0.06)			
Female model	1.10 (1.04,1.16)	0.12 (0.12,0.13)	0.07 (0.06,0.08)			
Tibia				0.059	0.996	0.006
Male model	0.82 (0.76,0.88)	0.04 (0.03,0.04)	0.02 (0.01,0.03)			
Female model	0.91 (0.86,0.96)	0.07 (0.06,0.07)	0.03 (0.02,0.04)			
Patella				0.064	0.993	0.010
Male model	0.56 (0.52,0.61)	0.04 (0.03,0.04)	0.00 (-0.01,0.01)			
Female model	0.77 (0.73,0.82)	0.07 (0.06,0.07)	0.01 (0.00,0.02)			
Overall (one group model)						
Femur	1.03 (0.99,1.07)	0.10 (0.09,0.10)	0.06 (0.06,0.07)	0.057	0.997	0.004
Tibia	0.87 (0.83,0.91)	0.05 (0.05,0.06)	0.03 (0.02,0.03)	0.047	0.997	0.006
Patella	0.69 (0.66,0.72)	0.05 (0.05,0.06)	0.01 (0.00,0.01)	0.043	0.996	0.007

One – group models given for comparisons

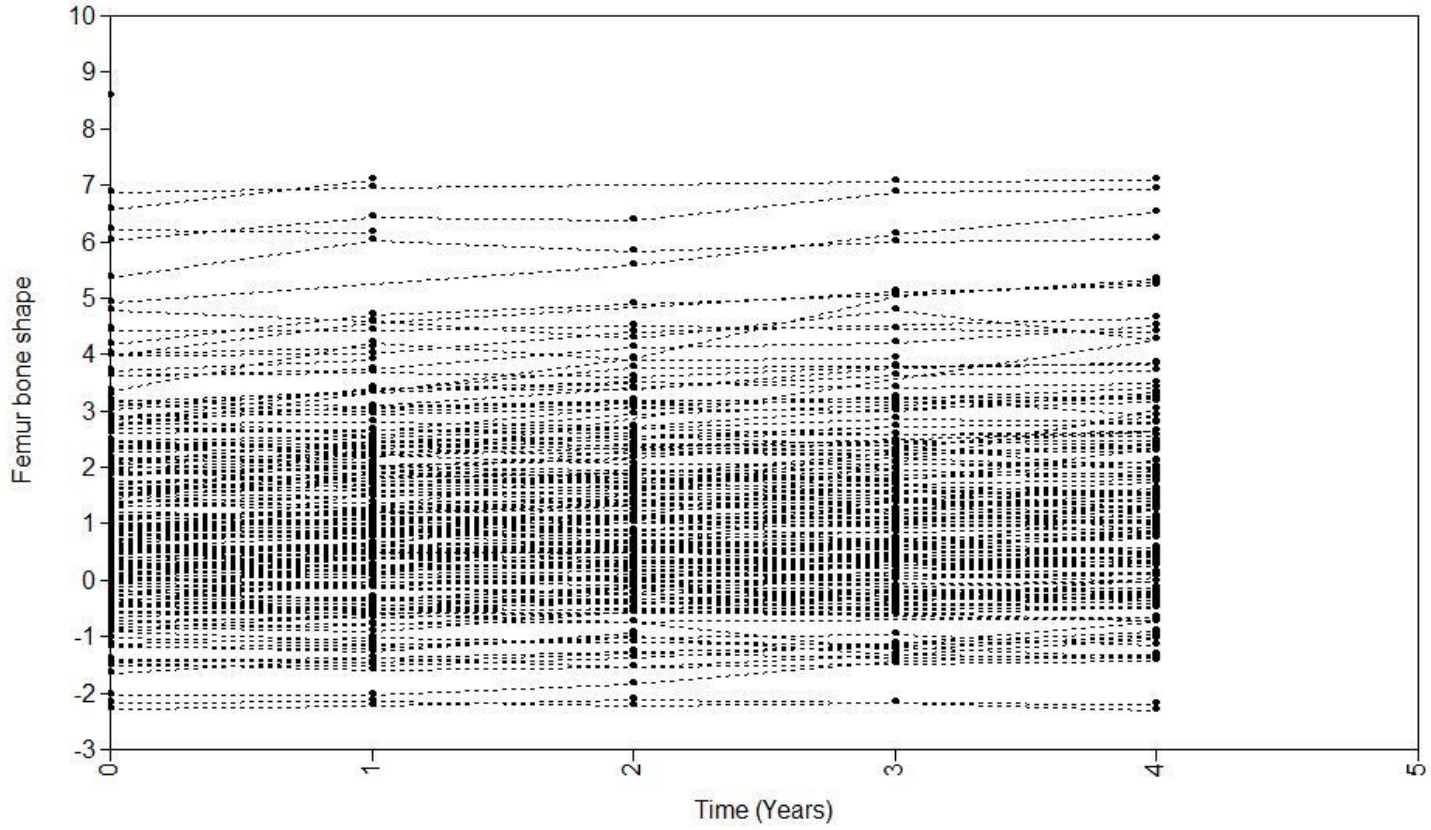


Figure 35. Graph to show actual values of femur bone shape from a random sample of 200 male participants

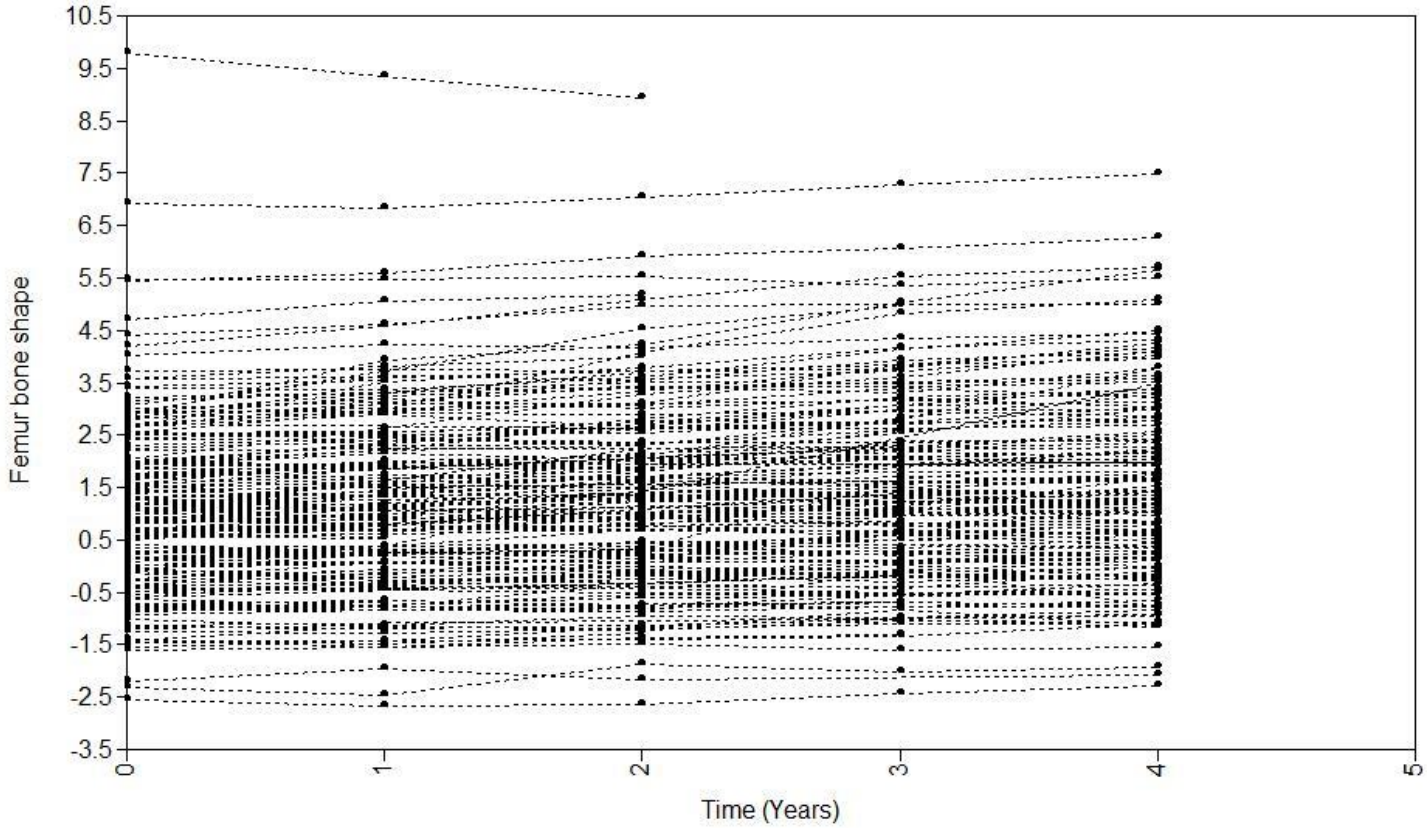


Figure 36. **Graph to show actual values of femur bone shape from a random sample of 200 female participants**

Table 24. Estimates of variability in growth factors

	Males			Females		
Parameter	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**
Femur						
Intercept variance	3.090	0.099	<0.001	3.267	0.088	<0.001
Slope variance	0.010	0.000	<0.001	0.020	0.001	<0.001
Tibia						
Intercept variance	2.759	0.089	<0.001	2.266	0.062	<0.001
Slope variance	0.007	0.000	<0.001	0.010	0.000	<0.001
Patella						
Intercept variance	1.501	0.050	<0.001	1.635	0.046	<0.001
Slope variance	0.005	0.001	<0.001	0.005	0.001	<0.001

6.4.3 Parallel process growth curve model of change in the three bones over time

The linear models for bone shape were selected for combination in a parallel process growth model stratified by gender (Figure 37). The resulting model showed adequate fit to the data, with fit indices as follows: chi-square = 1281, DF = 210, CFI = 0.993, RMSEA = 0.046, SRMR = 0.005.

6.4.3.1 Males

Table 25 gives the estimates obtained from parallel process models. As expected, linear slopes for all bones did not change or were within 0.01 of those estimated by the previous linear models of change in femur, tibia and patella separately shown in section 6.3.2. Differences in intercepts and slopes coefficients were also similar to those previously reported and furthermore, estimates of covariances between intercept and slope for all bones were within 0.1 of those reported for the respective singular models.

The covariances between the three bones` intercepts and slopes were important in establishing how the femur changed with tibia and patella, and also how the tibia changed with patella. All 3 intercepts varied significantly with each other (all p<0.001). The covariance between femur and tibia intercepts in male models was 2.03 (SE 0.08, p<0.001), which implied that baseline femur increased by 2.03 per unit of baseline tibia. Covariance between the femur and tibia slopes, was 0.007 (SE 0.00, p<0.001), so mean femur changed in the same direction (positive) as tibia, which represented worsening over time. In terms of magnitude this meant the femur rate of change

increased by 0.007 units for every 1 unit change in tibia slope. Increased baseline femur (or intercept), was associated with a 0.03 unit greater increase in tibia (or slope) over 4 years. This is likely because greater baseline femur was associated with greater baseline tibia and therefore increased potential increase in the slope value. The relationship between baseline tibia and rate of change in femur was indicated by the covariance between tibia intercept and femur slope, of 0.03 (SE 0.005, $p<0.001$), thus increased baseline tibia was associated with a greater increase in femur. The explanation for this is similar to that of the covariance between femur intercept and tibia slope, and could be because higher baseline tibia occurred in cases with higher baseline femur therefore increased the likelihood to increase (worsen) with time. While the patella slope did not significantly co-vary with any intercept growth factors of the other bones including its own (all $p >0.05$), there was significant covariance between the patella slope and tibia slope, (covariance = 0.02,SE = 0.005, $p<0.001$) and between the patella slope and femur slope, (covariance = 0.03,SE = 0.005, $p<0.001$) (Table 25).

6.4.3.2 Females

As seen for males, the estimates for females were as expected; linear slopes for all bones did not change or were with 0.01 of those estimated by the previous linear models of change in femur, tibia and patella separately shown in (section 6.3.2). Differences in intercepts and slopes cohorts were also similar to those previously reported for males and furthermore, estimates of covariances between intercept and slope for all bones were within 0.1 of those reported for the respective singular models (Table 25).

Similar to males all 3 intercepts varied significantly with each other (all $p<0.001$). The covariance between femur and tibia intercepts in female models was 1.87 (SE 0.06, $p<0.001$), which implied that baseline femur increased by 1.87 per unit of baseline tibia. Between the femur and tibia slopes, covariance was 0.013 (SE 0.00, $p<0.001$), so mean femur changed in the same direction (positive), which represented worsening over time. In terms of magnitude this meant that the femur rate of change increased by 0.013 units per 1-unit increase in tibia slope. Increased baseline femur (or intercept), was associated with a 0.04 unit greater increase in tibia (or slope) over 4 years for reasons suggested above in male models. The relationship between baseline tibia and rate of change in femur was indicated by the covariance between tibia intercept and femur slope of 0.04 (SE 0.005, $p<0.001$). Thus increased baseline tibia score was associated with a greater increase in femur shape. Patella results in females were different from males as all growth factors significantly co-varied with those of the femur and tibia (all $p<0.001$) (Table 25).

Table 25. Estimates from full parallel process model with 3 bones

	Males			Females		
	Estimate	**S.E**	**p-value**	**Estimate**	**S.E**	**p-value**
Parameter						
Mean femur intercept	0.93	0.040	<0.001	1.10	0.034	<0.001
Mean femur slope	0.07	0.003	<0.001	0.12	0.003	<0.001
Mean tibia intercept	0.82	0.038	<0.001	0.91	0.029	<0.001
Mean tibia slope	0.04	0.003	<0.001	0.07	0.003	<0.001
Mean patella intercept	0.56	0.028	<0.001	0.77	0.025	<0.001
Mean patella slope	0.04	0.003	<0.001	0.07	0.003	<0.001
Covariances between latent variables						
Femur intercept with femur slope	0.050	0.005	<0.001	0.068	0.006	<0.001
Femur intercept with tibia intercept	2.026	0.081	<0.001	1.868	0.063	<0.001
Femur intercept with patella intercept	1.127	0.056	<0.001	1.351	0.052	<0.001
Femur intercept with tibia slope	0.028	0.005	<0.001	0.043	0.005	<0.001
Femur intercept with patella slope	-0.001	0.005	0.814	0.017	0.005	<0.001
Femur slope with tibia slope	0.007	0.000	<0.001	0.013	0.000	<0.001
Femur slope with patella slope	0.003	0.000	<0.001	0.008	0.000	<0.001
Tibia intercept with tibia slope	0.019	0.004	<0.001	0.003	0.004	<0.001
Tibia intercept with patella intercept	0.859	0.051	<0.001	0.942	0.042	<0.001
Tibia intercept with patella slope	-0.004	0.005	0.480	0.011	0.004	0.007
Tibia intercept with femur slope	0.034	0.005	<0.001	0.043	0.005	<0.001
Tibia slope with patella slope	0.002	0.000	<0.001	0.005	0.000	<0.001
Patella intercept with patella slope	0.001	0.004	0.811	0.012	0.004	<0.001

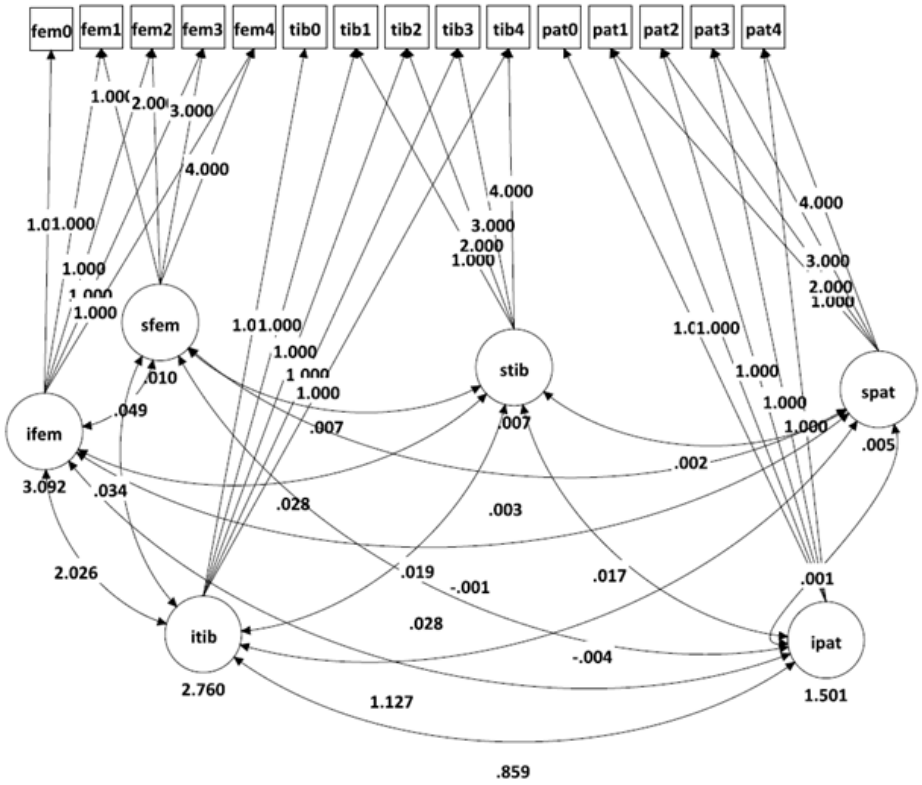


Figure 37. Graph shows parallel process growth curve model of change in all 3 bones.
fem0-fem4 = femur baseline to year 4; tib0-tib4 = tibia baseline to year 4 & pat0-pat4 = patella baseline to year 4.Values at the bottom of each growth factor represent variances of that growth factor. Prefix “i” before the growth factor represents intercept and prefix “s” represents slope e.g. ipat=patella intercept and spat=patella slope. Double edged arrows represent covariances between each growth factor and the covariance estimate is the value at the midpoint of each one

6.4.4 Sensitivity analyses

Results of the sensitivity analyses included repeating all models on the left knee, a parallel process model for each bone (left and right knee allowed to co-vary) and growth models using 8-year OAI data.

6.4.4.1 Latent growth models in the left knee

Parameter estimates from linear models applied on just left knees were very similar to those obtained from earlier models shown in Tables 22 and 23.However in terms of magnitude the estimates tended to be marginally smaller in the left knee, for example the femur shape had (Intercept = 0.89) in the left (Table 26) vs (Intercept = 0.93) in the right knee (Table 23). Fit indices were also very good for the left knee and were very similar to those in the right knee reported before. The resulting femur model in males showed good fit to the data, with fit indices as follows: chi-square = 914, DF = 98, CFI = 0.997, RMSEA = 0.052,SRMR = 0.005 (Table 26).

Table 26 Growth models in the 3 bones in the left knee

Model	Intercept (95% CI)	Slope (95% CI)	Intercept/slope covariance (95%CI)	RMSEA (<0.08)	CFI (>0.95)	SRMR (<0.08)
Multi group models						
Femur				0.052	0.997	0.005
Male model	0.89 (0.81,0.97)	0.08 (0.07,0.09)	0.06 (0.05,0.07)			
Female model	1.00 (0.93,1.06)	0.11 (0.11,0.12)	0.08 (0.07,0.09)			
Tibia				0.046	0.997	0.007
Male model	0.65 (0.57,0.72)	0.03 (0.02,0.03)	0.03 (0.02,0.04)			
Female model	0.76 (0.71,0.82)	0.05 (0.04,0.05)	0.03 (0.02,0.04)			
Patella				0.055	0.994	0.008
Male model	0.47 (0.42,0.53)	0.04 (0.03,0.05)	-0.00 (-0.01,0.01)			
Female model	0.59 (0.54,0.64)	0.06 (0.05,0.06)	0.01 (0.00,0.02)			
Overall (one group model)						
Femur	0.95 (0.90,1.00)	0.10 (0.09,0.10)	0.07 (0.07,0.08)	0.046	0.998	0.004
Tibia	0.71 (0.67,0.76)	0.04 (0.04,0.05)	0.03 (0.02,0.03)	0.037	0.998	0.006
Patella	0.54 (0.50,0.58)	0.05 (0.05,0.06)	0.01 (0.00,0.01)	0.039	0.997	0.006

One–group models provided for comparison

6.4.4.2 Latent growth models with 8-year data

Models using 8 year data showed parameter similar estimates to those seen with 4 year follow-up in both the overall and multi group models, for example in the female model patella intercept and slope were 0.78 and 0.06 respectively using 8-year data (Table 27) and 0.77 and 0.07 respectively in the 4-year models (Table 23). However in terms of model fit, there were significant differences between earlier models and 8-year models in favour of the 4 year models. RMSEA indices in all models were not statistically significant (all > 0.08) and although acceptable fit indices were observed for CFI and SRMR these were comparably much lower than those seen in the 4-year follow-up data shown previously in section 6.3.2.

6.4.4.2.1 Piece-wise growth models

As the linear model was not a good fit to the 8 year data and, because of known changes to the MRI system phantom, and acquisition protocol in the OAI after year 4, a piece-wise model was considered for the 8–year data. This allowed the slopes to vary after year 4 given that a correction factor had been applied on the MRI data after this point. As seen for the 8-year models before in the femur one-group model, there was poor fit as suggested by chi-square values = 1377, DF= 22, CFI = 0.983, RMSEA = 0.114,SRMR = 0.016 but this was an improvement from the linear models albeit still not adequate. The trend was similar in the multi-group models for the femur.

However differences were seen in the tibia and patella, with the respective piece-wise models all showing good fit (Table 28), a significant improvement compared to the previous 8-year linear models. Most notable differences were seen in the tibia, the tibia slope2 (year 6-8) in males was negative (slope2 = -0.05, 95% CI -0.06,-0.03) although confidence intervals were the widest for any models thus far for slopes. This would be interpreted as improvement after year 4 in the tibia of males. For the tibia in females there was worsening over time as expected and seen in previous models but the rate of change was slower (slope1 = 0.07 vs slope2 = 0.03). Importantly, there was also much higher variability associated with the second tibia slopes compared to year 0-4 slopes, 5 times higher for slope2 (slope2 variance in tibia = 0.033 vs 0.006 for tibia slope1) while in females this variance was lower for slope2 (slope1=0.07 vs slope2=0.03). Compared with linear models (0-4 year data) whose slope variances were ~ 0.005 in all models, the piece-wise model was justifiably not the best model and further strengthened the choice to keep the models to 5 time points due to the uncertainty around estimates (Figure 38-40) and measurement error after year 4,

including the application of a correction factor for MRI data. In the patella the trends were as seen before and the second slope in both genders showed worsening which represented worsening after year 4 (Table 28).

Table 27. Parameter estimates from Latent Growth Curve Model over 8 year follow-up

Model	Intercept (95% CI)	Slope (95% CI)	Intercept/slope covariance (95%CI)	RMSEA (<0.08)	CFI (>0.95)	SRMR (<0.08)
Overall (one group model)						
Femur	1.03 (0.98,1.08)	0.10 (0.09,0.10)	0.04 (0.04,0.05)	0.192	0.939	0.018
Tibia	0.90 (0.85,0.94)	0.04 (0.03,0.04)	0.01 (0.01,0.05)	0.156	0.953	0.018
Patella	0.69 (0.65,0.73)	0.05 (0.05,0.06)	0.01 (0.01,0.06)	0.106	0.968	0.026
Multi group models						
Femur				0.193	0.937	0.018
Male model	0.92 (0.84,1.00)	0.07 (0.07,0.08)	0.03 (0.02,0.04)			
Female model	1.10 (1.04,1.17)	0.11 (0.11,0.12)	0.05 (0.04,0.06)			
Tibia				0.156	0.952	0.019
Male model	0.85 (0.78,0.93)	0.01 (0.01,0.02)	0.02 (0.00,0.01)			
Female model	0.93 (0.87,0.99)	0.05 (0.05,0.06)	0.01 (0.01,0.02)			
Patella				0.113	0.962	0.028
Male model	0.56 (0.50,0.61)	0.04 (0.03,0.04)	0.00 (-0.02,0.03)			
Female model	0.78 (0.73,0.83)	0.06 (0.06,0.07)	0.00 (-0.01,0.01)			

Table 28. Parameter estimates from Piece-wise Latent Growth Curve model over 8 year follow-up

Model	Intercept (95% CI)	Slope 1 (95% CI)	Slope 2 (95%CI)	RMSEA (<0.08)	CFI (>0.95)	SRMR (<0.08)
Overall (one group model)						
Femur	1.04 (0.98,1.09)	0.10 (0.09,0.10)	0.06 (0.05,0.08)	0.114	0.983	0.016
Tibia	0.87 (0.82,0.91)	0.06 (0.05,0.06)	-0.01 (-0.01,0.01)	0.056	0.995	0.009
Patella	0.69 (0.65,0.73)	0.05 (0.05,0.06)	0.07 (0.05,0.08)	0.030	0.998	0.008
Multi group models						
Femur				0.116	0.982	0.017
Male model	0.94 (0.86,1.02)	0.07 (0.06,0.07)	0.06 (0.04,0.07)			
Female model	1.11 (1.04,1.17)	0.12 (0.11,0.12)	0.07 (0.05,0.09)			
Tibia				0.009	0.994	0.009
Male model	0.81 (0.74,0.89)	0.04 (0.04,0.05)	-0.05 (-0.06,-0.03)			
Female model	0.91 (0.85,0.96)	0.07 (0.06,0.07)	0.03 (0.02,0.05)			
Patella				0.053	0.993	0.010
Male model	0.56 (0.51,0.62)	0.04 (0.03,0.04)	0.07 (0.05,0.09)			
Female model	0.78 (0.73,0.83)	0.06 (0.06,0.07)	0.06 (0.05,0.08)			

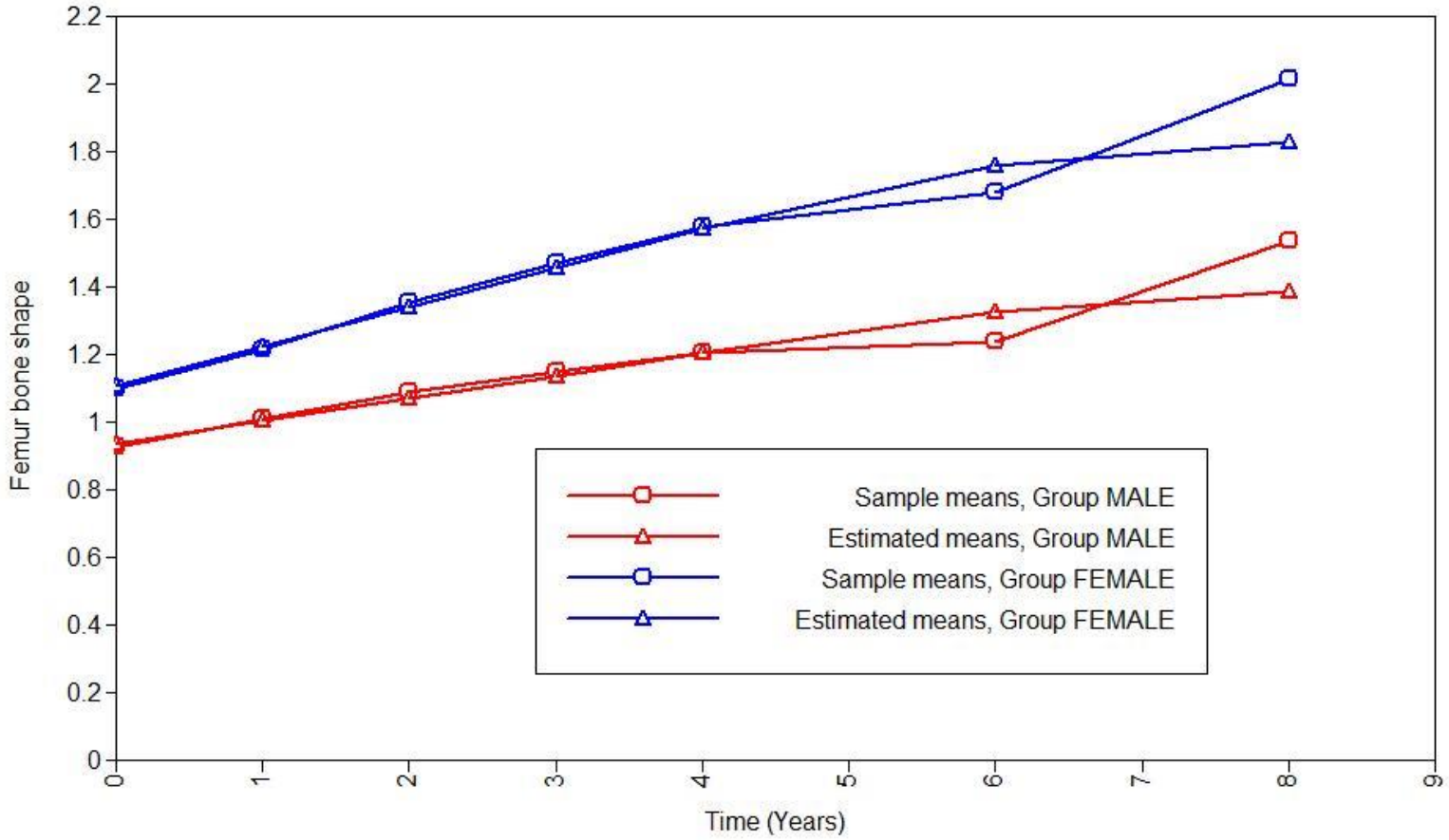


Figure 38.Graph to show how bone shape changed with time, compared to the means estimated by the piecewise growth curve model of change in femur.

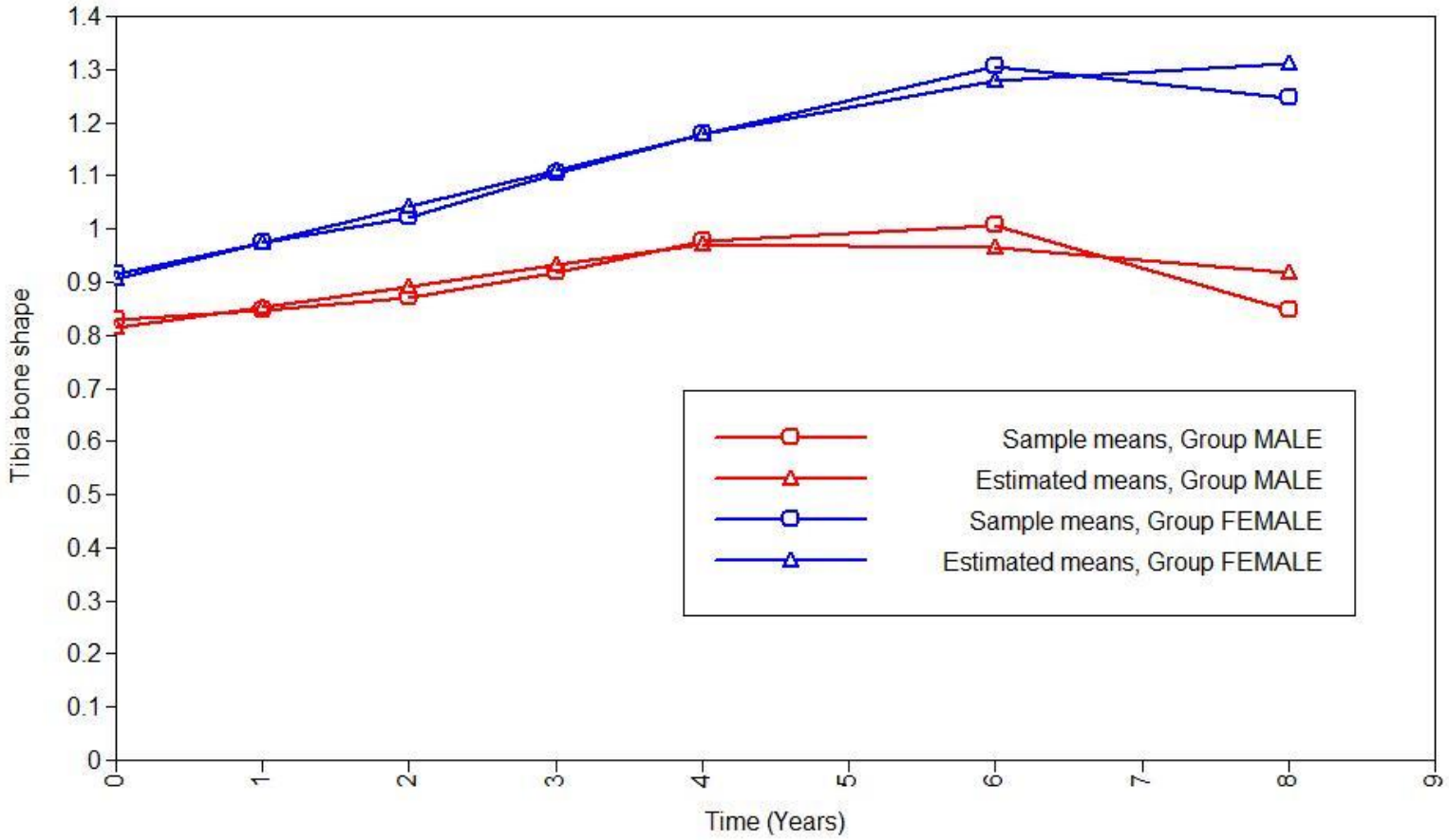


Figure 39. Graph to show how bone shape changed with time, compared to the means estimated by the piecewise growth curve model of change in tibia.

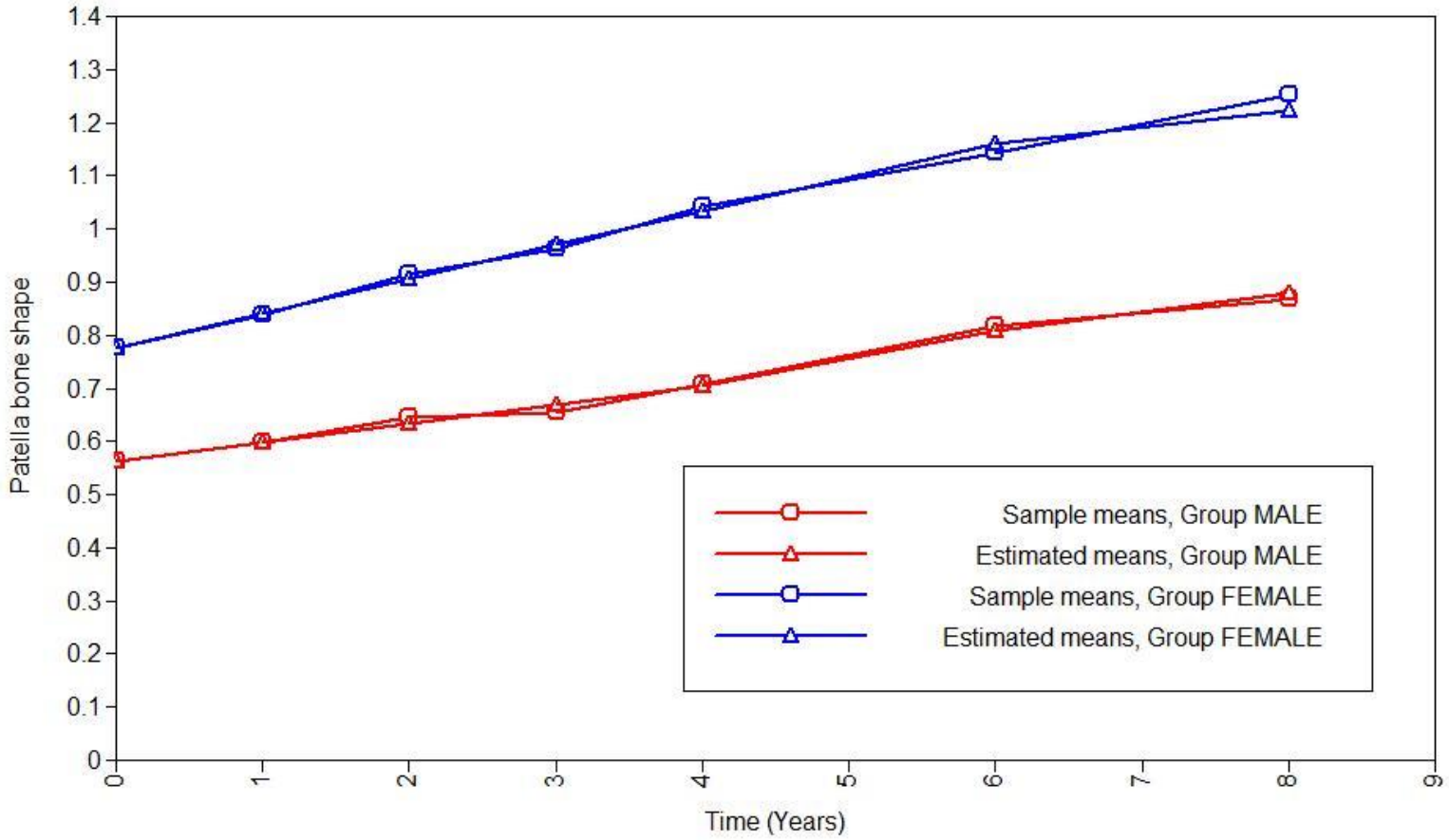


Figure 40. Graph to show how bone shape changed with time, compared to the means estimated by the piecewise growth curve model of change in patella.

6.4.4.3 Parallel process models for left and right knee

In each bone, the left and right knees were modelled in combination in a parallel process growth model stratified by gender (Table 29). The resulting femur model in males showed good fit to the data, with fit indices as follows: chi-square= 914, DF= 98, CFI = 0.993, RMSEA = 0.059,SRMR = 0.004.

Table 29 provides the estimates obtained. As expected, linear slopes for all bones did not change or were with 0.01 of those estimated by the previous linear models of change in the right and left knees separately for all bones. Differences in intercept and slope coefficients were also similar to those previously reported and furthermore, estimates of covariances between intercept and slope for all bones were within 0.1 of those reported for the respective singular models.

The covariances between the left knee and right knee were important in establishing how changes in the right knee changed with those of the left. All 3 intercepts (left vs right for each bone) varied significantly with each other (all $p<0.001$). The covariance between left femur and right femur intercepts in the male model was 2.02 (SE 0.08, $p<0.001$), which implied that baseline right femur increased by 2.02 per unit of baseline left femur. Between the left femur and right femur their slope covariance was 0.005 (SE 0.001, $p<0.001$), therefore both sides changed in the same direction (positive) which represents worsening over time in both sides. Increased right baseline femur (or intercept), was associated with a 0.03 unit greater increase in left femur (or slope) over 4 years. This is likely because greater baseline right femur was associated with greater baseline left femur and therefore increased potential increase in the slope value as seen previously in the full parallel process models with all bones. Similar findings were seen in the female models (Table 29).

The findings described for the left and right femur were similar in the tibia analyses and also consistent by gender (Table 29). The resulting tibia model, like the femur model before showed good fit to the data, with fit indices as follows: chi-square= 601, DF= 98, CFI = 0.995, RMSEA = 0.046,SRMR = 0.006.

Some differences when seen in the patella. While the intercepts significantly co-varied in both genders, as seen in the femur and tibia ($p<0.001$), the intercept/slope covariances were not statistically significant in males (Table 29) but were significant in females, and followed the same pattern as seen in the femur and tibia. The patella model also showed good fit to the data but as seen before the fit was better for the

femur and tibia models, with fit indices as follows: chi-square= 701, DF= 98, CFI = 0.992, RMSEA = 0.051,SRMR = 0.009.

Table 29. Estimates from parallel process model for left and right side

	Males			Females		
	Estimate	**S.E**	**p-value**	**Estimate**	**S.E**	**p-value**
Parameter						
Right femur intercept	0.93	0.040	<0.001	1.10	0.034	<0.001
Right femur slope	0.07	0.003	<0.001	0.12	0.003	<0.001
Left femur intercept	0.89	0.040	<0.001	1.00	0.033	<0.001
Left femur slope	0.08	0.003	<0.001	0.11	0.003	<0.001
Right tibia intercept	0.82	0.037	<0.001	0.91	0.029	<0.001
Right tibia slope	0.04	0.003	<0.001	0.07	0.003	<0.001
Left tibia intercept	0.65	0.037	<0.001	0.76	0.028	<0.001
Left tibia slope	0.03	0.003	<0.001	0.05	0.002	<0.001
Right patella intercept	0.57	0.028	<0.001	0.78	0.025	<0.001
Right patella slope	0.04	0.003	<0.001	0.07	0.003	<0.001
Left patella intercept	0.47	0.028	<0.001	0.59	0.024	<0.001
Left patella slope	0.04	0.003	<0.001	0.06	0.003	<0.001
Covariances						
Femur						
Right intercept with left intercept	2.024	0.084	<0.001	2.503	0.077	<0.001
Right slope with left slope	0.005	0.001	<0.001	0.010	0.001	<0.001
Right intercept with left slope	0.040	0.005	<0.001	0.066	0.006	<0.001
Left intercept with right slope	0.037	0.005	<0.001	0.059	0.006	<0.001
Tibia						
Right intercept with left intercept	1.834	0.074	<0.001	2.503	0.077	<0.001
Right slope with left slope	0.004	0.000	<0.001	0.010	0.001	<0.001
Right intercept with left slope	0.015	0.005	0.002	0.066	0.006	<0.001
Left intercept with right slope	0.016	0.004	<0.001	0.059	0.006	<0.001
Patella						
Right intercept with left intercept	1.148	0.044	<0.001	2.503	0.077	<0.001
Right slope with left slope	0.004	0.000	<0.001	0.010	0.001	<0.001
Right intercept with left slope	-0.002	0.004	0.638	0.066	0.006	<0.001
Left intercept with right slope	-0.005	0.004	0.191	0.059	0.006	<0.001

6.4.5 Addition of covariates

As shown previously in Table 24, there was significant variance in growth factors between individuals in particular the intercepts. By adjusting for covariates some of this variance was accounted for as shown in Tables 30-32 below. To recap, improvement fit was measured by calculating the difference in the model chi-square values between the 2 models and performing a likelihood ratio test between the nested models, and also assessing model fit indices as shown in Tables 30-32.One model was a null model with the covariates fixed to zero, and the subsequent model adjusting for the covariates and effect of covariates freed.

6.4.5.1 Model fit after adding covariates

Tables 30-32 show differences in fit between the fixed models, where covariates were firstly constrained to zero and then when this constraint was relaxed/freed. All freed models had much better fit than the constrained models, most of which showed evidence of misfit (most had SRMR values >0.008). Secondly the differences from chi-square tests between the constrained vs freed models were all statistically significant ($p<0.001$) showing that adjusting for covariates in these models significantly improved model fit and subsequently explained growth factor variances, importantly for the intercept which had earlier been shown to be large particularly for the intercepts.

Table 30. Model fit indices in femur growth models adjusting for covariates

	Model Chi- sq (df)	Change in chi-sq,df	CFI (>0.95)	RMSEA (<0.08)	SRMR (<0.08)	Males		Females	
						Intercept variance	Slope variance	Intercept variance	Slope variance
Age (constrained model)	378 (38)	45 (4) p<0.001	0.995	0.063	0.037	3.106	0.010	3.250	0.020
Age (freed model)	333 (34)		0.995	0.063	0.004	3.085	0.010	3.220	0.020
Weight(constrained model)	681 (38)	357 (4) p<0.001	0.990	0.090	0.119	3.113	0.010	3.212	0.020
Weight (freed model)	324 (34)		0.995	0.064	0.004	2.998	0.009	2.953	0.019
Pain (constrained model)	745 (38)	412 (4) p<0.001	0.989	0.092	0.123	3.098	0.010	3.251	0.020
Pain (freed model)	333 (34)		0.995	0.063	0.004	2.929	0.009	3.012	0.019
Surgery (constrained model)	517 (38)	182 (4) p<0.001	0.992	0.073	0.085	3.094	0.010	3.259	0.020
Surgery (freed model)	335 (34)		0.995	0.061	0.004	2.974	0.010	3.159	0.020
Ethnicity (constrained model)	429 (38)	88 (4) p<0.001	0.994	0.066	0.059	3.093	0.010	3.267	0.020
Ethnicity (freed model)	341 (34)		0.995	0.062	0.004	3.070	0.010	3.183	0.020
All covariates (constrained model)	1286 (78)	919 (20) p<0.001	0.981	0.086	0.152	3.116	0.009	3.204	0.020
All covariates (freed model)	367 (58)		0.995	0.050	0.003	2.683	0.009	2.651	0.018

Table 31. Model fit indices in tibia growth models adjusting for covariates

	Model Chi- sq (df))	Change in chi-sq,df	CFI (>0.95)	RMSEA (<0.08)	SRMR (<0.08)	Males		Females	
						Intercept variance	Slope variance	Intercept variance	Slope variance
Age (constrained model)	327 (38)	63 (4) p<0.001	0.995	0.058	0.047	2.785	0.007	2.292	0.010
Age (freed model)	264 (34)		0.996	0.055	0.006	2.754	0.007	2.260	0.010
Weight (constrained model)	556 (38)	292 (4) p<0.001	0.990	0.081	0.104	2.805	0.007	2.301	0.010
Weight (freed model)	264 (34)		0.996	0.057	0.006	2.698	0.006	2.180	0.009
Pain (constrained model)	587 (38)	320 (4) p<0.001	0.990	0.051	0.103	2.785	0.007	2.298	0.010
Pain (freed model)	267 (34)		0.996	0.056	0.006	2.693	0.006	2.181	0.009
Surgery (constrained model)	416 (38)	143 (4) p<0.001	0.993	0.065	0.076	2.761	0.007	2.258	0.010
Surgery (freed model)	273 (34)		0.996	0.054	0.006	2.688	0.007	2.192	0.010
Ethnicity (constrained model)	290 (38)	29(4) P<0.001	0.995	0.053	0.031	2.759	0.007	2.266	0.010
Ethnicity (freed model)	261 (34)		0.996	0.053	0.006	2.759	0.007	2.250	0.010
All covariates (constrained model)	1044 (78)	734 (20) P<0.001	0.982	0.077	0.130	2.805	0.007	2.293	0.010
All covariates (freed model)	310 (58)		0.995	0.045	0.005	2.494	0.006	1.994	0.009

Table 32.Model fit indices in patella growth models adjusting for covariates

	Model Chi- sq(df)	Change in chi-sq,df	CFI (>0.95)	RMSEA (<0.08)	SRMR (<0.08)	Males		Females	
						Intercept variance	Slope variance	Intercept variance	Slope variance
Age (constrained model)	408 (38)	105 (4) p<0.001	0.990	0.066	0.065	1.484	0.005	1.648	0.005
Age (freed model)	303 (34)		0.993	0.060	0.009	1.440	0.005	1.617	0.005
Weight (constrained model)	641 (38)	319 (4) p<0.001	0.983	0.087	0.118	1.489	0.005	1.659	0.005
Weight (freed model)	322 (34)		0.992	0.063	0.009	1.440	0.005	1.493	0.005
Pain (constrained model)	511 (38)	208 (4) p<0.001	0.987	0.075	0.093	1.486	0.005	1.651	0.005
Pain (freed model)	303 (34)		0.993	0.060	0.009	1.428	0.005	1.581	0.005
Surgery (constrained model)	344 (38)	46 (4) p<0.001	0.992	0.058	0.042	1.501	0.005	1.631	0.005
Surgery (freed model)	298 (34)		0.993	0.057	0.009	1.486	0.005	1.617	0.005
Ethnicity (constrained model)	407 (38)	104 (4) p<0.001	0.990	0.064	0.063	1.501	0.005	1.635	0.005
Ethnicity (freed model)	303 (34)		0.993	0.058	0.009	1.487	0.005	1.584	0.005
All covariates (constrained model)	1049 (78)	707 (20) p<0.001	0.973	0.077	0.134	1.490	0.005	1.654	0.005
All covariates (freed model)	342 (58)		0.992	0.048	0.007	1.309	0.005	1.370	0.005

6.4.5.2 Model estimates in growth models adjusted for covariates

6.4.5.2.1 Femur

In univariable models for females, the covariates that explained the most variance in the intercept were body weight, knee pain, and history of knee surgery (8.1%, 7.3% and 3.1% respectively) (Table 34), with body weight and knee pain also explaining variance in the slope (10% each in the male model). Female participants with pain scores one SD above the mean had higher intercepts (~0.148 units higher, 95% CI 0.128, 0.168) and marginally steeper slopes (~0.01 units higher, 95% CI 0.01,0.01) than those with mean pain. Women that reported having had knee surgery started off with higher femur scores (~0.88 units greater, 95% CI 0.70,1.07) than those without, and they also had steeper slopes (~0.05 units higher, 95% CI 0.03,0.07). Compared to non-whites, female Caucasians had lower intercepts (~0.68 units lower, 95% CI 0.52,0.82) and shallower slopes (~0.02 units lower, 95% CI 0.00,0.03).

In males the findings were similar to those seen in females as the same covariates explained the most variance in intercept and slope and both weight and knee pain explained slope variance. However there were differences in males as knee pain was more important in explaining intercept variance (5.5.% variance explained), unlike in females where it was body weight, followed by history of knee surgery (3.9% variance explained) and body weight (3.7% variance explained). Another difference seen in males was that slopes did not differ statistically by ethnicity. In both genders baseline age was not associated with the intercept or slope (Table 33).

When all covariates were included, compared to the model with all covariates fixed to zero, there were statistically significant differences between the 2 nested models (change in chi-square =919 (20 DF), $p<0.001$) and much better fit showing that adjusting for covariates significantly explained intercept and slope variance (SRMR in the free model =0.003 vs 0.152 in the model with covariates fixed to zero). In the fully adjusted female model all covariates explained 21.4% intercept variance and 10% of the slope variance while in males all covariates explained 13.9% variance but no overall slope variance. Table 35 shows the estimates for each covariate in the fully adjusted models, these did not vary significantly from the univariable analyses.

6.4.5.2.2 Tibia

The predictive covariates were very similar to those described for the femur above with the notable exception being ethnicity, which in males was not significantly associated with either the intercept or slope (Table 33). The covariates that explained the most variance in the intercept were body weight, knee pain, and history of knee surgery (5.2%, 5.1% and 3.0% respectively in females).As seen in the femur, both weight and pain explained slope variance (10% slope variance explained by each). The order of importance of these covariates in the tibia models was also consistent in both genders. Compared to non-whites, female Caucasians had lower intercepts (~0.30 units lower, 95% CI 0.16,0.42) and shallower slopes (~0.02 units lower, 95% CI 0.00,0.03).

In the fully adjusted tibia models, ethnicity was not predictive in both males and females but the estimates for all covariates were very similar between univariable and fully adjusted models. The fully adjusted tibia model explained 13.4% variance in the intercept and 10% slope variance in females and 11.2% and 14.3% in males respectively.

6.4.5.2.3 Patella

Unlike the femur and tibia, no covariates explained any variance in patella slopes in all univariable models for both genders (Table 33-35). As seen for the previous bones, body weight and knee pain explained the most variance in the intercept (10% and 4.2% variance explained respectively) but unlike the tibia, the association with ethnicity in the patella was similar to that seen for the femur (statistically lower intercepts in Caucasian males and females): in females (~0.53 units lower,95% CI 0.42,0.64), but with no slope associations in both genders as alluded to earlier.

In the fully adjusted tibia models, differences were seen for ethnicity as there were significant intercept differences by race in both genders unlike seen for the tibia before, however and as noted before, no slope effects. Estimates of the other covariates were also very similar to those from the univariable models. The fully adjusted patella model explained 17.4% variance in the intercept and 20% slope variance in females and 12.6% intercept variance in males.

Table 33. Effect of covariates on growth curves in males

	Male models			
Covariate	**Estimate Intercept (95% CI)**	**Estimate slope (95% CI)**	**Intercept variance explained (%)**	**Slope variance explained (%)**
Femur univariable				
Age	0.015 (0.007, 0.024)	0.000 (0.000, 0.001)	0.7	-
Weight	0.023 (0.018, 0.029	0.001(0.001, 0.001)	3.7	10.0
Ethnicity	-0.403 (-0.613, -0.194)	0.001 (-0.001, 0.016)	0.7	-
Pain	0.154 (0.125, 0.183)	0.008 (0.006, 0.010)	5.5	10.0
Surgery	0.746 (0.582, 0.910)	0.019 (0.007, 0.030)	3.9	-
All covariates (multivariable)	-	-	**13.9**	**-**
Tibia univariable				
Age	0.019 (0.011, 0.027)	0.000 (-0.001, 0.000)	-	-
Weight	0.023 (0.017, 0.001)	0.001(0.000, 0.001)	3.8	14.3
Ethnicity	-0.020 (-0.219, 0.179)	-0.003 (-0.017, 0.011)	-	-
Pain	0.114 (0.086, 0.142)	0.006 (0.004, 0.008)	3.3	14.3
Surgery	0.579 (0.423, 0.736)	0.011(0.000, 0.022)	2.6	-
All covariates (multivariable)	**-**	**-**	**11.2**	**14.3**
Patella univariable				
Age	0.022 (0.016, 0.028)	0.000 (-0.001, 0.001)	3.0	-
Weight	0.015 (0.011, 0.019)	0.000 (0.000, 0.000)	3.3	-
Ethnicity	-0.326 (-0.475, -0.178)	0.006 (-0.001, 0.023)	1.0	-
Pain	0.091 (0.070, 0.111)	0.001 (-0.001, 0.004)	3.9	-
Surgery	0.259 (0.141, 0.377)	-0.008 (-0.021, 0.005)	1.0	-
All covariates (multivariable)	**-**	**-**	**12.6**	**-**

Ethnicity reference = non-white group; Pain = WOMAC Knee pain; Surgery = history of knee surgery (includes meniscectomy, arthroscopy and ligament repair)

Table 34. Effect of covariates on growth curves in females

	Female models			
Covariate	**Estimate Intercept (95% CI)**	**Estimate slope (95% CI)**	**Intercept variance explained (%)**	**Slope variance explained (%)**
Femur univariable				
Age	0.019 (0.011, 0.027)	-0.001 (-0.001, 0.000)	0.9	-
Weight	0.035 (0.030, 0.039	0.002 (0.002, 0.003)	8.1	5.0
Ethnicity	-0.675 (-0.830, -0.520)	-0.018 (-0.033, -0.004)	2.6	-
Pain	0.148 (0.128, 0.168)	0.011 (0.009, 0.013)	7.3	5.0
Surgery	0.881 (0.696, 1.066)	0.052 (0.034, 0.069)	3.1	-
All covariates (multivariable)	**-**	**-**	**21.4**	**10.0**
Tibia univariable				
Age	0.020 (0.011, 0.027)	0.000 (-0.001, 0.000)	1.4	-
Weight	0.024 (0.020, 0.028)	0.002 (0.001, 0.002)	5.2	10.0
Ethnicity	-0.291 (-0.422, -0.161)	-0.020 (-0.032, -0.009)	0.7	-
Pain	0.103 (0.086, 0.142)	0.008 (0.086, 0.121)	5.1	10.0
Surgery	0.714 (0.423, 0.736)	0.027 (0.000, 0.022)	3.0	-
All covariates (multivariable)	**-**	**-**	**13.4**	**10.0**
Patella univariable				
Age	0.020 (0.014, 0.025)	0.000 (-0.001, 0.001)	1.7	-
Weight	0.028 (0.024, 0.031)	0.001 (0.000, 0.001)	10.0	-
Ethnicity	-0.528 (-0.639, -0.416)	0.007 (-0.006, 0.019)	3.1	-
Pain	0.080 (0.070, 0.111)	0.004 (-0.001, 0.004)	4.2	-
Surgery	0.321 (0.186, 0.456)	0.015 (0.000, 0.029)	1.4	-
All covariates (multivariable)	**-**	**-**	**17.4**	**20.0**

Ethnicity reference = non-white group; Pain = WOMAC Knee pain;

Surgery = history of knee surgery (includes meniscectomy, arthroscopy and ligament repair)

Table 35. Fully adjusted growth models for all 3 bones separately

	Male models				Female models			
Covariate	**Estimate Intercept (95% CI)**	**p-value**	**Estimate slope (95% CI)**	**p-value**	**Estimate Intercept (95% CI)**	**p-value**	**Estimate slope (95% CI)**	**p-value**
Femur								
Age	0.015 (0.007, 0.024)	<0.001	0.000 (0.000, 0.001)	0.091	0.034 (0.027, 0.042)	<0.001	0.000 (-0.001, 0.001)	0.882
Weight	0.021 (0.016, 0.027)	<0.001	0.001(0.001, 0.001)	<0.001	0.030 (0.025, 0.035)	<0.001	0.002 (0.001, 0.002)	<0.001
Ethnicity	-0.335 (-0.559, -0.112)	0.003	0.010 (-0.006, 0.025)	0.210	-0.280 (-0.445, -0.114)	0.001	0.011 (-0.004, 0.026)	0.162
Pain	0.121 (0.091, 0.183)	<0.001	0.007 (0.005, 0.009)	<0.001	0.098 (0.077, 0.119)	<0.001	0.009 (0.007, 0.011)	<0.001
Surgery	0.767 (0.599, 0.935)	<0.001	0.017 (0.006, 0.029)	0.003	0.832 (0.645, 1.018)	<0.001	0.035 (0.017, 0.052)	<0.001
Tibia								
Age	0.028 (0.020, 0.036)	<0.001	0.000 (-0.001, 0.000)	0.803	0.029 (0.022, 0.035)	<0.001	0.000 (0.000, 0.001)	0.720
Weight	0.022 (0.017, 0.027)	<0.001	0.001(0.000, 0.001)	0.001	0.022 (0.017, 0.026)	<0.001	0.001 (0.001, 0.002)	<0.001
Ethnicity	0.013 (-0.202, 0.229)	0.904	0.003 (-0.012, 0.018)	0.687	-0.021 (-0.165, 0.123)	0.774	0.000 (-0.012, -0.013)	0.939
Pain	0.094 (0.065, 0.123)	<0.001	0.006 (0.004, 0.008)	<0.001	0.077 (0.059, 0.096)	<0.001	0.007 (0.005, 0.009)	<0.001
Surgery	0.597 (0.435, 0.759)	<0.001	0.008 (-0.003, 0.019)	0.154	0.633 (0.471, 0.795)	<0.001	0.015 (0.001, 0.029)	0.042
Patella								
Age	0.029 (0.023, 0.034)	<0.001	0.000 (-0.001, 0.001)	0.853	0.032 (0.026, 0.037)	<0.001	0.000 (-0.001, 0.001)	0.139
Weight	0.015 (0.012, 0.019)	<0.001	0.000 (0.000, 0.000)	0.890	0.026 (0.023, 0.030)	<0.001	0.001 (0.000, 0.001)	0.084
Ethnicity	-0.244 (-0.403, -0.085)	0.003	0.010 (-0.009, 0.028)	0.296	-0.331 (-0.452, -0.210)	<0.001	0.016 (0.002, 0.030)	0.022
Pain	0.074 (0.053, 0.095)	<0.001	0.002 (-0.001, 0.004)	0.216	0.041 (0.026, 0.057)	<0.001	0.004 (0.002, 0.006)	<0.001
Surgery	0.274 (0.154, 0.393)	<0.001	-0.011 (-0.024, 0.002)	0.107	0.300 (0.163, 0.436)	<0.001	0.004 (-0.011, 0.019)	0.623

Ethnicity reference = non-white group; Pain = WOMAC Knee pain;

Surgery = history of knee surgery (includes meniscectomy, arthroscopy and ligament repair)

6.5 Discussion

This is the first study to assess the baseline and longitudinal associations between three knee bones measured quantitatively using 3D imaging technology, providing data on the possible natural history of 3D bone shape. The OAI cohort is well suited to assess OA progression over time, and therefore useful in work pertaining to development of new knee imaging biomarkers. This study found that all three bones were positively correlated with each other at baseline, with stronger associations between the femur and tibia, than between the tibia and patella or femur and patella. The most likely diseased bone combinations involved the femur and tibia bones and least likely involved patella.

It was established that all three bones changed linearly over the four year follow-up, and this was consistent when 8 year follow-up data was used as part of sensitivity analyses showing that once the knee bone was diseased it worsened continuously following a linear disease trajectory. The rate of change in the femur was approximately twice that of the tibia and patella, and females had twice the rate of change of males. A worse disease score at baseline was associated with more rapid structural worsening in all bones. When parallel process models were considered, statistically significant associations between baseline disease status and rate of change in all three bones were shown, suggesting that all bones started with similar disease status and all progressed in the same direction although the rate of progression was higher in the femur.

There was significant variation between individuals` starting points and rates of change which was then modelled using clinical covariates. In all bones, important predictors of baseline disease status and longitudinal change were consistently participant body weight, knee pain and a history of knee surgery; however these effects were smaller in the patella than in the femur and tibia, and furthermore none predicted its rate of change. The patella bone was less responsive and the variation in patella rate of change could not be explained using any of the covariates explored in this study. These findings were also consistent for both genders.

A strength of this work was that longitudinal change was measured using latent growth curve models, a more advanced analytical technique that offer several advantages compared to traditional analysis methods such ANOVA/ANCOVA and even multilevel approaches [533-536], including incorporating latent variables to assess change and simultaneously analysing parallel process growth models enabling the evaluation of the

relationships between intercept and slope factors of all three bones in one model, where all factors were allowed to co-vary.

The femur changed much more rapidly than both tibia and patella which is supported by work from the Leeds OA research group showing that the femur is more responsive longitudinally than both patella and tibia in Chapter 5 [405]. Analyses were stratified by gender because *Wise* and colleagues [402] have also shown, using 3D shape technology that the shapes of the distal femur and proximal tibia differ by gender. They have also shown more recently that the risk of knee OA was higher in females due to differences in femur and tibia bone shapes [403]. The findings from this study support those results and in addition to the cross-sectional associations, this study showed that over time females worsened at a faster rate than males. While this study and both studies by *Wise* et. al used 3D models derived from SSMs, theirs were derived from radiographs while this study used bone shape derived from MRI. It is also well recognised that females have a higher prevalence of OA, and an increased risk of knee replacement [53, 537, 538] which may be independent of the shape differences suggested.

Across all bones, age was not a strong predictor of the baseline OA status nor the rate of change. Age is thought to be one of the strongest risk factors of OA [531] but its direct effects are not clearly understood, it is possible the relationship between OA and age could be mediated by age-related factors such as increased muscle weakness, ligamentous laxity, decreased proprioception, cartilage thinning accompanied by poor anabolic response to growth factors and loss of chondrocytes [51, 52]. In this study only slight differences were seen between participants with OA bone shape vs those without (~ 2 year age difference at baseline between these groups) and age had minor effects on the intercept, with no age effects on all slopes. The concept that aging contributes to, but is not causally related to OA, is consistent with the multifactorial idea of OA and the knowledge that not all older adults develop OA and not all joints in the body are affected to the same degree. Also, radiographic changes especially osteophytes are prevalent in the aged population but symptoms of joint pain may be independent of radiographic severity in many older adults [49].

Unsurprisingly this study found that body weight and knee pain were moderately related to baseline bone shape scores consistently across all three bones and their rates of change over time in the femur and tibia. Being overweight is a clear risk factor for OA development and population based studies have consistently shown a link between being overweight or obese and knee OA. Results from the first National Health and Nutrition Examination Survey (NHANES I) showed that obese women had

approximately a 4-fold increase in the risk of knee OA compared with non-obese women; and similar comparisons in men showed the risk was nearly 5-fold greater in obese men [539]. In the Framingham study overweight individuals in their thirties who did not have knee OA were at greater risk of later developing the disease [84]. Another longitudinal study using radiography and with 12-year follow-up, found that being overweight significantly increased the risk of developing knee OA [540]. It is also appreciated that joint pain is strongly associated with body weight and this is likely due to increased loading on the knee which increases stress and may lead to structural damage [541]. Possible sources and causes of pain in patients with osteoarthritis include among others the synovial membrane, joint capsule, periosteum, and subchondral bone [541]. However the role of pain in OA patients may also be the outcome of a complex interplay between structural changes in the affected joint, peripheral and central pain mechanisms, and subjective differences in what constitutes pain influenced by culture, gender, and psychosocial factors [542].

Compared to participants reporting no knee surgery, those undergoing knee procedures had higher baseline scores and also had more rapid worsening over time (possibly due to altered joint biomechanics). A history of knee surgery in this study included meniscectomy, arthroscopy and ligament repair and this variable was likely acting as a proxy for meniscal and ligament damage in the knee. Meniscal damage is increasingly being appreciated as a major risk factor for the development of OA [297, 495, 499], results from the MOST study showed about a 6-fold increased risk in OA in participants with significant meniscal damage compared to those without [281]. Similarly anterior cruciate ligament (ACL) injury is also known to be a risk factor for the development of knee OA therefore this predictive association with surgery was expected and confirmed what is already known about the common OA risk factors. In one exploratory study ACL injury was associated with significant changes in articulating bone curvature over a 5 year period [543] and using 3D bone measures similar to this study, Bowes *et. al.* showed that rapid bone changes occurred after ACL injury similar to those seen in established OA [544]. Despite this, no associations were seen in the patella potentially due to the measurement error associated with segmenting the patella due to its smaller size or other factors currently unknown. It is however possible that as the patella is an extra-articular structure, due to its anatomical location may have no association with meniscal and ligament damage occurring in closer proximity to the femur and tibia bones.

Like previous studies, this study found that OA status and severity differed by ethnicity [539, 545, 546]. Non-whites in this study were more likely to have higher baseline

disease and slightly higher rates of change, and this finding was more pronounced in females. Interestingly no differences were seen for the tibia in males. Ethnicity did not explain any significant variance for any of the univariable male models but showed an association in all female univariable models, consistent with findings from the HANES I study that suggest ethnicity is more important in females [539]. Braga *et al* [546] suggested that racial differences in females could be explained by discrepancies in BMI but as males in their study had similar mean BMI scores between races they suggest that other factors such as genetics, bone mineral density and lifestyle factors, may be at play.

Viewed as a whole joint, there are similarities in baseline status and longitudinal change in all three bones over time. The sources of variation were very similar between the bones but not all variation could be explained by the models specified and this was more evident for the patella. One reason could be the assumption of homogeneity among the participants, and future work using growth mixture models which aim to uncover unobserved heterogeneity in a population, and to find substantively meaningful groups of people that are similar in their responses to measured variables, or have similar growth trajectories may be warranted. Also, some other important covariates such as muscle strength that may explain some of the variance may have been omitted although the analyses tried to encompass what are commonly cited as OA risk factors.

6.6 Limitations

There were some limitations to this work and readers should interpret the results cautiously. First, the proportion of missing data was high in some years due to the absence of MRI images, however the FIML missing data modelling techniques were implemented to address this and the chosen models problem and showed good fit. Second, the measurement error for patella bone measures was higher than femur and tibia because segmenting the patella is more difficult due its shape and size and may have contributed to the differences observed. The choice in picking the right knee was random and although results using the left knee were slightly different in terms of the magnitude of estimates, however the model fit and the conclusions using data from the left knee was the same as the right. Estimates of variance between the left knees and right knees were also similar. Also, although parallel process models used all three bones it was not specified *a priori* what the relationship between each bone was, and thus the only parameters modelled among the three were bones were just covariances. It is possible that one of the bones may act as “initiator” of the structural damage that occurs in the knee, influencing the other bones to then start deteriorating and therefore

these relationships could be modelled more accurately by regressing growth factors against each other to give more reliable estimates. However, as this is still unknown, these relationships were not be tested using this data-driven approach but this could be an interesting research question. While adjustment for clinical covariates reduced some of the intercept and slope variance, some unexplained residual variance still remained after these adjustments and that could be attributed to other unmeasured factors or measurement error in the measures.

6.7 Conclusion

This study has provided a hypothesis of what the natural history of bone shape looks like. Latent growth curve analysis has provided evidence that the direction of change is the same in all three knee bones and that they share common determinants, but vary by gender. This suggests that all are part of a single disease process, with the femur providing the greatest amount of change of the three bones.

Thus far all three knee bones have been shown to change linearly, with this change assumed homogenous among all participants and summarised by the group mean slope. However, the significant heterogeneity in intercept and slope but particularly in for intercepts was still observed even after adjusting for clinical covariates. This suggests that while mean change assumed for the whole group is useful to show the overall group change, it however does not reveal the subtle changes that may be occurring within this sub-sample. It is important to uncover the subtle patterns in growth that may exist because knowing which classes or trajectory groups exist could be an important step in optimally phenotyping individuals for clinical trials that measure structure. The existence of trajectory classes is explored in Chapter 7.

Chapter 7

Defining bone shape trajectories in three knee bones: the Osteoarthritis Initiative.

This chapter presents the results of the investigation into the trajectory of bone shape changes seen for the three knee bones (femur, tibia and patella) using latent class growth analysis. The three knee bones have been previously shown to change linearly but this change has been assumed homogenous among all participants. However results from LGCM (Chapter 6) suggest that significant variation in the growth parameters of all bones is present and it is important to uncover the subtle patterns in growth that exist. Knowing which classes or trajectory groups exist could be an important step in phenotyping individuals for clinical trials.

7.1 Introduction

Despite much research on cartilage and its favoured status as an FDA end point, its natural history as measured by JSW loss or novel 3D cartilage measures is still relatively unknown and, few risk factors have to date been consistently identified for knee OA progression [531]. There is, however, evidence that considerable variability exists in structural disease among patients with OA and this has been demonstrated in a few studies using both radiographic and MRI-based cartilage measures[531, 547]. As shown in Chapter 6, bone pathology is important in OA pathogenesis with potential to be a treatment and clinical trial target [111, 340, 393, 405]. However like cartilage, bone shape natural history is still not well studied. While 3D bone shape in three knee bones (femur, tibia and patella) was shown to change linearly over time (Chapter 6), it was also evident that significant variability exists with respect to where individuals start off in terms of structural damage, as well as their rates of change over time (Table 24, Chapter 6). This variability suggests that “OA” change is not homogenous and using a one-group mean intercept and slope does not reveal other important changes occurring within this OA sub-population.

Although OARSI guidelines exist [548], there are currently no universally accepted risk stratification criteria (based on structure) for selecting participants into OA targeted - treatment trials. This could be attributed to design or methodological/ statistical issues [549], or the use of measures that lack responsiveness longitudinally [4]. Bone shape is

now known to be a very responsive measure [339, 405] over follow-up times typical of those seen in OA trials although these studies have mainly focused on just the femur. As OA is a heterogeneous disease, identifying subgroups that could benefit more from targeted treatments is one of the promising advances in clinical research. The observation that variability or homogeneity exists in OA is not a new finding and has previously been suggested by other studies [550, 551]. As has been shown previously, the OAI provides a sufficiently large dataset with sufficient follow-up time to assess the longitudinal changes, enabling us to understand which bone might provide the most optimum stratification criteria to measure structural change.

7.2 Aims

This study sought to identify distinctive trajectories of 3D bone shape in the three knee bones, and assess their associated risk factors using data from the OAI. Another aim was to assess whether these trajectory groups were similar in all three bones and whether they were associated with OA clinical factors and structural end points longitudinally, using advanced statistical techniques.

7.3 Methods

7.3.1 Participants

This study included all 4796 participants using both knees per participant from baseline to 4 year follow-up (5 time points) as included previously in Chapter 6.

7.3.2 Bone shape measures

The same 3D bone shape measures as previously described in Chapter 6 were used for this study. Their derivation is as described in Chapter 3 (Methods).

7.3.3 Statistical analysis

Latent class growth analysis modelling (LCGA) was performed in Mplus software V.8.1 and descriptive analyses in STATA Corp V.13.1.Three bones (femur, tibia and patella) were analysed separately and all analyses stratified by gender. Having tested whether the longitudinal change in bone shape was best described by linear, quadratic or cubic trajectories, the linear model was found to best fit the data (Chapter 6) and it was also established that growth factors (intercept and slope) showed significant heterogeneity (Table 24, Chapter 6.).

LCGA was used to identify possible trajectories of bone shape change over time. The fit of the bone shape trajectories was tested for two to nine trajectory classes. The most optimal model was chosen based on a combination of fit indices and interpretability of the model. The following indices of fit were used: adjusted-Bayesian information criteria (a-BIC), Vuong–Lo–Mendell–Rubin likelihood ratio test (LRT) and bootstrap likelihood ratio test (BLRT). Significant LRT and BLRT for k groups ($p<0.05$) indicate that the fit of the specific model is an improvement over a model with k−1 groups. Entropy indices (ranging 0 –1) were checked to ensure the quality and reliability of the classification. An index close to 1 indicates a good classification. After determining the number of trajectories, the best possible/optimal model was also determined by considering the usefulness of the latent classes (ensuring at least > 5% in each class), inspecting the trajectory shapes for similarity, the number of individuals in each class, and also the posterior probabilities of group membership from each individual. The Guidelines for Reporting on Latent Trajectory Studies (GRoLTS)-Checklist guided the reporting of the trajectory analyses[552].

Next, multinomial logistic regression assessed the association of baseline risk factors with trajectory of bone shape change. These risk factors included demographic (age, race), disease severity (knee pain severity, history of knee surgery) and modifiable risk factors (obesity and comorbidity scores). Descriptively, the number of total knee replacements that occurred after the baseline visit and physical functioning using SF-12 scores were shown. As the initial analyses considered the three bones separately, the next step of analysis was to identify a group of participants that were classed into the same trajectory class for all 3 bones, and the multinomial logistic regression analyses above repeated as part of sensitivity analyses.

Each study subject contributed two knees to analyses however, LCGA methodology is unable to account for the within-person correlations. Therefore, the right knee from each participant was selected. This was justifiable as previous sensitivity analyses (Chapter 6) showed that the estimates using either knee were not different.

7.4 Results

7.4.1 Descriptive analysis

The study included the full OAI sample of 4796 participants and for LCGA only 4-year follow up was used. Descriptive summaries of this sample have been described previously. The results showed that 3 trajectories existed over 4 years (Tables 36 & 37, Figures 41-43,) in both genders and for all 3 bones. Trajectory 1 (slowest changers)

were characterised by having the lowest values at baseline and slowest rates of change, trajectory 2 (intermediate changers) had starting values higher than the slowest change group, had steeper slopes but less than trajectory 3 (fastest changers) who were characterised by the highest starting points and greatest range of change.

Tables 36 and 37 show the baseline characteristics for each trajectory group. As seen, participants in the fastest changer groups had higher prevalence of comorbidities, greater proportions in the higher KL grades, higher proportion in non-white ethnic group, and lower physical functioning scores (SF-12) compared to those in the slowest and intermediate changers groups. When assessed longitudinally the fastest changers recorded more incidents of total knee replacements during follow–up (right knee replacements, n = 119 in fastest changers vs n = 65 in intermediate changers vs n = 14 in the slowest changers group). These trends were similar for all bones and all genders but there was slightly better separation of trajectory classes, as measured using descriptive characteristics in women (Table 36).

A group of 971 female participants had consistent classification in all three bones (20% in the fastest group for all bones, 44% intermediate group and 36% slowest changers), the fastest group representing 7% of all females in the OAI. In males 667 participants had similar classification (17%, 46% and 37%) in the fastest, intermediate and slowest changers respectively, the fastest group representing 5.6% of the male OAI population. Characteristics of participants classified this way were similar to those shown in Tables 36 & 37 although the separation between trajectory groups was more evident for example the fastest group was composed of 40% non-white ethnicity in females vs 13% non-white ethnicity group in the slowest changers (Table 38).

Table 36. Baseline characteristics across trajectory groups in females

		Trajectory group		
Characteristic	**All females (N=2778)**	**Slowest changers (N=1140)**	**Intermediate group (N=1241)**	**Fastest changers (N=397)**
FEMUR				
Age, median (IQR)	61 (54-69)	60 (53-68)	61 (54-69)	63 (57-70)
BMI, median (IQR)	28.1 (24.4-32)	26 (23.2-30)	28.7 (25.1-32.6)	31.2 (27.4-34.8)
Ethnicity, non-white %	24	18	26	37
SF-12, physical summary	50.8 (42.5-56.0)	52.5 (46.0-56.1)	50.2 (42.4-55.5)	44.7 (36.6-52.4)
Comorbidity score, (0-10) %				
0	75	78	75	66
1	16	15	17	20
>=2	9	7	8	14
Knee pain (0-10), median (IQR)		1 (0-4)	3 (0-5)	4 (2-7)
KL grade (%)				
0	36	59	27	1
1	18	21	21	1
2	31	18	38	46
3 or 4	15	2	14	52
TKR, n after baseline	131	12	46	73
TIBIA				
Age, median (IQR)	61 (54-69)	59 (53-66)	62 (54-70)	63 (55-69)
BMI, median (IQR)	28.1 (24.4-32)	26.6 (23.5-30.7)	28.2 (24.5-32)	30.1 (26.5-34.1)
Ethnicity, non-white %	24	21	23	34
SF-12, physical summary	50.8 (42.5-56.0)	52.4 (45.1-56.1)	50.6 (43.0-55.9)	46.3 (38.0-53.4)
Comorbidity score, (0-10) %				
0	75	78	75	70
1	16	15	16	19
>=2	9	7	9	11
Knee pain (0-10), median (IQR)	3 (0-5)	2 (0-4)	2 (0-5)	4 (1-6)
KL grade (%)				
0	36	54	33	4
1	18	21	20	4
2	31	20	34	45
3 or 4	15	5	13	50
TKR, n after baseline	131	10	53	73
PATELLA				
Age, median (IQR)	61 (54-69)	59 (52-66)	62 (54-69)	63 (56-70)
BMI, median (IQR)	28.1 (24.4-32)	25.9 (23.2-29.8)	28.2 (24.5-31.8)	30.8 (26.9-34.6)
Ethnicity, non-white %	24	16	25	35
SF-12, physical summary	50.8 (42.5-56.0)	52.8 (46.4-56.7)	50.8 (42.7-55.9)	46.9 (38.1-53.0)
Comorbidity score, (0-10) %				
0	75	78	77	67
1	16	16	14	22
>=2	9	6	9	11
Knee pain (0-10), median (IQR)	3 (0-5)	1 (0-4)	3 (0-5)	4 (0-6)
KL grade (%)				
0	36	57	34	11
1	18	21	20	7
2	31	17	32	47
3 or 4	15	5	13	35
TKR, n after baseline	131	9	44	78

Table 37. Baseline characteristics across trajectory groups in males

		Trajectory group		
Characteristic	**All males (N=1992)**	**Slowest changers (N=792)**	**Intermediate group (N=907)**	**Fastest changers (N=271)**
FEMUR				
Age, median (IQR)	59 (53-70)	58 (52-68)	61 (53-70)	61 (54-69)
BMI, median (IQR)	28.5 (25.7-31.5)	27.2 (24.9-30.5)	28.8 (26.3-31.7)	30 (27.4-33.3)
Ethnicity, non-white %	16	13	18	22
SF-12, physical summary	52.1 (44.8-56.2)	53.2 (46.5-56.5)	51.9 (43.9-56.1)	49.6 (41.2-54.3)
Comorbidity score, (0-10) %				
0	75	79	74	72
1	14	12	15	17
>=2	11	9	11	11
Knee pain (0-10), median (IQR)	2 (0-4)	1 (0-4)	2 (0-4)	3 (1-5.5)
KL grade (%)				
0	39	60	33	1
1	18	21	20	2
2	23	16	29	26
3 or 4	20	3	18	71
TKR, n after baseline	67	2	19	46
TIBIA				
Age, median (IQR)	59 (53-70)	58 (52-67)	61 (54-70)	62 (54-70)
BMI, median (IQR)	28.5 (25.7-31.5)	27.7 (25.2-30.6)	28.8 (26.2-31.7)	29.7 (26.9-32.7)
Ethnicity, non-white %	16	16	16	20
SF-12, physical summary	52.1 (44.8-56.2)	53.2 (46.4-56.7)	51.6 (44.3-56.1)	49.7 (41.4-54.3)
Comorbidity score, (0-10) %				
0	75	79	73	76
1	14	12	17	12
>=2	11	9	10	12
Knee pain (0-10), median (IQR)	2 (0-4)	2 (0-4)	2 (0-5)	4 (1-6)
KL grade (%)				
0	39	59	32	6
1	18	21	19	6
2	23	13	32	21
3 or 4	20	7	17	67
TKR, n after baseline	67	4	26	37
PATELLA				
Age, median (IQR)	59 (53-70)	56 (51-65)	60 (53-70)	64 (55-71)
BMI, median (IQR)	28.5 (25.7-31.5)	27.4 (25-30.3)	28.6 (25.8-31.5)	29.5 (27-32.5)
Ethnicity, non-white %	16	13	16	23
SF-12, physical summary	52.1 (44.8-56.2)	52.9 (46.5-56.5)	52.3 (45.2-56.1)	50.2 (41.6-55.6)
Comorbidity score, (0-10) %				
0	75	79	75	72
1	14	12	14	15
>=2	11	9	11	13
Knee pain (0-10), median (IQR)	2 (0-4)	1 (0-4)	2 (0-4)	3 (0-5)
KL grade (%)				
0	39	57	40	15
1	18	18	21	12
2	23	18	25	30
3 or 4	20	7	14	43
TKR, n after baseline	67	2	23	42

Table 38. Descriptive analysis of participants classified in one group for all three bones

Females (N=971)	**Slowest changers (N=351)**	**Intermediate changers (N=423)**	**Fastest changers (N=197)**
Age, median (IQR)	59 (52-66)	62 (55-70)	63 (56-69)
BMI, median (IQR)	25.3 (22.9-28.9)	29.1 (25.3-32.7)	32.2 (28.6-35.3)
Ethnicity, non-white %	13	24	40
SF-12, physical summary	53.5 (46.9-57.5)	49.8 (41.4-54.5)	43.0 (36.4-51.9)
Comorbidity score, (0-10) %			
0	80	77	65
1	14	16	20
>=2	6	7	15
Knee pain (0-10), median (IQR)	0 (0-4)	3 (0-5)	4 (2-7)
KL grade (%)			
0	70	26	0
1	17	22	1
2	11	40	42
3 or 4	2	12	47
TKR, n after baseline	7	33	74
Males (N=667)	**Slowest changers (N=247)**	**Intermediate changers (N=307)**	**Fastest changers (N=113)**
Age, median (IQR)	55 (50-65)	62 (54-71)	64 (56-70)
BMI, median (IQR)	26.8 (24.4-30)	29 (26.3-31.7)	30.2 (26.9-33.4)
Ethnicity, non-white %	13	17	19
SF-12, physical summary	54.0 (48.8-57.5)	52.1 (44.7-56.1)	49.8 (41.3-54.7)
Comorbidity score, (0-10) %			
0	83	72	75
1	8	16	13
>=2	9	12	12
Knee pain (0-10), median (IQR)	1 (0-3)	2 (0-4)	3 (1-5)
KL grade (%)			
0	70	30	1
1	17	20	0
2	11	32	18
3 or 4	2	18	71
TKR, n after baseline	7	18	34

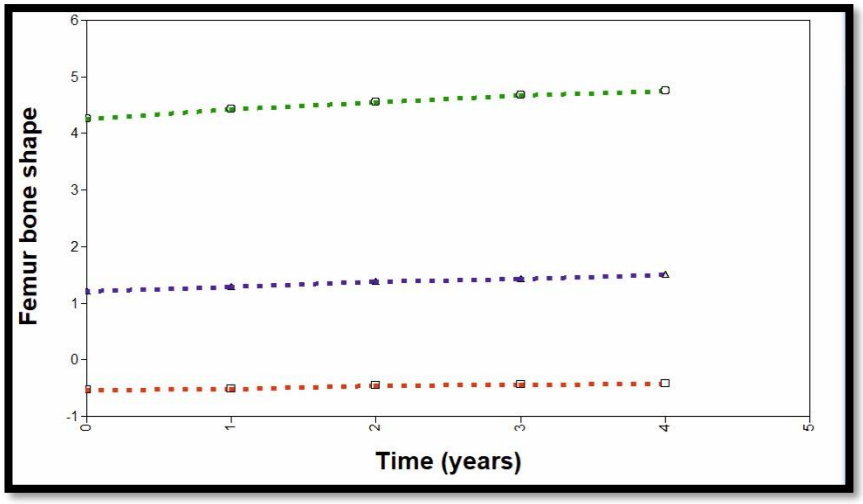


Figure 41. Bone shape trajectories in the femur

Red line = Class 1, 40% (slowest changers, Intercept = -0.53, Slope = 0.03)
Blue line = Class 2, 46% (intermediate changers, Intercept =1.21, Slope =0.07)
Green line = Class 3, 14% (fastest changers, Intercept = 4.28, slope =0.12)

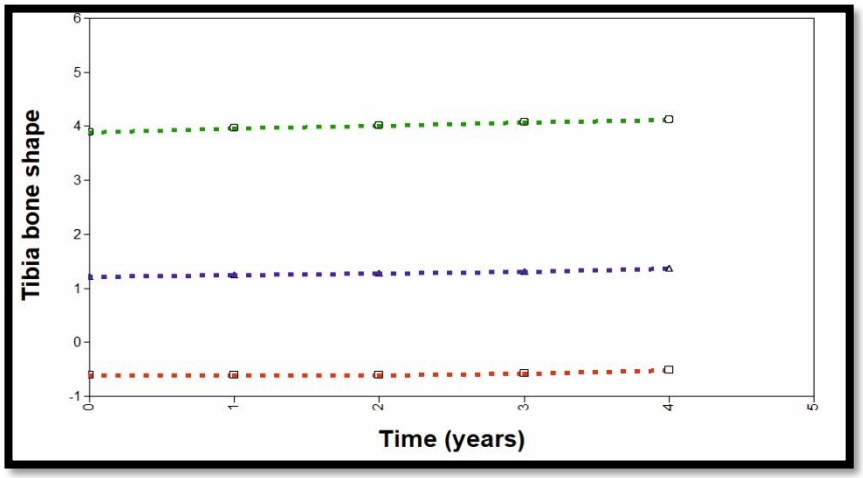


Figure 42. Bone shape trajectories in the tibia

Red line = Class 1, 40% (slowest changers, Intercept = -0.63, Slope = 0.02)
Blue line = Class 2, 47% (intermediate changers, Intercept =1.20, Slope =0.04)
Green line = Class 3, 13% (fastest changers, Intercept = 3.90, slope =0.06)

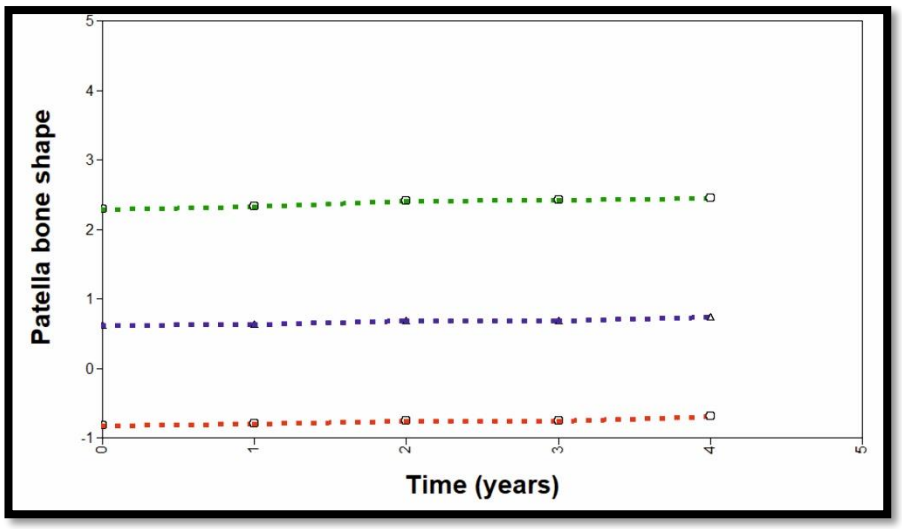


Figure 43.Bone shape trajectories in the patella

Red line = Class 1, 28% (slowest changers, Intercept = -0.83, Slope = 0.03)
Blue line = C ass 2, 51% (intermediate changers, Intercept =0.60, Slope =0.03)
Green line = Class 3, 21% (fastest changers, Intercept = 2.29, slope =0.04)

7.4.1.1 Classification quality

The posterior probability of allocating each participant into trajectories was mostly ≥ 0.95 (Table 39), indicating that there was a 95% probability on average of each trajectory class being correctly classified into the respective group trajectory. This was across both genders and for all three bones. Entropy values were also very high (ranging from 88.3% to 90.6%) in all bone shape models and for both genders, signifying good separation between classes, adequate model fit and correctly specified models. Model fit decision criteria was discussed in more detail in Chapter 3 and the indices for this study are displayed in Tables 40 and 41.

Table 39. Latent Class Probabilities for Most Likely Latent Class Membership

Females					Males			
	Class 1	**Class 2**	**Class 3**			**Class 1**	**Class 2**	**Class 3**
Femur					**Femur**			
Class 1	**96%**	0%	4%		Class 1	**95%**	5%	0%
Class 2	0%	**97%**	3%		Class 2	5%	**95%**	0%
Class 3	5%	1%	**94%**		Class 3	0%	2%	**98%**
Tibia					**Tibia**			
Class 1	**95%**	4%	1%		Class 1	**95%**	0%	5%
Class 2	4%	**96%**	0%		Class 2	0%	**98%**	2%
Class 3	3%	0%	**97%**		Class 3	4%	15	**95%**
Patella					**Patella**			
Class 1	**94%**	4%	2%		Class 1	**94%**	6%	0%
Class 2	5%	**95%**	0%		Class 2	3%	**95%**	2%
Class 3	4%	0%	**96%**		Class 3	0%	5%	**95%**

7.4.2 Bone shape trajectories

Parameter estimates were very similar across genders and the proportions of members in each class were also closely related. The 3-class solution was the most optimal solution for all bones and this was consistent in both genders. Although model fit improved in terms of lower AIC and BIC as well as slightly higher entropy values in the 4-class solution the LMR-LRT indices were not statistically significant (p>0.05) (Table 40 and 41).

7.4.2.1 Femur

In females trajectory 1 comprised 41% of females, had a baseline femur value (intercept) of -0.42 femur units (95% CI -0.54, -0.32) and statistically significant upward slope, 0.058 (95% CI 0.050, 0.066). Trajectory class 2, which comprised 45% of females, had intercept 1.46 (95% CI 1.29, 1.63) and slope 0.13 (95 % 0.12, 0.15) and trajectory latent class 3 comprised 14 % of females with intercept 4.38 (95% CI 4.06, 4.70) and the highest slope 0.18 (95% CI 0.15, 0.21). In general, the rates of progression were faster in women than in men. The proportions allocated into the 3 femur classes were similar in males (trajectory 1 = 46%, trajectory 2 = 46% and trajectory 3 =14%), see Table 2 for details on model fit and class membership details.

7.4.2.2 Tibia

In the tibia in females, the slope for trajectory class 3 (16% of the sample) was ~ 3.5 times the rate of the slowest changing group (38 %); for the intermediate group (46 % of the sample) the rate was ~ 1.7 times faster than the slowest group. Although the proportion of members in each group were similar in males and females, the rates of change were constantly higher in females (in trajectory class 3, slope ~ 0.11 in females vs 0.06 in males).

7.4.2.3 Patella

The proportions in each trajectory were slightly different in the patella. In females (trajectory 1 = 48%, trajectory 2 = 31% and trajectory 3 = 21%) while in males 51% were in trajectory 2 and 28% and 21% in trajectory 1 and 3 respectively. In females class 2 and 3 had very similar slopes with both ~ 2 times faster than the slowest changing group and their main difference being their baseline scores. In males the 3 trajectories were not well distinguished by the slopes (all slopes ~ 0.03)

Table 40. Trajectory classes and fit indices for 3D bone shape in females

Number of classes	Number of cases assigned			AIC	BIC	Adjusted -BIC	LMR-LRT p-value	BLRT p-value	Entropy
Femur									
	Trajectory	**N**	**%**						
1	Trajectory 1	2778	100	45 849	45 890	45 868			
2	Trajectory 1	2006	72	38 013	38 072	38 040	7525 <0.001	7842 <0.001	0.899
	Trajectory 2	772	28						
3	Trajectory 1	1140	41	32 988	33 065	33 024	4828 0.0086	5030 <0.001	0.897
	Trajectory 2	397	14						
	Trajectory 3	1241	45						
4	Trajectory 1	617	22	29 407	29 502	29 451	3442 0.24	3587 <0.001	0.906
	Trajectory 2	232	9						
	Trajectory 3	806	29						
	Trajectory 4	1123	40						
Tibia									
1	Trajectory 1	2778	100	41 873	41 915	41 892			
2	Trajectory 1	853	31	38 013	38 072	38 040	7196 <0.001	7498 <0.001	0.884
	Trajectory 2	1925	69						
3	Trajectory 1	1278	46	29 234	29 311	29 269	4946 <0.001	5153 <0.001	0.902
	Trajectory 2	1042	38						
	Trajectory 3	458	16						
4	Trajectory 1	658	24	26 263	26 359	26 308	2886 0.43	2975 <0.001	0.906
	Trajectory 2	705	25						
	Trajectory 3	1164	42						
	Trajectory 4	251	9						
Patella									
	Trajectory	N	%						
1	Trajectory 1	2778	100	38 308	38 349	38 327			
2	Trajectory 1	1731	62	31 211	31 270	31 239	6816 <0.001	7103 <0.001	0.861
	Trajectory 2	1047	38						
3	Trajectory 1	1337	48	26 872	26 949	26 908	4170 0.0001	4345 <0.001	0.883
	Trajectory 2	860	31						
	Trajectory 3	581	21						
4	Trajectory 1	1174	42	23 937	24 032	23 981	2822 0.20	2941 <0.001	0.906
	Trajectory 2	238	9						
	Trajectory 3	600	22						
	Trajectory 4	766	27						

Table 41.Trajectory classes and fit indices for 3D bone shape in males

Number of classes	Number of cases assigned			AIC	BIC	Adjusted -BIC	LMR-LRT p-value	BLRT p-value	Entropy
Tibia									
	Trajectory	**N**	**%**						
1	Trajectory 1	1970	100	32 225	32 264	32 241			
2	Trajectory 1	504	25	26 965	27 021	26 989	5044 0.0011	5265 <0.001	0.896
	Trajectory 2	1466	75						
3	Trajectory 1	792	40	23 167	23 239	23 198	3645 0.0007	3804 <0.001	0.903
	Trajectory 2	256	13						
	Trajectory 3	922	47						
4	Trajectory 1	891	45	20 728	20 817	20 766	2342 0.30	2445 <0.001	0.914
	Trajectory 2	119	6						
	Trajectory 3	535	27						
	Trajectory 4	425	22						
Femur									
1	Trajectory 1	1970	100	33 341	33 380	33 358			
2	Trajectory 1	1562	79	27 260	27 316	27 284	5830 <0.001	6086 <0.001	0.937
	Trajectory 2	408	21						
3	Trajectory 1	792	40	23 340	23 512	23 471	3665 <0.001	3826 <0.001	0.899
	Trajectory 2	907	46						
	Trajectory 3	271	14						
4	Trajectory 1	923	47	20 653	20 743	20 692	2675 0.29	2793 <0.001	0.919
	Trajectory 2	151	8						
	Trajectory 3	534	27						
	Trajectory 4	362	18						
Patella									
	Trajectory	N	%						
1	Trajectory 1	1970	100	27 402	27 442	27 419			
2	Trajectory 1	1192	61	22 553	22 609	22 577	4651 <0.001	4856 <0.001	0.849
	Trajectory 2	778	39						
3	Trajectory 1	554	28	19 451	19 523	19 482	2977 0.0052	3108 <0.001	0.885
	Trajectory 2	997	51						
	Trajectory 3	419	21						
4	Trajectory 1	210	11	17 457	17 546	17 496	1916 0.21	2000 <0.001	0.898
	Trajectory 2	802	41						
	Trajectory 3	342	17						
	Trajectory 4	616	31						

7.4.3 Association of demographic and disease factors with trajectory groups

Multivariable regression analyses (Table 42) showed that, after adjustment for all other baseline characteristics in the femur, participants in trajectory 3 (fastest changers), compared with participants in trajectory 1 (slowest changers), were characterised by: older age, being in the non-white ethnic group, obesity, higher pain scores and a history of surgery. For example being classified as obese was associated with odds ratio (3.91; 95% CI 2.99,5.17, p<0.001) and (1.91; 95% CI 2.99,5.17, p<0.001) of being in the fastest changers and intermediate changers trajectories respectively, compared with a non-obese individual. The corresponding odds ratios for being in the fastest changers and intermediate trajectories were 6.9, 1.9 respectively for history of knee surgery compared with knees not undergoing any surgery.

These findings were similar across genders and for all bones with differences in just the magnitudes of odds ratios. A few notable differences were seen for ethnicity where in the intermediate group no significant differences were seen for the femur and tibia in females, but in males ethnicity was significantly associated with the intermediate trajectory class (femur and patella bones), as well as the femur fastest changers.

In the group of participants whose bones were all classified into the fastest changers group, the trends seen were similar to those found for the individual bones mentioned above e.g. older age, being in the non-white ethnic group, obesity, higher pain scores and a history of surgery more likely to be associated with fastest changers (Table 43). However ethnicity in males did not seem to have a statistically significant effect suggesting that the different trajectory groups did not vary by ethnicity in men (Table 43).

Table 42. Association of demographic and OA risk factors with trajectory groups at baseline

	Slowest changes	Intermediate changers	Fastest changers
Females			
Femur			
Age 60-70 vs <60	1.0 (Reference)	1.26 (1.04,1.53)	2.29 (1.71,3.08)
Age> 70 vs <60	1.0 (Reference)	1.54 (1.22,1.95)	3.32 (2.32,4.74)
White vs non-white	1.0 (Reference)	**1.20 (0.97,1.50)**	1.53 (1.13,2.07)
Obese vs non-obese	1.0 (Reference)	1.91 (1.59,2.31)	3.93 (2.99,5.17)
Pain	1.0 (Reference)	1.08 (1.05,1.12)	1.25 (1.20,1.31)
Surgery yes vs none	1.0 (Reference)	1.88 (1.29,2.75)	6.91 (4.55,10.51)
Tibia			
Age 60-70 vs <60	1.0 (Reference)	1.55 (1.27,1.87)	2.00 (1.52,2.62)
Age> 70 vs <60	1.0 (Reference)	2.01 (1.58,2.55)	2.85 (2.05,3.97)
White vs non-white	1.0 (Reference)	**1.02 (0.82,1.28)**	1.33 (1.00,1.77)
Obese vs non-obese	1.0 (Reference)	1.45 (1.20,1.76)	2.20 (1.70,2.82)
Pain	1.0 (Reference)	1.06 (1.02,1.09)	1.19 (1.14,1.31)
Surgery yes vs none	1.0 (Reference)	**1.37 (0.95,1.99)**	4.36 (2.95,6.44)
Patella			
Age 60-70 vs <60	1.0 (Reference)	1.54 (1.26,1.89)	1.86 (1.43,2.43)
Age> 70 vs <60	1.0 (Reference)	2.18 (1.68,2.83)	4.00 (2.90,5.52)
White vs non-white	1.0 (Reference)	1.45 (1.13,1.85)	2.00 (1.50,2.67)
Obese vs non-obese	1.0 (Reference)	1.84 (1.50,2.27)	3.94 (3.06,5.06)
Pain	1.0 (Reference)	1.10 (1.06,1.14)	1.19 (1.14,1.24)
Surgery yes vs none	1.0 (Reference)	1.61 (1.10,2.37)	2.57(1.68,3.95)
Males			
Femur			
Age 60-70 vs <60	1.0 (Reference)	1.35 (1.06,1.72)	1.92 (1.34,2.74)
Age> 70 vs <60	1.0 (Reference)	1.83 (1.41,2.35)	2.39 (1.58,3.61)
White vs non-white	1.0 (Reference)	1.41 (1.06,1.88)	1.61 (1.07,2.41)
Obese vs non-obese	1.0 (Reference)	1.52 (1.23,1.89)	2.42 (1.77,3.31)
Pain	1.0 (Reference)	1.04 (1.00,1.09)	1.19 (1.13,1.27)
Surgery yes vs none	1.0 (Reference)	1.60 (1.20,2.13)	5.69 (4.01,8.06)
Tibia			
Age 60-70 vs <60	1.0 (Reference)	1.42 (1.12,1.80)	2.10 (1.46,3.00)
Age> 70 vs <60	1.0 (Reference)	1.83 (1.41,2.89)	2.88 (1.91,4.32)
White vs non-white	1.0 (Reference)	**0.91 (0.69,1.21)**	**1.10 (0.73,1.66)**
Obese vs non-obese	1.0 (Reference)	1.39 (1.12,1.71)	2.07 (1.51,2.83)
Pain	1.0 (Reference)	1.07 (1.02,1.11)	1.19 (1.12,1.26)
Surgery yes vs none	1.0 (Reference)	1.51 (1.14,1.99)	4.58 (3.23,6.50)
Patella			
Age 60-70 vs <60	1.0 (Reference)	1.87 (1.43,2.45)	2.68 (1.92,3.72)
Age> 70 vs <60	1.0 (Reference)	2.01 (1.49,2.71)	3.78 (2.63,5.43)
White vs non-white	1.0 (Reference)	1.54 (1.21,1.95)	1.94 (1.34,2.81)
Obese vs non-obese	1.0 (Reference)	1.54 (1.21,1.95)	2.23 (1.67,2.98)
Pain	1.0 (Reference)	**1.02 (0.97,1.07)**	1.10 (1.04,1.16)
Surgery yes vs none	1.0 (Reference)	2.01 (1.47,2.76)	2.69 (1.87,3.88)

Values are Odds ratio (95% CI).
Coefficients highlighted in bold were not statistically significant (p<0.05).

Table 43. Association of demographic and OA risk factors with trajectory groups from all 3 bone bones at baseline

Females (N=971)	**Slowest changers (N=351)**	**Intermediate changers (N=423)**	**Fastest changers (N=197)**
Age 60-70 vs <60	1.0 (Reference)	1.81 (1.27,2.58)	2.79 (1.72,4.52)
Age> 70 vs <60	1.0 (Reference)	3.31 (2.12,5.15)	5.90 (3.26,10.68)
White vs non-white	1.0 (Reference)	1.54 (1.00,2.38)	2.59 (1.55,4.33)
Obese vs non-obese	1.0 (Reference)	3.14 (2.19,4.51)	7.04 (4.46,11.10)
Pain	1.0 (Reference)	1.14 (1.07,1.21)	1.28 (1.19,1.39)
Surgery yes vs none	1.0 (Reference)	**1.24 (0.63,2.44)**	4.23 (2.09,8.56)
Males (N=667)	**Slowest changers (N=247)**	**Intermediate changers (N=307)**	**Fastest changers (N=113)**
Age 60-70 vs <60	1.0 (Reference)	2.50 (1.59,3.93)	4.17 (2.23,7.83)
Age> 70 vs <60	1.0 (Reference)	3.94 (2.38,6.52)	7.10 (3.50,14.41)
White vs non-white	1.0 (Reference)	**1.32 (0.77,2.25)**	**1.68 (0.82,3.40)**
Obese vs non-obese	1.0 (Reference)	2.00 (1.34,2.98)	3.43 (2.00,5.89)
Pain	1.0 (Reference)	**1.06 (0.98,1.14)**	1.17 (1.05,1.29)
Surgery yes vs none	1.0 (Reference)	3.17 (1.81,5.38)	11.96 (6.25,22.86)

Values are Odds ratio (95% CI).

Coefficients highlighted in bold were not statistically significant at ($p<0.05$).

7.5 Review of latent classes accounting for measurement noise

7.5.1 Defining measurement noise

Using “healthy controls” as previously described before from 885 participants (Chapter 6), cut-offs were established in individuals believed to be “healthy” throughout follow up. Briefly, this group comprised 493 women and 392 men who were KL grade 0 for 4 consecutive years in the chosen knee. The assumption was that this group of participants did not exhibit structural change and therefore any bone shape changes seen were likely due to measurement error in bone shape.

LCGM was performed on these participants to obtain individual slopes for each bone separately for each gender and the cut-offs for what was herein termed “greater than measurement noise” defined as that value above the 95th percentile of healthy knees. These cut–offs based on the 95th percentile are shown in Table 44.

Table 44. Cut-offs defining measurement error for bone shape

Bone	Male cut-off	Female cut-off
Femur	0.066	0.089
Tibia	0.045	0.068
Patella	0.059	0.053

7.5.2 Model fit in the healthy controls

Linear growth models separated by gender showed reasonable fit in both genders and for all bones, and as shown Table 45 and like previous LCGMs in the full sample, there was slightly better fit seen for the femur than tibia and patella (based on lower SRMR values). Across all bones, intercepts were close to approximately 0.0 bone shape units as expected for the "healthy controls" and slopes were on average about 0.03 across gender and bones indicating very slow rates of change.

7.5.2.1 Femur

Data from 885 participants were included in this analysis. Model fit in the femur model was good: (chi-square value = 76, DF = 28, CFI = 0.99, RMSEA = 0.091 and SRMR = 0.007). The average femur intercepts in males (0.04, *p* =0.59) and (–0.008, p=0.91) in females were not significantly different from zero indicating that they both started at about bone shape scores of zero or those not distinguishable from zero units which was expected and provided some validity on the choice of this group as "controls". Average slope in males (+0.029, p<0.001) and in females (+0.039, p<0.001), indicated that femur shape significantly increased over time (representing worsening over time) and, as seen previously, was higher for females. However, both means were significantly lower than the full group as was expected for "healthy individuals", indicating very slow change and their mean values were also lower than the cut-offs shown in Table 44.

7.5.2.2 Tibia

Model fit indices in the tibia were very good, although not as impressive as those seen in the femur: (chi-square value = 59, DF = 28, CFI = 0.99, RMSEA = 0.072 and SRMR =0.010) but worse than the model in the full sample. The average femur intercept (0.18, p = 0.36) in males and (0.04, p = 0.56) in females were not significantly different

from zero, indicating that their bone shape starting points were indistinguishable from zero units which represents the mean non-OA shape. The average slope in males (+0.004, *p* =0.36) and in females (+0.015, p<0.001) indicated that female bone shape increased over time, although these units were approximately 10X lower than slopes seen in the full model in the previous section which was not surprising given this group represent “healthy” individuals .

7.5.2.3 Patella

As seen for the femur and tibia, the patella models also showed good fit to the data: (chi-square value = 54, DF = 28, CFI = 0.99, RMSEA = 0.065 and SRMR =0.019). The average femur intercept in males (0.06, p = 0.42) and (0.09, p = 0.16) in females were not significantly different from zero indicating that they both started at about bone shape scores of zero. Average slope in males (0.028, p<0.001) and in females (0.039, p<0.001), indicated that femur shape significantly increased over time (representing worsening over time) and, as seen previously, was higher for females although both means were significantly lower than the full group as expected.

Table 45. Growth models in the 3 bones in the control group

Model	Intercept (95% CI)	Slope (95% CI)	RMSEA (<0.08)	CFI (>0.95)	SRMR (<0.08)
Femur			0.091	0.993	0.007
Male model	0.04 (-0.11,0.19)	0.03 (0.02,0.04)			
Female model	-0.01 (-0.14,0.13)	0.04 (0.03,0.05)			
Tibia			0.072	0.995	0.010
Male model	0.18 (-0.02,0.35)	0.004 (-0.005,0.013)			
Female model	0.04 (-0.09,0.18)	0.02 (0.01,0.02)			
Patella			0.065	0.994	0.019
Male model	0.06 (-0.09,0.21)	0.03 (0.02,0.04)			
Female model	0.09 (-0.04,0.24)	0.04 (0.03,0.05)			

7.5.3 Effect of measurement noise on previous LCGA classes

Having established cut offs for the slope measurement error, this information was used to assess the proportions of participants that changed more than measurement error longitudinally based on the previously established trajectory groups. The aim of this was a means of providing further clinical context to previous findings as to what "true" change might represent in the population. Secondly, separate LCGA was performed in only the participants changing greater than measurement noise to establish trajectory groups in those with "definite" change.

Proportions changing greater than measurement noise are shown in Table 46 below. In males, those changing greater than measurement noise were likely to be in the fastest changers group than the slowest changers i.e. patella (29% vs 10% respectively, chi square = 62,p<0.001), femur (80% vs 19%, chi square = 336,p<0.001) and tibia (64% vs 20%,chi square = 175,p<0.001).

Results were similar in females, those changing greater than measurement noise were likely to be in the fastest changing group derived from LCGA using the full sample; with better separation seen for the patella compared to the proportions seen in males. i.e. patella (84% vs 43% respectively, chi square =319, p<0.001), femur (82% vs 31%, chi square = 366, p<0.001) and tibia (69% vs 20%, chi square = 357, p<0.001).

Table 46.Proportions in the different classes changing greater than measurement error

	Slowest changers	Intermediate changers	Fastest changers
Males			
Femur	154/792 (19)	412/907 (45)	218/271 (80)
Tibia	156/792 (20)	318/922 (47)	163/256 (64)
Patella	55/554 (10)	155/997 (16)	120/419 (29)
Females			
Femur	353/1140 (31)	736/1241 (59)	324/397 (82)
Tibia	204/1042 (20)	563/1278 (44)	318/458 (69)
Patella	373/860 (43)	987/1337 (74)	488/581 (84)

Values in parentheses are percentages

7.5.4 Latent classes in the group showing changes greater than measurement error

In males, 784 participants changed more than measurement noise in the femur, 637 in the tibia and 330 in the patella. In females these values were 1413,1085 and 1848 respectively for femur, tibia and patella. The average intercept and slope estimates for the femur assuming no significant classes in this sub-population was: in females

intercept = 1.68, slope ~ 0.20 and in males intercept = 1.41, slope ~ 0.09 (Tables 47 and 48), those of the tibia and patella are also shown in the respective tables. These findings were consistent with earlier findings of steeper slopes in females however the intercept estimates were much closer than previous estimates.

7.5.4.1 Bone shape trajectories

Parameter estimates were very similar across genders and the proportions of members in each class were also closely related. The 2-class solution was chosen to be the most optimal solution for all bones and this was consistent in both genders. Although model fit improved in terms of lower AIC and BIC as well as slightly higher entropy values in the 3-class solution for patella in males and femur in females, for reasons of parsimony the 2-class solution was favoured after inspecting the slope estimates and observing that these were the same between class 2 and class 3 (Table 47 and 48). While the AIC and BIC were reduced in the 4-class models the LMR-LRT indices were not statistically significant ($p>0.05$). Full details of latent classes and model fit indices are shown in Table 47 and 48.

7.5.4.2 Femur

In females trajectory 1 comprised 68% of females, baseline femur value (intercept) of 0.65 femur units (95% CI 0.43, 0.77) and statistically significant upward slope, 0.19 (95% CI 0.10, 0.23). Trajectory class 2, which comprised 32% of females, had intercept 3.77 (95% CI 3.29, 3.98) and slope 0.24 (95% 0.19, 0.26). In general and as previously shown in Section 7.3, the rates of progression were faster in women than in men. The proportions allocated into the 2 femur classes were similar in males (trajectory 1 = 65%, trajectory 2 = 35%).

7.5.4.3 Tibia

In the tibia in females, the slope for trajectory class 2 (35% of the sample) was ~ 1.6 times the rate of the slower changing group (38%). Although the proportion of members in each group were similar in males and females, the rates of change were constantly higher in females (in trajectory class 2, slope ~ 0.20 in females vs 0.12 in males). However in males differences in the slope estimates were very small (0.118 vs 0.123) in contrast to the LCGA models in the full sample where clearer separation was evident for slopes, although like previous models in the full sample the classes derived had clear intercept differences.

7.5.4.4 Patella

The proportions in each trajectory were slightly different in the patella. In females (trajectory 1 = 60%, trajectory 2 = 40%) while in males 52% were in trajectory 1 and 48% in trajectory 2. In females class 1 and 2 showed very similar slopes (0.11 vs 0.13), their main difference being their baseline scores. In males the 2 trajectories were similarly not well distinguished by the slopes (class 1 slope = 0.20 vs class 2 slope = 0.18) however their starting points were very distinct (Table 47 and 48).

Table 47. Results of LCGA in just the participants showing change in females

Number of classes	Number of cases assigned			AIC	BIC	Adjusted BIC	LMR-LRT p-value	BLRT p-value	Entropy
Tibia	**Trajectory**	**N**	**%**						
1	Trajectory 1 Intercept =1.479 Slope =0.116	1085	100	16089	16124	16102			
2	**Trajectory 1** **Intercept = 0.551** **Slope =0.123**	702	65	13093	13143	13111	1982 <0.001	5265 <0.001	0.854
	Trajectory 2 **Intercept = 3.144** **Slope =0.198**	383	35						
3	Trajectory 1 Intercept = 0.061 Slope =0.142	429	40	11494	11559	11518	1531 0.32	1604 <0.001	0.882
	Trajectory 2 Intercept =1.836 Slope =0.179	473	42						
	Trajectory 3 Intercept =3.992 Slope =0.187	183	18						
Femur									
1	Trajectory 1 Intercept =1.676 Slope =0.196	1413	100	21646	21683	21661			
2	**Trajectory 1** **Intercept =0.646** **Slope =0.192**	958	68	18218	18181	18149	3368 <0.001	3523 <0.001	0.867
	Trajectory 2 **Intercept =3.773** **Slope =0.244**	455	32						
3	Trajectory 1 Intercept = -0.034 Slope =0.166	540	38	15799	15847	15806	2251 <0.001	2355 <0.001	0.876
	Trajectory 2 Intercept =1.948 Slope =0.244	645	45						
	Trajectory 3 Intercept =4.699 Slope =0.243	238	17						
Patella									
1	Trajectory 1 Intercept= 1.002 Slope =0.109	1848	100	23353	23392	23370			
2	**Trajectory 1** **Intercept = 0.19** **Slope 0.113**	1103	60	18988	19043	19011	4186 <0.001	4371 <0.001	0.853
	Trajectory 2 **Intercept =2.169** **Slope 0.130**	745	40						
3	Trajectory 1 Intercept =-0.234 Slope= 0.114	664	36	16449	16521	16480	2436 <0.001	2544 <0.001	0.862
	Trajectory 2 Intercept = 1.193 Slope= 0.131	859	46						
	Trajectory 3 Intercept = 2.903 Slope= 0.124	325	18						

Table 48. Results of LCGA in just the participants showing change in males

Number of classes	Number of cases assigned			AIC	BIC	Adjusted BIC	LMR-LRT p-value	BLRT p-value	Entropy
Tibia	**Trajectory**	**N**	**%**						
1	Trajectory 1 Intercept =1.403 Slope =0.085	637	100	11094	11139	11106			
2	**Trajectory 1** **Intercept = 0.352** **Slope =0.118**	416	65	9009	9054	9022	1982 <0.001	5265 <0.001	0.913
	Trajectory 2 **Intercept = 3.439** **Slope =0.123**	221	35						
3	Trajectory 1 Intercept = -0.249 Slope =0.123	257	40	7800	7858	7817	1156 0.62	1216 <0.001	0.910
	Trajectory 2 Intercept =1.720 Slope =0.123	259	41						
	Trajectory 3 Intercept =4.263 Slope =0.143	121	19						
Femur									
1	Trajectory 1 Intercept 1.697, Slope 0.131	784	100	13077	13109	13807			
2	**Trajectory 1** **Intercept =0.603** **Slope =0.133**	539	69	10676	10722	10691	2292 <0.001	2407 <0.001	0.908
	Trajectory 2 **Intercept =3.984** **Slope =0.178**	245	31						
3	Trajectory 1 Intercept =0.05 Slope =0.116	341	43	9359	9420	9379	1259 0.10	1322 <0.001	0.889
	Trajectory 2 Intercept =2.146 Slope =0.190	132	17						
	Trajectory 3 Intercept =4.963 Slope =0.180	311	40						
Patella									
1	Trajectory 1 Intercept= 0.788 Slope =0.176	330	100	5161	5188	5166			
2	**Trajectory 1** **Intercept = - 0.210** **Slope 0.196**	170	52	4178	4216	4184	936 0.003	989 <0.001	0.889
	Trajectory 2 **Intercept =1.808** **Slope 0.183**	160	48						
3	Trajectory 1 Intercept =-0.585 Slope= 0.195	106	32	3603	3652	3611	581 0.26	581 <0.001	0.920
	Trajectory 2 Intercept = 0.955 Slope= 0.196	161	49						
	Trajectory 3 Intercept = 2.657 Slope= 0.169	63	19						

Bold represents optimal class chosen.

7.6 Discussion

This is the first study to use quantitative bone shape obtained from SSMs to identify heterogeneous trajectories of structural progression in the three knee bones. This study found that bone shape change was linear, and that for all bones, three distinct growth patterns existed: slowest changers, intermediate changers and fastest changers. Slowest changers were characterised by having the lowest bone shape values at baseline and slowest rates of change, intermediate changers had starting values higher than the slowest change group, had steeper slopes but less than fastest changers who were characterised by the highest starting points and greatest range of change.

Participants in the fastest change groups had higher prevalence of comorbidities and lower physical functioning scores compared to those in the slowest and intermediate changing groups. When assessed longitudinally, the fastest changers recorded more incidents of TKRs compared to the slowest changing group. These trends were similar for all bones and all genders but there was clearer separation between trajectory groups in women. However, the rates of change differed the highest rates were seen in the femur followed by tibia and then patella.

The study also identified a group of about 15% of participants whose femur and tibia bone shapes changed about 2-3 times faster than the slowest changing group, representing worsening structural decline. Obesity, knee pain, non-white ethnicity and history of knee surgery were strongly associated with being in this trajectory group. The slowest and intermediate groups showed less rapid change and at baseline had less structural disease as measured by bone shape.

When described for all three bones, this study found a group that were classified as fastest changers for all 3 bones (20% of females and 17% males). This group were more strongly associated with OA risk factors and approximately 70% of those that underwent knee replacements at follow-up came from this group. Individually, classification in the bones was consistent, with similar proportions in each trajectory class and better separation of classes (in terms of rate of change) seen in the femur and tibia than for the patella. The finding of the same number of trajectories for all bones and similar associations with OA risk factors and outcomes also adds to existing knowledge on OA pathophysiology, and suggests that all three bones are likely part of a single disease process as previously hypothesised in Chapter 6.

A strength of this work was the use of repeated measures and a more advanced analytical technique (LCGA) that offers several advantages compared to traditional clustering techniques such as latent class analysis, including the ability to model heterogeneity by classifying individuals into groupings with similar patterns using longitudinal data [487, 488]. Second, FIML missing data techniques were applied to minimise estimate biases as compared to list-wise and pair-wise deletion methods under the missing data at random assumptions [536]. Furthermore, sensitivity analyses were performed by applying measurement error cut-offs in order to provide more clinical evidence on whether the classes found were robust and represented true change. By applying these cut-offs it was shown that 80% of males and 82% of females classified as fastest changers before, had actually shown longitudinal change greater than measurement noise and could be considered to be evidently changing significantly.

Findings of distinct bone shape trajectories suggest that OA is likely not a homogenous disease where one growth trajectory or mean shape change defines the population at risk, but that subtle patterns do exist when the inclusion criteria such as that of the OAI contains a good case-mix of at risk and diseased individuals. The trajectories from this study were clinically meaningful and confirmed what is already known about OA i.e. those in the fastest group were associated with known OA risk factors [30, 530-532], and went on to have important end points such as TKR. Previous work from our research group in Leeds found that 3 trajectories existed for femur and tibia bone shapes however, unlike this study where all OAI participants were included, those studies were restricted to participants with no radiographic OA at baseline, with one restricted to just females with no radiographic OA, and both did not assess the patella [553, 554]. One study using 3D bone shape also found three trajectories of femur and tibia bone shape, and that these were associated with gender and radiographic OA[404]. However, unlike this study their bone shape measures were derived from radiographs and while this study stratified all analyses by gender, their study did not.

Other studies using similar statistical techniques to assess trajectories of structural change in OA have shown that there is significant variability in trajectories of structural progression, and support findings of heterogeneity in OA demonstrated in the present study. These studies have however all looked at cartilage; using 3D cartilage thickness *Deveza* et.al [531]found distinct trajectories of cartilage loss in the OAI cohort while another study assessing radiographic joint space width found seven trajectories of change [547] and similar heterogeneity revealed for joint space width by *Neogi* et. al [555] and *Kwoh* et.al [556]. In common with this study, the aforementioned studies also

found consistencies in relating structural trajectories to important OA risk factors or outcomes, as those in the “fastest changers” in our case a proxy for “worst structural deterioration” was more strongly related to OA risk factors.

The Arthritis Research UK clinical studies group previously suggested a recommended core set of data to be collected to assess knee structural progression including KL-grade and previous knee surgery [251]. Findings from this study will add to the current knowledge on risk stratification for structural progression in OA clinical trials through the use of a novel MRI structural measures. This study found that a history of knee surgery was an important predictor of being in the fastest changing group consistently across all bones, and provided some face validity for bone measures through the demonstration of an association between worsening structural damage (as measured with bone shape) with other OA risk factors such as obesity and non-white ethnicity .

Bone shape measures once fully validated have the potential to be included in this set of measures and would provide useful structural measures worth considering for OA trials. While the need to provide more useful stratification factors is clear [557] and would help draw consistent conclusion from clinical trials to develop new OA disease – modifying drugs, very few studies assessing trajectories of structure have been undertaken so far and these have mostly studied cartilage with varying results [531, 547]. This area still remains relatively less studied and more studies looking directly at structural changes are warranted.

Findings in the sub-sample of individuals thought to have change greater than measurement error (sensitivity analyses) were not surprising. Since this group was more homogenous than the full OAI sample, obtaining distinct classes based on slopes was unlikely hence the main differences seen between classes was due to intercept differences/variation. This finding also suggests that in terms of participants thought to exhibit true structural change, OA follows two trajectories of change but as seen in this study the distinction between the slopes was not very clear and it could be argued that in these groups they generally follow the same trajectory (one) based on their slopes. Caution should however be applied on these findings since the sample composition makes a difference here. This finding could have implications when conducting such analyses in future in that the research questions need to be framed appropriately. If the desire is to show that trajectory classes exist in a sample, then LCGA methods could be appropriately applied for this, and would provide useful results provided substantial variation exists in intercepts and slopes after exploring this using latent curve growth modelling. However to strengthen findings from those analyses an application of measurement error would refine these classes and optimise the desired groups better

as done by this study. A fundamental of performing LCGA is that substantial variation exists in intercepts and slopes, therefore applying this technique in homogenous groups as alluded to would be inefficient.

In summary, the shapes of the distal femur, proximal tibia and patella change linearly over time and divide into separate/distinct trajectory groups over time. These trajectory subgroups are associated with knee OA risk factors. Knowing which participants are likely to change most rapidly could be useful in patient selection for future OA trials that wish to demonstrate structural progression.

7.7 Limitations

In terms of limitations the proportion of missing MRI data was high in some years, although FIML missing data modelling techniques were applied to address the problem. While LCGA is an established method for analysing longitudinal data, decisions regarding the number and shape of trajectories can be somewhat subjective, and although three trajectories were identified, there was variation in bone shape in each trajectory. However, the goodness of fit indices and probability of correct classification was very good (>95% accuracy).Selecting the model with the “correct” number of classes can be heavily influenced by the method used to parameterize the structure of the random effects, and a common approach which is one used in this study was to constrain growth factor variances of all latent classes to be zero. Other choices include estimating the variances and covariances of the growth factors separately for each latent class, or specifying homoscedastic or heteroscedastic models by constraining or freely estimating the latent error variances across time and classes. However another problem is that having too many freely estimated parameters leads to model convergence issues. All these issues need careful consideration and importantly the choice of what parametrisation to use should be driven by an understanding of the underlying data, associated theory and the research question. Some of the constraints mentioned earlier although necessary for model estimation, may not accurately reflect the underlying growth process and could have led to wrong conclusions however upon inspection of the variation in each class using graphical methods, there was no noticeable pattern highlighting significant variability to warrant freeing the variances across classes. The measurement error for patella measures was high mainly because of the difficulty in segmenting the patella due to its size and may have contributed to differences seen for this bone.

7.8 Conclusion

Latent class growth analysis has provided evidence that in the OAI, participants cluster into 3 distinct patterns of structural change suggesting that OA is not a homogenous disease. The trajectory classes observed are largely similar in all three knee bones and in both genders. This provides further evidence that all 3 knee bones are part of a single disease process whose change shares similar predictors. Bone shape may therefore enable better understanding of the "natural history" of knee OA. The three trajectories of change observed were clinically relevant, and for future OA structure-based trials seeking optimal recruitment these trajectories may be a useful consideration to optimise this recruitment process. The trajectory-based approaches may also have additional benefits of further probing the risk factors for OA progression. Further work to reduce the amount of measurement error inherent in imaging studies is still warranted to establish what true change looks like.

Chapter 8 Discussion, future directions and conclusions

8.1 Introduction

This chapter addresses the main findings of the Thesis, including novel 3D meniscal measures as potentially feasible measures for knee OA trials in Section 1.2. Findings from the Thesis are put in the context of the recent literature in Section 1.3 and directions for further research are proposed throughout and expanded in Section 1.4.

8.2 Thesis synopsis

This thesis aimed to characterise potentially novel knee OA imaging biomarkers and specifically describe the responsiveness of novel meniscal and bone shape measures, assess 3D knee bone shape against established measures of OA structural progression (cross-sectionally and longitudinally) and characterise the longitudinal change in three knee bones (represented by bone shape) using advanced statistical techniques.

The hypothesis underlying this thesis was that novel quantitative, multi-tissue imaging biomarkers of bone and meniscus would provide valid measures for use in knee OA clinical trials. As part of biomarker validation, the hypothesis was that bone and meniscal measures would demonstrate good reliability, responsiveness, feasibility, demonstrate construct validity, be associated with known OA risk factors and be useful in classifying participants for OA trial recruitment.

The main findings from Chapters 4, 5, 6 and 7 in this thesis are as follows:

Chapter 4:*Where does meniscal damage progress most rapidly? An analysis using three-dimensional shape models on data from the Osteoarthritis Initiative*

The aim of this study was to apply quantifiable novel 3D image analysis on meniscus in a cohort typical to that included in clinical trials, in order to determine the spatial distribution of change, and determine the meniscal pathologies most associated with change during 1-year of OA progression. This would provide construct and face validity for these measures and provide evidence on their feasibility for use in future trials. Reliability of meniscus measures was also assessed and the group-level internal responsiveness of various meniscal measures calculated using SRMs to determine the most responsive to 1-year change.This study concluded that two meniscal measures

(medial tibial coverage and thickness) were the most responsive measures of change, with change as measured by SRMs comparable to other MRI outcomes such as cartilage thickness and better than radiographic JSN. The spatial location of damage was predominantly in the posterior subregion of the medial meniscus providing construct validity for these novel measures as other studies have consistently shown change in this region. The repeatability indices for all meniscal measures were very good, therefore enhancing their feasibility for use in future meniscal studies. Visual appraisal of images from this study confirmed the heterogeneity of meniscal pathologies providing face validity for these novel measures. This study provided preliminary evidence that meniscal measures have the potential for use in clinical trials and should now be investigated for their ability to add discriminatory power in OA progression assessment either as surrogate measures, part of composite measures or for OA phenotyping .

Chapter 5: The relationship between two different measures of osteoarthritis bone pathology: bone marrow lesions and 3D bone shape

The aims of this chapter were to investigate the relationship between a potential imaging biomarker (bone shape) with a well-studied bone pathology (BMLs), and to compare their responsiveness and construct validity in a well-defined cohort selected to investigate OA biomarkers. To allow effective comparisons, SQ BML MOAKS scores (assessing both size and numbers) were scored in a manner that allowed comparisons between femur and tibia bone regions to be as closely matched as possible. A range of descriptive statistics including correlations and comparisons of proportions were applied to compare the two measures. Linear regression was then performed to assess the association between bone shape and presence or absence of BMLs (binary) at baseline. Longitudinal analyses evaluated the relationship between change in BMLs and concurrent change in bone shape using multilevel regression analyses. Secondly, responsiveness between the two measures over 2-year follow up was assessed by means of SRMs.

These analyses demonstrated a moderate positive correlation between 3D bone shape and BMLs regardless of BML measure used (BML size or BML total number). Linear regression revealed modest baseline associations between BMLs and bone shape while multilevel analyses showed that worsening OA status, as assessed by femur 3D bone shape, was associated with worsening bone pathology as scored using BMLs, and once bone shape started to change its progression seemed to follow a linear trajectory but this could not be confirmed given the relatively short follow up. Based on these findings construct validity for bone shape was established. Also, given that

worsening OA status measured using bone shape has been show to relate to incidence radiographic OA and its progression, and BMLs are known to be related to OA prevalence and progression, these findings suggest a relationship between these two pathologies exists. As evidenced by much higher SRMs, bone shape measures demonstrated much greater responsiveness than the different BML measures applied in this study, the femur showing greater responsiveness than the tibia. The high responsiveness indices seen for 3D bone shape measures suggests they could be candidates for future OA trials targeting the bone. The potentially linear trajectory seen for 3D bone shape in this study was then explored in the following Chapter.

Chapter 6: Determinants of osteoarthritis 3D bone shape and its change in the three knee bones; a latent growth modelling approach on 37,583 MR images from the Osteoarthritis Initiative

The previous chapter provided evidence of construct validity for bone shape (in particular for the femur), while previous work by our research group has also highlighted predictive validity of bone shape for TKR and also knee pain, especially for the femur bone shape. The aim of this chapter was to establish the relationship between the three knee bones (femur, tibia and patella) both in cross section and longitudinally, so as to establish whether responsiveness indices seen thus far were only femur-specific or similar across all three knee bones. Descriptive analyses for the three knee bones were assessed via correlations amongst the three bones and comparisons of proportions of bone shape measures classified as outside healthy limits and referred to as “OA bone”. The definition of OA bone was those bones which fell above the 95th percentile of the group mean from individuals thought to be free from OA (a sample of 885 healthy OAI participants). Latent growth curve analyses modelled all three bones separately initially, to test whether change over time was linear or followed a non-linear pattern (quadratic, etc.). Having established that changes were linear for all knee bones, parallel process growth models were applied in order to evaluate all three bones at the same time, and as no prior hypothesis on which bone likely drives change, the intercepts and slopes of all three bones were allowed to simply co-vary amongst themselves. Lastly in order to explain variance in starting points and rates of change latent models were performed adjusting for known OA risk factors.

Descriptive analyses showed that all three bones correlated well, with greater correlation seen between the femur and tibia then followed by femur and patella. In the knees classed as having OA bone as defined earlier, the most frequent pattern involved having all three bones classed as OA. Latent growth models provided

evidence that the direction of change is linear and the same in all three knee bones over time; the greatest rate of change occurring in the femur. Parallel process models revealed that once a knee bone was diseased it was likely the other bones were on the same disease trajectory also. The effect of covariates was similar in all three bones but varied by gender. This suggests that based on this model in a non-population based sample the three knee bones are part of a single disease process, with the femur providing the greatest amount of change of the three bones. This could have implications on the choice of tissue for biomarker development. However, testing in a post-traumatic cohort should be considered to allow a wider range of the OA disease spectrum to be considered.

Chapter 7: Defining bone shape trajectories in three knee bones: the Osteoarthritis Initiative.

Chapter 6 showed that the femur, tibia and patella changed linearly over time, and this change was assumed homogenous among all OAI participants included in the study based on the mean starting points and mean rates of change. However, significant variation in the growth factors (intercept and slope) was observed and while covariates were added to explain some of this, significant unexplained variation in bone shape growth still remained. The aim of this chapter was to identify distinct trajectory groups of change for each bone using growth mixture models, a type of latent class growth analysis (LCGA) using the entire 4796 OAI cohort, and then modelling the probability of group membership using established OA risk factors.

This study found evidence that heterogeneous growth trajectory groups based on structure exist in OA. Three distinct trajectories were consistently identified by LCGA in each knee bone; obesity, knee pain, non-white ethnicity and history of knee surgery were strongly associated with classification into the fastest trajectory group. This was consistent across all three bones with the effects of covariates on these trajectory classes also the same. The study also identified a group of participants that showed the fastest rate of change in all three bones in whom clinical covariates mentioned before, showed stronger associations compared to those consistently in the slowest trajectory group. This ability to classify participants more accurately based on structure may be useful in enriching OA clinical trials through improved baseline stratification and more optimal participant selection. This study also provided supporting evidence that while OA is thought to be largely heterogeneous, subtle patterns in growth exist suggesting that homogenous structural trajectories exist in OA. However, the finding that all three bones had the same trajectory classes and were influenced by covariates in the same manner strengthens the findings from Chapter 6 which concluded that the

three bones are part of the same disease process. Knowledge on which classes or trajectory groups exist based on structure as shown in this study could be an important step in better phenotyping individuals for clinical trials.

8.2.1 Overall summary

Findings from this thesis include evidence for the utility of meniscal measures in OA knee trials, the validity of bone shape for use in clinical trials and evidence of symptom-structure relationships between bone shape and known OA clinical risk factors as shown in Chapters 6 and 7. This thesis also added new knowledge on the natural history of bone shape change, by showing that this change was largely linear throughout 8 years of follow-up, using univariable models separately for femur, tibia and patella and also in parallel growth models incorporating all 3 bones (described in Chapter 6). These findings provide evidence in support of the hypotheses set out by this thesis, that bone and meniscal measures would demonstrate good reliability, responsiveness and feasibility. There was also evidence supporting the hypothesis that bone and meniscal measures would demonstrate construct validity, be related to previously studied OA risk factors and be useful in classifying participants for OA trial recruitment. Specific examples of how this was achieved are outlined below.

Work on novel meniscal measures has provided evidence supporting the biomarker validation domain of discrimination, shown that novel 3D meniscal measures were highly reliable, demonstrated greater responsiveness than JSN, and provided construct validity by showing consistent spatial distribution of meniscal change in regions previously shown to change in other studies. Bone shape measures demonstrated construct validity as measures of structural OA progression and were related to BMLs, showed predictive validity through longitudinal association with important OA endpoints including TKR, and concurrent validity by association with knee OA symptoms such as pain in cross-section and longitudinally.

Also, meniscal and 3D bone shape biomarkers could fulfil the OMERACT domain of feasibility. The novel quantification methods used in this thesis demonstrated excellent accuracy and reliability in assessing meniscus and bone (low SDD values and high intra-reader reliability indices). Automated segmentation of the meniscus is possible although this still needs further validation work, but at present manual segmentation in cohorts well-defined to demonstrate meniscal change appears feasible. For bone shape automated segmentation in large trials is now potentially feasible and will allow for rapid analysis of large datasets. This will provide a consistent measurement metric scalar measure. As a continuous measure (compared to SQ scores like BMLs), bone

shape permits the use of more powerful statistical methods for analysis as were used in this thesis.

This thesis also demonstrated greater precision in phenotyping at-risk individuals. By following up participants for 8 years and using longitudinal analyses that applied advanced statistical techniques (parallel process models and growth mixture models) valid conclusions supported by the robust analyses were reached. Findings from growth models applied in this thesis can be translated to interventional knee OA clinical trials to help select potential recruits or for enriching existing trials. While further sensitivity analyses may be needed, the current work supports the existence of structural trajectory classes.

8.3 Thesis findings and recent literature

The following section discusses work from this thesis in context of recent relevant literature, including work that was done after the thesis literature review was completed. A narrative review was undertaken in Chapter 2 inclusive of studies up to September 2017. Subsequent literature has been reviewed to provide a more recent update with focus on imaging biomarker status (specifically meniscus and bone),updates on how the OA field has advanced with regards wet biomarkers and other tissues such as cartilage.

8.3.1 Recent developments in meniscal measures

8.3.1.1 Responsiveness

Findings from this thesis that demonstrated meniscal measures were more sensitive and responsive than radiographic JSW were similar to those found by Roth et. al recently using 3D measures of the meniscus in a similar cohort [558]. Roth et. al. using 35 OAI participants with confirmed TKR, assessed whether 3D meniscal measures had similar responsiveness to cartilage thickness over two year follow up by measuring this at 2 time points preceding the TKR. Inclusion criteria into the Roth study was knees with KL≤ 2 at baseline which differed from Chapter 4 of this thesis whose inclusion criteria was knees with KL ≥2. They found that meniscus measures of position and morphology showed responsive changes over 2-year follow up. However, differences with this thesis were seen for the type of measures showing significant change. In common with Roth et. al., Chapter 4 of this thesis also reported statistically significant medial tibia coverage measures but non-significant meniscal volume changes. However unlike Roth et. al who concluded that meniscal thickness did not change

significantly, findings from this thesis suggested that meniscal thickness was one of the promising meniscal measures in terms of responsiveness.

This difference for meniscal thickness could be attributed to differences in the way the measurements were derived. In their study meniscal thickness was measured in just one region (by averaging two bi-directional Euclidean distance transforms between tibia and femur) while as shown in Chapter 4, this thesis used three regions using the tibia as reference. Use of sub-regions appear to provided promising measures of change based on their responsiveness indices. Another difference was their finding of higher SRMs for meniscal extrusion measures unlike the thesis study, but this may have been related to a follow up for a year longer than used in this thesis.

While Roth et al study described found greater magnitude of change for extrusion, their SRM values were only indicative of only moderate change, and less than what is considered clinical important (>3mm, thought to be associated with cartilage degeneration as suggested by Lerer and colleagues[559]). When compared with extrusion findings shown in Chapter 4 of this thesis, this suggests that meniscal measures by Roth et.al. were subject to greater variability and therefore any changes seen were possibly subject to higher measurement error, and without data on the smallest detectable difference for these measures it is difficult to assess if “true change” did occur, which could limit findings this study. Despite differences in magnitude, with the measurement error issues highlighted, it could be argued the SRMs for extrusion findings from this thesis and the study by Roth et.al are similar. Moreover, their sample size was less than half that used in this thesis and may also be a limitation.

8.3.1.2 Reliability and repeatability

Okazaki and colleagues assessed the advantages of 2D over 3D MRI measures of the meniscus in terms of precision and reliability. They hypothesized that 3D MRI would provide the precise length, width, and height of the meniscus. Their secondary aims were to assess if meniscal volume and l extrusion differed between knees that had meniscal root tears and controls.

Okazaki and colleagues showed that meniscal size was more precisely measured in 3D [560]. In their study, 2D MRI referred to images of usually single thin slices, with a non-zero gap between them while 3D MRI images represented a stack of contiguous slices i.e. with zero gap between them. Their study used similar measures to those in Chapter 4 of this thesis but they employed a different study design as they compared participants selected for meniscal root tears vs controls with no tears, unlike Chapter 4 where the main focus was responsiveness of novel 3D meniscal measures. Similar to

this thesis, their study highlighted the importance of accurate, reliable measurement as they demonstrated the advantages of 3D measures in distinguishing healthy and unhealthy tissue. Their 3D measures showed high repeatability indices (ICCs > 0.85) and therefore classified as excellent, similar to those found in the reliability sub-study in Chapter 4.

The implications from these reliability findings are that highly repeatable meniscal measures were found in Chapter 4 which is important for their consideration as potential clinical trial outcomes. While not related to clinical trials directly, Okazaki et. al`s study has important implications for improved understanding of the biomechanical changes that happen following meniscal root tears, and thus aid their repair by surgeons. A limitation of their study was the small sample size of 32 participants.

8.3.1.3 Validity and utility of novel meniscal measures

More recently quantitative measures of meniscal position and morphology were shown by Roth et.al to be associated with subsequent TKR in knees with baseline KL grade ≤ 2 and OARSI JSN not greater than zero in a case-control study of 35 case-knees from the OAI [561]. Cases were defined as any individual having undergone a TKR between the 36 month and 60 month OAI follow-up visit while controls had no TKR throughout the study and were matched for age, sex and KL grade. Roth and colleagues found that two meniscal measures with greatest responsiveness from this thesis, medial tibial coverage and thickness were associated with TKR but in addition also found that extrusion and meniscal width demonstrated significant changes prior to TKR. As shown in Chapter 4 meniscal volume did not change significantly which was similar to findings by Roth et.al [561]. These findings from Roth et. al highlighted that structural changes in the meniscus are related to important OA clinical outcomes which adds to evidence on their predictive validity. Due to their superior reliability indices compared to 2D measures as highlighted before, these novel measures have the potential to be used as new outcome measures. However, these findings should be interpreted with caution as they may have been hampered by small sample size (n = 35 cases) for valid predictive inferences to be made.

Preliminary work by Sharma et. al. provided further evidence for the utility of 3D meniscal measures by showing that meniscal measures for position and size could discriminate between OA progressors and non-progressors as defined by cartilage loss over 2 years [562]. By comparing baseline meniscal measures between cases and matched controls, they showed that 3D meniscal thickness was responsive as shown in Chapter 4 but in addition also found that meniscal extrusion was responsive unlike this thesis. However their study may have been hampered by a small sample size,

n=37. It should also be noted that the study designs were different, the study by Sharma et. al was a case-control design assessing differences between groups while this thesis sought to establish the most responsive meniscal measures. However, their findings importantly highlight the potential for 3D meniscal measures as additions to a suite of potential imaging biomarkers that include bone measures, and as discussed in Chapter 2 and 4 could add to the discriminatory capacity of knee OA trial measures.

Further validity for meniscal measures was also shown by Roth and colleagues using OAI data. They found that meniscal measures explained significant variance in JSW, with measures of meniscal morphology explaining most of fixed location JSW and meniscal position explaining most of mJSW variance. This was done in two groups, one study used the OAI control group and then a separate study using a group selected for OA progression [558, 563]. Analyses involved multiple linear regression analyses of JSW against meniscal measures. Meniscal parameters alone accounted for up to 48% of the change in JSW parameters in a group selected for knee OA progression with similar findings using the control group. These results suggest that longitudinal change in JSW represents a composite measure of progression involving both cartilage and meniscus and not just cartilage as also discussed in Chapter 4. However, their results like most literature on novel meniscal measures should be treated with caution given they performed regression analyses on just *n*=35 participants and may have been underpowered for such analyses.

8.3.2 Potential for bone as an imaging biomarker

8.3.2.1 Quantification with 3D bone shape

This thesis has shown that 3D bone shape obtained from ML-derived quantification provides a highly accurate, novel measure that is specific for OA and highly responsive over time in OA progression. Newer studies using 3D bone shape such as that of Bowes et. al [564] using 4796 OAI participants examined the relationship between bone shape, KL, current and future pain and function, and total knee replacement (TKR) up to 8 years. Bowes et.al demonstrated that bone shape was responsive, associated with symptoms (knee pain and function) and predicted knees at risk of progressing to TKR which supports its utility as a viable imaging biomarker. In addition they showed that bone shape (termed B-score in that study) provides reader-independent quantification, providing unambiguous classification of OA status and a continuous metric measure, which, as shown in Chapters 5,6 and 7 of this thesis, is an important characteristic for any potential imaging biomarker. Without this precise

quantification it would have been inefficient to use bone shape in advanced modelling techniques as was used in Chapter 6 and 7.

There is however a paucity of data on studies directly comparing the construct validity of 3D bone shape and its responsiveness against other potential imaging biomarkers. An attempt to address this was performed in Chapter 5 which found that bone shape was associated with a more established bone biomarker, BMLs and longitudinally was more responsive than SQ BMLs. Studies comparing OA imaging biomarkers are still lacking. Hunter and colleagues using the FNIH cohort, as similarly used in Chapter 5 showed that 3D bone area and 3D bone shape were independently associated with both radiographic and pain progression but did not compare these with any other imaging biomarker[209].

8.3.2.2 Measuring progression and composite measures

In terms of progression, LCGM showed that bone shape changed linearly over time in all three knee bones and a composite of all three bones did not result in improved model fit compared to just the individual knee bones. Collins et al.[215] explored the effect of a combination of joint structures on OA progression (as defined in the FNIH cohort in Chapter 3) and while they did not use 3D quantification such as used in this thesis, reported that changes in BMLs for example were not significant predictors of OA progression in models that already included cartilage, meniscus, and effusion markers but were associated with progression when modelled independently. In their study, the predictive model with just cartilage as imaging feature had an area under the curve (AUC) of 0.71 and after addition of four additional imaging features (meniscal extrusion, meniscal morphology, effusion-synovitis and Hoffa-synovitis) the AUC increased by just 3 percent to 0.74.This could have implications for development of composite measures as potential biomarkers as those that contribute little to the precision or efficiency of the measures would be excluded. This also further highlights how unresponsive SQ measures of OA have been.

In Chapter 7 trajectory analyses confirmed earlier findings from growth curve analysis, that a composite of all three bones (in this case participants in the fastest trajectory in all 3 knee bones) did not result in significantly improved model fit compared to just the femur or tibia individual models. In terms of progression however, those in the worst trajectory group were more strongly associated with progression to TKR and more strongly with clinical covariates in models assessing single tissues and then more strongly in the composite trajectory that included all three bones. There is however, limited literature to currently compare these trajectory findings, as is highlighted later in section 7.3.3 with only two studies considered. Moreover, those two studies [531, 565]

assessed cartilage and not bone like this thesis and secondly did not assess composite measures as in Chapter 7. However, in common with Chapter 7, Deveza et. al [531] used 3D quantification and showed that OA structural progression was linear in nature and those in the worst trajectory were associated with clinical covariates using similar statistical techniques and from a sub-sample of the OAI.

As detailed in Chapter 3 this thesis was concerned with imaging biomarkers of bone and meniscus and did not assess traditional “wet” bone biomarkers. There is currently weak evidence of an association between systemic bone markers and OA imaging biomarkers, which could be explained by the lack of specificity inherent in systemic OA markers as these are markers of systemic processes involving all OA joints in the body including the spine. In an attempt to assess composites between biochemical and imaging biomarkers, Deveza et.al [524] performed head-to-head comparisons between bone biochemical biomarkers and bone imaging biomarkers including bone shape used in this thesis, and found weak associations in cross-section and no associations longitudinally. CTX-2 which was found most consistently associated with BMLs, only explained 3% of the variance in baseline bone shape when modelled independently and under 1% of the variance in 24-month bone shape. This suggests that a combination of bone biochemical and bone imaging biomarkers are not feasible as composite measures for clinical trials. Also, the practical use of biochemical markers as predictors of bone features seems limited as demonstrated by these weak associations. Further studies are therefore still need in order to characterise relationships between different bone biomarkers and include other tissues so as to add to existing knowledge.

8.3.2.3 Bone quantification using 2D and SQ measures

Currently the best studied bone pathologies are BMLs, which contribute to both development and progression of radiographic disease, and to pain. However, as shown in this study BMLs were far less responsive over time, due to their SQ nature, and the advantages of bone shape as a metric measure are further highlighted by its utility in complex statistical models such as LCGM and GMMs performed here that were useful to elucidate the natural history of bone shape as well trajectories of change, which would not be possible with ordinal BML measures. As alluded to earlier in this discussion, 3D bone shape provides a better measurement “ruler”. Development of more reliable 3D BML measures are currently ongoing to address this gap as shown by Bowes and colleagues [353] recently using novel BML measures.

More recently, Perry and colleagues [566] as part of the UK-VIDEO trial applied quantitative segmentation techniques to derive novel 2D BML measures of volume but

in contrast to many OA studies found no association between pain and total subchondral BML volume, and attributed this difference to their more reliable quantitative BML measures unlike previous studies using SQ BML measures. However, a major limitation could be that they had a small sample size (N=50) and possibly underpowered to detect true changes.

Chapter 6 provided preliminary evidence for the natural history of bone shape, showing that bone shape changed in a linear manner based on a reasonable length of follow-up and advanced statistical methodology. 3D bone shape used in this thesis incorporates both spreading of bone and osteophytic changes. Jones and colleagues recently studied the natural history of another subchondral bone pathology, MRI-detected osteophytes showing that these worsened over time [567]. They also showed that other baseline structural abnormalities in the knee (BMLs, cartilage defects, meniscal extrusion, synovitis) were associated with worsening MRI-detected osteophytes over 2.6 year follow-up [567] by assessing change in osteophytes as outcome and baseline imaging features as predictors. The major limitations however were the short duration of follow-up of 2.6 years compared to this thesis with 8 year follow-up. Also, their study design was not optimal to assess natural history as they only described percentages of individuals reporting change at follow-up as their endpoint for change. Moreover, measurement of osteophytes using SQ measures which are known to be less responsive than quantitative ones could have masked longitudinal finings.

8.3.2.4 Breakthroughs in structure modification

Bone continues to provide promising opportunities as structural targets. Conaghan and colleagues showed structure-modifying effects of MIV-711 (a Cathepsin K inhibitor) on bone area [568], a continuous bone measure very similar to bone shape and also derived from SSMs like 3D bone shape used in this thesis. Bone area is a continuous measure that encompasses the complete 3D bone surface but differs from 3D bone shape as the latter is scaled into a vector based on how it varies with the mean shape of “healthy controls”. In the MIV-711 trial Conaghan et. al. recruited 240 participants aged 40-80 with KL 2 or 3 knees from six European countries who were randomised to placebo and 2 doses of MIV-711. They demonstrated statistically significant attenuation of MRI bone area progression and reduction of cartilage thickness in the active arms compared with placebo, consistent with the mechanism of action of MIV-711.

While the therapy was not associated with pain reduction in the study by Conaghan et.al, that study highlighted the importance of the need for sensitive biomarkers as bone area measures were responsive to change over time, showing that the current lack of responsive markers may be one of the reasons for the slow breakthroughs with

structure-modifying clinical trials. However, the study has limitations. It is possible that due to the mechanism of action of MIV-711 (reduction of bone remodelling and resorption), the drug may alter bone phenotypes independent of OA. The value of bone area as an imaging marker in such trials would need further consideration. One way to improve its use in such trials could be to devise a new measure of bone that is directly attributable to OA. This “OA-attributable” bone area would then be assessed as a structural measure in trials using bone agents such as MIV-711.

As discussed in Chapter 2, observational studies assessing structural change in OA have relied largely on indirect cartilage measures using radiography and do not assess bone morphology. However, while these quantitative measures aid the detection of significant structural changes within a relatively short study period, data from the OAI indicates that changes in bone shape over a period of 24 months are related to progression of pain over 48 months [209]. Thus, the duration of this study may have been another limitation to demonstrating structural improvements associated with statistically significant reductions in pain.

8.3.2.5 Further insights into pathogenesis

The understanding of bone pathogenesis has improved through the use of new bone measures and SSM technology. Chapter 5 showed how two different bone measures related to each other while Chapters 6 and 7 highlighted the natural history of bone shape. Zhong et. al.[569] in another study found statistically significant differences post-ACL injury between cases and controls using bone shape measures derived from SSMs. They found that tibial plateau area increased over time in injured knees and noted other bone changes including femur sphericity and notch width. Zhong also found correlations between changes in bone and changes in cartilage post ACL injury. All this adds to the continuing body of evidence on bone pathogenesis. However due to the very small sample size (N=30), the results should be treated with caution.

Another study assessing ACL injuries (the KANON study) [544], revealed rapid 3D bone shape changes after acute ACL injury, and an acceleration of this change post-surgical intervention. The KANON trial employed the same SSM techniques used in this thesis to derive bone shape and followed participants for 5-years to compare changes in bone between participants with ACL-injuries and suitably matched controls. The findings from the KANON study are in line with findings from this thesis as shown in Chapter 6 where knee surgery was associated with increased slopes of bone shape change. However, this explained only 4% slope variance and the definition of knee surgery variable used in Chapter 6 was broad and not just confined to ACL repair.

8.3.2.6 Bone and radiography

Subchondral bone texture may be a useful quantitative imaging biomarker for use in clinical trials, particularly those with interventions targeting subchondral bone. Using a matched study, MacKay and colleagues showed that subchondral bone texture (medial femoral and tibial subchondral bone) was modestly predictive of radiographic OA progression, defined as ≥ 0.7 mm, mJSW loss. However, the best-performing model which combined both tibia and femur only had a modest area under the curve of only 0.68 which does not represent good discrimination. Despite this, that study provided some signals that suggest bone texture could still be a useful addition to the suite of imaging biomarkers available for further OA imaging research. An advantage is that due to multidimensional data output of texture analysis, is it better placed to interact with machine-learning-based approaches to image interpretation [570]. However, unlike 3D bone shape which is a scaled unit that provides an objective structural measure direct interpretation of the texture scores are challenging.

8.3.3 Applicability of SSM measures from this thesis

Previous models of the knee have been generated from either simplified mathematical descriptors of the knee anatomy or directly from 3D image-based models such as SSMs. In both cases, there is a level of uncertainty in the accuracy of the geometric representation of structures such as the meniscus used in this thesis.

Previous computational models of the knee have mainly used the finite-element (FE) method due to its ability to represent the complex geometry. The accuracy of the FE model predictions is however affected by geometry, material properties and boundary conditions. An advantage of computational methods is the ease of generation of tissue geometry using software drawing tools, and the resulting shapes readily meshed for FE analysis, and requires less time to develop than image-based models. The 3D image-based models as used in this thesis have the advantage that detailed geometric features can be captured. However, the accuracy of the geometry obtained from this method is limited by the resolution of the MRI and accuracy of the segmentation method as well as smoothing algorithms applied. Errors in the 3D reconstruction from MR images have however been reported to be as high as two pixels in some cases. MRI commonly uses a slice thickness larger than the in-plane resolution which may add additional uncertainties in the direction of the slice thickness. Furthermore, because menisci are viscoelastic, further uncertainty is introduced by the level of pre-deformation of the tissues when the joint is scanned [571].

Some level of geometric uncertainty will therefore inevitably exist in the 3D model of the knee. If the effects of these uncertainties are not evaluated, the level of reliability of

the model predictions cannot be determined. In terms of rolling out these results to a larger population or clinical trials, using meniscus for example in lateral wedge insole trials considerations are worth noting.

Where multiple users may perform segmentation, to minimise variability, variation between paired images should be set to a pre-specified threshold. In terms of feasibility manual segmentation would be challenging for large cohorts because of time and associated costs, therefore the current measures explored in this thesis would need automated segmentation to allow their utility. Furthermore, assessment of how repeatable the automated methods is then warranted. One study found that the prediction of knee contact mechanics was less sensitive to the small variations in the inner and outer radius of the meniscus than the height of the meniscus. This finding has an important implication for reducing the uncertainties caused by the resolution of the MR images when image-based models are created. The height of the meniscus should always be captured by the highest resolution in the settings. Moreover, because meniscus exhibit viscoelastic characteristics, the initial loading conditions of the joint need careful consideration before imaging.

Other challenges that need to be overcome include the capture and representation of appropriate geometry and material properties, representation of appropriate motions and loads and establishment of relevant outputs and their levels of uncertainty [572]. In their present form, 3D meniscal measures from this thesis are still in development and will need validation and calibration in different cohorts or settings possibly using upright, weight bearing MRIs and also cohorts from younger age groups as well as in settings applying varying inclusion criteria. These models will also require verification, some of this work was covered in this thesis and numerical outputs that were clinically relevant were obtained. However, further verification work would be needed for example in assessing and altering the smoothing algorithms if needed to ensure automated models reach convergence. However no changes are thought to be needed to the methods deriving these 3D measures as their repeatability and the reliability indices were quite good. In terms of measurement accuracy, these measures showed they were very precise and would result in reduced sample sizes and costs in clinical trials.

8.3.4 Longitudinal data analysis with LCGA

8.3.4.1 Measurement issues

Latent growth curve models described change in bone shape over time and LCGA was used to identify trajectories of change within the OAI. Although these techniques of analysing longitudinal data with repeated, and therefore correlated measures of the same variable have been in use particularly in the social sciences for decades now,

their application to OA research is emerging. The strengths of longitudinal analysis versus cross sectional techniques were outlined in Chapter 2 and 3. In Chapter 7 LCGA was chosen to assess trajectories in preference to latent class-GMM because of its simplicity and flexibility. There appears to be no evidence of superiority of either approach.

Differences however exist, for LCGA, all cases within a given trajectory are presumed to follow that trajectory and the variance is therefore zero, whilst latent class-GMM allows for inter-individual variation, and variance of trajectory groups is provided within the results. Among the different classification methods assessed by Twisk and Hoekstra [573], including these 2 techniques, they observed that LCGA was preferable to latent class-GMM when the data followed linear trajectories, but neither performed well when the trajectories were quadratic. This finding strengthens the choice of LCGA as used in this thesis as it was demonstrated in Chapter 6 that all three bones followed a linear trajectory before LCGA was performed. However, whilst longitudinal methods appear superior to cross sectional techniques, LCGA is likely to estimate a greater number of trajectories compared to latent class-GMM; however, the analysis of bone shape identified 3 trajectories in all bones and only 2 trajectories when this was restricted to participants showing change greater than measurement error, which is a relatively small number and unlikely to be greater than that estimated by an alternative method.

Choosing the ‘correct’ model requires an understanding of the context in which data are generated to model variation correctly. Gilthorpe and colleagues [574] highlighted that failure to model random structure of growth outcomes carefully within a GMM framework can result in mis-specified models. Model parameterisation should therefore be driven by an understanding of underlying data generation processes, associated theory and the research question.

Given the increasing popularity of LCGA methods, it is likely that many applications will adopt constraints for the random structure to promote model parsimony or attain model convergence. These constraints may affect the number of classes obtained, creating a challenge of determining which model parameterisation is ‘correct’. This problem normally arises when there is more between-subject than within-subject heterogeneity and this can resolved by freely estimating the growth factor variances and covariances; however, as noted before, such free estimation causes convergence issues. An alternative, more parsimonious approach is to model explicitly the emergent autocorrelation structure[534].

8.3.4.2 Latent growth models in OA

In the OA field, there have been very few published LCGA studies assessing structure, with the vast majority applying these methods to assess trajectories of pain. In the literature searches conducted within this thesis, in the last 2-3 years only 2 studies applied LCGA techniques to assess structural trajectories in knee OA. However, neither assessed bone, ML techniques; or parallel process growth models. Deveza [531] found 3 trajectories for cartilage loss although one of the classes contained less than 5% participants, which based on statistical recommendations would suggest only 2 classes existed. A major limitation of that study was that follow up was restricted to only 3 time-points thereby limiting the growth patterns that could be tested to just linear growth models and the existence of quadratic growth patterns in this group cannot be ruled out. Another difference was that Deveza et.al. selected only about 25% of the OAI sub-population for their analyses and the trajectory class containing only 2.2% participants may have been less robust and is contrary to recommendations of at least 5% per class previously discussed in this thesis.

More recently Collins [565] assessed trajectories of fixed location JSW and also found that 3 trajectory classes existed. What is common in these two studies and this thesis is that they provide consistent evidence towards the “inertia hypothesis” in knee OA, that knees that have begun to progress are likely to experience further worsening while baseline differences in structural severity were explained by the “horse racing effect”, (knees that have already started progressing are likely to be “out in front” (i.e., have less joint space in the case of the Collins study and more positive bone shape values in this thesis) at baseline, because they were in a worsening trajectory before the start of the study. A consistent finding in the two studies and this thesis was that the effect of clinical and demographic covariates was very similar across all three.

Despite the application of advanced statistical methods, including parallel process growth models to assess three knee bones, a limitation of the methods used in Chapters 6 and 7 is that these methods would be inefficient, and interpretation of models too complex in cases where data is highly dimensional or when several measures need to be assed simultaneously in the same model.

8.3.5 Strengths and limitations of this thesis

There are several strengths from this thesis. The use of novel 3D measures of bone and meniscus that showed high reliability and responsiveness indices strengthens findings from this thesis, as results from such imaging data are unlikely to be masked by large measurement errors. Another strength was the (necessary) use of advanced statistical analyses to efficiently study the natural history of bone shape and investigate

longitudinal associations. The type of analyses used in this thesis were made possible by the metric nature of the novel 3D measures unlike the commonly used ordinal and SQ measures studied in OA. Another strength of this thesis was the use of data from a well-designed longitudinal cohort, the OAI, with 4796 individuals and considered one of the largest databases designed for knee OA to date, with a reasonable follow–up and standardised measurement of MRIs annually in all participants.

Analyses using 3D meniscal measures provided much richer information and better quantification than currently exists and these measures were more responsive than the current FDA-approved structural endpoint (JSN). The suitability of the bone shape biomarker described in this thesis (3D bone shape) for use as an outcome measure in disease modification trials is supported by its responsiveness and the predictive validity for TKR, described in Chapter 5 and 7, the concurrent validity for knee pain in Chapter 6 and 7 and the work done in parallel with this thesis. This work done in parallel indicated that 3D bone shape provides a measure of OA status across the whole range of severity including early disease including its predictive validity for pain and function.

However limitations also exist. The majority of the OAI cohort was Caucasian with smaller numbers from other ethnic groups. Therefore conclusions cannot be readily generalised to non-Caucasian groups. Also, as this was a USA-based cohort consisting of volunteers who may not necessarily be reflective of the USA general population themselves, conclusions from this thesis are not easily generalisable to the US population nor beyond. Body mass index was one risk factor considered as part of the inclusion criteria into the OAI. This may have influenced the magnitude of the associations described in this analysis and could introduce confounding and bias. Furthermore, the OAI recruited participants aged 45-80 and findings from this thesis may not be generalisable to a younger population or post-traumatic OA cohorts where certain adjustments may be needed for the bone and meniscal measures to allow comparability.

The feasibility of using meniscal measures derived from manual segmentation would be difficult and costly due to the time taken to segment each menisci, therefore automated segmentation methods would need to be developed to allow for their practical use in larger cohorts. This may limit the utility of meniscal measures as part of composite OA measures in their present form. Another limitation was that no coronal views were obtained and only sagittal DESS images, which offer the best compromise for identification of multiple OA tissues (here meniscus and bone) were used, but may not be the optimal sequence for detecting particular meniscal pathologies. For bone, the DESS-we MR images were used in this study and although it has been previously

demonstrated that the SSM technology is applicable to similar MRI sequences for bone, the methods would need validation in other MRI sequences.

In terms of the latent class growth models some limitations are worth noting. The model estimation algorithms seek to find the best representation of the data, given the specifications of the models and the derivation of classes may be subjective. Secondly, the trajectory groups observed and the differences among groups provided may not necessarily represent true processes that generated the data. Thus, it is important to ensure the estimates and number of classes makes clinical sense. As such, careful definition of the research questions and knowledge about the underlying data generation processes is needed when applying such techniques.

Other specific strengths and limitations for each study have been covered previously in the relevant chapters.

8.4 Future Directions

Much research on novel imaging biomarkers has been conducted with an ultimate goal of characterising their use for structure-based trials, and development within this area is ongoing. With this target in mind, the work in this Thesis has raised several areas of potential focus for future study. This section discusses further work required to determine the validity of 3D meniscus and bone shape as surrogate measures, and the potential to segment all of the joint tissues using AAMs to provide quantitative measures of whole joint OA-tissue pathology. This section also describes the potential advantages of using more precisely defined OA phenotypes from well-characterised measures. Finally, the implications of applying novel imaging analysis technology used in this thesis to other joints and clinical scenarios beyond the knee is discussed.

As discussed in Section 8.3.4 once validated and tested, future work using 3D measures from this thesis could be tested in trials such as the lateral wedge insole trials that have demonstrated reduced extrusion in OA patients before. Such work would test whether extrusion measured in 3D is influenced in the same way as SQ measures, by the intervention. The advantages of these 3D measures are their superior reliability and responsiveness which would lead to reduced sample size requirements making them attractive for new structure-based trials in the future. Future work will also involve repeating the same work as in Chapter 4 in a larger cohort subject to adequate numbers of meniscal segmentations and assessing if results are similar to this thesis.

One of the limitations of the BML study in Chapter 5 was that comparisons were made between bone shape (a continuous measure) and BMLs (measured semi-

quantitatively) therefore these comparisons were not direct. Future work should perform a head-to–head comparison between 3D bone shapes and segment BMLs to their quantitative 3D BML equivalents to perform a more direct comparison. This could be done in the FINH cohort of the OAI that already has bone and BML data collected. For BMLs future work may also consider subdividing BMLs on the basis of their location and association with adjacent tissues and then testing whether the relationship with bone shape is improved as the weak associations seen in Chapter 5 may have been due to such anatomic variations. It is possible that BMLs occurring at the ligaments and meniscal attachments may show a different response. While this thesis has provided evidence for the construct validity of bone shape and parallel work shown that 3D bone shape is highly responsive and associated with knee OA pain progression and TKR, further analyses of their validity will be required. Future validation of bone biomarkers may also include a composite measure of BMLs and bone shape to enhance the validity of the potential biomarker. This could be designed to assess the concurrent and predictive validity against knee OA symptoms and also assess whether this composite is more responsive. Such a composite measure may then be expanded in future to include cartilage, synovitis and meniscus thus employing a truly “whole organ” approach.

Chapter 6 and 7 highlighted the importance of understanding the natural history of potential biomarkers leading to improved definition and phenotyping for OA trials. Using the OAI and similar methods to this thesis, this work could be extended to assess LCGM models in cartilage and synovitis and determine the longitudinal patterns in both tissues as done for bone shape. As done for bone shape, future work would assess how these two issues are influenced by covariates with subsequent trajectory analyses performed to assess if participants can be grouped into classes based on multiple structure. This work would involve as a sub-study, parallel process growth models that incorporate bone shape, 3D cartilage measures, 3D synovitis and meniscus subject to availability of adequate numbers of tissue segmentations. Such a model could be useful to inform which structure initiates structural OA, and how this influences the rest of the knee structures and in what order.

Other themes on future work are explored in the following sections.

8.4.1 Quantitative imaging biomarkers

This thesis employed novel 3D imaging techniques that have shown great promise; however some of these are still in their testing and development stage and are not yet ready for use in clinical trials. ML technology used in this thesis can be adapted to other tissues such as cartilage and synovitis and also transferable to other regions for

example the hip and hand to improve measures for hip and hand OA progression respectively.

Bowes and colleagues [575] have demonstrated that a novel automated cartilage segmentation method using ML is feasible and comparable to manual segmentation in term of reliability and precision and additionally was more responsive than manual methods. Recently, Kingsbury et. al [576] demonstrated pain reduction with oral methotrexate in the PROMOTE study, a multi-centre RCT comparing methotrexate to placebo. Importantly in that study, they were able to assess as a secondary outcome, quantifiable synovitis changes derived from ML-techniques showing that ML measurement methods for synovitis are precise and are feasible for future trials.

Future work should explore 3D cartilage measures and assess its natural history and determinants, and also investigate if different trajectories of structural change exist as was done in this thesis for bone. For synovitis, Perry and colleagues applied a novel semi-automated assessment method using ML methods for the quantification of synovial tissue volume (STV) and found it was as accurate and reliable, but much quicker than manual segmentation for assessment of STV. This method may help increase efficiency of image assessment in large imaging studies.

8.4.1.1 Application of ML in other joints

In the hip, novel 3D imaging measures have been useful in assessing the relationship between bone shape and cartilage, and suggestions are that 3D MRI-based bone shape could be a promising biomarker of early hip joint degeneration [577]. Hip cohorts such as the Leeds OA Hip Cohort have been set up recently to characterise novel hip measures using ML techniques. ML in the hip joint will address issues of hip shape; such as whether Cam-type and pincer-type femoro-acetabular impingement are merely variations on a continuum or consistently pathological. Additionally, 3D shape modes in the hip will help elucidate the pattern of OA structural changes that cannot be readily appreciated on 2D imaging e.g. osteophyte development in the infero-posterior and posterolateral hip joint.

In hand RA the use of MR images with OMERACT RAMRIS SQ scoring has greatly improved the way that DMARD clinical trial image data is analysed through analysis of soft tissue changes and inflammation compared to radiographic scoring. More recently RAMRIQ, a quantitative analogue of the RAMRIS scoring method developed from ML has also been developed. This new technique has offered improved sensitivity and responsiveness and due to automation, reduced image analysis time compared to RAMRIS [578]. As such, RAMRIQ is already being employed in numerous

retrospective and prospective studies of clinical trial image data [579, 580]. In hand OA, RAMRIQ should be easily adapted to produce a hand OA diagnostic tool. In addition, the continuous variables produced would allow the use of more powerful statistical techniques than those used with the categorical scoring of RAMRIS.

Back pain is the most commonly reported musculoskeletal disorder, however its diagnosis is challenging and the cause is often unclear with most patients often characterised as having non-specific back pain. While screening tools such as the StarT-back[581], have improved the assessment of back pain, such tools still lack objective structural measures. Application of ML technology to the spine will result in more precise assessment of back pathology and result in more objective and precise assessment of the spine thereby greatly improving diagnostic accuracy.

8.4.1.2 Implications for understanding pathogenesis

Improved structure measurement

Structure measurements are challenging due a multiple of factors. There are varying reliability indices in the imaging techniques and these also vary by tissue involved. The interpretation of images also varies due to the different scoring systems available and user preferences and notwithstanding the measurement errors are inherent in imaging studies [315]. Future work should look to improve reliability scores and precision across all tissues and also standardise scoring through the use of uniform quantification techniques. Such improvements are possible with the use of ML measurements which should ensure much better structure assessment in future.

Any study evaluating different imaging modalities requires the assessment of reliability. A strength of the work in this thesis is that all measurements used herein were carefully obtained and reliability tested for each one. Reliability indices were also reported for the primary MRI source, meniscal and bone shape measures; however this is not consistently reported for structural measures in other OA studies and there needs to be improvement in the transparency and reporting across the OA field. Over time such improvements in reporting would result in more precise measures becoming available.

While the advantages of ML techniques have been demonstrated in this thesis, their uptake in OA is still slow. This may be partly due to feasibility issues for example lack of expertise regarding their use by other research groups, but also the general lack of understanding of the relationship between structural pathologies and pain or other structures. Another reason is that there are only a few OA structure trials currently ongoing for reasons highlighted in Chapter 3. An understanding of these relationships

could provide better knowledge on pathogenesis and a need for more appropriate, targeted interventions.

Better understanding of structure-pain and structure-structure relationships

Although a large part of this thesis investigated the responsiveness, construct validity and predictive ability of 3D measures rather than direct causal relationships between structure and pain, the importance of understanding these associations should be considered in light of findings from other knee imaging studies. There has been much work in the area of structural pathology in OA and relationship to symptoms but less assessing structure-structure associations. In knees, understanding such relationships is difficult in part due to the complexity of measuring pain, and the variability in measuring structure. Pain is a multifaceted and subjective experience as noted in Chapter 2, and measuring pain experience is difficult due to the various modifiers of pain for example intensity, location, mood and beliefs. As discussed later in section 8.4.2 of this Chapter, methodological issues including confounding assessment also contribute to problems with accurately assessing OA structure-pain concepts.

Whilst there is evidence supporting the contributory effect of structural pathology with pain, the particular structural pathologies responsible have yet to be elucidated fully. Areas such as subchondral bone, periosteum, synovium, ligaments and peri-articular muscle contain nociceptive receptors, and represent potential targets [582]. Further work in the knee, therefore, needs to involve deeper understanding of the relationship between the different structural pathologies and also with pain to enable appropriately targeted therapies. MRI has enabled the evaluation of more pathologies than would otherwise be detectable. Future work should now involve a more detailed assessment of temporal and spatial relationships of all structural elements of OA using these more accurate measures. A limitation of this thesis is that while the structural changes over time have been assessed, the causal associations between structural changes and pain were not fully explored but limited to just predictive associations.

8.4.1.3 Improved clinical trial outcomes and stratification for clinical trials

Accurate automated segmentation of structures should provide rapid reliable outcomes for both epidemiological studies and clinical trials. ML has made possible the development of a quantitative measures of OA status; four measures of 3D meniscal

pathology and 3D bone shape in three knee bones. Using a large observational cohort, meniscal measures demonstrated superior responsiveness while bone shape, a measure of structural disease status was associated with clinically important outcomes, providing construct validity with the existing radiographic standard and was more responsive.

Findings from recent studies highlighted in Section 8.3 and that of this thesis have implications for future clinical trial measures as they highlight that cartilage assessment that is currently measured indirectly as JSW is not reflective of the true underlying structural pathology (hyaline cartilage thickness). Therefore by exploiting the potential of ML, meniscal measures should be further evaluated and characterised so they can be incorporated into existing cartilage measures to give more reliable composites given how much variance in JSW they explain. In clinical trials, bone shape would provide a reliable stratification tool, and has already shown to be a sensitive outcome measure [215].

Implications for clinical practice require further consideration, and at present may improve assessment of prognosis more than selection of therapy (given the limited non-surgical therapeutic options available). However, bone shape may initially provide clinical usefulness in situations where MRI is already commonly performed (e.g. sporting injuries or ‘possible early OA’).

8.4.2 Improved methodology and study design

To better understand the structure-pain and structure-structure relationships, certain methodological and statistical challenges need to be overcome. The ‘natural history’ of structural lesions and symptom development needs clarity. This was attempted in Chapters 6 and 7 of this thesis and has led to new understanding of structural changes in subchondral bone tissue changes over time. Currently, it is unknown how much of the variance in pain is accounted for by structural change and assessing the causal effects of the various pathologic features in OA to the pain experience remains difficult. However, there is still scope to improve how the models are specified so as to obtain more accurate estimates of this variance. Chapter 6 of this thesis assessed some commonly cited risk factors for OA and it was shown that all these jointly contributed to 20% of the variance in structure at best, and only 5% was attributable to knee pain. Greater variance was seen at baseline while over longitudinally risk factors had less influence on the rate of change in structure. This was uniform across all three knee bones.

8.4.2.1 Improved assessment of confounding

Some pitfalls in the current methodologies in OA studies were highlighted in Chapter 2. A commonly occurring methodological problem is that of including all structural lesions in a single statistical model to obtain “independent associations” by interpreting the coefficients as mutually adjusted effects of various structural pathologies, and has been dubbed the “Table 2 Fallacy” [233]. Presentation of the main exposure and confounder effect estimates from a single model may lead to difficulties in interpretation of direct- and total-effects of the covariates, as this table presenting multiple estimated effect measures from the same model that encourages the reader to interpret all these estimates in the same way, typically as total-effect estimates. However, interpretation of a confounder-effect estimate may be different than for the exposure-effect estimate.

It may be possible in some cases that the effect estimates of covariates may also be confounded even though the estimate for the main exposure is not confounded. Interpretation of models can be further confounded by heterogeneity of the exposure effect measure across covariate levels. Two critical issues in multiple linear regression models are multi-collinearity and variable selection strategies. When there is multi-collinearity the resulting regression coefficient estimates can be very unstable, and the standard errors large. Another important issue in multiple linear regression analysis is variable selection, especially when the number of potential predictors is large. Strategies including forward selection, backward elimination, and all-subsets regression, and each has with its strengths and limitations. Moreover, clinical or biologic considerations also need to be taken into account in model building.

While solutions such as use DAGs [583] or causal diagrams have been proposed to address issues to do with confounding and better address causal associations, their uptake in the OA field has been slow. DAGs are useful as they are designed to reduce different biases including those caused by confounding in order to disentangle the effects of individual pathologies on pain. The use of DAGs has improved the understanding of how the different individual pathologies interact as part of a complex causal framework. For example an appreciation of what confounders, mediators (that could dilute or attenuate the effect of the pathology of interest and should be avoided when estimating total causal effects) and competing exposures (useful in the model to improve precision of model estimates) are, and their effects on the studied exposure. Improvements in this regard have continued including the availability of resources such as DAGitty [584], an online resource tool that aids in the selection of DAGS and can help identify the appropriate adjustment sets to include in the modelling. DAGs are a simple tool, semi parametric in nature while regression models are parametric. Many

different DAGs will be compatible with different data structures and several data structures compatible with a DAG.

The importance of DAGS is that they help us identify our assumptions of how variables are causally related, by ensuring the study explicitly states what assumptions are being made which is an advancement from the post-hoc approaches, by allowing transparency at the onset. However limitations still remains as interpretation of the effect sizes from different studies can be subject to variations as a result of the different causal and interrelated pathways between covariates in each setting as well as study designs [189, 232]. Another issue is that even after conditioning, there is always some residual confounding remaining because one can never capture and measure the concept perfectly. Another limitation of DAGs is that the accuracy of estimated causal effects is conditional of the accuracy of the corresponding DAG.

While the importance of DAGs in model selection is important, improvements in the modelling framework need to be considered carefully for each confounder before each analyses. Different ways by which confounders can be treated include restriction (estimating the effect of the confounder at a fixed level of such a confounder), stratification (by estimating the effect within strata of the confounder), covariate adjustment (estimating the effect conditional on values of the confounder) and matching (estimating the effect in clusters with matching values of the confounders). In summary, to estimate total causal effects it is important to condition on confounders to block confounding paths, avoiding conditioning on mediators as these may block true causal paths or worse result in collider bias and optionally condition on competing exposures to improve the precision of estimates.

8.4.2.2 Improved study design and analysis

The difficulty in understanding the degree to which structural pathology accounts for pain and the casual contributions of different structural pathologies may in part be due to assessing knee pain in the late stages thus the need for large, longitudinal prospective studies of patients with limited initial pathologies, to understand the natural history of pain and structure-pain relationships. The OAI is a useful resource used in this thesis and one of the largest knee cohorts to date but has its own limitations as robustly exploring structure-structure associations was not possible due to lack of measurements from other structures such as the meniscus, and the OAI inclusion criteria for pain hampers robust structure-pain assessment. It is thus imperative to develop highly responsive, reliable and sensitive measures as part of solutions to limit the need for very large longitudinal cohorts.

While the causal assessment of the structure-pain relationships was not possible in this thesis due to the nature of the OAI study design, this has does not affect the validity of the statistical analyses carried out. Growth curve analysis and latent class growth analysis are robust methods and have a role in phenotyping individuals more precisely for knee OA clinical trials. Participants with more severe pain were found to be consistently in the latent classes with the most structural severity based on bone shape. The strength of this work includes very large patient numbers, but there are limitations. There was no attempt to explore longitudinal change or relationship to cartilage as focus, was on characterising and adding validation data to what is already known about this novel bone measures. However, clear relationships with clinically important outcomes were demonstrated. As only one structural pathology (bone) was considered, it would be of interest to understand how other pathologies may perform when treated similarly.

Therefore, further work is required to elucidate if bone shape can improve understanding of structure-symptom relationships. Further work is also needed to understand how different tissues would be specified in latent growth parallel process models. The bone shape vector revealed here may not hold for very late stages of the disease, where fewer patient numbers were available in this study. Also, when osteophytes begin to carry load directly, they are likely to remodel, and may produce shape changes that are less systematic than those reported here.

8.4.3 Application of machine learning imaging technology

OA research has revealed several imaging findings, serum biomarkers and symptoms that are now considered important for understanding OA progression. This thesis has shown that use of accurate, reliable and highly reproducible measurement techniques such as 3D when applied in a large dataset can result useful novel imaging tools. This provides proof of concept for ML quantification.

ML approaches (unsupervised and deep learning) are relatively new to OA research but could possibly provide new approaches especially in enabling better characterisation of structure and subsequently stratification of affected individuals. Determining the relative importance of such variables can lead to identification of new OA phenotypes that represent different pathways. A useful method is to define clinically important outcomes and subsequently employ ML algorithms to identify the relative contributions of each variable for each one of them. This is particularly important now, with the advent of big data from various sources such as MRI imaging, serum analysis, genome sequencing, and electronic medical records.

Using the OAI dataset, Nelson and colleagues have shown that ML techniques can feasibly assess the relative importance of each pre-specified variable in the FNIH progression groups for example [585], the same dataset assessed in Chapter 5 of this thesis. They employed ML techniques that are useful in the analysis of high dimensional datasets [586, 587], and one such technique recently used to phenotype and subgroup hip analysis in an OA hip cohort [588].

8.4.3.1 Better understanding of pathogenesis

Using ML-derived measures Chapter 7 demonstrated the existence of homogenous sub-groups based on structural change which were clinically meaningful as they related to known OA risk factors, although further work on this is still needed. In OA, ML techniques have recently been evaluated for their ability to discriminate between normal vs pathological cartilage from MR images with high accuracy demonstrating their potential use in the clinical detection and grading of OA [589].

Imaging advances continue to improve the understanding of multiple tissues in knee OA. It is possible that deep learning techniques such as 2D and 3D convolutional neural networks (CNN) may be useful in future OA imaging studies. The 3D-CNN modelling approach has been shown to be well-suited for performing rapid and accurate and detailed tissue segmentation of the knee joint [590]. Deep learning-based segmentation methods have promising potential applications in musculoskeletal imaging.

Pedoia and colleagues provided proof of concept of a fully automated deep-learning techniques that can identify the presence of meniscal and patellar cartilage lesions. They were able to detect meniscus lesions with a sensitivity of about 90% and specificity of 82%. This work has the potential for more in-depth examinations of lesions for multiclass prediction and severity staging [591]. There is already emerging data that these techniques may be feasible, although much work in their validation is still necessary.

Norman et.al. analysed how automatic segmentation using neural networks compared with manual segmentation in terms of accuracy and precision by applying this technology on the meniscus [592]. They showed that neural networks demonstrated efficacy and precision in quickly generating accurate segmentations that can be used to extract relaxation times and morphologic characterization, and values that can be used in the monitoring and diagnosis of OA. An advantage of these techniques for OA is improved measurement due to highly reliable measures but also increasing the speed and accuracy of the work flows that use MRI. The models from this study averaged 5 seconds to generate automatic segmentations [592].

This thesis used DESS-we MR images and while it has been previously demonstrated that the method is applicable to similar MRI sequences [544], the method would need further validation for other MRI sequences. Although the ML method for bone shape determination used in this study is proprietary, several methods for bone shape measurement have been published, and the measurement of bone shape is actively being pursued by multiple groups. ML technology can almost certainly be applied to cheaper imaging methods such as computerised tomography.

8.4.3.2 Improved understanding of structure-pain

With increased big data availability and development of novel algorithms, including the recent emergence of quantum ML algorithms [593], there is the potential to successfully define unexpected patient's phenotypes linked with responsiveness to precise therapies in the near future. All these ML improvements have the potential to improve current tools for clinical and imaging diagnostics, hopefully supporting more robust study designs and have the potential to make a step-change in improving patients outcomes by distinguishing the responders from the non-responders in clinical trials.

Understanding patterns of symptoms and trajectories of pain and functional decline in OA where fluctuations of symptoms are common is important. With improved methods and software now able to handle multivariate longitudinal data ML has potential to identify variables which are important for each patient-group in a specific time scale [594]. This thesis has investigated one such method using data derived from ML techniques and has provided evidence of the longitudinal change in a potential imaging biomarker and also shown trajectories of this change over time.ML techniques employed on big data will furthermore enable hypothesis-driven analyses investigating overlap between different known phenotypes, for example the potential overlap between inflammatory and metabolic OA phenotypes [594].

To date, most studies using ML have largely been employed to define biomarker panels from ‘omics’ and multi-dimensional imaging data due to their size and number of variables [589, 595]. The 'omics' (genomics, epigenomics, transcriptomics, proteomics, metabolomics and lipidomics) are poised to make significant contributions to the identification of novel biomarkers. Applying ‘omics’ results in the generation of large datasets that are suitable for bioinformatic analysis using ML, to extract important information is part of a growing body of work [596]. With further research into ML techniques it is expected that such type of analysis will enable the identification of specific phenotypes with a higher likelihood of precision.

8.4.3.3 Provision of better outcome measures and improved stratification for clinical trials (better phenotyping)

The growth of ML techniques should help improve the subgrouping of clinically important groups through clustering of important variables. As OA is commonly thought to be a heterogeneous disease, identifying new biomarkers might help select individuals more optimally and make trials in well-defined subgroups feasible.

It would be worthwhile to better define OA phenotypes by using a broad range of potential phenotype determinants. This includes symptom and demographic data, MRI imaging biomarkers, wet biomarkers, along with epigenetic data. Kraus and colleagues, and also Kerkhof have used similar methods [176, 209]. Together these can be used to establish more homogenous classes through growth mixture models in a similar manner to Wesseling and colleagues [597] as was demonstrated by this thesis. Knee OA phenotypes and their prognosis have previously been more precisely described by stratifying individuals by their trajectories of structural and symptomatic progression as described by Felson and colleagues [598].

8.5 Conclusion

The lack of valid, predictive and responsive imaging biomarkers still hampers development of structure modifying therapies in OA. Improved outcome measures, such as those derived from novel machine-learning techniques should provide valid and responsive measures for understanding OA pathogenesis and also result in better stratification of individuals for knee OA clinical trials. A longitudinal study identified new 3D meniscus measures which were highly responsive and better than radiographic JSN and this thesis provides an important contribution to work on the responsiveness of meniscal biomarkers. This thesis has demonstrated a relationship between a novel bone biomarker and a well-established one providing validity for 3D bone measures and justifies the need for further work on this. Using advanced latent growth modelling methodology, the natural history of bone shape was reported for the first time and characterised more precisely showing that any of the three knee bones were feasible biomarkers of bone, although evidence suggests the femur may be a more optimal choice for clinical trials based on its superior responsiveness. A subsequent latent class growth analysis confirmed that different trajectory groups exist in terms of structural change which may have implications for future knee OA trials through optimised selection of participants.

In conclusion, Machine Learning coupled with advanced statistical techniques has enabled the characterisation of novel meniscus and bone imaging biomarkers

representing OA status. These novel measures have demonstrated clear relationships with clinically important outcomes in a disease where development of new therapies has been hampered by poor understanding of structure-symptom relationships.. These findings should aid the identification of both known and novel knee OA phenotypes, potentially improving patient selection for specific interventions, and providing insight into pathophysiology in this heterogeneous condition. The lack of valid and responsive biomarkers currently slows therapeutic advances in OA and results in increased costs in conducting clinical trials. The attraction of integrating valid and responsive biomarkers in the development process is that the costs can be optimised and less promising projects can be stopped much earlier. There is therefore a public health benefit in developing OA therapies in a shorter time and using validated biomarkers other than radiographic JSN.

Chapter 9 References

1. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015; 386: 743-800.
2. Felson DT, Lawrence RC, Hochberg MC, McAlindon T, Dieppe PA, Minor MA, et al. Osteoarthritis: New insights - Part 2: Treatment approaches. Annals of Internal Medicine 2000; 133: 726-737.
3. Hunter DJ, Losina E, Guermazi A, Burstein D, Lassere MN, Kraus V. A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials. Current Drug Targets 2010; 11: 536-545.
4. Hunter DJ, Nevitt M, Losina E, Kraus V. Biomarkers for osteoarthritis: Current position and steps towards further validation. Best Practice & Research Clinical Rheumatology 2014; 28: 61-71.
5. Richter WS. Imaging biomarkers as surrogate endpoints for drug development. Eur J Nucl Med Mol Imaging 2006; 33 Suppl 1: 6-10.
6. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. J Health Econ 2003; 22: 151-185.
7. Mazzuca SA, Brandt KD, Buckwalter KA, Lequesne M. Pitfalls in the accurate measurement of joint space narrowing in semiflexed, anteroposterior radiographic imaging of the knee. Arthritis Rheum 2004; 50: 2508-2515.
8. Hunter DJ, Zhang YQ, Tu X, Lavalley M, Niu JB, Amin S, et al. Change in joint space width: hyaline articular cartilage loss or alteration in meniscus? Arthritis Rheum 2006; 54: 2488-2495.
9. Conaghan PG, Hunter DJ, Maillefert JF, Reichmann WM, Losina E. Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group. Osteoarthritis Cartilage 2011; 19: 606-610.
10. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. Ieee Transactions on Pattern Analysis and Machine Intelligence 2001; 23: 681-685.
11. Williams TG, Vincent G, Bowes M, Cootes T, Balamoody S, Hutchinson C, et al. AUTOMATIC SEGMENTATION OF BONES AND INTER-IMAGE ANATOMICAL CORRESPONDENCE BY VOLUMETRIC STATISTICAL MODELLING OF KNEE MRI. 2010 7th Ieee International Symposium on Biomedical Imaging: From Nano to Macro 2010: 432-435.
12. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis and cartilage 2013; 21: 1145-1153.
13. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012; 380: 2163-2196.
14. Chen A, Gupte C, Akhtar K, Smith P, Cobb J. The Global Economic Cost of Osteoarthritis: How the UK Compares. Arthritis 2012; 2012: 698709.

15. Geusens PP, van den Bergh JP. Osteoporosis and osteoarthritis: shared mechanisms and epidemiology. Curr Opin Rheumatol 2016; 28: 97-103.
16. Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol 2000; 27: 1513-1517.
17. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957; 16: 494-502.
18. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007; 15 Suppl A: A1-56.
19. Bedson J, Croft PR. The discordance between clinical and radiographic knee osteoarthritis: a systematic search and summary of the literature. BMC Musculoskelet Disord 2008; 9: 116.
20. 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; 52: 3152-3159.
21. Guermazi A, Niu J, Hayashi D, Roemer FW, Englund M, Neogi T, et al. Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population based observational study (Framingham Osteoarthritis Study). BMJ 2012; 345: e5339.
22. Baker K, Grainger A, Niu J, Clancy M, Guermazi A, Crema M, et al. Relation of synovitis to knee pain using contrast-enhanced MRIs. Ann Rheum Dis 2010; 69: 1779-1783.
23. Roemer FW, Kassim Javaid M, Guermazi A, Thomas M, Kiran A, Keen R, et al. Anatomical distribution of synovitis in knee osteoarthritis and its association with joint effusion assessed on non-enhanced and contrast-enhanced MRI. Osteoarthritis Cartilage 2010; 18: 1269-1274.
24. D'Agostino MA, Conaghan P, Le Bars M, Baron G, Grassi W, Martin-Mola E, et al. EULAR report on the use of ultrasonography in painful knee osteoarthritis. Part 1: prevalence of inflammation in osteoarthritis. Ann Rheum Dis 2005; 64: 1703-1709.
25. Belo JN, Berger MY, Koes BW, Bierma-Zeinstra SM. The prognostic value of the clinical ACR classification criteria of knee osteoarthritis for persisting knee complaints and increase of disability in general practice. Osteoarthritis Cartilage 2009; 17: 1288-1292.
26. Zhang W, Doherty M, Peat G, Bierma-Zeinstra MA, Arden NK, Bresnihan B, et al. EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. Ann Rheum Dis 2010; 69.
27. Braun HJ, Gold GE. Diagnosis of osteoarthritis: imaging. Bone 2012; 51: 278-288.
28. Keen HI, Wakefield RJ, Grainger AJ, Hensor EM, Emery P, Conaghan PG. Can ultrasonography improve on radiographic assessment in osteoarthritis of the hands? A comparison between radiographic and ultrasonographic detected pathology. Ann Rheum Dis 2008; 67: 1116-1120.
29. Arden N, Nevitt MC. Osteoarthritis: epidemiology. Best Pract Res Clin Rheumatol 2006; 20: 3-25.
30. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, et al. Osteoarthritis: new insights. Part 1: the disease and its risk factors. Ann Intern Med 2000; 133: 635-646.

31. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014; 73: 1323-1330.
32. Guccione AA, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. Am J Public Health 1994; 84: 351-358.
33. Palazzo C, Ravaud J-F, Papelard A, Ravaud P, Poiraudeau S. The burden of musculoskeletal conditions. PloS one 2014; 9.
34. Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 1998; 41.
35. Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 1998; 41: 778-799.
36. Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. Ann Rheum Dis 2001; 60: 91-97.
37. van Saase JL, van Romunde LK, Cats A, Vandenbroucke JP, Valkenburg HA. Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. Ann Rheum Dis 1989; 48: 271-280.
38. DeFrances CJ, Podgornik MN. 2004 National Hospital Discharge Survey. Adv Data 2006: 1-19.
39. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis Rheum 1987; 30: 914-918.
40. Duncan RC, Hay EM, Saklatvala J, Croft PR. Prevalence of radiographic osteoarthritis–it all depends on your point of view. Rheumatology (Oxford) 2006; 45.
41. Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, et al. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. Ann Rheum Dis 1998; 57: 649-655.
42. Jinks C, Jordan K, Ong BN, Croft P. A brief screening tool for knee pain in primary care (KNEST). 2. Results from a survey in the general population aged 50 and over. Rheumatology (Oxford) 2004; 43: 55-61.
43. Ladouceur M, Rahme E, Pineau CA, Joseph L. Robustness of prevalence estimates derived from misclassified data from administrative databases. Biometrics 2007; 63: 272-279.
44. Oliveria SA, Felson DT, Reed JI, Cirillo PA, Walker AM. INCIDENCE OF SYMPTOMATIC HAND, HIP, AND KNEE OSTEOARTHRITIS AMONG PATIENTS IN A HEALTH MAINTENANCE ORGANIZATION. Arthritis and Rheumatism 1995; 38: 1134-1141.
45. Wilson MG, Michet CJJ, Ilstrup DM, Melton LJ, III. IDIOPATHIC SYMPTOMATIC OSTEOARTHRITIS OF THE HIP AND KNEE A POPULATION-BASED INCIDENCE STUDY. Mayo Clinic Proceedings 1990; 65: 1214-1221.

46. Yu D, Peat G, Bedson J, Jordan KP. Annual consultation incidence of osteoarthritis estimated from population-based health care data in England. Rheumatology (Oxford) 2015; 54: 2051-2060.
47. Palazzo C, Nguyen C, Lefevre-Colau MM, Rannou F, Poiraudeau S. Risk factors and burden of osteoarthritis. Ann Phys Rehabil Med 2016.
48. Neogi T, Zhang Y. Epidemiology of osteoarthritis. Rheum Dis Clin North Am 2013; 39: 1-19.
49. Anderson AS, Loeser RF. Why is Osteoarthritis an Age-Related Disease? Best practice & research. Clinical rheumatology 2010; 24: 15.
50. Dillon CF, Rasch EK, Gu Q, Hirsch R. Prevalence of knee osteoarthritis in the United States: arthritis data from the Third National Health and Nutrition Examination Survey 1991-94. J Rheumatol 2006; 33: 2271-2279.
51. Knoop J, van der Leeden M, van der Esch M, Thorstensson CA, Gerritsen M, Voorneman RE, et al. Association of lower muscle strength with self-reported knee instability in osteoarthritis of the knee: results from the Amsterdam Osteoarthritis cohort. Arthritis Care Res (Hoboken) 2012; 64: 38-45.
52. Pai YC, Rymer WZ, Chang RW, Sharma L. Effect of age and osteoarthritis on knee proprioception. Arthritis Rheum 1997; 40: 2260-2265.
53. Srikanth VK, Fryer JL, Zhai G, Winzenberg TM, Hosmer D, Jones G. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis Cartilage 2005; 13: 769-781.
54. Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan JL, Protheroe J, Jordan KP. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. Osteoarthritis Cartilage 2015; 23: 507-515.
55. Spector TD, Campion GD. Generalised osteoarthritis: a hormonally mediated disease. Ann Rheum Dis 1989; 48: 523-527.
56. Zhang Y, McAlindon TE, Hannan MT, Chaisson CE, Klein R, Wilson PW, et al. Estrogen replacement therapy and worsening of radiographic knee osteoarthritis: the Framingham Study. Arthritis Rheum 1998; 41: 1867-1873.
57. Wluka AE, Davis SR, Bailey M, Stuckey SL, Cicuttini FM. Users of oestrogen replacement therapy have more knee cartilage than non-users. Annals of the Rheumatic Diseases 2001; 60: 332-336.
58. Nevitt MC, Felson DT. Sex hormones and the risk of osteoarthritis in women: epidemiological evidence. Ann Rheum Dis 1996; 55: 673-676.
59. Nevitt MC, Felson DT, Williams EN, Grady D. The effect of estrogen plus progestin on knee symptoms and related disability in postmenopausal women: The Heart and Estrogen/Progestin Replacement Study, a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2001; 44: 811-818.
60. Nelson AE, Braga L, Renner JB, Atashili J, Woodard J, Hochberg MC, et al. Characterization of individual radiographic features of hip osteoarthritis in African American and White women and men: the Johnston County Osteoarthritis Project. Arthritis Care Res (Hoboken) 2010; 62: 190-197.
61. Nevitt MC, Xu L, Zhang YQ, Lui LY, Yu W, Lane NE, et al. Very low prevalence of hip osteoarthritis among Chinese elderly in Beijing, China,

compared with whites in the United States - The Beijing Osteoarthritis study. Arthritis and Rheumatism 2002; 46: 1773-1779.
62. Foley B, Cleveland RJ, Renner JB, Jordan JM, Nelson AE. Racial differences in associations between baseline patterns of radiographic osteoarthritis and multiple definitions of progression of hip osteoarthritis: the Johnston County Osteoarthritis Project. Arthritis Research & Therapy 2015; 17.
63. Spector TD, MacGregor AJ. Risk factors for osteoarthritis: genetics. Osteoarthritis Cartilage 2004; 12 Suppl A: S39-44.
64. Fernandez-Moreno M, Rego I, Carreira-Garcia V, Blanco FJ. Genetics in osteoarthritis. Curr Genomics 2008; 9: 542-547.
65. Panoutsopoulou K, Zeggini E. Advances in osteoarthritis genetics. J Med Genet 2013; 50: 715-724.
66. Hochberg MC, Yerges-Armstrong L, Yau M, Mitchell BD. Genetic epidemiology of osteoarthritis: recent developments and future directions. Current Opinion in Rheumatology 2013; 25: 192-197.
67. Barr AJ, Campbell TM, Hopkinson D, Kingsbury SR, Bowes MA, Conaghan PG. A systematic review of the relationship between subchondral bone features, pain and structural pathology in peripheral joint osteoarthritis. Arthritis Res Ther 2015; 17: 228.
68. Lo GH, Zhang Y, McLennan C, Niu J, Kiel DP, McLean RR, et al. The ratio of medial to lateral tibial plateau bone mineral density and compartment-specific tibiofemoral osteoarthritis. Osteoarthritis Cartilage 2006; 14: 984-990.
69. Bruyere O, Dardenne C, Lejeune E, Zegels B, Pahaut A, Richy F, et al. Subchondral tibial bone mineral density predicts future joint space narrowing at the medial femoro-tibial compartment in patients with knee osteoarthritis. Bone 2003; 32: 541-545.
70. Dore D, Quinn S, Ding C, Winzenberg T, Cicuttini F, Jones G. Subchondral bone and cartilage damage: a prospective study in older adults. Arthritis Rheum 2010; 62: 1967-1973.
71. Buckland-Wright JC, Lynch JA, Macfarlane DG. Fractal signature analysis measures cancellous bone organisation in macroradiographs of patients with knee osteoarthritis. Ann Rheum Dis 1996; 55: 749-755.
72. Nevitt MC, Zhang Y, Javaid MK, Neogi T, Curtis JR, Niu J, et al. High systemic bone mineral density increases the risk of incident knee OA and joint space narrowing, but not radiographic progression of existing knee OA: the MOST study. Ann Rheum Dis 2010; 69: 163-168.
73. Jacobsen S, Jensen TW, Bach-Mortensen P, Hyldstrup L, Sonne-Holm S. Low bone mineral density is associated with reduced hip joint space width in women: results from the Copenhagen Osteoarthritis Study. Menopause-the Journal of the North American Menopause Society 2007; 14: 1025-1030.
74. Li B, Aspden RM. Composition and mechanical properties of cancellous bone from the femoral head of patients with osteoporosis or osteoarthritis. J Bone Miner Res 1997; 12: 641-651.
75. McAlindon T, LaValley M, Schneider E, Nuite M, Lee JY, Price LL, et al. Effect of vitamin D supplementation on progression of knee pain and cartilage volume loss in patients with symptomatic osteoarthritis: a randomized controlled trial. JAMA 2013; 309: 155-162.

76. Peregoy J, Wilder FV. The effects of vitamin C supplementation on incident and progressive knee osteoarthritis: a longitudinal study. Public Health Nutr 2011; 14: 709-715.
77. McAlindon TE, Biggee BA. Nutritional factors and osteoarthritis: recent developments. Curr Opin Rheumatol 2005; 17: 647-652.
78. Felson DT, Niu J, Clancy M, Aliabadi P, Sack B, Guermazi A, et al. Low levels of vitamin D and worsening of knee osteoarthritis: results of two longitudinal studies. Arthritis Rheum 2007; 56: 129-136.
79. McAlindon TE, Jacques P, Zhang YQ, Hannan MT, Aliabadi P, Weissman B, et al. Do antioxidant micronutrients protect against the development and progression of knee osteoarthritis? Arthritis and Rheumatism 1996; 39: 648-656.
80. Lane NE, Gore LR, Cummings SR, Hochberg MC, Scott JC, Williams EN, et al. Serum vitamin D levels and incident changes of radiographic hip osteoarthritis: a longitudinal study. Study of Osteoporotic Fractures Research Group. Arthritis Rheum 1999; 42: 854-860.
81. Moreno-Reyes R, Mathieu F, Boelaert M, Begaux F, Suetens C, Rivera MT, et al. Selenium and iodine supplementation of rural Tibetan children affected by Kashin-Beck osteoarthropathy. Am J Clin Nutr 2003; 78: 137-144.
82. Canter PH, Wider B, Ernst E. The antioxidant vitamins A, C, E and selenium in the treatment of arthritis: a systematic review of randomized clinical trials. Rheumatology (Oxford) 2007; 46: 1223-1233.
83. Cooper C, Inskip H, Croft P, Campbell L, Smith G, McLaren M, et al. Individual risk factors for hip osteoarthritis: obesity, hip injury, and physical activity. Am J Epidemiol 1998; 147: 516-522.
84. Felson DT, Anderson JJ, Naimark A, Walker AM, Meenan RF. Obesity and knee osteoarthritis. The Framingham Study. Ann Intern Med 1988; 109: 18-24.
85. Oliveria SA, Felson DT, Cirillo PA, Reed JI, Walker AM. Body weight, body mass index, and incident symptomatic osteoarthritis of the hand, hip, and knee. Epidemiology 1999; 10: 161-166.
86. Gushue DL, Houck J, Lerner AL. Effects of childhood obesity on three-dimensional knee joint biomechanics during walking. J Pediatr Orthop 2005; 25: 763-768.
87. Messier SP. Osteoarthritis of the knee and associated factors of age and obesity: effects on gait. Med Sci Sports Exerc 1994; 26: 1446-1452.
88. Messier SP, Gutekunst DJ, Davis C, DeVita P. Weight loss reduces knee-joint loads in overweight and obese older adults with knee osteoarthritis. Arthritis Rheum 2005; 52: 2026-2032.
89. DeVita P, Hortobagyi T. Obesity is not associated with increased knee joint torque and power during level walking. J Biomech 2003; 36: 1355-1362.
90. Lago F, Dieguez C, Gomez-Reino J, Gualillo O. The emerging role of adipokines as mediators of inflammation and immune responses. Cytokine Growth Factor Rev 2007; 18: 313-325.
91. Visser M. Higher levels of inflammation in obese children. Nutrition 2001; 17: 480-481.
92. Fain JN. Release of interleukins and other inflammatory cytokines by human adipose tissue is enhanced in obesity and primarily due to the nonfat cells. Vitam Horm 2006; 74: 443-477.

93. Ding C, Stannus O, Cicuttini F, Antony B, Jones G. Body fat is associated with increased and lean mass with decreased knee cartilage loss in older adults: a prospective cohort study. Int J Obes (Lond) 2013; 37: 822-827.
94. Dumond H, Presle N, Terlain B, Mainard D, Loeuille D, Netter P, et al. Evidence for a key role of leptin in osteoarthritis. Arthritis Rheum 2003; 48: 3118-3129.
95. Simopoulou T, Malizos KN, Iliopoulos D, Stefanou N, Papatheodorou L, Ioannou M, et al. Differential expression of leptin and leptin's receptor isoform (Ob-Rb) mRNA between advanced and minimally affected osteoarthritic cartilage; effect on cartilage metabolism. Osteoarthritis Cartilage 2007; 15: 872-883.
96. Pottie P, Presle N, Terlain B, Netter P, Mainard D, Berenbaum F. Obesity and osteoarthritis: more complex than predicted! Annals of the Rheumatic Diseases 2006; 65: 1403-1405.
97. Zheng H, Chen C. Body mass index and risk of knee osteoarthritis: systematic review and meta-analysis of prospective studies. BMJ Open 2015; 5: e007568.
98. Rossignol M, Leclerc A, Allaert FA, Rozenberg S, Valat JP, Avouac B, et al. Primary osteoarthritis of hip, knee, and hand in relation to occupational exposure. Occup Environ Med 2005; 62: 772-777.
99. Zhang Y, Xu L, Nevitt MC, Aliabadi P, Yu W, Qin M, et al. Comparison of the prevalence of knee osteoarthritis between the elderly Chinese population in Beijing and whites in the United States: The Beijing Osteoarthritis Study. Arthritis Rheum 2001; 44: 2065-2071.
100. McWilliams DF, Leeb BF, Muthuri SG, Doherty M, Zhang W. Occupational risk factors for osteoarthritis of the knee: a meta-analysis. Osteoarthritis Cartilage 2011; 19: 829-839.
101. Spector TD, Harris PA, Hart DJ, Cicuttini FM, Nandra D, Etherington J, et al. Risk of osteoarthritis associated with long-term weight-bearing sports: a radiologic survey of the hips and knees in female ex-athletes and population controls. Arthritis Rheum 1996; 39: 988-995.
102. Kujala UM, Kettunen J, Paananen H, Aalto T, Battie MC, Impivaara O, et al. Knee osteoarthritis in former runners, soccer players, weight lifters, and shooters. Arthritis Rheum 1995; 38: 539-546.
103. Lane NE, Michel B, Bjorkengren A, Oehlert J, Shi H, Bloch DA, et al. The risk of osteoarthritis with running and aging: a 5-year longitudinal study. J Rheumatol 1993; 20: 461-468.
104. Sharma L, Chmiel JS, Almagor O, Felson D, Guermazi A, Roemer F, et al. The role of varus and valgus alignment in the initial development of knee cartilage damage by MRI: the MOST study. Ann Rheum Dis 2013; 72: 235-240.
105. Felson DT, Niu J, Gross KD, Englund M, Sharma L, Cooke TD, et al. Valgus malalignment is a risk factor for lateral knee osteoarthritis incidence and progression: findings from the Multicenter Osteoarthritis Study and the Osteoarthritis Initiative. Arthritis Rheum 2013; 65: 355-362.
106. 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; 46: 2632-2636.

107. Felson DT, McLaughlin S, Goggins J, LaValley MP, Gale ME, Totterman S, et al. Bone marrow edema and its relation to progression of knee osteoarthritis. Ann Intern Med 2003; 139: 330-336.
108. Muthuri SG, McWilliams DF, Doherty M, Zhang W. History of knee injuries and knee osteoarthritis: a meta-analysis of observational studies. Osteoarthritis Cartilage 2011; 19: 1286-1293.
109. Lohmander LS, Ostenberg A, Englund M, Roos H. High prevalence of knee osteoarthritis, pain, and functional limitations in female soccer players twelve years after anterior cruciate ligament injury. Arthritis Rheum 2004; 50: 3145-3152.
110. Roos EM, Ostenberg A, Roos H, Ekdahl C, Lohmander LS. Long-term outcome of meniscectomy: symptoms, function, and performance tests in patients with or without radiographic osteoarthritis compared to matched controls. Osteoarthritis Cartilage 2001; 9: 316-324.
111. Neogi T, Bowes MA, Niu J, De Souza KM, Vincent GR, Goggins J, et al. Magnetic resonance imaging-based three-dimensional bone shape of the knee predicts onset of knee osteoarthritis: data from the osteoarthritis initiative. Arthritis Rheum 2013; 65: 2048-2058.
112. Gregory JS, Waarsing JH, Day J, Pols HA, Reijman M, Weinans H, et al. Early identification of radiographic osteoarthritis of the hip using an active shape model to quantify changes in bone morphometric features: can hip shape tell us anything about the progression of osteoarthritis? Arthritis Rheum 2007; 56: 3634-3643.
113. Haverkamp DJ, Schiphof D, Bierma-Zeinstra SM, Weinans H, Waarsing JH. Variation in joint shape of osteoarthritic knees. Arthritis Rheum 2011; 63: 3401-3407.
114. Golightly YM, Allen KD, Renner JB, Helmick CG, Salazar A, Jordan JM. Relationship of limb length inequality with radiographic knee and hip osteoarthritis. Osteoarthritis Cartilage 2007; 15: 824-829.
115. Harvey WF, Yang M, Cooke TD, Segal NA, Lane N, Lewis CE, et al. Association of leg-length inequality with knee osteoarthritis: a cohort study. Ann Intern Med 2010; 152: 287-295.
116. Raja R, Dube B, Hensor EMA, Hogg SF, Conaghan PG, Kingsbury SR. The clinical characteristics of older people with chronic multiple-site joint pains and their utilisation of therapeutic interventions: data from a prospective cohort study. BMC Musculoskeletal Disorders 2016; 17: 194.
117. Brandt KD, Heilman DK, Slemenda C, Katz BP, Mazzuca SA, Braunstein EM, et al. Quadriceps strength in women with radiographically progressive osteoarthritis of the knee and those with stable radiographic changes. J Rheumatol 1999; 26: 2431-2437.
118. Amin S, Baker K, Niu J, Clancy M, Goggins J, Guermazi A, et al. Quadriceps strength and the risk of cartilage loss and symptom progression in knee osteoarthritis. Arthritis and rheumatism 2009; 60: 189-198.
119. Slemenda C, Brandt KD, Heilman DK, Mazzuca S, Braunstein EM, Katz BP, et al. Quadriceps weakness and osteoarthritis of the knee. Ann Intern Med 1997; 127: 97-104.
120. Sharma L, Dunlop DD, Cahue S, Song J, Hayes KW. Quadriceps strength and osteoarthritis progression in malaligned and lax knees. Ann Intern Med 2003; 138: 613-619.

121. Chaisson CE, Zhang Y, Sharma L, Kannel W, Felson DT. Grip strength and the risk of developing radiographic hand osteoarthritis: results from the Framingham Study. Arthritis Rheum 1999; 42: 33-38.
122. Whittle R, Jordan KP, Thomas E, Peat G. Average symptom trajectories following incident radiographic knee osteoarthritis: data from the Osteoarthritis Initiative. RMD open 2016; 2: e000281-e000281.
123. Cross M, Dubouis L, Mangin M, Hunter DJ, March L, Hawker G, et al. Defining Flare in Osteoarthritis of the Hip and Knee: A Systematic Literature Review — OMERACT Virtual Special Interest Group. The Journal of Rheumatology 2017; 44: 1920.
124. Zhang W, Doherty M, Peat G, Bierma-Zeinstra MA, Arden NK, Bresnihan B, et al. EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. Ann Rheum Dis 2010; 69: 483-489.
125. Woolf AD, Akesson K. Primer: history and examination in the assessment of musculoskeletal problems. Nat Clin Pract Rheum 2008; 4: 26-33.
126. Keenan AM, Tennant A, Fear J, Emery P, Conaghan PG. Impact of multiple joint problems on daily living tasks in people in the community over age fifty-five. Arthritis Rheum 2006; 55: 757-764.
127. Axford J, Heron C, Ross F, Victor CR. Management of knee osteoarthritis in primary care: pain and depression are the major obstacles. J Psychosom Res 2008; 64.
128. Pereira D, Severo M, Barros H, Branco J, Santos RA, Ramos E. The effect of depressive symptoms on the association between radiographic osteoarthritis and knee pain: a cross-sectional study. BMC Musculoskeletal Disorders 2013; 14: 1-9.
129. Botha-Scheepers S, Riyazi N, Kroon HM, Scharloo M, Houwing-Duistermaat JJ, Slagboom E, et al. Activity limitations in the lower extremities in patients with osteoarthritis: the modifying effects of illness perceptions and mental health. Osteoarthritis and Cartilage 2006; 14: 1104-1110.
130. Stubbs B, Aluko Y, Myint PK, Smith TO. Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis. Age Ageing 2016; 45: 228-235.
131. Kadam UT, Jordan K, Croft PR. Clinical comorbidity in patients with osteoarthritis: a case-control study of general practice consulters in England and Wales. Ann Rheum Dis 2004; 63: 408-414.
132. Caporali R, Cimmino MA, Sarzi-Puttini P, Scarpa R, Parazzini F, Zaninelli A, et al. Comorbid conditions in the AMICA study patients: effects on the quality of life and drug prescriptions by general practitioners and specialists. Semin Arthritis Rheum 2005; 35: 31-37.
133. Tuominen U, Blom M, Hirvonen J, Seitsalo S, Lehto M, Paavolainen P, et al. The effect of co-morbidities on health-related quality of life in patients placed on the waiting list for total joint replacement. Health Qual Life Outcomes 2007; 5: 16.
134. Kadam UT, Croft PR. Clinical comorbidity in osteoarthritis: associations with physical function in older patients in family practice. J Rheumatol 2007; 34: 1899-1904.
135. Hubertsson J, Petersson IF, Thorstensson CA, Englund M. Risk of sick leave and disability pension in working-age women and men with knee osteoarthritis. Ann Rheum Dis 2013; 72: 401-405.

136. Hiligsmann M, Cooper C, Arden N, Boers M, Branco JC, Luisa Brandi M, et al. Health economics in the field of osteoarthritis: an expert's consensus paper from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). Semin Arthritis Rheum 2013; 43: 303-313.
137. Berger A, Hartrick C, Edelsberg J, Sadosky A, Oster G. Direct and indirect economic costs among private-sector employees with osteoarthritis. J Occup Environ Med 2011; 53: 1228-1235.
138. DiBonaventura Md, Gupta S, McDonald M, Sadosky A. Evaluating the health and economic impact of osteoarthritis pain in the workforce: results from the National Health and Wellness Survey. BMC Musculoskeletal Disorders 2011; 12: 83.
139. Brandt KD, Dieppe P, Radin EL. Commentary: Is It Useful to Subset “Primary” Osteoarthritis? A Critique Based on Evidence Regarding the Etiopathogenesis of Osteoarthritis. Seminars in Arthritis and Rheumatism 2009; 39: 81-95.
140. Dieppe PA, Lohmander LS. Pathogenesis and management of pain in osteoarthritis. The Lancet; 365: 965-973.
141. Burr DB, Gallant MA. Bone remodelling in osteoarthritis. Nat Rev Rheumatol 2012; 8: 665-673.
142. Lories RJ, Luyten FP. The bone-cartilage unit in osteoarthritis. Nat Rev Rheumatol 2011; 7: 43-49.
143. Eriksen EF. Treatment of bone marrow lesions (bone marrow edema). BoneKEy reports 2015; 4: 755-755.
144. Zhang Y, Nevitt M, Niu J, Lewis C, Torner J, Guermazi A, et al. Fluctuation of knee pain and changes in bone marrow lesions, effusions, and synovitis on magnetic resonance imaging. 2011; 63: 691-699.
145. Laslett LL, Dore DA, Quinn SJ, Boon P, Ryan E, Winzenberg TM, et al. Zoledronic acid reduces knee pain and bone marrow lesions over 1 year: a randomised controlled trial. Ann Rheum Dis 2012; 71: 1322-1328.
146. 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; 134: 541-549.
147. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part II. Arthritis & Rheumatism 2008; 58: 26-35.
148. Sellam J, Berenbaum F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. Nat Rev Rheumatol 2010; 6: 625-635.
149. Ayral X, Pickering EH, Woodworth TG, Mackillop N, Dougados M. Synovitis: a potential predictive factor of structural progression of medial tibiofemoral knee osteoarthritis -- results of a 1 year longitudinal arthroscopic study in 422 patients. Osteoarthritis Cartilage 2005; 13: 361-367.
150. Sokolove J, Lepus CM. Role of inflammation in the pathogenesis of osteoarthritis: latest findings and interpretations. Ther Adv Musculoskelet Dis 2013; 5: 77-94.
151. Sharma L, Eckstein F, Song J, Guermazi A, Prasad P, Kapoor D, et al. Relationship of meniscal damage, meniscal extrusion, malalignment, and joint laxity to subsequent cartilage loss in osteoarthritic knees. Arthritis Rheum 2008; 58: 1716-1726.

152. Gale DR, Chaisson CE, Totterman SMS, Schwartz RK, Gale ME, Felson D. Meniscal subluxation: association with osteoarthritis and joint space narrowing. Osteoarthritis and Cartilage 1999; 7: 526-532.
153. Edd SN, Giori NJ, Andriacchi TP. The role of inflammation in the initiation of osteoarthritis after meniscal damage. J Biomech 2015; 48: 1420-1426.
154. Harkey MS, Luc BA, Golightly YM, Thomas AC, Driban JB, Hackney AC, et al. Osteoarthritis-related biomarkers following anterior cruciate ligament injury and reconstruction: a systematic review. Osteoarthritis Cartilage 2015; 23: 1-12.
155. Lohmander LS, Englund PM, Dahl LL, Roos EM. The Long-term Consequence of Anterior Cruciate Ligament and Meniscus Injuries: Osteoarthritis. The American Journal of Sports Medicine 2007; 35: 1756-1769.
156. Hill CL, Seo GS, Gale D, Totterman S, Gale ME, Felson DT. Cruciate ligament integrity in osteoarthritis of the knee. Arthritis Rheum 2005; 52: 794-799.
157. McDougall JJ. Arthritis and Pain. Neurogenic origin of joint pain. Arthritis Research & Therapy 2006; 8: 220.
158. Hines AE, Birn H, Teglbjaerg PS, Sinkjaer T. Fiber type composition of articular branches of the tibial nerve at the knee joint in man. Anat Rec 1996; 246: 573-578.
159. Felson DT. The sources of pain in knee osteoarthritis: Editorial review. Current Opinion in Rheumatology 2005; 17: 624-628.
160. Suri S, Gill SE, De Camin SM, Wilson D, McWilliams DF, Walsh DA. Neurovascular invasion at the osteochondral junction and in osteophytes in osteoarthritis. Annals of the Rheumatic Diseases 2007; 66: 1423-1428.
161. Malfait A-M, Schnitzer TJ. Towards a mechanism-based approach to pain management in osteoarthritis. Nature Reviews Rheumatology 2013; 9: 654-664.
162. Woolf CJJP. Central sensitization: implications for the diagnosis and treatment of pain. 2011; 152: S2-S15.
163. Bajaj P, Bajaj P, Graven-Nielsen T, Arendt-Nielsen L. Osteoarthritis and its association with muscle hyperalgesia: an experimental controlled study. Pain 2001; 93: 107-114.
164. Case R, Thomas E, Clarke E, Peat G. Prodromal symptoms in knee osteoarthritis: a nested case-control study using data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2015; 23: 1083-1089.
165. Hawker GA, Stewart L, French MR, Cibere J, Jordan JM, March L, et al. Understanding the pain experience in hip and knee osteoarthritis--an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008; 16: 415-422.
166. Schuelert N, McDougall JJ. Involvement of Nav 1.8 sodium ion channels in the transduction of mechanical pain in a rodent model of osteoarthritis. Arthritis Res Ther 2012; 14: R5.
167. Strickland IT, Martindale JC, Woodhams PL, Reeve AJ, Chessell IP, McQueen DS. Changes in the expression of NaV1.7, NaV1.8 and NaV1.9 in a distinct population of dorsal root ganglia innervating the rat knee joint in a model of chronic inflammatory joint pain. Eur J Pain 2008; 12: 564-572.
168. Miller RJ, Jung H, Bhangoo SK, White FA. Cytokine and chemokine regulation of sensory neuron function. In: Sensory nerves: Springer 2009:417-449.

169. Brenn D, Richter F, Schaible HG. Sensitization of unmyelinated sensory fibers of the joint nerve to mechanical stimuli by interleukin-6 in the rat: an inflammatory mechanism of joint pain. Arthritis Rheum 2007; 56: 351-359.
170. Woolf CJ, Safieh-Garabedian B, Ma Q-P, Crilly P, Winter JJN. Nerve growth factor contributes to the generation of inflammatory sensory hypersensitivity. 1994; 62: 327-331.
171. Barthel C, Yeremenko N, Jacobs R, Schmidt RE, Bernateck M, Zeidler H, et al. Nerve growth factor and receptor expression in rheumatoid arthritis and spondyloarthritis. Arthritis research & therapy 2009; 11: R82-R82.
172. Halliday DA, Zettler C, Rush RA, Scicchitano R, McNeil JD. Elevated nerve growth factor levels in the synovial fluid of patients with inflammatory joint disease. Neurochem Res 1998; 23: 919-922.
173. Shelton DL, Zeller J, Ho WH, Pons J, Rosenthal A. Nerve growth factor mediates hyperalgesia and cachexia in auto-immune arthritis. Pain 2005; 116: 8-16.
174. Bedson J, Jordan K, Croft P. How do GPs use x rays to manage chronic knee pain in the elderly? A case study. Ann Rheum Dis 2003; 62: 450-454.
175. Hart DJ, Doyle DV, Spector TD. Incidence and risk factors for radiographic knee osteoarthritis in middle-aged women: The Chingford Study. Arthritis & Rheumatism 1999; 42: 17-24.
176. Kerkhof HJM, Bierma-Zeinstra SMA, Arden NK, Metrustry S, Castano-Betancourt M, Hart DJ, et al. Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors. Annals of the Rheumatic Diseases 2014; 73: 2116-2121.
177. Javaid MK, Lynch JA, Tolstykh I, Guermazi A, Roemer F, Aliabadi P, et al. Pre-radiographic MRI findings are associated with onset of knee symptoms: the most study. Osteoarthritis Cartilage 2010; 18: 323-328.
178. Hunter DJ, Arden N, Conaghan PG, Eckstein F, Gold G, Grainger A, et al. Definition of osteoarthritis on MRI: results of a Delphi exercise. Osteoarthritis and Cartilage 2011; 19: 963-969.
179. Lethbridge-Cejku M, Scott WW, Jr., Reichle R, Ettinger WH, Zonderman A, Costa P, et al. Association of radiographic features of osteoarthritis of the knee with knee pain: data from the Baltimore Longitudinal Study of Aging. Arthritis Care Res 1995; 8: 182-188.
180. Cicuttini FM, Baker J, Hart DJ, Spector TD. Association of pain with radiological changes in different compartments and views of the knee joint. Osteoarthritis Cartilage 1996; 4: 143-147.
181. Claessens AA, Schouten JS, van den Ouweland FA, Valkenburg HA. Do clinical findings associate with radiographic osteoarthritis of the knee? Ann Rheum Dis 1990; 49: 771-774.
182. Lanyon P, O'Reilly S, Jones A, Doherty M. Radiographic assessment of symptomatic knee osteoarthritis in the community: definitions and normal joint space. Ann Rheum Dis 1998; 57: 595-601.
183. Odding E, Valkenburg HA, Algra D, Vandenouweland FA, Grobbee DE, Hofman A. Associations of radiological osteoarthritis of the hip and knee with locomotor disability in the Rotterdam Study. Ann Rheum Dis 1998; 57: 203-208.

184. Davis MA, Ettinger WH, Neuhaus JM, Barclay JD, Segal MR. CORRELATES OF KNEE PAIN AMONG UNITED-STATES ADULTS WITH AND WITHOUT RADIOGRAPHIC KNEE OSTEOARTHRITIS. Journal of Rheumatology 1992; 19: 1943-1949.
185. Hochberg MC, Lawrence RC, Everett DF, Cornonihuntley J. EPIDEMIOLOGIC ASSOCIATIONS OF PAIN IN OSTEO-ARTHRITIS OF THE KNEE - DATA FROM THE NATIONAL-HEALTH AND NUTRITION EXAMINATION SURVEY AND THE NATIONAL-HEALTH AND NUTRITION EXAMINATION-I EPIDEMIOLOGIC FOLLOW-UP SURVEY. Seminars in Arthritis and Rheumatism 1989; 18: 4-9.
186. Neogi T, Felson D, Niu J, Nevitt M, Lewis CE, Aliabadi P, et al. Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. BMJ 2009; 339.
187. Wenham CY, Conaghan PG. Imaging the painful osteoarthritic knee joint: what have we learned? Nat Clin Pract Rheumatol 2009; 5: 149-158.
188. Lachance L, Sowers M, Jamadar D, Jannausch M, Hochberg M, Crutchfield M. The experience of pain and emergent osteoarthritis of the knee. Osteoarthritis Cartilage 2001; 9: 527-532.
189. Hunter DJ, Guermazi A, Roemer F, Zhang Y, Neogi T. Structural correlates of pain in joints with osteoarthritis. Osteoarthritis Cartilage 2013; 21: 1170-1178.
190. Yusuf E, Kortekaas MC, Watt I, Huizinga TWJ, Kloppenburg M. Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review. Annals of the Rheumatic Diseases 2011; 70: 60.
191. Hunter DJ, Zhang W, Conaghan PG, Hirko K, Menashe L, Li L, et al. Systematic review of the concurrent and predictive validity of MRI biomarkers in OA. Osteoarthritis and Cartilage 2011; 19: 557-588.
192. Zanetti M, Bruder E, Romero J, Hodler J. Bone Marrow Edema Pattern in Osteoarthritic Knees: Correlation between MR Imaging and Histologic Findings. Radiology 2000; 215: 835-840.
193. 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; 67: 206-211.
194. Felson DT, Niu J, Guermazi A, Roemer F, Aliabadi P, Clancy M, et al. Correlation of the development of knee pain with enlarging bone marrow lesions on magnetic resonance imaging. 2007; 56: 2986-2992.
195. Link TM, Steinbach LS, Ghosh S, Ries M, Lu Y, Lane N, et al. Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. Radiology 2003; 226: 373-381.
196. Kornaat PR, Bloem JL, Ceulemans RYT, Riyazi N, Rosendaal FR, Nelissen RG, et al. Osteoarthritis of the Knee: Association between Clinical Features and MR Imaging Findings. Radiology 2006; 239: 811-817.
197. Barker K, Lamb SE, Toye F, Jackson S, Barrington S. Association between radiographic joint space narrowing, function, pain and muscle power in severe osteoarthritis of the knee. Clin Rehabil 2004; 18.
198. Foong YC, Khan HI, Blizzard L, Ding C, Cicuttini F, Jones G, et al. The clinical significance, natural history and predictors of bone marrow lesion change over eight years. Arthritis Research & Therapy 2014; 16: R149.

199. Driban JB, Price L, Lo GH, Pang J, Hunter DJ, Miller E, et al. Evaluation of bone marrow lesion volume as a knee osteoarthritis biomarker--longitudinal relationships with pain and structural changes: data from the Osteoarthritis Initiative. Arthritis Res Ther 2013; 15: R112.
200. Dore D, Quinn S, Ding C, Winzenberg T, Zhai G, Cicuttini F, et al. Natural history and clinical significance of MRI-detected bone marrow lesions at the knee: a prospective study in community dwelling older adults. Arthritis Research & Therapy 2010; 12: R223.
201. Kornaat PR, Kloppenburg M, Sharma R, Botha-Scheepers SA, Le Graverand MP, Coene LN, et al. Bone marrow edema-like lesions change in volume in the majority of patients with osteoarthritis; associations with clinical features. Eur Radiol 2007; 17: 3073-3078.
202. Moisio K, Eckstein F, Chmiel JS, Guermazi A, Prasad P, Almagor O, et al. Denuded subchondral bone and knee pain in persons with knee osteoarthritis. Arthritis Rheum 2009; 60: 3703-3710.
203. Zhai G, Blizzard L, Srikanth V, Ding C, Cooley H, Cicuttini F, et al. Correlates of knee pain in older adults: Tasmanian Older Adult Cohort Study. Arthritis Rheum 2006; 55: 264-271.
204. Sharma L, Chmiel JS, Almagor O, Dunlop D, Guermazi A, Bathon JM, et al. Significance of preradiographic magnetic resonance imaging lesions in persons at increased risk of knee osteoarthritis. Arthritis Rheumatol 2014; 66: 1811-1819.
205. Lo GH, McAlindon TE, Niu J, Zhang Y, Beals C, Dabrowski C, et al. Bone marrow lesions and joint effusion are strongly and independently associated with weight-bearing pain in knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2009; 17: 1562-1569.
206. Stefanik JJ, Gross KD, Guermazi A, Felson DT, Roemer FW, Zhang Y, et al. The relation of MRI-detected structural damage in the medial and lateral patellofemoral joint to knee pain: the Multicenter and Framingham Osteoarthritis Studies. Osteoarthritis and cartilage 2015; 23: 565-570.
207. Everhart JS, Siston RA, Flanigan DC. Tibiofemoral subchondral surface ratio (SSR) is a predictor of osteoarthritis symptoms and radiographic progression: data from the Osteoarthritis Initiative (OAI). Osteoarthritis Cartilage 2014; 22: 771-778.
208. Ochiai N, Sasho T, Tahara M, Watanabe A, Matsuki K, Yamaguchi S, et al. Objective assessments of medial osteoarthritic knee severity by MRI: new computer software to evaluate femoral condyle contours. Int Orthop 2010; 34: 811-817.
209. Hunter D, Nevitt M, Lynch J, Kraus VB, Katz JN, Collins JE, et al. Longitudinal validation of periarticular bone area and 3D shape as biomarkers for knee OA progression? Data from the FNIH OA Biomarkers Consortium. Annals of the Rheumatic Diseases 2016; 75: 1607.
210. Barr AJ, Dube B, Hensor EM, Kingsbury SR, Peat G, Bowes MA, et al. The relationship between knee symptoms and 3D MRI knee bone shape: Data from the Osteoarthritis Initiative. Osteoarthritis and Cartilage 2016; 24: S278-S279.
211. Hayashi D, Roemer FW, Katur A, Felson DT, Yang SO, Alomran F, et al. Imaging of synovitis in osteoarthritis: current status and outlook. Semin Arthritis Rheum 2011; 41: 116-130.

212. 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; 19: 990-1002.
213. Felson DT, Niu J, Neogi T, Goggins J, Nevitt MC, Roemer F, et al. Synovitis and the risk of knee osteoarthritis: the MOST Study. Osteoarthritis Cartilage 2016; 24: 458-464.
214. Roemer FW, Kwoh CK, Hannon MJ, Hunter DJ, Eckstein F, Fujii T, et al. What comes first? Multitissue involvement leading to radiographic osteoarthritis: magnetic resonance imaging-based trajectory analysis over four years in the osteoarthritis initiative. Arthritis Rheumatol 2015; 67: 2085-2096.
215. Collins JE, Losina E, Nevitt MC, Roemer FW, Guermazi A, Lynch JA, et al. Semiquantitative Imaging Biomarkers of Knee Osteoarthritis Progression: Data From the Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheumatol 2016; 68: 2422-2431.
216. Neogi T, Guermazi A, Roemer F, Nevitt MC, Scholz J, Arendt-Nielsen L, et al. Association of Joint Inflammation With Pain Sensitization in Knee Osteoarthritis: The Multicenter Osteoarthritis Study. Arthritis Rheumatol 2016; 68: 654-661.
217. Hill CL, Gale DG, Chaisson CE, Skinner K, Kazis L, Gale ME, et al. Knee effusions, popliteal cysts, and synovial thickening: association with knee pain in osteoarthritis. The Journal of Rheumatology 2001; 28: 1330.
218. Hill CL, Hunter DJ, Niu J, Clancy M, Guermazi A, Genant H, et al. Synovitis detected on magnetic resonance imaging and its relation to pain and cartilage loss in knee osteoarthritis. Annals of the Rheumatic Diseases 2007; 66: 1599.
219. Hunter DJ, McDougall JJ, Keefe FJ. The Symptoms of Osteoarthritis and the Genesis of Pain. Rheumatic Disease Clinics of North America 2008; 34: 623-643.
220. Callaghan MJ, Parkes MJ, Hutchinson CE, Gait AD, Forsythe LM, Marjanovic EJ, et al. A randomised trial of a brace for patellofemoral osteoarthritis targeting knee pain and bone marrow lesions. Ann Rheum Dis 2015; 74: 1164-1170.
221. Arroll B, Goodyear-Smith F. Corticosteroid injections for osteoarthritis of the knee: meta-analysis. Bmj 2004; 328: 869.
222. Wenham CYJ, Conaghan PG. The Role of Synovitis in Osteoarthritis. Therapeutic Advances in Musculoskeletal Disease 2010; 2: 349-359.
223. Sowers MF, Hayes C, Jamadar D, Capul D, Lachance L, Jannausch M, et al. Magnetic resonance-detected subchondral bone marrow and cartilage defect characteristics associated with pain and X-ray-defined knee osteoarthritis. Osteoarthritis and Cartilage 2003; 11: 387-393.
224. Wluka AE, Wolfe R, Stuckey S, Cicuttini FM. How does tibial cartilage volume relate to symptoms in subjects with knee osteoarthritis? Annals of the Rheumatic Diseases 2004; 63: 264-268.
225. Hunter DJ, March L, Sambrook PN. The association of cartilage volume with knee pain. Osteoarthritis and Cartilage 2003; 11: 725-729.
226. Moisio K, Eckstein F, Song J, Cahue S, Marshall M, Dunlop D. The relationship of denuded subchondral bone area to knee pain severity and incident frequent knee pain. Arthritis Rheum 2008; 58: S237-S238.

227. Ashraf S, Wibberley H, Mapp PI, Hill R, Wilson D, Walsh DA. Increased vascular penetration and nerve growth in the meniscus: A potential source of pain in osteoarthritis. Annals of the Rheumatic Diseases 2011; 70: 523-529.
228. Englund M, Guermazi A, Gale D, Hunter DJ, Aliabadi P, Clancy M, et al. Incidental Meniscal Findings on Knee MRI in Middle-Aged and Elderly Persons. New England Journal of Medicine 2008; 359: 1108-1115.
229. Bhattacharyya T, Gale D, Dewire P, Totterman S, Gale ME, McLaughlin S, et al. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. J Bone Joint Surg Am 2003; 85: 4-9.
230. Englund M, Niu J, Guermazi A, Roemer FW, Hunter DJ, Lynch JA, et al. Effect of meniscal damage on the development of frequent knee pain, aching, or stiffness. Arthritis and Rheumatism 2007; 56: 4048-4054.
231. Hill CL, Gale DR, Chaisson CE, Skinner K, Kazis L, Gale ME, et al. Periarticular lesions detected on magnetic resonance imaging: prevalence in knees with and without symptoms. Arthritis Rheum 2003; 48: 2836-2844.
232. Zhang Y, Niu J, Felson DT, Choi HK, Nevitt M, Neogi T. Methodologic challenges in studying risk factors for progression of knee osteoarthritis. 2010; 62: 1527-1532.
233. Westreich D, Greenland S. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. American Journal of Epidemiology 2013; 177: 292-298.
234. Ranstam J, Cook JA. Causal relationship and confounding in statistical models. 2016; 103: 1445-1446.
235. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology 2003; 14: 300-306.
236. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC Medical Research Methodology 2008; 8: 70.
237. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods 2013; 18: 137-150.
238. Collins JE, Katz JN, Dervan EE, Losina E. Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2014; 22: 622-630.
239. Wesseling J, Bierma-Zeinstra SM, Kloppenburg M, Meijer R, Bijlsma JW. Worsening of pain and function over 5 years in individuals with 'early' OA is related to structural damage: data from the Osteoarthritis Initiative and CHECK (Cohort Hip & Cohort Knee) study. Ann Rheum Dis 2015; 74: 347-353.
240. Dimitroulas T, Duarte RV, Behura A, Kitas GD, Raphael JH. Neuropathic pain in osteoarthritis: a review of pathophysiological mechanisms and implications for treatment. Semin Arthritis Rheum 2014; 44: 145-154.
241. Abadie E, Ethgen D, Avouac B, Bouvenot G, Branco J, Bruyere O, et al. Recommendations for the use of new methods to assess the efficacy of disease-modifying drugs in the treatment of osteoarthritis. Osteoarthritis and Cartilage 2004; 12: 263-268.

242. Pham T, Van Der Heijde D, Lassere M, Altman RD, Anderson JJ, Bellamy N, et al. Outcome variables for osteoarthritis clinical trials: The OMERACT-OARSI set of responder criteria. J Rheumatol 2003; 30: 1648-1654.
243. Vina ER, Richardson D, Medvedeva E, Kent Kwoh C, Collier A, Ibrahim SA. Does a Patient-centered Educational Intervention Affect African-American Access to Knee Replacement? A Randomized Trial. Clin Orthop Relat Res 2016.
244. Kurtz SM, Ong KL, Lau E, Widmer M, Maravic M, Gómez-Barrena E, et al. International survey of primary and revision total knee replacement. International Orthopaedics 2011; 35: 1783-1789.
245. Lingard EA, Sledge CB, Learmonth ID. Patient expectations regarding total knee arthroplasty: differences among the United States, United Kingdom, and Australia. J Bone Joint Surg Am 2006; 88: 1201-1207.
246. Kloppenburg M, Boyesen P, Visser AW, Haugen IK, Boers M, Boonen A, et al. Report from the OMERACT Hand Osteoarthritis Working Group: Set of Core Domains and Preliminary Set of Instruments for Use in Clinical Trials and Observational Studies. J Rheumatol 2015; 42: 2190-2197.
247. Rolfson O, Wissig S, van Maasakkers L, Stowell C, Ackerman I, Ayers D, et al. Defining an International Standard Set of Outcome Measures for Patients With Hip or Knee Osteoarthritis: Consensus of the International Consortium for Health Outcomes Measurement Hip and Knee Osteoarthritis Working Group. Arthritis Care Res (Hoboken) 2016; 68: 1631-1639.
248. Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) (Western Ontario and McMaster Universities Osteoarthritis Index) instrument in the measurement of outcomes after knee replacement surgery. J Rheumatol 1995; 22: 1193-1196.
249. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010; 63: 1179-1194.
250. Khanna D, Krishnan E, Dewitt EM, Khanna PP, Spiegel B, Hays RD. The future of measuring patient-reported outcomes in rheumatology: Patient-Reported Outcomes Measurement Information System (PROMIS). Arthritis Care Res (Hoboken) 2011; 63 Suppl 11: S486-490.
251. Kingsbury SR, Corp N, Watt FE, Felson DT, O'Neill TW, Holt CA, et al. Harmonising data collection from osteoarthritis studies to enable stratification: recommendations on core data collection from an Arthritis Research UK clinical studies group. Rheumatology (Oxford) 2016; 55: 1394-1402.
252. Neogi T, Frey-Law L, Scholz J, Niu J, Arendt-Nielsen L, Woolf C, et al. Sensitivity and sensitisation in relation to pain severity in knee osteoarthritis: trait or state? Annals of the Rheumatic Diseases 2013.
253. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug

therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988; 15: 1833-1840.
254. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998; 28: 88-96.
255. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care & Research 2011; 63: S240-S252.
256. Freynhagen R, Baron R, Gockel U, Tolle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin 2006; 22: 1911-1920.
257. Bennett MI, Smith BH, Torrance N, Potter J. The S-LANSS score for identifying pain of predominantly neuropathic origin: validation for use in clinical and postal research. J Pain 2005; 6: 149-158.
258. Klassbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. Scand J Rheumatol 2003; 32: 46-51.
259. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. Arthritis Care Res 1999; 12: 331-335.
260. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013; 21: 1042-1052.
261. Bennell K, Dobson F, Hinman R. Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. Arthritis Care Res (Hoboken) 2011; 63 Suppl 11: S350-370.
262. Davis AM. Osteoarthritis year in review: rehabilitation and outcomes. Osteoarthritis Cartilage 2012; 20: 201-206.
263. Visser AW, Boyesen P, Haugen IK, Schoones JW, van der Heijde DM, Rosendaal FR, et al. Instruments Measuring Pain, Physical Function, or Patient's Global Assessment in Hand Osteoarthritis: A Systematic Literature Search. J Rheumatol 2015; 42: 2118-2134.
264. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989; 28: 193-213.
265. Douglass AB, Bornstein R, Nino-Murcia G, Keenan S, Miles L, Zarcone VP, Jr., et al. The Sleep Disorders Questionnaire. I: Creation and multivariate structure of SDQ. Sleep 1994; 17: 160-167.
266. Kosinski M, Janagap CC, Gajria K, Schein J. Psychometric testing and validation of the Chronic Pain Sleep Inventory. Clin Ther 2007; 29 Suppl: 2562-2577.

267. Hawker GA, Gignac MA, Badley E, Davis AM, French MR, Li Y, et al. A longitudinal study to explain the pain-depression link in older adults with osteoarthritis. Arthritis Care Res (Hoboken) 2011; 63.
268. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002; 52: 69-77.
269. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement 1977; 1: 385-401.
270. Rathbun AM, Yau MS, Shardell M, Stuart EA, Hochberg MC. Depressive symptoms and structural disease progression in knee osteoarthritis: data from the Osteoarthritis Initiative. Clin Rheumatol 2017; 36: 155-163.
271. Power JD, Badley EM, French MR, Wall AJ, Hawker GA. Fatigue in osteoarthritis: a qualitative study. BMC Musculoskelet Disord 2008; 9: 63.
272. Hawker GA, French MR, Waugh EJ, Gignac MA, Cheung C, Murray BJ. The multidimensionality of sleep quality and its relationship to fatigue in older adults with painful osteoarthritis. Osteoarthritis Cartilage 2010; 18: 1365-1371.
273. Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. J Rheumatol 2005; 32: 811-819.
274. Stein KD, Martin SC, Hann DM, Jacobsen PB. A multidimensional measure of fatigue for use with cancer patients. Cancer Pract 1998; 6: 143-152.
275. 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; 237: 998-1007.
276. Hernandez-Molina G, Neogi T, Hunter DJ, Niu J, Guermazi A, Reichenbach S, et al. The association of bone attrition with knee pain and other MRI features of osteoarthritis. Ann Rheum Dis 2008; 67: 43-47.
277. Torres L, Dunlop DD, Peterfy C, Guermazi A, Prasad P, Hayes KW, et al. The relationship between specific tissue lesions and pain severity in persons with knee osteoarthritis. Osteoarthritis Cartilage 2006; 14: 1033-1040.
278. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. Lancet 2011; 377: 2115-2126.
279. Eckstein F, Boudreau R, Wang Z, Hannon MJ, Duryea J, Wirth W, et al. Comparison of radiographic joint space width and magnetic resonance imaging for prediction of knee replacement: A longitudinal case-control study from the Osteoarthritis Initiative. Eur Radiol 2016; 26: 1942-1951.
280. Peterfy CG, Gold G, Eckstein F, Cicuttini F, Dardzinski B, Stevens R. MRI protocols for whole-organ assessment of the knee in osteoarthritis. Osteoarthritis Cartilage 2006; 14 Suppl A: A95-111.
281. Englund M, Guermazi A, Roemer FW, Aliabadi P, Yang M, Lewis CE, et al. Meniscal tear in knees without surgery and the development of radiographic osteoarthritis among middle-aged and elderly persons: The Multicenter Osteoarthritis Study. Arthritis Rheum 2009; 60: 831-839.

282. Peterfy CG, 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; 12: 177-190.
283. Kornaat PR, Ceulemans RY, 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; 34: 95-102.
284. 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; 67.
285. Crema MD, Hunter DJ, Burstein D, Roemer FW, Li L, Eckstein F, et al. Association of changes in delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) with changes in cartilage thickness in the medial tibiofemoral compartment of the knee: a 2 year follow-up study using 3.0 T MRI. Ann Rheum Dis 2014; 73: 1935-1941.
286. Brandt KD, Mazzuca SA, Conrozier T, Dacre JE, Peterfy CG, Provvedini D, et al. Which is the best radiographic protocol for a clinical trial of a structure modifying drug in patients with knee osteoarthritis? J Rheumatol 2002; 29: 1308-1320.
287. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997; 24: 799-802.
288. Lane NE, Nevitt MC, Genant HK, Hochberg MC. Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. J Rheumatol 1993; 20: 1911-1918.
289. Lane NE, Kremer LB. Radiographic indices for osteoarthritis. Rheum Dis Clin North Am 1995; 21: 379-394.
290. Salaffi F, Carotti M, Stancati A, Grassi W. Radiographic assessment of osteoarthritis: analysis of disease progression. Aging Clin Exp Res 2003; 15: 391-404.
291. Altman RD, Hochberg M, Murphy WA, Jr., Wolfe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. Osteoarthritis Cartilage 1995; 3 Suppl A: 3-70.
292. Felson DT, McAlindon TE, Anderson JJ, Naimark A, Weissman BW, Aliabadi P, et al. Defining radiographic osteoarthritis for the whole knee. Osteoarthritis Cartilage 1997; 5: 241-250.
293. McAlindon TE, Snow S, Cooper C, Dieppe PA. Radiographic patterns of osteoarthritis of the knee joint in the community: the importance of the patellofemoral joint. Ann Rheum Dis 1992; 51: 844-849.
294. Kurz B, Lemke AK, Fay J, Pufe T, Grodzinsky AJ, Schunke M. Pathomechanisms of cartilage destruction by mechanical injury. Ann Anat 2005; 187: 473-485.
295. Hashimoto S, Creighton-Achermann L, Takahashi K, Amiel D, Coutts RD, Lotz M. Development and regulation of osteophyte formation during experimental osteoarthritis. Osteoarthritis Cartilage 2002; 10: 180-187.
296. Teichtahl AJ, Wluka AE, Wijethilake P, Wang Y, Ghasem-Zadeh A, Cicuttini FMJAR, et al. Wolff's law in action: a mechanism for early knee osteoarthritis. 2015; 17: 207.

297. Hunter DJ, Zhang YQ, Niu JB, Tu X, Amin S, Clancy M, et al. The association of meniscal pathologic changes with cartilage loss in symptomatic knee osteoarthritis. Arthritis Rheum 2006; 54: 795-801.
298. Kraus VB, Feng S, Wang S, White S, Ainslie M, Brett A, et al. Trabecular morphometry by fractal signature analysis is a novel marker of osteoarthritis progression. Arthritis Rheum 2009; 60: 3711-3722.
299. Kraus VB, Feng S, Wang S, White S, Ainslie M, Graverand MHL, et al. Subchondral bone trabecular integrity predicts and changes concurrently with radiographic and magnetic resonance imaging-determined knee osteoarthritis progression. Arthritis Rheum 2013; 65: 1812-1821.
300. Podsiadlo P, Nevitt MC, Wolski M, Stachowiak GW, Lynch JA, Tolstykh I, et al. Baseline trabecular bone and its relation to incident radiographic knee osteoarthritis and increase in joint space narrowing score: directional fractal signature analysis in the MOST study. Osteoarthritis Cartilage 2016; 24: 1736-1744.
301. Podsiadlo P, Cicuttini FM, Wolski M, Stachowiak GW, Wluka AE. Trabecular bone texture detected by plain radiography is associated with an increased risk of knee replacement in patients with osteoarthritis: a 6 year prospective follow up study. Osteoarthritis Cartilage 2014; 22: 71-75.
302. Nagaosa Y, Mateus M, Hassan B, Lanyon P, Doherty M. Development of a logically devised line drawing atlas for grading of knee osteoarthritis. Ann Rheum Dis 2000; 59: 587-595.
303. Crema MD, Nevitt MC, Guermazi A, Felson DT, Wang K, Lynch JA, et al. Progression of cartilage damage and meniscal pathology over 30 months is associated with an increase in radiographic tibiofemoral joint space narrowing in persons with knee OA--the MOST study. Osteoarthritis Cartilage 2014; 22: 1743-1747.
304. Eckstein F, Wirth W, Hunter DJ, Guermazi A, Kwoh CK, Nelson DR, et al. Magnitude and regional distribution of cartilage loss associated with grades of joint space narrowing in radiographic osteoarthritis--data from the Osteoarthritis Initiative (OAI). Osteoarthritis Cartilage 2010; 18: 760-768.
305. Altman RD, Fries JF, Bloch DA, Carstens J, Cooke TD, Genant H, et al. Radiographic assessment of progression in osteoarthritis. Arthritis Rheum 1987; 30: 1214-1225.
306. Spector TD, Hart DJ, Byrne J, Harris PA, Dacre JE, Doyle DV. Definition of osteoarthritis of the knee for epidemiological studies. Ann Rheum Dis 1993; 52: 790-794.
307. Ravaud P, Giraudeau B, Auleley GR, Drape JL, Rousselin B, Paolozzi L, et al. Variability in knee radiographing: implication for definition of radiological progression in medial knee osteoarthritis. Ann Rheum Dis 1998; 57: 624-629.
308. Ravaud P, Giraudeau B, Auleley GR, Edouard-Noel R, Dougados M, Chastang C. Assessing smallest detectable change over time in continuous structural outcome measures: application to radiological change in knee osteoarthritis. J Clin Epidemiol 1999; 52: 1225-1230.
309. Ravaud P, Giraudeau B, Auleley GR, Chastang C, Poiraudeau S, Ayral X, et al. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. J Rheumatol 1996; 23: 1756-1764.

310. Gossec L, Jordan JM, Mazzuca SA, Lam MA, Suarez-Almazor ME, Renner JB, et al. Comparative evaluation of three semi-quantitative radiographic grading techniques for knee osteoarthritis in terms of validity and reproducibility in 1759 X-rays: report of the OARSI-OMERACT task force. Osteoarthritis Cartilage 2008; 16: 742-748.
311. Hart DJ, Spector TD. Definition and epidemiology of osteoarthritis of the hand: a review. Osteoarthritis Cartilage 2000; 8 Suppl A: S2-7.
312. Felson DT, Nevitt MC, Yang M, Clancy M, Niu J, Torner JC, et al. A new approach yields high rates of radiographic progression in knee osteoarthritis. J Rheumatol 2008; 35: 2047-2054.
313. Schiphof D, Boers M, Bierma-Zeinstra SM. Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. Ann Rheum Dis 2008; 67: 1034-1036.
314. Reichmann WM, Maillefert JF, Hunter DJ, Katz JN, Conaghan PG, Losina E. Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: A systematic review. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 2011; 19: 550-556.
315. Hunter DJ, Zhang W, Conaghan PG, Hirko K, Menashe L, Reichmann WM, et al. Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence. Osteoarthritis Cartilage 2011; 19: 589-605.
316. Duryea J, Neumann G, Niu J, Totterman S, Tamez J, Dabrowski C, et al. Comparison of radiographic joint space width with magnetic resonance imaging cartilage morphometry: analysis of longitudinal data from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2010; 62: 932-937.
317. Bruyere O, Richy F, Reginster JY. Three year joint space narrowing predicts long term incidence of knee surgery in patients with osteoarthritis: an eight year prospective follow up study. Ann Rheum Dis 2005; 64: 1727-1730.
318. Wirth W, Duryea J, Hellio Le Graverand MP, John MR, Nevitt M, Buck RJ, et al. Direct comparison of fixed flexion, radiography and MRI in knee osteoarthritis: responsiveness data from the Osteoarthritis Initiative. Osteoarthritis and Cartilage 2013; 21: 117-125.
319. Conaghan PG, Felson D, Gold G, Lohmander S, Totterman S, Altman R. MRI and non-cartilaginous structures in knee osteoarthritis. Osteoarthritis Cartilage 2006; 14 Suppl A: A87-94.
320. Sharma L, Nevitt M, Hochberg M, Guermazi A, Roemer FW, Crema M, et al. Clinical significance of worsening versus stable preradiographic MRI lesions in a cohort study of persons at higher risk for knee osteoarthritis. Ann Rheum Dis 2016; 75: 1630-1636.
321. Hayashi D, Guermazi A, Roemer FW. MRI of osteoarthritis: the challenges of definition and quantification. Semin Musculoskelet Radiol 2012; 16: 419-430.
322. Burstein D, Gray M, Mosher T, Dardzinski B. Measures of molecular composition and structure in osteoarthritis. Radiol Clin North Am 2009; 47: 675-686.
323. Allen RG, Burstein D, Gray ML. Monitoring glycosaminoglycan replenishment in cartilage explants with gadolinium-enhanced magnetic resonance imaging. J Orthop Res 1999; 17: 430-436.

324. Loeuille D, Sauliere N, Champigneulle J, Rat AC, Blum A, Chary-Valckenaere I. Comparing non-enhanced and enhanced sequences in the assessment of effusion and synovitis in knee OA: associations with clinical, macroscopic and microscopic features. Osteoarthritis Cartilage 2011; 19: 1433-1439.
325. Pelletier JP, Raynauld JP, Abram F, Haraoui B, Choquette D, Martel-Pelletier J. A new non-invasive method to assess synovitis severity in relation to symptoms and cartilage volume loss in knee osteoarthritis patients using MRI. Osteoarthritis Cartilage 2008; 16 Suppl 3: S8-13.
326. Englund M, Guermazi A, Roemer FW, Yang M, Zhang Y, Nevitt MC, et al. Meniscal pathology on MRI increases the risk for both incident and enlarging subchondral bone marrow lesions of the knee: the MOST Study. Ann Rheum Dis 2010; 69: 1796-1802.
327. Lo GH, Hunter DJ, Nevitt M, Lynch J, McAlindon TE, Group OAII. Strong association of MRI meniscal derangement and bone marrow lesions in knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2009; 17: 743-747.
328. Roemer FW, Neogi T, Nevitt MC, Felson DT, Zhu Y, Zhang Y, et al. Subchondral bone marrow lesions are highly associated with, and predict subchondral bone attrition longitudinally: the MOST study. Osteoarthritis Cartilage 2010; 18: 47-53.
329. Neogi T, Felson D, Niu J, Lynch J, Nevitt M, Guermazi A, et al. Cartilage loss occurs in the same subregions as subchondral bone attrition: a within-knee subregion-matched approach from the Multicenter Osteoarthritis Study. Arthritis Rheum 2009; 61: 1539-1544.
330. 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; 69: 841-844.
331. Hunter DJ, Zhang Y, Niu J, Tu X, Amin S, Goggins J, et al. Structural factors associated with malalignment in knee osteoarthritis: the Boston osteoarthritis knee study. J Rheumatol 2005; 32: 2192-2199.
332. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014; 22: 363-388.
333. Knoop J, Steultjens MP, Roorda LD, Lems WF, van der Esch M, Thorstensson CA, et al. Improvement in upper leg muscle strength underlies beneficial effects of exercise therapy in knee osteoarthritis: secondary analysis from a randomised controlled trial. Physiotherapy 2015; 101: 171-177.
334. Raynauld JP, Pelletier JP, Roubille C, Dorais M, Abram F, Li W, et al. Magnetic Resonance Imaging-Assessed Vastus Medialis Muscle Fat Content and Risk for Knee Osteoarthritis Progression: Relevance From a Clinical Trial. Arthritis Care Res (Hoboken) 2015; 67: 1406-1415.
335. Nakamura M, Sumen Y, Sakaridani K, Exham H, Ochi M. Relationship between the shape of tibial spurs on X-ray and meniscal changes on MRI in early osteoarthritis of the knee. Magn Reson Imaging 2006; 24: 1143-1148.
336. Panzer S, Augat P, Atzwanger J, Hergan K. 3-T MRI assessment of osteophyte formation in patients with unilateral anterior cruciate ligament injury and reconstruction. Skeletal Radiol 2012; 41: 1597-1604.

337. de Lange-Brokaar BJ, Bijsterbosch J, Kornaat PR, Yusuf E, Ioan-Facsinay A, Zuurmond AM, et al. Radiographic progression of knee osteoarthritis is associated with MRI abnormalities in both the patellofemoral and tibiofemoral joint. Osteoarthritis Cartilage 2016; 24: 473-479.
338. Bowes MA, De Souza K, Vincent GR, Conaghan PG. OA may not be as structurally heterogeneous as expected: Shape analysis of all knees from the Osteoarthritis Initiative reveals a consistent pattern of bone shape change over 8 years. Osteoarthritis and Cartilage 2016; 24: S254-S255.
339. Bowes MA, Vincent GR, Wolstenholme CB, Conaghan PG. A novel method for bone area measurement provides new insights into osteoarthritis and its progression. Ann Rheum Dis 2015; 74: 519-525.
340. Barr AJ, Dube B, Hensor EM, Kingsbury SR, Peat G, Bowes MA, et al. The relationship between three-dimensional knee MRI bone shape and total knee replacement-a case control study: data from the Osteoarthritis Initiative. Rheumatology (Oxford) 2016.
341. Ding C, Garnero P, Cicuttini F, Scott F, Cooley H, Jones G. Knee cartilage defects: association with early radiographic osteoarthritis, decreased cartilage volume, increased joint surface area and type II collagen breakdown. Osteoarthritis Cartilage 2005; 13: 198-205.
342. Crema MD, Guermazi A, Sayre EC, Roemer FW, Wong H, Thorne A, et al. The association of magnetic resonance imaging (MRI)-detected structural pathology of the knee with crepitus in a population-based cohort with knee pain: the MoDEKO study. Osteoarthritis Cartilage 2011; 19: 1429-1432.
343. Berthiaume MJ, Raynauld JP, Martel-Pelletier J, Labonte F, Beaudoin G, Bloch DA, et al. Meniscal tear and extrusion are strongly associated with progression of symptomatic knee osteoarthritis as assessed by quantitative magnetic resonance imaging. Ann Rheum Dis 2005; 64: 556-563.
344. Guermazi A, Roemer FW, Hayashi D, Crema MD, Niu J, Zhang Y, et al. Assessment of synovitis with contrast-enhanced MRI using a whole-joint semiquantitative scoring system in people with, or at high risk of, knee osteoarthritis: the MOST study. Ann Rheum Dis 2011; 70: 805-811.
345. Eckstein F, Burstein D, Link TM. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. NMR Biomed 2006; 19: 822-854.
346. Eckstein F, Wirth W. Quantitative cartilage imaging in knee osteoarthritis. Arthritis 2011; 2011: 475684.
347. Buck RJ, Wyman BT, Le Graverand MP, Wirth W, Eckstein F. An efficient subset of morphological measures for articular cartilage in the healthy and diseased human knee. Magn Reson Med 2010; 63: 680-690.
348. Hunter DJ, Altman RD, Cicuttini F, Crema MD, Duryea J, Eckstein F, et al. OARSI Clinical Trials Recommendations: Knee imaging in clinical trials in osteoarthritis. Osteoarthritis Cartilage 2015; 23: 698-715.
349. Roemer FW, Guermazi A, Felson DT, Niu J, Nevitt MC, Crema MD, et al. Presence of MRI-detected joint effusion and synovitis increases the risk of cartilage loss in knees without osteoarthritis at 30-month follow-up: the MOST study. Ann Rheum Dis 2011; 70: 1804-1809.
350. Edwards MH, Parsons C, Bruyere O, Petit Dop F, Chapurlat R, Roemer FW, et al. High Kellgren-Lawrence Grade and Bone Marrow Lesions

Predict Worsening Rates of Radiographic Joint Space Narrowing; The SEKOIA Study. J Rheumatol 2016; 43: 657-665.
351. Eckstein F, Collins JE, Nevitt MC, Lynch JA, Kraus VB, Katz JN, et al. Brief Report: Cartilage Thickness Change as an Imaging Biomarker of Knee Osteoarthritis Progression: Data From the Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheumatol 2015; 67: 3184-3189.
352. Niu J, Felson DT, Neogi T, Nevitt MC, Guermazi A, Roemer F, et al. Patterns of Coexisting Lesions Detected on Magnetic Resonance Imaging and Relationship to Incident Knee Osteoarthritis: The Multicenter Osteoarthritis Study. 2015; 67: 3158-3165.
353. Bowes MA, McLure SW, Wolstenholme CB, Vincent GR, Williams S, Grainger A, et al. Osteoarthritic bone marrow lesions almost exclusively colocate with denuded cartilage: a 3D study using data from the Osteoarthritis Initiative. Ann Rheum Dis 2016; 75: 1852-1857.
354. Doran SJ, Charles-Edwards L, Reinsberg SA, Leach MO. A complete distortion correction for MR images: I. Gradient warp correction. Phys Med Biol 2005; 50: 1343-1361.
355. Plante E, Turkstra L. Sources of error in the quantitative analysis of MRI scans. Magnetic Resonance Imaging 1991; 9: 589-595.
356. Despotović I, Goossens B, Philips W. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. Computational and Mathematical Methods in Medicine 2015; 2015: 450341.
357. Bruno F, Barile A, Arrigoni F, Laporta A, Russo A, Carotti M, et al. Weight-bearing MRI of the knee: a review of advantages and limits. Acta bio-medica : Atenei Parmensis 2018; 89: 78-88.
358. Cahoy PM, Orwin JF, Tuite MJ. Evaluation of post-exercise magnetic resonance images of the rotator cuff. Skeletal Radiol 1996; 25: 739-741.
359. Baillet A, Gaujoux-Viala C, Mouterde G, Pham T, Tebib J, Saraux A, et al. Comparison of the efficacy of sonography, magnetic resonance imaging and conventional radiography for the detection of bone erosions in rheumatoid arthritis patients: a systematic review and meta-analysis. Rheumatology (Oxford) 2011; 50: 1137-1147.
360. Razek AA, El-Basyouni SR. Ultrasound of knee osteoarthritis: interobserver agreement and correlation with Western Ontario and McMaster Universities Osteoarthritis. Clin Rheumatol 2016; 35: 997-1001.
361. Conaghan PG, D'Agostino MA, Le Bars M, Baron G, Schmidely N, Wakefield R, et al. Clinical and ultrasonographic predictors of joint replacement for knee osteoarthritis: results from a large, 3-year, prospective EULAR study. Ann Rheum Dis 2010; 69: 644-647.
362. Keen HI, Hensor EM, Wakefield RJ, Mease PJ, Bingham CO, 3rd, Conaghan PG. Ultrasound assessment of response to intra-articular therapy in osteoarthritis of the knee. Rheumatology (Oxford) 2015; 54: 1385-1391.
363. Okano T, Filippucci E, Di Carlo M, Draghessi A, Carotti M, Salaffi F, et al. Ultrasonographic evaluation of joint damage in knee osteoarthritis: feature-specific comparisons with conventional radiography. Rheumatology (Oxford) 2016; 55: 2040-2049.
364. Iagnocco A, Ceccarelli F, Perricone C, Valesini G. The role of ultrasound in rheumatology. Semin Ultrasound CT MR 2011; 32: 66-73.

365. Nogueira-Barbosa MH, Gregio-Junior E, Lorenzato MM, Guermazi A, Roemer FW, Chagas-Neto FA, et al. Ultrasound assessment of medial meniscal extrusion: a validation study using MRI as reference standard. AJR Am J Roentgenol 2015; 204: 584-588.
366. Kloppenburg M, Maheu E, Kraus VB, Cicuttini F, Doherty M, Dreiser RL, et al. OARSI Clinical Trials Recommendations: Design and conduct of clinical trials for hand osteoarthritis. Osteoarthritis Cartilage 2015; 23: 772-786.
367. Boegard T, Rudling O, Dahlstrom J, Dirksen H, Petersson IF, Jonsson K. Bone scintigraphy in chronic knee pain: comparison with magnetic resonance imaging. Ann Rheum Dis 1999; 58: 20-26.
368. 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; 52: 557-563.
369. Turmezei TD, Treece GM, Gee AH, Fotiadou AF, Poole KE. Quantitative 3D analysis of bone in hip osteoarthritis using clinical computed tomography. Eur Radiol 2016; 26: 2047-2054.
370. Torigian DA, Zaidi H, Kwee TC, Saboury B, Udupa JK, Cho ZH, et al. PET/MR imaging: technical aspects and potential clinical applications. Radiology 2013; 267: 26-44.
371. Lee IS, Jin YH, Hong SH, Yang SO. Musculoskeletal applications of PET/MR. Semin Musculoskelet Radiol 2014; 18: 203-216.
372. Hong YH, Kong EJ. (18F)Fluoro-deoxy-D-glucose uptake of knee joints in the aspect of age-related osteoarthritis: a case-control study. BMC Musculoskelet Disord 2013; 14: 141.
373. Kobayashi N, Inaba Y, Tateishi U, Ike H, Kubota S, Inoue T, et al. Comparison of 18F-fluoride positron emission tomography and magnetic resonance imaging in evaluating early-stage osteoarthritis of the hip. Nucl Med Commun 2015; 36: 84-89.
374. Nakamura H, Masuko K, Yudoh K, Kato T, Nishioka K, Sugihara T, et al. Positron emission tomography with 18F-FDG in osteoarthritic knee. Osteoarthritis Cartilage 2007; 15: 673-681.
375. Kobayashi N, Inaba Y, Tateishi U, Yukizawa Y, Ike H, Inoue T, et al. New application of 18F-fluoride PET for the detection of bone remodeling in early-stage osteoarthritis of the hip. Clin Nucl Med 2013; 38: e379-383.
376. Kobayashi N, Inaba Y, Yukizawa Y, Ike H, Kubota S, Inoue T, et al. Use of 18F-fluoride positron emission tomography as a predictor of the hip osteoarthritis progression. Mod Rheumatol 2015; 25: 925-930.
377. Stahl R, Jain SK, Lutz J, Wyman BT, Le Graverand-Gastineau MP, Vignon E, et al. Osteoarthritis of the knee at 3.0 T: comparison of a quantitative and a semi-quantitative score for the assessment of the extent of cartilage lesion and bone marrow edema pattern in a 24-month longitudinal study. Skeletal Radiol 2011; 40: 1315-1327.
378. Burgkart R, Glaser C, Hyhlik-Durr A, Englmeier KH, Reiser M, Eckstein F. Magnetic resonance imaging-based assessment of cartilage loss in severe osteoarthritis: accuracy, precision, and diagnostic value. Arthritis Rheum 2001; 44: 2072-2077.
379. Graichen H, von Eisenhart-Rothe R, Vogl T, Englmeier KH, Eckstein F. Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging: technical validation for use in

analysis of cartilage volume and further morphologic parameters. Arthritis Rheum 2004; 50: 811-816.
380. Cicuttini F, Wluka A, Hankin J, Wang Y. Longitudinal study of the relationship between knee angle and tibiofemoral cartilage volume in subjects with knee osteoarthritis. Rheumatology (Oxford) 2004; 43: 321-324.
381. Eckstein F, Mc Culloch CE, Lynch JA, Nevitt M, Kwoh CK, Maschek S, et al. How do short-term rates of femorotibial cartilage change compare to long-term changes? Four year follow-up data from the osteoarthritis initiative. Osteoarthritis Cartilage 2012; 20: 1250-1257.
382. Raynauld J-P, Martel-Pelletier J, Berthiaume M-J, Beaudoin G, Choquette D, Haraoui B, et al. Long term evaluation of disease progression through the quantitative magnetic resonance imaging of symptomatic knee osteoarthritis patients: correlation with clinical symptoms and radiographic changes. Arthritis Research & Therapy 2006; 8: R21-R21.
383. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Labonte F, Beaudoin G, de Guise JA, et al. Quantitative magnetic resonance imaging evaluation of knee osteoarthritis progression over two years and correlation with clinical symptoms and radiologic changes. Arthritis Rheum 2004; 50: 476-487.
384. Cotofana S, Wyman BT, Benichou O, Dreher D, Nevitt M, Gardiner J, et al. Relationship between knee pain and the presence, location, size and phenotype of femorotibial denuded areas of subchondral bone as visualized by MRI. Osteoarthritis Cartilage 2013; 21: 1214-1222.
385. Cicuttini FM, Jones G, Forbes A, Wluka AE. Rate of cartilage loss at two years predicts subsequent total knee arthroplasty: a prospective study. Ann Rheum Dis 2004; 63: 1124-1127.
386. Raynauld JP, Martel-Pelletier J, Haraoui B, Choquette D, Dorais M, Wildi LM, et al. Risk factors predictive of joint replacement in a 2-year multicentre clinical trial in knee osteoarthritis using MRI: results from over 6 years of observation. Ann Rheum Dis 2011; 70: 1382-1388.
387. Eckstein F, Boudreau RM, Wang Z, Hannon MJ, Wirth W, Cotofana S, et al. Trajectory of cartilage loss within 4 years of knee replacement--a nested case-control study from the osteoarthritis initiative. Osteoarthritis Cartilage 2014; 22: 1542-1549.
388. Pelletier JP, Raynauld JP, Berthiaume MJ, Abram F, Choquette D, Haraoui B, et al. Risk factors associated with the loss of cartilage volume on weight-bearing areas in knee osteoarthritis patients assessed by quantitative magnetic resonance imaging: a longitudinal study. Arthritis Research & Therapy 2007; 9.
389. Bashir A, Gray ML, Burstein D. Gd-DTPA2- as a measure of cartilage degradation. Magn Reson Med 1996; 36: 665-673.
390. Bashir A, Gray ML, Boutin RD, Burstein D. Glycosaminoglycan in articular cartilage: in vivo assessment with delayed Gd(DTPA)(2-)-enhanced MR imaging. Radiology 1997; 205: 551-558.
391. Vandecasteele J, De Deene Y. On the validity of 3D polymer gel dosimetry: I. reproducibility study. Phys Med Biol 2013; 58: 19-42.
392. Owman H, Ericsson YB, Englund M, Tiderius CJ, Tjornstrand J, Roos EM, et al. Association between delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) and joint space narrowing and osteophytes: a

cohort study in patients with partial meniscectomy with 11 years of follow-up. Osteoarthritis Cartilage 2014; 22: 1537-1541.
393. Neogi T, Felson DT. Osteoarthritis: Bone as an imaging biomarker and treatment target in OA. Nat Rev Rheumatol 2016; 12: 503-504.
394. Carballido-Gamio J, Majumdar S. Atlas-based knee cartilage assessment. Magn Reson Med 2011; 66: 574-583.
395. Williams TG, Holmes AP, Waterton JC, Maciewicz RA, Hutchinson CE, Moots RJ, et al. Anatomically corresponded regional analysis of cartilage in asymptomatic and osteoarthritic knees by statistical shape modelling of the bone. IEEE Trans Med Imaging 2010; 29: 1541-1559.
396. Williams TG, Holmes AP, Bowes M, Vincent G, Hutchinson CE, Waterton JC, et al. Measurement and visualisation of focal cartilage thickness change by MRI in a study of knee osteoarthritis using a novel image analysis tool. Br J Radiol 2010; 83: 940-948.
397. Ledingham J, Regan M, Jones A, Doherty M. Factors affecting radiographic progression of knee osteoarthritis. Annals of the rheumatic diseases 1995; 54: 53-58.
398. Goldring SR. The role of bone in osteoarthritis pathogenesis. Rheum Dis Clin North Am 2008; 34: 561-571.
399. Chen JH, Liu C, You L, Simmons CA. Boning up on Wolff's Law: mechanical regulation of the cells that make and maintain bone. J Biomech 2010; 43: 108-118.
400. Agricola R, Reijman M, Bierma-Zeinstra SM, Verhaar JA, Weinans H, Waarsing JH. Total hip replacement but not clinical osteoarthritis can be predicted by the shape of the hip: a prospective cohort study (CHECK). Osteoarthritis Cartilage 2013; 21: 559-564.
401. Lynch JA, Parimi N, Chaganti RK, Nevitt MC, Lane NE, Study of Osteoporotic Fractures Research G. The association of proximal femoral shape and incident radiographic hip OA in elderly women. Osteoarthritis and cartilage 2009; 17: 1313-1318.
402. Wise BL, Liu F, Kritikos L, Lynch JA, Parimi N, Zhang Y, et al. The association of distal femur and proximal tibia shape with sex: The Osteoarthritis Initiative. Semin Arthritis Rheum 2016; 46: 20-26.
403. Wise BL, Niu J, Zhang Y, Liu F, Pang J, Lynch JA, et al. Bone shape mediates the relationship between sex and incident knee osteoarthritis. BMC Musculoskeletal Disorders 2018; 19: 331.
404. Wise BL, Niu JB, Zhang YQ, Liu F, Pang J, Lynch JA, et al. Trajectories of Knee Bone Shape Change Are Associated with Sex and Osteoarthritis. Arthritis & Rheumatology 2017; 69.
405. Dube B, Bowes MA, Hensor EMA, Barr A, Kingsbury SR, Conaghan PG. The relationship between two different measures of osteoarthritis bone pathology, bone marrow lesions and 3D bone shape: data from the Osteoarthritis Initiative. Osteoarthritis and Cartilage 2018.
406. Bowes MA, McLure SWD, Wolstenholme CBH, Vincent GR, Williams S, Grainger A, et al. Osteoarthritic bone marrow lesions almost exclusively colocate with denuded cartilage: a 3D study using data from the Osteoarthritis Initiative. Annals of the Rheumatic Diseases 2016; 75: 1852.
407. Bloecker K, Wirth W, Guermazi A, Hitzl W, Hunter DJ, Eckstein F. Longitudinal change in quantitative meniscus measurements in knee osteoarthritis-data from the Osteoarthritis Initiative. Eur Radiol 2015.

408. Dube B, Bowes MA, Kingsbury SR, Hensor EMA, Muzumdar S, Conaghan PG. Where does meniscal damage progress most rapidly? An analysis using three-dimensional shape models on data from the Osteoarthritis Initiative. Osteoarthritis and Cartilage 2018; 26: 62-71.
409. Altman DG. Statistics in medical journals: some recent trends. Statistics in Medicine 2000; 19: 3275-3289.
410. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ (Clinical research ed.) 2006; 332: 1080-1080.
411. Bishop PA, Herron RL. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. International journal of exercise science 2015; 8: 297-302.
412. Knapp TR. Treating ordinal scales as interval scales: an attempt to resolve the controversy. Nurs Res 1990; 39: 121-123.
413. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. International Journal of Epidemiology 2013; 42: 1511-1519.
414. Turkiewicz A, Luta G, Hughes HV, Ranstam J. Statistical mistakes and how to avoid them – lessons learned from the reproducibility crisis. Osteoarthritis and Cartilage 2018; 26: 1409-1411.
415. Liu C, Cripe TP, Kim MO. Statistical issues in longitudinal data analysis for treatment efficacy studies in the biomedical sciences. Mol Ther 2010; 18: 1724-1730.
416. Garcia TP, Marder K. Statistical Approaches to Longitudinal Data Analysis in Neurodegenerative Diseases: Huntington's Disease as a Model. Current neurology and neuroscience reports 2017; 17: 14-14.
417. Emrani PS, Katz JN, Kessler CL, Reichmann WM, Wright EA, McAlindon TE, et al. Joint space narrowing and Kellgren-Lawrence progression in knee osteoarthritis: an analytic literature synthesis. Osteoarthritis Cartilage 2008; 16: 873-882.
418. Doganay Erdogan B, Leung YY, Pohl C, Tennant A, Conaghan PG. Minimal Clinically Important Difference as Applied in Rheumatology: An OMERACT Rasch Working Group Systematic Review and Critique. J Rheumatol 2016; 43: 194-202.
419. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evidence Based Medicine 2016; 21: 125.
420. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emergency Medicine Journal 2003; 20: 54.
421. Barr AJ, Dube B, Hensor EM, Kingsbury SR, Peat G, Bowes MA, et al. The relationship between clinical characteristics, radiographic osteoarthritis and 3D bone area: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2014; 22: 1703-1709.
422. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000; 11: 550-560.
423. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. Epidemiology 2009; 20: 18-26.
424. Tan X, Shiyko MP, Li R, Li Y, Dierker L. A time-varying effect model for intensive longitudinal data. Psychological methods 2012; 17: 61-77.
425. McCulloch CE, Neuhaus JM. Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. Statistical Science 2011; 26: 388-402.

426. Curran PJ. Have Multilevel Models Been Structural Equation Models All Along? Multivariate Behav Res 2003; 38: 529-569.
427. National Clinical Guideline C. National Institute for Health and Clinical Excellence: Guidance. In: Osteoarthritis: Care and Management in Adults London: National Institute for Health and Care Excellence (UK)

Copyright (c) National Clinical Guideline Centre, 2014. 2014.
428. Zhang W, Doherty M, Leeb BF, Alekseeva L, Arden NK, Bijlsma JW, et al. EULAR evidence-based recommendations for the diagnosis of hand osteoarthritis: report of a task force of ESCISIT. Ann Rheum Dis 2009; 68: 8-17.
429. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: The chronic osteoarthritis management initiative of the U.S. bone and joint initiative. Semin Arthritis Rheum 2014; 43: 701-712.
430. Block JA. OA guidelines: improving care or merely codifying practice? Nature Reviews Rheumatology 2014; 10: 324.
431. Barr A, Conaghan P. Osteoarthritis: recent advances in diagnosis and management. Prescriber 2014; 25: 26-34.
432. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P, Conaghan PG, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis 2013; 72: 1125-1135.
433. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2003; 62.
434. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SMA, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis and Cartilage 2019; 27: 1578-1589.
435. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. 2020; 72: 220-233.
436. Warsi A, LaValley MP, Wang PS, Avorn J, Solomon DH. Arthritis self-management education programs: a meta-analysis of the effect on pain and disability. Arthritis Rheum 2003; 48: 2207-2213.
437. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2015; 1: Cd004376.
438. Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. Exercise for osteoarthritis of the hip. Cochrane Database Syst Rev 2014: Cd007912.
439. Deyle GD, Allison SC, Matekel RL, Ryder MG, Stang JM, Gohdes DD, et al. Physical therapy treatment effectiveness for osteoarthritis of the knee: a randomized comparison of supervised clinical exercise and manual therapy procedures versus a home exercise program. Phys Ther 2005; 85: 1301-1317.
440. Christensen R, Bartels EM, Astrup A, Bliddal H. Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a

systematic review and meta-analysis. Ann Rheum Dis 2007; 66: 433-439.
441. Messier SP, Mihalko SL, Legault C, Miller GD, Nicklas BJ, DeVita P, et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. Jama 2013; 310: 1263-1273.
442. Duivenvoorden T, Brouwer RW, van Raaij TM, Verhagen AP, Verhaar JA, Bierma-Zeinstra SM. Braces and orthoses for treating osteoarthritis of the knee. Cochrane Database Syst Rev 2015: Cd004020.
443. Brouwer RW, Jakma TS, Verhagen AP, Verhaar JA, Bierma-Zeinstra SM. Braces and orthoses for treating osteoarthritis of the knee. Cochrane Database Syst Rev 2005: Cd004020.
444. van Raaij TM, Reijman M, Brouwer RW, Bierma-Zeinstra SM, Verhaar JA. Medial knee osteoarthritis treated by insoles or braces: a randomized trial. Clin Orthop Relat Res 2010; 468: 1926-1932.
445. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012; 64: 465-474.
446. Manheimer E, Cheng K, Linde K, Lao L, Yoo J, Wieland S, et al. Acupuncture for peripheral joint osteoarthritis. Cochrane Database Syst Rev 2010: Cd001977.
447. Bjordal JM, Klovning A, Ljunggren AE, Slordal L. Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: A meta-analysis of randomised placebo-controlled trials. Eur J Pain 2007; 11: 125-138.
448. Makris UE, Kohler MJ, Fraenkel L. Adverse effects of topical nonsteroidal antiinflammatory drugs in older adults with osteoarthritis: a systematic literature review. J Rheumatol 2010; 37: 1236-1243.
449. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010; 18: 476-499.
450. Towheed TE, Maxwell L, Judd MG, Catton M, Hochberg MC, Wells G. Acetaminophen for osteoarthritis. Cochrane Database Syst Rev 2006: Cd004257.
451. Machado GC, Maher CG, Ferreira PH, Pinheiro MB, Lin CW, Day RO, et al. Efficacy and safety of paracetamol for spinal pain and osteoarthritis: systematic review and meta-analysis of randomised placebo controlled trials. Bmj 2015; 350: h1225.
452. Zhang W, Jones A, Doherty M. Does paracetamol (acetaminophen) reduce the pain of osteoarthritis? A meta-analysis of randomised controlled trials. Ann Rheum Dis 2004; 63: 901-907.
453. Rahme E, Barkun A, Nedjar H, Gaugris S, Watson D. Hospitalizations for upper and lower GI events associated with traditional NSAIDs and acetaminophen among the elderly in Quebec, Canada. Am J Gastroenterol 2008; 103: 872-882.

454. Clegg DO, Reda DJ, Harris CL, Klein MA, O'Dell JR, Hooper MM, et al. Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. N Engl J Med 2006; 354: 795-808.
455. Bannuru RR, Natov NS, Obadan IE, Price LL, Schmid CH, McAlindon TE. Therapeutic trajectory of hyaluronic acid versus corticosteroids in the treatment of knee osteoarthritis: a systematic review and meta-analysis. Arthritis Rheum 2009; 61: 1704-1711.
456. Bartels EM, Bliddal H, Schondorff PK, Altman RD, Zhang W, Christensen R. Symptomatic efficacy and safety of diacerein in the treatment of osteoarthritis: a meta-analysis of randomized placebo-controlled trials. Osteoarthritis Cartilage 2010; 18: 289-296.
457. Hochberg MC, Wohlreich M, Gaynor P, Hanna S, Risser R. Clinically relevant outcomes based on analysis of pooled data from 2 trials of duloxetine in patients with knee osteoarthritis. J Rheumatol 2012; 39: 352-358.
458. Avouac J, Gossec L, Dougados M. Efficacy and safety of opioids for osteoarthritis: a meta-analysis of randomized controlled trials. Osteoarthritis Cartilage 2007; 15: 957-965.
459. Kirkley A, Birmingham TB, Litchfield RB, Giffin JR, Willits KR, Wong CJ, et al. A Randomized Trial of Arthroscopic Surgery for Osteoarthritis of the Knee. 2008; 359: 1097-1107.
460. Brouwer RW, Raaij van TM, Bierma-Zeinstra SM, Verhagen AP, Jakma TS, Verhaar JA. Osteotomy for treating knee osteoarthritis. Cochrane Database Syst Rev 2007: Cd004019.
461. Carr AJ, Robertsson O, Graves S, Price AJ, Arden NK, Judge A, et al. Knee replacement. Lancet 2012; 379: 1331-1340.
462. Hunter DJ. Pharmacologic therapy for osteoarthritis—the era of disease modification. Nature Reviews Rheumatology 2011; 7: 13-22.
463. Zou K, Wong J, Abdullah N, Chen X, Smith T, Doherty M, et al. Examination of overall treatment effect and the proportion attributable to contextual effect in osteoarthritis: meta-analysis of randomised controlled trials. Annals of the Rheumatic Diseases 2016; 75: 1964.
464. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. DEVELOPMENT OF CRITERIA FOR THE CLASSIFICATION AND REPORTING OF OSTEOARTHRITIS - CLASSIFICATION OF OSTEOARTHRITIS OF THE KNEE. Arthritis and Rheumatism 1986; 29: 1039-1049.
465. Peterfy CG, Schneider E, Nevitt M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. Osteoarthritis and cartilage 2008; 16: 1433-1441.
466. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001; 69: 89-95.
467. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino M-A, et al. Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. Journal of Clinical Epidemiology 2014; 67: 745-753.
468. Gossec L, Hawker G, Davis AM, Maillefert JF, Lohmander LS, Altman R, et al. OMERACT/OARSI initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. J Rheumatol 2007; 34: 1432-1435.

469. Lesko LJ, Atkinson AJ, Jr. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. Annu Rev Pharmacol Toxicol 2001; 41: 347-366.
470. Goodsaid FM, Frueh FW, Mattes W. Strategic paths for biomarker qualification. Toxicology 2008; 245: 219-223.
471. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. Journal of Clinical Epidemiology 2000; 53: 459-468.
472. Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. Journal of Chronic Diseases 1987; 40: 171-178.
473. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. Controlled Clinical Trials 1991; 12: S142-S158.
474. Kropmans TJ, Dijkstra PU, Stegenga B, Stewart R, de Bont LG. Smallest detectable difference in outcome variables related to painful restriction of the temporomandibular joint. J Dent Res 1999; 78: 784-789.
475. Liaw LJ, Hsieh CL, Lo SK, Chen HM, Lee S, Lin JH. The relative and absolute reliability of two balance performance measures in chronic stroke patients. Disabil Rehabil 2008; 30: 656-661.
476. Kaplan D. Structural Equation Modeling. In: International Encyclopedia of the Social & Behavioral Sciences, Smelser NJ, Baltes PB Eds. Oxford: Pergamon 2001:15215-15222.
477. Tarka P. An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. Quality & Quantity 2018; 52: 313-354.
478. Weston R, Gore Jr PA. A Brief Guide to Structural Equation Modeling. The Counseling Psychologist 2006; 34: 719-751.
479. Bollen KA. Latent variables in psychology and the social sciences. Annu Rev Psychol 2002; 53: 605-634.
480. Kline RB. Assumptions in structural equation modeling. In: Handbook of structural equation modeling. New York, NY, US: Guilford Press 2012:111-125.
481. Martens MP. The Use of Structural Equation Modeling in Counseling Psychology Research. The Counseling Psychologist 2005; 33: 269-298.
482. Steiger JH. Structural Model Evaluation and Modification: An Interval Estimation Approach. Multivariate Behav Res 1990; 25: 173-180.
483. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990; 107: 238-246.
484. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999; 6: 1-55.
485. Hamilton J, E. Gagne P, Hancock G. The Effect of Sample Size on Latent Growth Models2003.
486. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychological Methods 1996; 1: 130-149.
487. Jung T, Wickrama KAS. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. Social and Personality Psychology Compass 2008; 2: 302-317.

488. Nylund KL, Asparouhov T, Muthén BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Structural Equation Modeling: A Multidisciplinary Journal 2007; 14: 535-569.
489. Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. Journal of Classification 1996; 13: 195-212.
490. Hipp JR, Bauer DJ. Local solutions in the estimation of growth mixture models. Psychol Methods 2006; 11: 36-53.
491. Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika 1974; 61: 215-231.
492. Wirth W, Frobell RB, Souza RB, Li X, Wyman BT, Le Graverand MP, et al. A three-dimensional quantitative method to measure meniscus shape, position, and signal intensity using MR images: a pilot study and preliminary results in knee osteoarthritis. Magn Reson Med 2010; 63: 1162-1171.
493. Bowers ME, Tung GA, Fleming BC, Crisco JJ, Rey J. Quantification of meniscal volume by segmentation of 3T magnetic resonance images. J Biomech 2007; 40: 2811-2815.
494. Siorpaes K, Wenger A, Bloecker K, Wirth W, Hudelmaier M, Eckstein F. Interobserver reproducibility of quantitative meniscus analysis using coronal multiplanar DESS and IWTSE MR imaging. Magn Reson Med 2012; 67: 1419-1426.
495. Englund M, Roemer FW, Hayashi D, Crema MD, Guermazi A. Meniscus pathology, osteoarthritis and the treatment controversy. Nat Rev Rheumatol 2012; 8: 412-419.
496. Fox AJS, Bedi A, Rodeo SA. The Basic Science of Human Knee Menisci: Structure, Composition, and Function. Sports Health 2012; 4: 340-351.
497. Brindle T, Nyland J, Johnson DL. The Meniscus: Review of Basic Principles With Application to Surgery and Rehabilitation. Journal of Athletic Training 2001; 36: 160-169.
498. Makris EA, Hadidi P, Athanasiou KA. The knee meniscus: structure-function, pathophysiology, current repair techniques, and prospects for regeneration. Biomaterials 2011; 32: 7411-7431.
499. Englund M. The role of biomechanics in the initiation and progression of OA of the knee. Best Pract Res Clin Rheumatol 2010; 24: 39-46.
500. Krishnan N, Shetty SK, Williams A, Mikulis B, McKenzie C, Burstein D. Delayed gadolinium-enhanced magnetic resonance imaging of the meniscus: an index of meniscal tissue degeneration? Arthritis Rheum 2007; 56: 1507-1511.
501. Englund M, Felson DT, Guermazi A, Roemer FW, Wang K, Crema MD, et al. Risk factors for medial meniscal pathology on knee MRI in older US adults: a multicentre prospective cohort study. Ann Rheum Dis 2011; 70: 1733-1739.
502. Guermazi A, Roemer FW, Haugen IK, Crema MD, Hayashi D. MRI-based semiquantitative scoring of joint pathology in osteoarthritis. Nat Rev Rheumatol 2013; 9: 236-251.
503. Wenger A, Wirth W, Hudelmaier M, Noebauer-Huhmann I, Trattnig S, Bloecker K, et al. Meniscus body position, size, and shape in persons with and persons without radiographic knee osteoarthritis: quantitative

analyses of knee magnetic resonance images from the osteoarthritis initiative. Arthritis Rheum 2013; 65: 1804-1811.
504. Bloecker K, Guermazi A, Wirth W, Benichou O, Kwoh CK, Hunter DJ, et al. Tibial coverage, meniscus position, size and damage in knees discordant for joint space narrowing - data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2013; 21: 419-427.
505. Tack A, Mukhopadhyay A, Zachow S. Knee menisci segmentation using convolutional neural networks: data from the Osteoarthritis Initiative. Osteoarthritis and Cartilage 2018; 26: 680-688.
506. Bloecker K, Englund M, Wirth W, Hudelmaier M, Burgkart R, Frobell RB, et al. Revision 1 size and position of the healthy meniscus, and its correlation with sex, height, weight, and bone area- a cross-sectional study. BMC Musculoskelet Disord 2011; 12: 248.
507. Bloecker K, Wirth W, Hudelmaier M, Burgkart R, Frobell R, Eckstein F. Morphometric differences between the medial and lateral meniscus in healthy men - a three-dimensional analysis using magnetic resonance imaging. Cells Tissues Organs 2012; 195.
508. Eckstein F, Hudelmaier M, Wirth W, Kiefer B, Jackson R, Yu J, et al. Double echo steady state magnetic resonance imaging of knee articular cartilage at 3 Tesla: a pilot study for the Osteoarthritis Initiative. 2006; 65: 433-441.
509. Hughes SW, D'Arcy TJ, Maxwell DJ, Saunders JE, Ruff CF, Chiu WS, et al. Application of a new discreet form of Gauss' theorem for measuring volume. Phys Med Biol 1996; 41: 1809-1821.
510. Bloecker K, Wirth W, Hunter DJ, Duryea J, Guermazi A, Kwoh CK, et al. Contribution of regional 3D meniscus and cartilage morphometry by MRI to joint space width in fixed flexion knee radiography-A between-knee comparison in subjects with unilateral joint space narrowing. European Journal of Radiology 2013; 82: E832-E839.
511. Adams JG, McAlindon T, Dimasi M, Carey J, Eustace S. Contribution of meniscal extrusion and cartilage loss to joint space narrowing in osteoarthritis. Clinical Radiology 1999; 54: 502-506.
512. Wang Y, Wluka AE, Pelletier JP, Martel-Pelletier J, Abram F, Ding C, et al. Meniscal extrusion predicts increases in subchondral bone marrow lesions and bone cysts and expansion of subchondral bone in osteoarthritic knees. Rheumatology (Oxford) 2010; 49: 997-1004.
513. Bruns K, Svensson F, Turkiewicz A, Wirth W, Guermazi A, Eckstein F, et al. Meniscus body position and its change over four years in asymptomatic adults: a cohort study using data from the Osteoarthritis Initiative (OAI). BMC Musculoskeletal Disorders 2014; 15: 32.
514. Jung KA, Lee SC, Hwang SH, Yang KH, Kim DH, Sohn JH, et al. High frequency of meniscal hypertrophy in persons with advanced varus knee osteoarthritis. Rheumatol Int 2010; 30.
515. Guermazi A, Hayashi D, Jarraya M, Roemer FW, Zhang Y, Niu J, et al. Medial posterior meniscal root tears are associated with development or worsening of medial tibiofemoral cartilage damage: the multicenter osteoarthritis study. Radiology 2013; 268: 814-821.
516. Stone KR, Freyer A, Turek T, Walgenbach AW, Wadhwa S, Crues J. Meniscal sizing based on gender, height, and weight. Arthroscopy 2007; 23: 503-508.

517. Eckstein F, Kwoh CK, Boudreau RM, Wang Z, Hannon MJ, Cotofana S, et al. Quantitative MRI measures of cartilage predict knee replacement: a case–control study from the Osteoarthritis Initiative. Annals of the Rheumatic Diseases 2013; 72: 707.
518. Neogi T. Clinical significance of bone changes in osteoarthritis. Therapeutic Advances in Musculoskeletal Disease 2012; 4: 259-267.
519. Driban JB, Lo GH, Lee JY, Ward RJ, Miller E, Pang J, et al. Quantitative bone marrow lesion size in osteoarthritic knees correlates with cartilage damage and predicts longitudinal cartilage loss. BMC Musculoskeletal Disorders 2011; 12: 217.
520. Ornetti P, Brandt K, Hellio-Le Graverand MP, Hochberg M, Hunter DJ, Kloppenburg M, et al. OARSI-OMERACT definition of relevant radiological progression in hip/knee osteoarthritis. Osteoarthritis and cartilage 2009; 17: 856-863.
521. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol 2002; 29: 131-138.
522. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001; 45: 384-391.
523. Campbell TM, Churchman SM, Gomez A, McGonagle D, Conaghan PG, Ponchel F, et al. Mesenchymal Stem Cell Alterations in Bone Marrow Lesions in Patients With Hip Osteoarthritis. Arthritis Rheumatol 2016; 68: 1648-1659.
524. Deveza LA, Kraus VB, Collins JE, Guermazi A, Roemer FW, Bowes M, et al. Association Between Biochemical Markers of Bone Turnover and Bone Changes on Imaging: Data From the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2017; 69: 1179-1191.
525. Aitken D, Ding C, Pelletier JP, Martel-Pelletier J, Cicuttini F, Jones G. Responsiveness of magnetic resonance imaging-derived measures over 2.7 years. J Rheumatol 2014; 41: 2060-2067.
526. Roemer FW, Frobell R, Hunter DJ, Crema MD, Fischer W, Bohndorf K, et al. MRI-detected subchondral bone marrow signal alterations of the knee joint: terminology, imaging appearance, relevance and radiological differential diagnosis. Osteoarthritis Cartilage 2009; 17: 1115-1131.
527. 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; 54: 1529-1535.
528. Hunter DJ, Bowes MA, Eaton CB, Holmes AP, Mann H, Kwoh CK, et al. Can cartilage loss be detected in knee osteoarthritis (OA) patients with 3-6 months' observation using advanced image analysis of 3T MRI? Osteoarthritis Cartilage 2010; 18: 677-683.
529. Mahfouz M, Badawi A, Merkl B, Fatah EE, Pritchard E, Kesler K, et al. Patella sex determination by 3D statistical shape models and nonlinear classifiers. Forensic Sci Int 2007; 173: 161-170.
530. Felson DT, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. Arthritis Rheum 1998; 41: 1343-1355.

531. Deveza LA, Downie A, Tamez-Pena JG, Eckstein F, Van Spil WE, Hunter DJ. Trajectories of femorotibial cartilage thickness among persons with or at risk of knee osteoarthritis: development of a prediction model to identify progressors. Osteoarthritis Cartilage 2019; 27: 257-265.
532. Mahfouz M, Abdel Fatah EE, Bowers LS, Scuderi G. Three-dimensional morphology of the knee reveals ethnic differences. Clinical orthopaedics and related research 2012; 470: 172-185.
533. Tu YK, D'Aiuto F, Baelum V, Gilthorpe MS. An introduction to latent growth curve modelling for longitudinal continuous data in dental research. Eur J Oral Sci 2009; 117: 343-350.
534. Tu Y-K, Tilling K, Sterne JAC, Gilthorpe MS. A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. International Journal of Epidemiology 2013; 42: 1327-1339.
535. Burant CJ. Latent Growth Curve Models: Tracking Changes Over Time. The International Journal of Aging and Human Development 2016; 82: 336-350.
536. Enders CK, Bandalos DL. The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. Structural Equation Modeling: A Multidisciplinary Journal 2001; 8: 430-457.
537. Skousgaard SG, Skytthe A, Moller S, Overgaard S, Brandt LP. Sex differences in risk and heritability estimates on primary knee osteoarthritis leading to total knee arthroplasty: a nationwide population based follow up study in Danish twins. Arthritis Res Ther 2016; 18: 46.
538. Wise BL, Niu J, Yang M, Lane NE, Harvey W, Felson DT, et al. Patterns of compartment involvement in tibiofemoral osteoarthritis in men and women and in whites and African Americans. Arthritis Care & Research 2012; 64: 847-852.
539. Anderson JJ, Felson DT. Factors associated with osteoarthritis of the knee in the first national Health and Nutrition Examination Survey (HANES I). Evidence for an association with overweight, race, and physical demands of work. Am J Epidemiol 1988; 128: 179-189.
540. Schouten JS, van den Ouweland FA, Valkenburg HA. A 12 year follow up study in the general population on prognostic factors of cartilage loss in osteoarthritis of the knee. Ann Rheum Dis 1992; 51: 932-937.
541. Creamer P, Hochberg MC. Osteoarthritis. Lancet 1997; 350: 503-508.
542. Hochberg MC, Creamer P. Why does osteoarthritis of the knee hurt--sometimes? Rheumatology 1997; 36: 726-728.
543. Hunter DJ, Lohmander LS, Makovey J, Tamez-Peña J, Totterman S, Schreyer E, et al. The effect of anterior cruciate ligament injury on bone curvature: exploratory analysis in the KANON trial. Osteoarthritis and Cartilage 2014; 22: 959-968.
544. Bowes MA, Lohmander LS, Wolstenholme CBH, Vincent GR, Conaghan PG, Frobell RB. Marked and rapid change of bone shape in acutely ACL injured knees - an exploratory analysis of the Kanon trial. Osteoarthritis Cartilage 2019; 27: 638-645.
545. Jordan JM, Helmick CG, Renner JB, Luta G, Dragomir AD, Woodard J, et al. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston

County Osteoarthritis Project. The Journal of Rheumatology 2007; 34: 172.
546. Braga L, Renner JB, Schwartz TA, Woodard J, Helmick CG, Hochberg MC, et al. Differences in radiographic features of knee osteoarthritis in African-Americans and Caucasians: the Johnston County Osteoarthritis Project. Osteoarthritis and Cartilage 2009; 17: 1554-1561.
547. Bartlett SJ, Ling SM, Mayo NE, Scott SC, Bingham III CO. Identifying common trajectories of joint space narrowing over two years in knee osteoarthritis. 2011; 63: 1722-1728.
548. McAlindon TE, Driban JB, Henrotin Y, Hunter DJ, Jiang GL, Skou ST, et al. OARSI Clinical Trials Recommendations: Design, conduct, and reporting of clinical trials for knee osteoarthritis. Osteoarthritis and Cartilage 2015; 23: 747-760.
549. Jordan JM, Sowers MF, Messier SP, Bradley J, Arangio G, Katz JN, et al. Methodologic issues in clinical trials for prevention or risk reduction in osteoarthritis. Osteoarthritis and cartilage 2011; 19: 500-508.
550. Eckstein F, Buck R, Wirth W. Location-independent analysis of structural progression of osteoarthritis-Taking it all apart, and putting the puzzle back together makes the difference. Semin Arthritis Rheum 2017; 46: 404-410.
551. Driban JB, Sitler MR, Barbe MF, Balasubramanian E. Is osteoarthritis a heterogeneous disease that can be stratified into subsets? Clin Rheumatol 2010; 29: 123-131.
552. van de Schoot R, Sijbrandij M, Winter SD, Depaoli S, Vermunt JK. The GRoLTS-Checklist: Guidelines for Reporting on Latent Trajectory Studies. Structural Equation Modeling: A Multidisciplinary Journal 2017; 24: 451-467.
553. Neogi T, Bowes MA, Zhang Y, Niu J, Conaghan PG. Effect of bisphosphonate use on trajectories of MRI-based three-dimensional bone shape of the knee over four years. Osteoarthritis and Cartilage 2017; 25: S58.
554. Neogi T, Conaghan PG, Niu J, Zhang Y, Bowes MA. Ttrajectories of MRI-based 3-dimensional bone shape of the knee over 6 years: Insights into knee osteoarthritis progression. Osteoarthritis and Cartilage 2016; 24: S57.
555. Neogi T, Niu J, Duryea J, Lynch J, Zhang Y. IDENTIFYING TRAJECTORIES OF MEDIAL JOINT-SPACE WIDTH LOSS AND ASSOCIATED RISK FACTORS. Osteoarthritis and Cartilage 2012; 20: S182-S183.
556. Kwoh CK, Ran D, Ashbeck EL, Duryea J. Distinct Trajectories of Medial Fixed Joint Space Width Loss over Four Years of Follow-up Among Knees with and at Risk for Knee Osteoarthritis. Arthritis & Rheumatology 2017; 69.
557. Herrero-Beaumont G, Roman-Blas JA, Largo R, Castañeda S. Osteoarthritis: a progressive disease with changing phenotypes. Rheumatology 2013; 53: 1-3.
558. Roth M, Emmanuel K, Wirth W, Kwoh CK, Hunter DJ, Eckstein F. Sensitivity to change and association of three-dimensional meniscal measures with radiographic joint space width loss in rapid clinical progression of knee osteoarthritis. Eur Radiol 2018; 28: 1844-1853.

559. Lerer DB, Umans HR, Hu MX, Jones MH. The role of meniscal root pathology and radial meniscal tear in medial meniscal extrusion. Skeletal Radiol 2004; 33: 569-574.
560. Okazaki Y, Furumatsu T, Yamaguchi T, Kodama Y, Kamatsuki Y, Masuda S, et al. Medial meniscus posterior root tear causes swelling of the medial meniscus and expansion of the extruded meniscus: a comparative analysis between 2D and 3D MRI. Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA 2019.
561. Roth M, Emmanuel K, Wirth W, Kwoh CK, Hunter DJ, Hannon MJ, et al. Changes in Medial Meniscal 3D Position and Morphology Predict Knee Replacement in Rapidly Progressing Knee Osteoarthritis - Data from the Osteoarthritis Initiative (OAI). Arthritis Care Res (Hoboken) 2020.
562. Sharma K, Eckstein F, Wirth W, Emmanuel K. Meniscus position and size in knees with versus without structural knee osteoarthritis progression - data from the osteoarthritis initiative. Osteoarthritis and Cartilage 2019; 27: S467-S468.
563. Roth M, Wirth W, Emmanuel K, Culvenor AG, Eckstein F. The contribution of 3D quantitative meniscal and cartilage measures to variation in normal radiographic joint space width-Data from the Osteoarthritis Initiative healthy reference cohort. Eur J Radiol 2017; 87: 90-98.
564. Bowes MA, Kacena KA, Alabas O, Brett AD, Dube B, Bodick N, et al. A novel structural status assessment of oa is strongly associated with all patient important outcomes - an mri study of 9,433 knees from the osteoarthritis initiative. Osteoarthritis and Cartilage 2020; 28: S47-S48.
565. Collins JE, Neogi T, Losina E. Trajectories of Structural Disease Progression in Knee Osteoarthritis. Arthritis Care Res (Hoboken) 2020.
566. Perry TA, Parkes MJ, Hodgson RJ, Felson DT, Arden NK, O'Neill TW. Association between Bone marrow lesions & synovitis and symptoms in symptomatic knee osteoarthritis. Osteoarthritis and Cartilage 2020; 28: 316-323.
567. Zhu Z, Ding C, Han W, Zheng S, Winzenberg T, Cicuttini F, et al. MRI-detected osteophytes of the knee: natural history and structural correlates of change. Arthritis Research & Therapy 2018; 20: 237.
568. Conaghan PG, Bowes MA, Kingsbury SR, Brett A, Guillard G, Rizoska B, et al. Disease-Modifying Effects of a Novel Cathepsin K Inhibitor in Osteoarthritis: A Randomized Controlled Trial. Annals of Internal Medicine 2020; 172: 86-95.
569. Zhong Q, Pedoia V, Tanaka M, Neumann J, Link TM, Ma B, et al. 3D bone-shape changes and their correlations with cartilage T1ρ and T2 relaxation times and patient-reported outcomes over 3-years after ACL reconstruction. Osteoarthritis and Cartilage 2019; 27: 915-921.
570. MacKay JW, Kapoor G, Driban JB, Lo GH, McAlindon TE, Toms AP, et al. Association of subchondral bone texture on magnetic resonance imaging with radiographic knee osteoarthritis progression: data from the Osteoarthritis Initiative Bone Ancillary Study. European Radiology 2018; 28: 4687-4695.
571. Meng Q, Fisher J, Wilcox R. The effects of geometric uncertainties on computational modelling of knee biomechanics. R Soc Open Sci 2017; 4: 170670.

572. Cooper RJ, Wilcox RK, Jones AC. Finite element models of the tibiofemoral joint: A review of validation approaches and modelling challenges. Medical Engineering & Physics 2019; 74: 1-12.
573. Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: a comparison between methods. J Clin Epidemiol 2012; 65: 1078-1087.
574. Gilthorpe MS, Dahly DL, Tu YK, Kubzansky LD, Goodman E. Challenges in modelling the random structure correctly in growth mixture models and the impact this has on model mixtures. Journal of developmental origins of health and disease 2014; 5: 197-205.
575. Bowes MA, Guillard GA, Vincent GR, Brett AD, Wolstenholme CBH, Conaghan PG. Precision, Reliability, and Responsiveness of a Novel Automated Quantification Tool for Cartilage Thickness: Data from the Osteoarthritis Initiative. J Rheumatol 2020; 47: 282-289.
576. Kingsbury SR, Tharmanathan P, Keding A, Corbacho B, Watt FE, Scott DL, et al. Significant pain reduction with oral methotrexate in knee osteoarthritis; results from the promote randomised controlled phase iii trial of treatment effectiveness. Osteoarthritis and Cartilage 2019; 27: S84-S85.
577. Pedoia V, Samaan MA, Inamdar G, Gallo MC, Souza RB, Majumdar S. Study of the interactions between proximal femur 3d bone shape, cartilage health, and biomechanics in patients with hip Osteoarthritis. J Orthop Res 2018; 36: 330-341.
578. Bowes MA, Guillard G, Gill E, Vincent GR, Hensor E, Freeston JE, et al. Novel quantification of MRI provides a more sensitive outcome measure than Ramris. 2014; 66: abstract 1178.
579. Conaghan PG, Østergaard M, Bowes MA, Wu C, Fuerst T, van der Heijde D, et al. Comparing the effects of tofacitinib, methotrexate and the combination, on bone marrow oedema, synovitis and bone erosion in methotrexate-naive, early active rheumatoid arthritis: results of an exploratory randomised MRI study incorporating semiquantitative and quantitative techniques. Annals of the rheumatic diseases 2016; 75: 1024-1033.
580. Conaghan PG, Østergaard M, Troum O, Bowes MA, Guillard G, Wilkinson B, et al. Very early MRI responses to therapy as a predictor of later radiographic progression in early rheumatoid arthritis. Arthritis Research & Therapy 2019; 21: 214.
581. Robinson HS, Dagfinrud H. Reliability and screening ability of the StarT Back screening tool in patients with low back pain in physiotherapy practice, a cohort study. BMC Musculoskelet Disord 2017; 18: 232.
582. Neogi T. Structural correlates of pain in osteoarthritis. Clin Exp Rheumatol 2017; 35 Suppl 107: 75-78.
583. Merchant AT, Pitiphat W. Directed acyclic graphs (DAGs): an aid to assess confounding in dental research. Community Dent Oral Epidemiol 2002; 30: 399-404.
584. Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. International Journal of Epidemiology 2017; 45: 1887-1894.
585. Nelson AE, Fang F, Arbeeva L, Cleveland RJ, Schwartz TA, Callahan LF, et al. A machine learning approach to knee osteoarthritis

phenotyping: data from the FNIH Biomarkers Consortium. Osteoarthritis and Cartilage 2019; 27: 994-1001.
586. Marron JS, Todd MJ, Ahn J. Distance-Weighted Discrimination. Journal of the American Statistical Association 2007; 102: 1267-1271.
587. Wei S, Lee C, Wichers L, Marron JJJoC, Statistics G. Direction-projection-permutation for high-dimensional hypothesis tests. 2016; 25: 549-569.
588. An H, Marron JS, Schwartz TA, Renner JB, Liu F, Lynch JA, et al. Novel statistical methodology reveals that hip shape is associated with incident radiographic hip osteoarthritis among African American women. Osteoarthritis and Cartilage 2016; 24: 640-646.
589. Ashinsky BG, Coletta CE, Bouhrara M, Lukas VA, Boyle JM, Reiter DA, et al. Machine learning classification of OARSI-scored human articular cartilage using magnetic resonance imaging. Osteoarthritis and Cartilage 2015; 23: 1704-1712.
590. Zhou ZY, Zhao GY, Kijowski R, Liu F. Deep convolutional neural network for segmentation of knee joint anatomy. Magnetic Resonance in Medicine 2018; 80: 2759-2770.
591. Pedoia V, Norman B, Mehany SN, Bucknor MD, Link TM, Majumdar S. 3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. Journal of Magnetic Resonance Imaging 2019; 49: 400-410.
592. Norman B, Pedoia V, Majumdar S. Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry. Radiology 2018; 288: 177-185.
593. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd SJN. Quantum machine learning. 2017; 549: 195-202.
594. Kluzek S, Mattei TA. Machine-learning for osteoarthritis research. Osteoarthritis and Cartilage 2019; 27: 977-978.
595. Swan AL, Stekel DJ, Hodgman C, Allaway D, Alqahtani MH, Mobasheri A, et al. A machine learning heuristic to identify biologically relevant and minimal biomarker panels from omics data. 2015; 16: S2.
596. Sun CS, Markey MK. Recent advances in computational analysis of mass spectrometry for proteomic profiling. J Mass Spectrom 2011; 46: 443-456.
597. Wesseling J, Bastick AN, ten Wolde S, Kloppenburg M, Lafeber FP, Bierma-Zeinstra SM, et al. Identifying Trajectories of Pain Severity in Early Symptomatic Knee Osteoarthritis: A 5-year Followup of the Cohort Hip and Cohort Knee (CHECK) Study. J Rheumatol 2015; 42: 1470-1477.
598. Felson D, Niu J, Sack B, Aliabadi P, McCullough C, Nevitt MC. Progression of osteoarthritis as a state of inertia. Annals of the rheumatic diseases 2013; 72: 924-929.