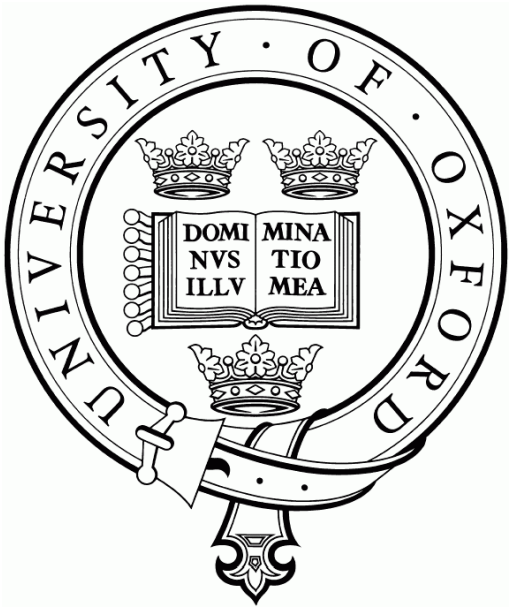
The association between radiographic knee osteoarthritis and pain: an epidemiological analysis of a twenty-year community-based cohort



Kirsten Leyland

Wolfson College

University of Oxford

Supervised by

Prof. Nigel Arden, Dr. Kassim Javaid and Dr. Andy Judge

Submitted: Michaelmas Term 2012

This thesis is submitted to the Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Kirsten Leyland, Wolfson College, University of Oxford

Submission for the degree of Doctor of Philosophy

Michaelmas Term 2012

ABSTRACT

Title: The association between radiographic knee osteoarthritis and pain: an epidemiological analysis of a twenty-year community-based cohort

Background: Knee osteoarthritis (OA) is one of the leading musculoskeletal burdens in the world, causing both structural damage and pain to the joint. Radiographs are the most common imaging method for diagnosing OA, however the relationship between radiographic changes (ROA) and symptoms is not well understood. This thesis will establish the natural history of twenty-year ROA, compare diagnostic methods of ROA assessment, evaluate the cross-sectional relationship between ROA and pain, and will determine the long-term predictive validity of features of ROA with future knee replacements.

Methods: Data from the Chingford women's study, a twenty-year prospective UK-based cohort was used for the analysis. Risk factors included atlas-based ROA scoring methods and quantitative joint space width, which were analysed against pain and TKR outcome measures. A novel method for assessing joint space on low-contrast x-rays was developed which had high reproducibility and validity.

Results: The twenty-year natural history of ROA showed relatively low levels of annual incidence (3.8%), progression (3.6%), and worsening (4.5%), and emphasised the involvement of both medial and lateral osteophytes. Severe JSN had the best construct validity with pain, while any size of lateral tibial osteophyte had good construct validity. Medial quantitative JSW had good construct validity, but no predictive validity with future TKRs. Lateral JSN and osteophytes had the best predictive validity for future TKR.

Conclusion: This research demonstrates that radiographic scoring methods have strong construct and predictive validity with symptomatic knee OA. These results support the use of x-rays to identify early disease changes which indicate future joint failure.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors without whose support and encouragement this thesis would not have been possible. To Professor Nigel Arden for helping me maintain focus and keep my eye on the big picture, to Dr. Kassim Javaid whose creativity in research inspired me to push my work further, and to Dr. Andrew Judge whose constant guidance and advice were invaluable.

I would also like to thank all of the members of the epidemiology research group (G24), who have generously given their time to help me with my research and encouraged me every step of the way. A special thank you to Dr. David Hunter for his amazing work on KneeMorf.

I would like to sincerely thank my family, both old and new, for loving and supporting me throughout this process. To my brother Jim, my Aunt Susan and Uncle Louis, The Leylands, Jeff Stranford and finally my mom whose encouragement and belief in me was unwavering.

And last, but certainly not least, my husband Craig. Without his support I would not have had the courage to even begin this project. Thank you.

LIST OF TABLES

LIST OF FIGURES

COMMON ABBREVIATIONS

AP	anteroposterior
BMI	body mass index
CC	correlation coefficient
CI	confidence intervals
GEE	general estimating equation
ICC	intra-class correlations
IMD	inter-margin distance
IRD	inter-rim distance
JSA	joint space area
JSN	joint space narrowing
JSW	joint space width
K/L	Kellgren and Lawrence
meanJSW	mean joint space width
minJSW	minimum joint space width
MRI	magnetic resonance imaging
MTP	metatarsophalangeal
NHANES	national health and nutrition examination survey
OA	osteoarthritis
OR	odds ratio
RA	Rheumatoid arthritis
ROA	radiographic OA
SE	standard error
TKR	total knee replacement

PUBLICATIONS ARISING FROM THIS THESIS

Journal Article

Leyland, K. M., Hart, D. J., Javaid, M. K., Judge, A., Kiran, A., Soni, A., et al. (2012). The natural history of radiographic knee osteoarthritis: a fourteen-year population-based cohort study. Arthritis Rheum, 64(7), 2243-2251.

Conference Papers

Leyland, K.M.; Hunter, D.; Judge, A.; Bottomley, N.; Hart, D.; Gill, R.; Javaid, M.K.; Arden, N.K. 'Bezier curves for measuring joint space on knee radiographs – reproducibility and validity'. *Osteoarthritis and Cartilage* vol. 19 September, 2011. p. S180-S181

Leyland, K.; Bottomley, N.; Judge, A.; Spector, T.; Hart, D.; Javaid, M.K.; Arden, N.K. 'A comparison of four radiographic scoring methods for knee osteoarthritis – short and medium term reproducibility *Osteoarthritis and Cartilage* vol. 19 September, 2011. p. S189

White, K.M; Soni, A.; Kiran, A.; Goulston, L.; Hart, D.; Spector, T.; Javaid, K.; Arden, N.K.; 'Mild radiographic features predict structural progression of knee osteoarthritis'. BSR Concurrent Oral 11, April 2010

1 INTRODUCTION

1.1 Rationale

Knee osteoarthritis is a major world health burden, and radiographs are the most commonly used imaging technology for diagnosis. However, research into the relationship between radiographic disease and clinical symptoms is both confusing and incomplete. Due to a lack of standardised diagnostic methods and the limited availability of long-term population-based cohorts, it has been difficult to determine the true prevalence of disease over its full natural course, to assess the relationship between pain and radiographic OA and to determine the long term risk of having radiographic knee osteoarthritis. All of these analyses are essential for identifying risk factors for severe OA validate the diagnostic methods for ROA. This thesis aims to address these fundamental issues surrounding both the natural history of knee osteoarthritis as well as determine the construct and predictive validity of common and novel radiographic diagnostic methods to pain and knee replacement in a population-based cohort.

Osteoarthritis affects the tissue of diarthrodial joints, the most common of these joints being the hip, knee, hand, foot and spine. OA is a complex disease process involving both bone and soft tissue changes which ultimately can lead to a complete loss of articular cartilage in severe cases. Pain is the primary symptom of OA which causes patients to seek out medical care such as pain management, physiotherapy and, in severe cases, joint replacement surgery. Knee osteoarthritis can be diagnosed by both its structural and symptomatic components, either independently or jointly, which makes determining overall disease prevalence within and between populations complicated.

Research cohorts focusing on osteoarthritis have used x-rays as the primary imaging method over the past 50 years and therefore long-term information about structural OA and natural history is retained on radiographs. Although there is a move within the

research community in favour of the use of MRI, its advantage over radiographic assessment has not yet been proven (Hannan 2000 et al.). Even if MRI has diagnostic benefits over x-ray, due to the lack of long-term data, it cannot be used to answer questions about the natural history of structural OA or the relationship of early structural changes to future pain and the need for joint replacements over long periods of time.

The percentage of people with knee osteoarthritis has been predicted to rise sharply in the future due to the ageing population in addition to the on-going obesity epidemic, both of which are major risk factors for knee osteoarthritis (Felson et al. 1987; Jinks et al. 2006; Blagojevic et al. 2009). The economic burden of osteoarthritis is considerable, taking into account GP visits, physiotherapy, pain management and surgery (Dakin et al. 2012). Knee joint replacements are expensive procedures costing the NHS £7458 per patient with 60,000 procedures each year in England and Wales (Dakin et al. 2012). Temporal trends for knee replacements have shown a three-fold increase between 1991 and 2006 in the UK (Culliford et al. 2010). Due to these projections, there is a major need to identify early risk factors of OA and the natural history of ROA in order to prevent this increasing burden and subsequently the need for knee replacements.

Despite the known disease burden of knee osteoarthritis, research has been limited for assessing general populations without evidence of established disease**.** This is primarily due to OA being a relatively slow progression disease and the cost involved in following up subjects over a long period of time through incidence and progression. Therefore, the majority of studies have focused on subjects who already have clinical or radiographic evidence of knee OA, have common risk factors for OA or have focussed on the cross-sectional analysis of pain, radiographic OA and other risk factors. While these analyses are extremely important in understanding knee osteoarthritis, the complete picture

of disease is still lacking both for the risk of incidence for normal subjects, the long-term disease course and the predictive validity of current radiographic diagnostic methods.

This thesis will address several major issues surrounding both the confusion in existing ROA and pain data. It will contribute new information to help understand the long-term disease process as well as identify early radiographic risk factors for severe knee osteoarthritis leading to knee replacements. The Chingford study, a population-based cohort of women followed for 20 years, is the ideal study to look at the long-term natural history of ROA in both symptomatic and asymptomatic subjects. Using this cohort, this research will focus on improving the understanding of the relationship between radiographic OA and pain by assessing common and novel methods of diagnosis, determining the natural history of radiographic OA over twenty years, examining the construct validity of ROA scoring methods with pain, and evaluating the predictive validity of radiographic risk factors for future total knee replacements.

1.2 Objectives

- To determine the best way to evaluate and measure radiographic knee osteoarthritis, if there is one, by assessing reproducibility and validity, and comparing individual scores from each of the methods against one another

- To develop a novel method for measuring quantitative joint space width on older, low-contrast x-rays while maintaining comparability with modern edge-detection methods

- To describe the natural history of ROA over 20 years in order to better understand disease prevalence and incidence in a normal population including subjects with little or no disease

- To examine the construct validity of established and novel radiographic scoring methods by assessing their cross-sectional relationship between ROA and pain at the year 20 visit

- To compare and assess the predictive validity of baseline ROA and five-year change of ROA scoring methods for future knee replacements over the next twenty years

2 LITERATURE REVIEW

This chapter will review the current literature focusing on the epidemiology of knee osteoarthritis, the most common diagnostic methods for assessing radiographic knee osteoarthritis (OA) (both atlas-based methods and quantitative scoring), the established natural history of radiographic OA (ROA), and the known relationship between ROA and pain/knee replacements.

2.1 Epidemiology of knee osteoarthritis

2.1.1 Background

Knee osteoarthritis is the one of leading health burdens in the world, costing 14.6 billion dollars in the US for knee replacements in 2004 in the US alone (Kim et al. 2008). This figure does not address the additional expense of pain management, loss of work due to disability, and various treatment options such as physiotherapy and revision surgery. The economic burden of OA is increasing, with 54% more knee replacements performed in 2004 than four years earlier, with this number estimated to rise to 1.4 million by the year 2015 (Kim 2008). The trend has been further substantiated in a long-term study based in the UK, where the rate of knee replacements tripled between 1991 and 2006 (Culliford et al. 2010). Because of the increasing health burden due to the ageing population and a projected cumulative prevalence of developing symptomatic knee OA of 45%, there is an urgent need to understand the natural course of knee osteoarthritis in order to target preventative therapies and reduce known risk factors for both incidence and progression (Murphy et al. 2008).

Knee osteoarthritis is an extremely complex disease involving many aspects of the synovial joint and a variety of responses from the tissues involved. It is generally

recognised that OA is a progressive disease that may stem from an attempt by the joint to repair damage, such as trying to stabilise the joint by growing bone spurs (osteophytes) (Pottenger et al. 1990; Nagaosa et al. 2002) or repairing loading damage by strengthening subchondral bone (subchondral sclerosis) (Ding et al. 2005; Dore et al. 2006). Other tissues involved at different stages of disease include articular cartilage, meniscus, synovium and ligaments (Peterfy et al. 2004). Early disease changes include the fibrillation and focal-loss of cartilage, the thickening and remodeling of subchondral bone and the occurrence of excess bone growth (osteophytes) (Garstang and Stitik 2006; Hough 2007; Lane et al. 2001). End-stage disease is marked by complete destruction of the cartilage combined with increasing numbers and/or size of osteophytes, leading to clinical symptoms of pain, stiffness and disability (Hough 2007; Lane et al. 2011). Although osteoarthritis is generally considered a non-inflammatory disease, the inflammation of the synovial membrane has been linked to other on-going disease changes within the joint (Garstang and Stitik 2006). Pain is generally the symptom that drives patients to seek out a doctor and further investigation into the joint.

Criteria were developed by the American College of Rheumatology to determine clinical osteoarthritis, defining it as a combination of at least three of the following features: age>50; crepitus on motion; morning stiffness (less than30 minutes); tenderness of bony margins of the joint; presence of bony enlargement on examination; and lack of palpable warmth of the synovium (Altman et al. 1986). However, the most sensitive definition included some of these features in addition to the presence of an osteophyte found on x-ray (Altman et al. 1986). In population-based research, where clinical evaluations are not always possible, osteoarthritis-specific pain can be defined by either questions about frequency of pain in the last month (NHANES) or by a series of

functional, pain, and disability questions (WOMAC) (NHANES 1979; Bellamy et al. 1986).

2.1.2 Defining knee osteoarthritis

Knee osteoarthritis can be defined in one of three ways: by the presence of structural changes using imaging methods (x-ray or MRI), by the presence of symptoms, or by a combination of both symptoms and structural features.

Plain film radiography is the most common diagnostic imaging technique used to evaluate knee OA; although other imaging modalities such as MRI are being assessed within the research community, their advantage over radiographic assessment in clinical practice remains uncertain (Hannan et al. 2000). Radiographs are inexpensive, have a low dose of radiation, and are a relatively simple and quick diagnostic tool to use.

The primary issue with using radiography as the gold-standard diagnostic tool for knee osteoarthritis is its perceived poor correlation with pain and disability. A significant association has been found between radiographic knee OA and the presence of pain (McAlindon et al. 1992; Lethbridge-Cejku et al. 1995; Cicuttini et al. 1996) as well as between pain and increasing radiographic scores (Neogi et al. 2009; Cho et al. 2010). The reported prevalences, however, range between 15% and 76% depending on the study (Bedson and Croft 2008), which does not inspire confidence in the reliability or usefulness of this association. These differences between studies could be explained by the variation in study populations, radiographic protocols and views, scoring methods, and measures of pain and/or disability. A US-based cohort study looked at the differences in prevalence using the different definitions; 43% of the subjects had symptomatic knee OA, 28% had radiographic OA and 16% had a combination of both symptomatic and radiographic OA (Jordan et al. 2007).

2.1.3 Mechanisms and risk factors for disease

The potential causes for knee osteoarthritis can be divided into local factors (biomechanical) and systemic factors (genetic and biochemical) (table 2.1). Biomechanical factors linked to the onset and progression of OA include abnormal joint loading, muscle weakness, knee alignment, trauma or injury to the joint, and obesity (Felson et al. 2000; Sharma et al. 2001; Englund et al. 2004; Roos et al. 2005; Garstang et al. 2006; Goulston et al. 2011; Lane et al. 2011). Varus alignment has been shown to cause four-fold increase of medial narrowing over 18 months (Sharma et al. 2001), while previous knee injury in young adults has been shown to increase risk of subsequent OA in 12 years by over five times compared with those without injury (Roos et al. 2005). Abnormal joint loading is associated with an increased risk of knee OA, possibly related to previous joint trauma or a loss of muscle function (Roos 2005). Obesity is associated with an increased prevalence of symptomatic knee OA, and it has been suggested that reduction in BMI to the normal range would reduce the need for surgery by 24% (Coggon et al. 2001; Goulston et al. 2011). Physical activity in the form of sport or occupation has been found to be associated with knee osteoarthritis. Subjects with a history of regular sports participation has been linked to an increased risk of incident ROA, and occupations requiring manual labour (agricultural, forestry, fishing, etc.) have been found to have a significantly higher risk of OA than more sedentary occupations (clerical, technical, etc.) (Muraki et al. 2011). Biomechanical factors are extremely important, because unlike many of the systemic and/or genetic factors, there is a possibility of modification which could reduce pain and improve function (Lane et al. 2011).

Systemic factors for increased risk of knee osteoarthritis include age, sex, ethnicity, genetic and psychosocial factors (Neame et al. 2004; Garstang et al. 2006; Hawker et al. 2011). Obesity may also be associated with systemic factors as part of a generalised

metabolic syndrome (including hypertension, impaired glucose tolerance, etc.), which has been found to have a dose-response relationship with the risk of knee osteoarthritis based on the number of metabolic factors present (Yoshimura et al.2012). Women have a higher prevalence of knee OA than men after age 50, and an overall association between age and increasing OA prevalence has been identified (Van Sasse et al. 1989; Felson et al. 1987). The index to ring finger ratio (2D:4D) is generally a sexually dimorphic trait, with ratios associated with higher testosterone levels. It has been identified as an independent risk factor for knee OA (Zhang et al. 2008). The presence of Heberden's nodes, indicating generalised OA, has been shown to predict progressive knee osteoarthritis (Cooper et al. 2000). Sibilings have been found to have an increased risk of prevalent OA, even after adjustment for relevant environmental risk factors, indicating a likely genetic component of OA (Neame et al. 2004). Ethnicity is also associated with prevalence of knee OA, with higher rates found among African-Americans, Chinese and Japanese populations compared to European/Caucasian subjects (Zhang et al. 2001; Jordan et al. 2007; Muraki et al. 2009). Both nutrition and depression have been indicated in studies as influencing the course of the disease process and have been linked to increasing OA related pain (McAlindon et al. 1996; Hawker et al. 2011).

While this distinction between systemic (genetic) and local (biomechanical) factors is useful for characterising risk factors, the interaction between these factors is complicated and they do not likely occur in isolation. For example, the genetic influence of ethnicity or sex may affect joint shape, which in turn may be more susceptible to the effects of mal-alignment or abnormal joint loading increasing the risk of OA.

Table 2.1 Commonly reported risk factors for knee osteoarthritis

Type	Risk Factor	References
Systemic Factors	Age	Cooper 2000; Hart 1999; Van Sasse 1989
	Sex	Felson1998; Felson 2000
	Ethnicity	Zhang 2001; Jordan 2007; Muraki 2009
	Heberden's Nodes (hand OA)	Cooper 2000; Hart 1999
	Metabolic Syndrome (incl. hypertension)	Yoshimura 2012; Hart 1995
	Nutritional Factors	McAlindon 1996
	Genetics	Neame 2004; Spector 1996
	2D:4D Finger Ratio	Zhang 2008
	Psychosocial (depression/fatigue)	Hawker 2011
Local Factors	Obesity	Goulston 2011; Felson 1997; Coggon 2001
	Joint Injury	Roos 2005; Lau 2000
	Occupation	Cooper 2000; Lau 2000
	Meniscectomy	Englund 2004
	Muscle Weakness	Felson 2000; Slemenda 1998
	Joint Biomechanics	Sharma 2001; Lane 2011
	Joint Alignment	Sharma 2001
	Sports and Physical Activity	Felson 1997; Cooper 2000; Hart 1999

2.1.4 Summary

Knee osteoarthritis is a major health burden affecting a significant portion of the older population of the world. The disease process is complex, involving changes to bone, soft-tissue and the presence of clinical symptoms. One of the key questions, is which of these factors initiates the disease process. This research focuses on the bony structural changes, how they relate cross-sectionally to pain, and whether they can predict future complete joint failure.

2.2 Radiographic scoring methods

Once clinical symptoms of knee OA have been identified, radiography is used to establish a final diagnosis of OA, to rule out other diseases, and to monitor disease progression (Buckland-Wright et al. 2003). Clinical trials often use radiographic evidence

as inclusion criteria for subjects (Hochberg et al. 2010), while cohort studies rely heavily on it to establish incidence and progression within a population (Felson et al. 1995; Pavelka et al. 2000). In spite of this reliance on radiographs for all aspects of disease evaluation, the relationship between ROA and pain and disability is not well understood. New diagnostic scoring methods have been developed in order to improve this relationship, as well as to improve sensitivity to disease progression and the general understanding of the natural history of radiographic knee osteoarthritis.

2.2.1 Background of atlas-based radiographic scoring methods

Over half a dozen methods have been developed to score the presence and severity of knee OA found on plain-film x-rays. These categorical methods can be grouped into two categories, the first which evaluates all known radiographic features together as a 'global' score and the second which scores each feature individually. They provide images of each level of disease severity by grade in the form of an atlas, which is a tool used to standardise training and reproducibility between graders. Commonly evaluated features include osteophytes, joint space narrowing, subchondral cysts, subchondral sclerosis, tibial spiking, and attrition.

The oldest method, developed in 1957 by Kellgren and Lawrence (K/L), is a global scoring method which has been widely used since its creation and is still the most common diagnostic score for radiographic knee OA (Kellgren and Lawrence 1957; Schiphof et al. 2011). More recently, scoring methods have been developed that focus on individual radiographic features, primarily joint space narrowing and osteophytes. The other common features listed above are also sometimes included in these scoring methods. The major benefit of these methods is that they allow for analysis of the natural history of individual structural features and knee compartments, and are able to assess them individually against

pain, stiffness and disability. While these methods show promise, they continue to share some limitations with K/L grades as highlighted below.

So many different scoring systems have been developed, often being used only for a single cohort, that they are not able to be accurately compared against one another. Most have been developed using arbitrarily defined grading intervals where the interval between grades 1 and 2 is not equivalent to the interval between grades 2 and 3 (OARSI, Chingford, Framingham, etc.) (Burnett et al. 1994; Felson et al. 1997; Altman et al. 2007). One of the most recent scoring methods, the Line Drawing atlas, has been mathematically derived to have equal intervals between each grade (Nagaosa et al. 2000). It has also emphasised the differences found in joint space between men and women by having sex-specific joint space grades, the only atlas to do so (Nagaosa et al. 2000). The OARSI atlas is the most popular method in use to grade individual features and compartments, and was put forward by Osteoarthritis Research International as an attempt at standardisation, but it has not yet been universally adopted (Altman et al. 1995; Altman et al. 2007).

2.2.2 Kellgren and Lawrence

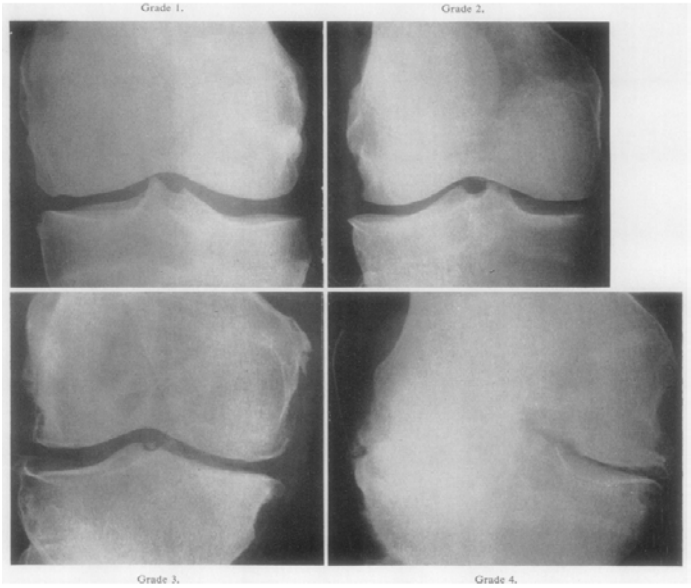
2.2.2.1 Formal grading system

The most widely accepted system of grading was developed by Kellgren and Lawrence (K/L) over a series of articles from the late 1950s through the 1970s. In 1957, they published a landmark article providing the features used to identify osteoarthritis, a grading system based on 'severity' of OA, and a series of x-ray images illustrating osteoarthritic severity for eleven joints (Kellgren and Lawrence 1957). Radiologic features considered to be important for grading included; osteophytes on joint margins or tibial spines, periarticular ossicles, narrowing joint space with subchondral sclerosis, and subchondral cysts. The basic grading system outlined was divided into five grades, from

'none' to 'severe', without any description of which features made up each grade (Kellgren and Lawrence 1957).

Kellgren *et al* formalised this grading system in 1963 with an 'atlas of standard radiographs' and provided further descriptions for the radiographic images (see figure 2.1 below) (Kellgren and Lawrence 1963). They described grade 1 as representing 'doubtful narrowing of joint space and possible osteophytic lipping', grade 2 as 'definite osteophytes and possible narrowing of joint space', grade 3 as 'moderate multiple osteophytes, definite narrowing of joint space, some sclerosis and possible deformity of bone ends', and grade 4 as having 'large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends' (Kellgren and Lawrence 1963).

Figure 2.1 Kellgren and Lawrence atlas images (1957)



Lawrence used a slightly different description of the features of each grade in a later work, with grade 1 being a 'minute osteophyte of doubtful significance', grade 2 as a 'definite osteophyte and unimpaired joint space', grade 3 given as a 'moderate diminution

of the joint space', and grade 4 as having 'greatly impaired joint space and subchondral sclerosis' (Lawrence 1977).

The differences between these two descriptions may not, at first, appear to be significant, but the variation in wording would likely affect which patients are included in each grade. The first notable difference is in grade 1, where the presence of a 'minute' osteophyte is required in Lawrence's 1977 score, while only 'possible' osteophytic lipping is required for Kellgren and Lawrence's 1963 grade 1. Scoring for grade 2 has an even greater discrepancy with 'unimpaired' joint space in Lawrence's description and 'possible' narrowing in the 1963 article. Sclerosis is a requirement in Kellgren and Lawrence's 1963 grade 3, while it is not mentioned until Lawrence's grade 4. These variations may have had an unknown effect on results, and may have hindered the comparability and reproducibility between studies on prevalence, incidence and progression.

2.2.2.2 Alternate definitions

Schiphof et al. looked at the impact of the many variations of Kellgren and Lawrence scoring used in studies (figure 2.2) (Schiphof et al. 2009; Schiphof et al. 2011). They found that in addition to the two versions discussed above, at least three other permutations existed in studies claiming to use the 'original' K/L scoring (Schiphof et al. 2011). The authors even noted discrepancies in different articles using the same cohort (Schiphof et al. 2011). The table below from this article outlines the variations in wording used for each of the K/L grades.

Figure 2.2 Kellgren and Lawrence variations from Schiphof et al. 2011

Definition grades	Original	Alternative 1	Alternative 2	Alternative 3	Alternative 4
Grade 0: No osteoarthritis	No osteoarthritis	No osteoarthritis	No osteoarthritis	No osteoarthritis	No osteoarthritis
Grade 1: Doubtful	Doubtful narrowing of joint space and possible osteophytic lipping	Possible osteophytes	Minute osteophyte, doubtful significance	Possible osteophytes only	Possible osteophytic lipping
Grade 2: Mild	Definite osteophytes and possible narrowing of joint space	Definite osteophytes	Definite osteophytes, unimpaired joint space	Definite osteophytes and possible JSN	Definite osteophytes and possible JSN
Grade 3: Moderate	Multiple osteophytes, definite narrowing of joint space and some sclerosis and possible deformity of bone ends	Osteophytes and JSN	Moderate diminution of joint space (with osteophyte)	Moderate osteophytes and/or definite narrowing	Moderate multiple osteophytes, definite JSN, some sclerosis and possible bone contour deformity (bony attrition)
Grade 4: Severe	Large osteophyte, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends	Large osteophytes, marked narrowing of joint space and definite deformity	Joint space greatly impaired with sclerosis of subchondral bone	Large osteophytes, severe JSN and/or bony sclerosis	Large osteophytes, marked JSN, severe sclerosis and definite bony contour deformity (bony attrition)

*Alternative 1 is used in Williams 2004; alternative 2 is in Jordan 2003, Cooper 1996, Hart 1991; alternative 3 in Zhang 2001; alternative 4 in Wilder 2002 and Williams 2005

Felson et al. compared a modified scale which was more heavily reliant on joint space narrowing than the original K/L and found a high level of agreement in prevalence between the two methods (Felson et al. 1997). This result implies that there also may not be a large difference between those purporting to use the 'traditional' K/L and the 'alternate' definitions, because although the descriptions differ, the accompanying x-ray images used for grading remain the same.

Limitations of the Kellgren and Lawrence scoring system include inconsistency in application, lack of official patellofemoral compartment grading, absence of sex-specific atlas images, its continued weak association with symptomatic OA (that is, not all knees with pain have high K/L scores), relative insensitivity to disease change, and an assumption of a non-validated disease natural history. Applying K/L to both semi-flexed and fully flexed knees can also result in different grades, as the system was not devised for a specific radiographic patient position (Schiphof et al. 2008). Despite these limitations, it has been accepted as the current 'gold standard' by the World Health Organisation for epidemiological studies (Symmons et al. 2002).

2.2.3 Ahlback scoring method

This method was developed by Ahlback in 1968 and is often used in eastern European studies (Ahlback 1968; Petersson et al. 1997). It focuses primarily on joint space narrowing with an expected progression to attrition. The scoring does not capture mild to moderate disease well, but focuses on severe disease through complete joint destruction. Petersson et al. (Petersson et al. 1997) described an Ahlback grade of I as corresponding to a K/L grade of 2 to 3, and of his grade II to a K/L grade 3 to 4. When the two grading systems were used on the same population of subjects with knee pain, 15.1% were identified as having ROA using a K/L grade ≥2 definition, while 10.8% were identified using the Ahlback scoring method with a grade of I or greater (Petersson et al. 1997).

This scoring method shares several of the same limitations as Kellgren and Lawrence, such as evaluation of the whole joint and an assumption of disease progression that has not been validated. It is also only useful for studies looking at moderate to severe disease, as it would not capture any isolated osteophytes present before joint space narrowing.

Figure 2.3 Ahlback scoring method description

Grade I	Joint space narrowing (<3 mm)
Grade II	Joint space obliteration
Grade III	Minor bone attrition (0-5 mm)
Grade IV	Moderate bone attrition (5-10 mm)
Grade V	Severe bone attrition (>10 mm)

2.2.4 Chingford atlas

The Chingford atlas was developed at a time when no definitive atlases existed for analysing individual radiographic features. Ahlback had recommended noting the presence and location of osteophytes and grading size from zero to five, but did not provide images for comparison (Petersson et al. 1997). The Chingford atlas focuses on four individual features; osteophytes, joint space narrowing, subchondral sclerosis and tibial spiking. Osteophytes and joint space narrowing are graded on a scale between zero and three, while subchondral sclerosis and tibial spiking is marked as present or absent (Burnett et al. 1994; Hart and Spector 1995). Unlike K/L, reference images include the patellofemoral compartment with both lateral and skyline views in addition to the medial and lateral compartments (Burnett et al. 1994).

Limitations of the Chingford atlas include: a lack of features including subchondral cysts; the lack of sex-specific atlases for joint space narrowing (as it was developed for use on a cohort of women); and the fact that it is not as widely used as the OARSI atlas which was developed several years later.

Figure 2.4 Osteophytes from the Chingford atlas (Burnett 1993)

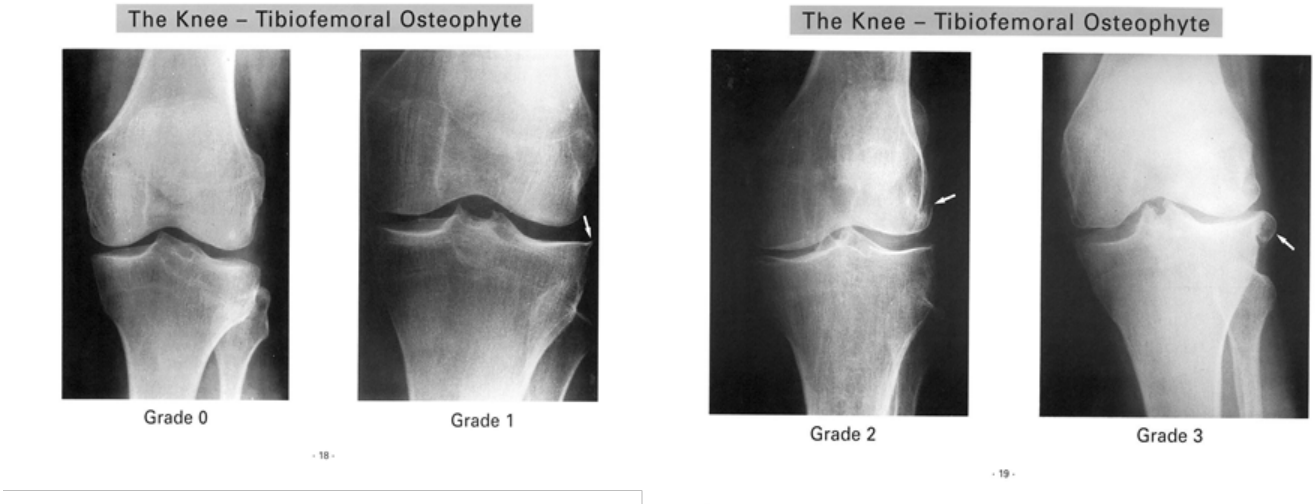
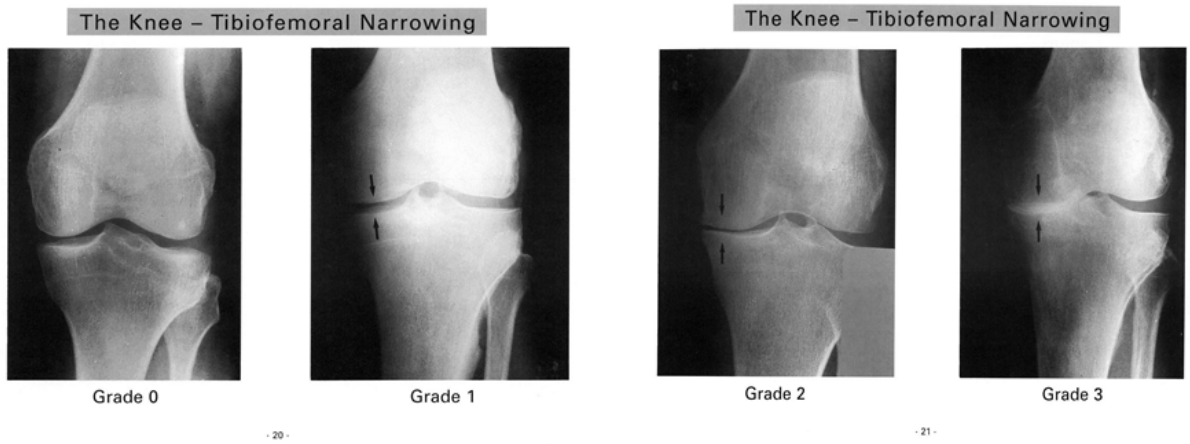


Figure 2.3 Joint space narrowing from the Chingford atlas (Burnett 1993)

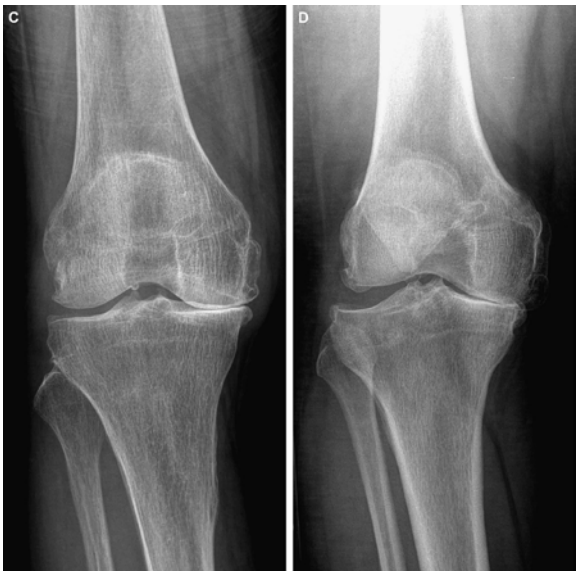


2.2.5 Osteoarthritis International atlas (OARSI)

The OARSI atlas was developed in 1995, for similar reasons to the Chingford atlas, in order to isolate features and compartments to be graded separately, and to create a system usable by all studies. Distinguishing each feature independently was also meant to address the trend in medical research to give more detail and description of the disease process rather than just a simple presence/absence statement (Altman et al. 2007).

The selection of radiographs of the knee includes four areas for marginal osteophytes: medial femoral condyle, medial tibial plateau, lateral femoral condyle, and the lateral tibial plateau. Joint space narrowing is scored from 0 to 3 in both medial and lateral compartments, while attrition is scored only as present or absent in the medial compartment. The final feature, sclerosis, is scored present/absent in both compartments. The original 1995 atlas also included views of the patellofemoral compartment. The OARSI atlas was redone in 2001 due to the first atlas being difficult to locate and the demand by researchers. Some issues with this method include lack of sex-specific atlases, missing subchondral cysts as a gradable feature, no ‘standardised’ definition of OA based on the scoring, and the lack of a patellofemoral view in the current atlas.

Figure 2.6 Osteophyte image from the OARSI atlas



2.2.6 Logically Derived Line Drawing atlas

The Logically Derived Line Drawing atlas was developed in 2000 by Nottingham University in response to issues found while using the OARSI atlas to analyse individual radiographic features. The main problems identified by the authors with the OARSI atlas were: the use of ordinal grades for joint space narrowing; non-interval grading for osteophytes; lack of medial and lateral femoral trochlea views in skyline; magnification differences in radiographic images; and that several features are visible at the same time in images (Nagaosa et al. 2000). They contend that the OARSI atlas is difficult to use, resulting in poor reproducibility between observers (see table 2.2) (Nagaosa et al. 2000).

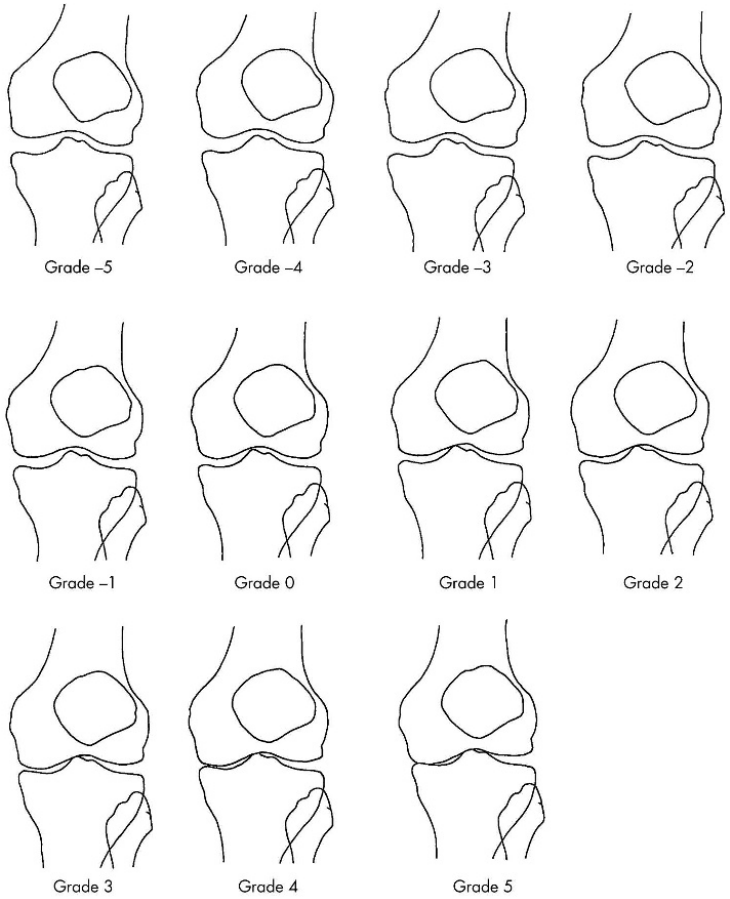
Two features, osteophytes and joint space narrowing (JSN), were selected for representation in the atlas based on observer reproducibility, their identification generally as the two main features of radiographic osteoarthritis, and their ease of being reproduced as line drawings. ‘Normal’ joint space narrowing was identified by viewing several hundred radiographs of asymptomatic men and women. It was noted that the mean joint space varied significantly between men and women, prompting sex-specific joint space line drawings (Nagaosa et al. 2000). JSN was graded from 0 to 3, with 0 being ‘normal’

and three as ‘bone-on-bone’, and each compartment was drawn individually. The other grades were developed in an interval fashion, where grade 1 is two-thirds of grade 0 and grade 2 is one-third of ‘normal’ joint space.

Osteophytes were selected in a similar fashion, where a sample of over seven hundred radiographs of symptomatic patients was analysed for the largest osteophytes as well as the most common shape and direction at each site. Two sites (the lateral tibial plateau and medial femoral trochlea) had more than one common shape/direction of osteophytes, so a second set of drawings was included for these sites (Nagaosa et al. 2000). The largest osteophyte identified was represented as grade 3, with grades 1 and 2 being one-third and two-thirds of grade 3 respectively.

Due to the high level of reproducibility and validity of the pilot version of the atlas (see Nagaosa 2000 results in table 2.2), a modified version incorporating additional grades was tested (Wilkinson et al. 2005). Negative grades were added in for joint space to capture pseudo-widening, bringing the possible grades to between -5 and 5, and osteophyte scores were increased to have six possible grades (0 to 5). Inter-observer reproducibility was found to be the same despite the additional grades (see Wilkinson 2005 results in table 2.2), and identifying joint space using the expanded negative grades was found to be quicker than with the previous versions of the atlas (Wilkinson et al. 2005).

Figure 2.7 Joint space narrowing image from the Line Drawing atlas **(**Wilkinson 2005)



2.2.7 Less commonly used scoring methods

In 1991, Brandt developed a six point, semi-quantitative scale where each grade represented a percentage of 'normal' joint space; i.e. grade 0 equalled less than 25% of normal joint space narrowing. The system also took other radiographic features into account by including an alternate definition where either joint space narrowing *or* less joint space narrowing plus a secondary feature (subchondral sclerosis, osteophytes, etc.) needed to be present (Brandt et al. 1991).

An atlas of individual radiographic features was developed for specific use in the Framingham knee osteoarthritis study (Felson et al. 1995; Felson et al. 1997). Anteroposterior and lateral images were graded for four features; osteophytes, JSN, and sclerosis on a 0 to 3 scale with subchondral cysts marked as present or absent (Felson et al.

1997). Images for AP joint space narrowing and osteophytes were published, but a full atlas was never made widely available, leading to a restricted use of this system. It was, however, noted by the authors that their grade 2 osteophytes correspond to the grade 1 osteophytes found in both the Chingford and OARSI atlases (Felson et al. 1997).

2.2.8 Reproducibility and validity of the scoring methods

In order for the grading systems to be useful diagnostic tools, they must be easy to use, have high inter- and intra-observer reproducibility, be sensitive to disease progression, provide construct validity for pain and function, be valid for a variety of radiographic protocols, identify different modalities of disease, diagnose disease within each compartment (medial, lateral and patellofemoral), be gender specific, and ultimately associate with end-stage pathology and disability. It is no wonder, with such a list of desirable traits, that the perfect scoring method has yet to be found despite half a century of attempts.

The most basic and important requirement of any diagnostic tool is to be reproducible; that is, if a measurement is repeated using a scoring system, it must give the same result regardless of the person measuring or the time interval between measurements. Otherwise, the differences found between patients or between years will be the result of measurement bias rather than indicating a meaningful change in disease status.

Being reproducible provides a precise measurement or score, but does not necessarily imply that the score provides any useful information to a researcher. The tool must also be accurate in identifying disease prevalence, progression or incidence. This involves having both construct validity and predictive validity. Construct validity evaluates whether the scoring method correlates to external variables (such as pain and disability), and predictive validity evaluates whether the method correlates to a ‘gold-standard’ outcome measure after a follow-up period (Reijman et al. 2004). Without construct

validity with the disease being measured, precision is meaningless. For most knee osteoarthritis radiographic scoring methods, construct validity was tested using accepted clinical aspects of osteoarthritis such as the presence of pain and disability.

Most of the scoring methods previously evaluated solidly address the above two issues of reproducibility and construct validity; however, the predictive validity of radiographic methods has not been comprehensively evaluated. The reproducibility and construct validity of each of the methods is described below and summarised in table 2.1.

2.2.8.1 Kellgren and Lawrence

Despite the many criticisms aimed at the K/L score, it has been found to have very high intra- and inter-observer reproducibility (Cooper et al. 2000; Neame et al. 2004; Thorstensson et al. 2008) and continues to be the single most-used radiographic scoring system for osteoarthritis.

Validation studies have compared K/L grades to pain, the amount of cartilage loss seen on MRI, evidence of arthroscopic surgery, and whether there is a correlation between K/L grades and future knee replacements. While there remains a large gap between the radiographic presence of OA and the occurrence of symptomatic OA, higher K/L grades have been shown to be correlated with pain (Duncan et al. 2007; Muraki et al. 2009; Neogi et al. 2009). It has also been validated against both cartilage loss and defects in magnetic resonance imaging (Hayes et al. 2005; Hernandez-Molina et al. 2008).

2.2.8.2 Chingford and OARSI atlases

Intra- and inter-observer reproducibility was found to be good for all features (kappa≥0.6) except inter-observer joint space which was only acceptable (kappa≥0.5) (Spector et al. 1993). The feature with the strongest association with pain was medial osteophytes (53.4%). The features with the weakest construct validity with pain were the joint space measurements, particularly in the lateral compartment (Spector et al. 1993).

The OARSI atlas is widely used and is generally good for both intra- (kappa≥0.6) and inter-observer (kappa≥0.5) reproducibility (table 2.1). Validity was assessed for OARSI features by looking at their association with pain. It was found that increasing osteophyte grade and JSN grade are both associated with increasing knee pain (Neogi et al. 2009; Lanyon et al. 1998).

2.2.8.3 Line Drawing atlas

Reproducibility for both osteophytes and JSN was found to be good using the Line Drawing atlas, with patellar osteophytes as the feature with the worst reproducibility overall (Nagaosa et al. 2000). The Line Drawing atlas has been tested for both face validity and content validity (Wilkinson et al. 2005). Predictive validity was examined in a retrospective cohort where knees with baseline osteophytes≥1 had a hazard ratio of 2.5 and JSN≥1 had a hazard ratio of 2.2 for knee pain 12 years later compared to subjects without osteophytes and JSN at baseline (Ingham et al. 2011).

Table 2.2 Published reproducibility of atlas-based ROA scoring methods

Method	Studies	N	Population	X-ray View	Intra-Observer (JSN/Ost)	Inter-Observer (JSN/Ost)
K/L	*Kellgren /Lawrence 1957*	85	age 55-64	extended	0.83[+]	0.83[+]
	Spector 1993	100	women 45-65	extended	>0.66*	>0.56*
	Gunther 1999	100	ortho. clinic	extended	>0.85[‡]	>0.76[‡]
	Cooper 2000	20	age >55	extended	0.71	0.58
	Neame 2004	20	age>40	extended	>0.69*	
	Gossec 2008	50	symptomatic	semi-flexed	0.50	
	Gossec 2008	50	symptomatic	extended	0.61	0.56
	Neogi 2009	n/a	age 50-79	fixed-flexion		0.90
OARSI	Lanyon 1998	40	age >40	extended		>0.7/>0.7*
	Nagaosa 2000	50	age>40	extended	>0.75/>0.57	>0.49/>0.49
	Gossec 2008	50	symptomatic	semi-flexed	0.67/	
	Gossec 2008	50	symptomatic	extended	0.71/	0.48/
Chingford	*Spector 1993*	100	women 45-65	extended	>0.74/0.84	>0.49/>0.64
	Cicuttini 1996	50	twin study	extended	>0.88/>0.94	>0.60/>0.88
	Gunther 1999	100	ortho. clinic	extended	>0.86/>0.81[‡]	>0.24/>0.65[‡]
	Cooper 2000	20	age >55	extended	0.65/0.71	0.60/0.56
Line Drawing	*Nagaosa 2000*	50	age>40	extended	>0.76/>0.60	>0.56/>0.46
	Neame 2004	20	age>40	extended	>0.69/>0.69*	
	Wilkinson 2005	121	age>40	extended	>0.81/>0.71	>0.65/>0.64

All methods used weighted kappa statistic unless otherwise indicated: *unweighted kappa, + correlation coefficient, ‡ Intra-class correlation. Studies in italics indicate the 'official' reported reproducibility

2.3 Quantitative joint space

Joint space width on radiographs is often used as a surrogate outcome marker for knee cartilage loss as a result of osteoarthritis. The x-rays show the delimitations of the bony edges of the joint and any space in between is assumed to represent cartilage (or lack thereof).

Quantitative methods measure the actual space between the articular surfaces of the bones in millimetres. The measurement is made independently in each compartment from the distal edge of the femur to the sclerotic line indicating the floor of the tibial plateau.

This measurement is referred to as the inter-bone distance. Previously this measurement was made by hand on plain-film x-rays using either specially designed callipers or rulers (Lequesne et al 1994; Ravaud et al 1996), but more recently semi-automated methods designed for use on digitised plain-film x-rays or digital x-rays have been developed. These methods are described in detail in chapter 4.

The measurement of quantitative joint space is extremely important, as change in minimum joint space is currently one of the primary outcomes for testing disease-modifying osteoarthritis drugs (Abadie et al. 2004), and has also been suggested as an early indicator for pain and future knee replacements (Bruyere et al. 2005).

2.4 Radiographic joint space narrowing as a marker of cartilage

Several issues exist with the use of joint space width as a surrogate for knee cartilage loss in the progression of osteoarthritis. First, the use of this measurement as a surrogate assumes that articular cartilage is the only soft tissue that influences joint space width. Research has shown that meniscal subluxation results in joint space narrowing in cases where cartilage loss is not evident on MRI (Gale et al. 1999). Joint space loss greater than 3mm due to meniscal subluxation was seen in 81% of cases (with knee osteoarthritis) and 64% of the controls in research comparing MRIs to plain radiographs (Gale et al. 1999). While meniscal subluxation has been shown to be correlated with osteoarthritis, this research demonstrates that cartilage loss is not the only factor in joint space narrowing.

An additional issue with the use of joint space width or narrowing is the effect of patient positioning during the radiograph. The extended knee position is one of the older standardised methods, and is still used for some ongoing cohort studies. Some research has shown this method to lead to an overestimation of joint space due to the variability of knee positioning (Buckland-Wright 1999). The two methods currently used to correct somewhat

for these issues are fluoroscopic semi-flexed positioning and fixed flexion. Magnification error due to variable patient positioning is an important issue for quantitative measurement of minimum joint space and will cause inaccuracies for all measurements made without having a standard reference point (e.g. a standardised steel ball) in all radiographs. Differences in joint space width have also been found between weight-bearing and non-weight-bearing x-rays, particularly in the lateral compartment (Buckland-Wright et al. 1995(B)).

2.5 Natural history of radiographic knee osteoarthritis

2.5.1 Introduction

The natural history of radiographic knee OA has been assessed by approximately a dozen studies (table 2.3) with various study designs, inclusion criteria and lengths of follow-up. The studies can be divided into two main groups; those that used some aspect of pain as inclusion criteria (or recruited from rheumatology clinics) and those that used community-based cohorts. The former, as would be expected, show a higher percentage of prevalence, incidence and progression of ROA than the latter. The majority of these studies use Kellgren and Lawrence (K/L) as the primary method of radiographic scoring excepting two, one of which uses the Ahlback scoring method (Hernborg 1976), and one which uses only individual features such as osteophytes and joint space narrowing (Hart et al. 1999).

2.5.2 Natural history of ROA in the community

The longest natural history study conducted using a community-based cohort to date had 12 year follow-up (Schouten et al. 1992) and many of these studies have relatively elderly cohorts at baseline (mean age 70 or above) (Felson et al. 1995; Felson et

al. 1997; Cooper et al. 2000). There is a marked lack of long-term natural history studies that examine the incidence of mild/moderate ROA in younger subjects, regardless of pain status or known risk factors for pain.

2.5.3 Natural history of ROA in painful knees

Analyses conducted on primarily symptomatic cohorts have found a range of figures for incidence, progression and worsening ROA (essentially defined as incidence plus progression). Rates were 7.2% for incident ROA (Thorstensson et al. 2008), 4-8.8% for progression (Pavelka et al. 2000; Thorstensson et al. 2008), and between 3.3 and 7.7% for yearly worsening (Massardo et al. 1989; Spector et al. 1992). As would be expected, these yearly rates were found to be smaller in population-based cohort studies and case-control studies which included both symptomatic and non-symptomatic subjects in the analysis. The results were between 1.9 and 3.6% for yearly incidence (Felson et al. 1995; Felson et al. 1997; Cooper et al. 2000), 3.5-4% for yearly progression (Felson et al. 1995; Cooper et al. 2000) and 4.4% for worsening (Cooper et al. 2000).

Table 2.3 Studies of the natural history of radiographic knee osteoarthritis

Author	Year	Length	N	Age	Inclusion	Methods	Risk Factor	Results
Thorstensson	2008	12 years	143	35-54	Knee pain >3 months	Extended at baseline, Semi-flexed at year 12	OA= KL>1	85.5% had incident TF OA 97.0% had progressive TF OA 31.0% had incident PF OA
Lachance	2002	3 years	679	42-52	Population-based >=40 years	Extended Blinded to series	OA= KL>1 OA=KL>2	Baseline grade 1s OR 2.2 to have incident ROA than grade 0. Age and BMI risk factors for progression
Felson	1995	8-10 years	869	mean 71	Population-based	Extended, weight bearing	OA=KL>2	Incident OA was 1.7 times higher in women then men Progression was 1.4 times higher In women, 2% yearly incident ROA, 1% incident symptomatic 4% had progressive ROA
Schouten	1992	12 years	142	46-68	Population-based >=45 years	Extended, weight bearing	OA=K/L>2 JSN= binary loss (yes/no)	34% lost cartilage over 12 years Women had cartilage loss less often than men, but when present was more severe
Spector	1992	11 years	63	mean 58	Rheum Clinic Clinical OA of hand or knee	AP, non-weight bearing	OA=K/L>1 Osteo>1 JSN>1 Global -4 to 4	19-42% showed progression depending on scoring method Global score=50% worsening Same progression between baseline K/L 0,1 and 2 More progression in subject with pain except for JSN

Pavelka	2000	5	139	mean 59.1	Age >= 40 Pain + JSN and/or Osteo and/or Scler.	weight-bearing AP	OA=K/L>2 Quant. JSW	Subjects with K/L 2+ at baseline showed higher JSW progression JSW change is phasic - quickest over 1st year then slower 19% of subjects progressed in K/L over 5 years JSW more sensitive to change then K/L
Hornberg	1976	10-18	71	mean 63	TF sclerosis Osteophytes	AP weight-bearing Lateral with 45° flexion	Ahlback	Only 71 patients out of 2195 had sclerosis but no osteophytes Of 80 knees, 60 remained medial Only knees w/both med and lateral were worse
Massardo	1989	8	31	mean 71.7	Symptomatic knee	AP	OA=JSN + Sclerosis	Medial OA showed most change 38.5% were unchanged 61.5% had more signs of OA 2 subjects with improved symptoms had more structural change
Felson	1997	8.1	598	mean 70.5	ROA free	Weight-bearing AP lateral 45° flexion	OA=modified K&L 2+ Osteo>1 JSN>2	15.6% developed incident OA Women OR of 1.8 BMI OR 1.6 per 5 kilo increase Weight change OR 1.4 per 10 lb High physical activity OR 3.3
Hart	1999	5	715	54.1 +-5.9	Age>45	Weight-bearing AP	OA=JSN>1 and/or Osteo>1	3.1% yearly incidence of JSN 3.3% yearly incidence of osteo. Subjects who developed incident osteophytes were more likely to be older, heavier, have hand OA and knee symptoms Subjects in top tertile of BMI had OR 2.38 for incident osteophytes

Cooper	2000	5	354	mean 70.7	Case-control Pain and pain free	AP weight-bearing	OA=K&L 1+ and 2+	Incidence and progression were 2.5% and 3.6% yearly Risk of incident ROA: Highest tertile BMI OR 18.3 Previous Injury OR 4.8 Sport participation OR 3.2 ROA progression: Knee pain at baseline OR 2.4 Heberden's nodes OR 2.0

2.6 Outcome assessment methods – pain and total knee replacements

2.6.1 Introduction

While the presence of pain is a key aspect in the disease process of knee osteoarthritis, the cause of pain is not well understood. The major impediment to this understanding is that the primary tissue involved in the disease process, cartilage, is not innervated with pain nociceptive fibres. Other features involved in the disease pathogenesis, such as osteophytes and sclerosis, also do not have pain fibres, yet have also been frequently found to have an association with pain. Osteophytes may be inducing pain by impinging on other joint structures (Sofat et al. 2011). MR imaging has identified other features with more possibilities of a direct relationship with pain, such as bone marrow lesions (Felson et al. 2007), effusion, bursitis, meniscal lesions and synovitis (Keen et al. 1997; Sofat et al. 2011).

Osteoarthritis pain is difficult to assess due to its episodic nature and worsening with movement and weight-bearing. A few of the possible sources of osteoarthritic pain include periosteal damage by osteophytes, raised intraosseous pressure, microfractures, ligament damage, capsular tension, meniscal injury and synovitis (Creamer 2000). Anxiety, depression and other co-morbidities are also thought to be related to the presence and/or severity of OA pain (Dieppe et al. 2000; Hawker et al. 2008). Alternatively, subjects who report more and/or frequent knee pain may be influenced by central sensitisation (Soni et al. 2012).

Pain is the primary feature of clinical knee osteoarthritis. Some other common features assessed clinically (and specifically mentioned as part of the ACR criteria), are stiffness lasting less than 30 minutes, bony enlargement and local effusion. For research

purposes, pain is often used as the only non-radiological feature to distinguish symptomatic OA from radiographic OA.

2.6.2 NHANES

The most commonly used assessment in identifying painful knees for research comes from a large U.S. cohort study, the National Health and Nutrition Examination Survey (NHANES) which began in the 1960s and included comprehensive questionnaires and physical examinations on all aspects of health. The 1970s examinations included the knee pain question, 'have you ever had pain in or around a knee on most days for at least a month?' (NHANES 1979). This question was then used in a major U.S. prospective cohort, the Framingham study (Felson et al. 1987). In 1999, for NHANES III, an additional question was asked; 'have you had pain in or around your knee for at least a month in the last year?' This question, or a variation of it, is now practically a standard question for cohort studies, and is often used to divide subjects into pain and pain-free groups (Dillon et al. 2006).

The use of a simple binary pain question has proved useful, and therefore variations of this question have developed, often to modify the length of recall, from the more general 'have you ever had pain for more than fifteen days?' (Lethbridge-Cejku et al. 1995) to the more restrictive, 'have you had pain for most days in the last three months?' (Petersson et al. 1997). A study in 1996 looked at how these questions affect the resulting reported pain prevalence (O'Reilly et al. 1996). The authors compared three different questions asked to the same subjects and found that reported pain levels varied by 10%, depending on which question was asked. Even repeating the same question within the same population within a two week time span revealed a 10% response difference (Felson et al. 1987). The use of these questions about knee pain is the simplest and most commonly used method of determining symptomatic subjects within a cohort. Trying to

capture incident pain and worsening pain in longitudinal cohorts is extremely difficult with these types of questions, so further scoring methods have been developed.

2.6.3 Total knee replacements

Knee replacements have been described as the 'single most relevant outcome in knee osteoarthritis' (Bruyere et al. 2005). They represent complete joint failure at the end of the disease process where both the symptoms and structural changes are judged to be the most severe. Because of its usefulness as a 'hard' outcome, change in joint space over 3 years has been recommended as a surrogate marker for having a total knee replacement (TKR), as TKRs are rare and require much longer follow-up (Abadie et al. 2004).

Only a handful of studies have evaluated baseline ROA or worsening of ROA against the risk of having a future TKR. Bruyere et al. found that 3-year change of minimum joint space width but not mean joint space width was highly predictive of having a TKR within the next five years (Bruyere et al. 2005). In a post-hoc analysis of a clinical trial, neither baseline nor medial minJSW was found to be predictive of a TKR within the next 5 years (Raynauld et al. 2011).

The limitations of using TKRs as a clinical outcome (rather than pain) include the variability in access and desire for surgery, and that it is highly dependent on local medical and surgical practices.

2.7 Relationship between ROA and pain

A systematic review published in 2008 looked at the discordance between symptomatic and radiographic knee OA and found that studies reported that between 15 and 76% of painful knees had ROA and between 15 and 81% of ROA knees had pain (Bedson and Croft 2008). They identified three major issues which they felt contributed to

the extreme variation of disease prevalence between studies. First, the number of subjects used as well as the number of radiographic views (i.e. AP only); second, the definition of pain and radiographic OA that was used; and third, the type of study population (Bedson and Croft 2008).

The table below (table 2.4) displays all of the known studies that examine the relationship between pain and radiographic knee OA, cross-tabulated by radiographic scoring method across the top and pain assessment down the left side. From this table, the variation of pain assessment and radiographic scoring used by these studies is clear.

The most commonly used definition for radiographic OA continues to be K/L, with 19 out of the 23 studies using K/L. Roughly half of these employed an additional scoring method using individual feature grades (i.e. OARSI). The pain definitions are more difficult to group, with the most common question, NHANES III 'have you had knee pain for the most days of a month in the last year?' used by only seven of the studies. The next most commonly used pain assessments also come from NHANES, with 'have you had pain for the most days of one month ever?' and '…most days in the last month?' A few of the more recent studies have used the WOMAC score in addition to/ instead of the variations of NHANES (Brandt et al. 2000; Williams et al. 2004; Duncan et al. 2007; Neogi et al. 2009).

A major limitation of many of these studies is the lack of either a skyline or lateral radiographic view for the patellofemoral compartment. Work as far back as the early 1990s (McAlindon et al. 1992) has emphasised the association between pain and radiographic patellofemoral osteoarthritis. To emphasise the necessity for the inclusion of this compartment, Duncan et al. looked at 777 subjects with knee pain and imaged them using AP, skyline and lateral views. They found a prevalence of 68% ROA when all three views were used (Duncan et al. 2006). When only the subjects with ROA (in any view)

were analysed (n=531), 57% only had ROA in the AP view, while 87% were identified with the addition of either a skyline or lateral view (Duncan et al. 2006).

As indicated previously, a slight wording difference in a single pain question can cause a variation in up to 10% (O'Reilly et al. 1996), so it is no surprise that utilising different pain assessment methods as well as different radiographic scoring methods could cause such disparity in results.

An additional limitation is that the majority of studies assess the cross-sectional relationship between ROA and pain only. A 12-year longitudinal study of a general population found that subjects with a K/L grade$\geq$ 1 at baseline were over three times more likely to develop pain over the course of the study compared to those with a K/L grade 0 at baseline (Ingham et al. 2011).

Table 2.4 Comparison of ROA and pain discordance literature

	K&L*‡	Modified K&L	OARSI	Line Drawing	Spector	Ahlback
Any Last Month	Odding 1998‡ Lachance 2001					
15+ Last Month					Cicuttini 1996	
Most Days Last Month	Lethbridge-Cejku 1995 Muraki 2009 Cooper 2000 Hart 1991	Neogi 2009	Neogi 2009 Muraki 2009			
Most Days Last 3 Months	Petersson 1997					Petersson 1997
Any Last Year	Duncan 2007 Peat 2007		Peat 2007		Peat 2007	
15+ (1 mo.) Last Year					Cicuttini 1996	
Most Days (1 mo.) Last Year - NHANES III	Lethbridge-Cejku 1995 Maruki 2009 Cooper 2000 Felson 1997 McAlindon 1992 Davis 1992*	McAlindon 1992	Maruki 2009 Lanyon 1998		Cooper 2000	McAlindon 1992
15+ Ever	Lethbridge-Cejku 1995					
Most Days (1 mo.) Ever - NHANES I	Lanyon 1998 Hannan 2000	LaValley2001 McAlindon 1992	Lanyon 1998	Ingham 2011		McAlindon 1992
Current Pain	Dieppe 1997 Claessens 1990					
WOMAC (Pain)	Williams 2004 Duncan 2007 Brandt 2000	Neogi 2009	Neogi 2009			

*Used K&L 1+ as additional definition
‡Used K&L 1+ as additional definition

2.8 Summary

This literature review highlighted important areas of research that need to be explored in the relationship between radiographic knee osteoarthritis and pain. Although different definitions of pain have been directly compared to one another, the same has not yet been done for structural definitions of knee osteoarthritis. The lack of research into the long-term natural history of radiographic OA is noticeable especially in regards to population-based cohorts and/or asymptomatic subjects, with the longest reported study being 12 years in subjects with long-term knee pain. The long-term association between change in joint space and future pain or the need for future knee replacement also needs to be assessed.

3 METHODS AND MATERIALS

3.1 Introduction

The methods and study designs for the four major research objectives of this thesis (reproducibility, natural history, construct validity and predictive validity of radiographic scoring methods) are described in this section. It will outline the Chingford study, the previous data collected over the duration of the cohort, the new data collected specifically for this research, and the study design and statistical methods used in each analysis.

The Chingford Women's study is a population-based UK cohort uniquely positioned to fully address all objectives outlined in this research, particularly looking at early or mild radiographic predictors of knee pain and joint replacement due to OA in a normal population. This cohort is unique in that it followed a middle-aged group of women rather than an elderly population of the kind usually targeted for this type of research. It also has an extremely high retention rate of over 50% over the 19-year follow-up. This cohort is one of the few with the type of data able to assess radiographic natural history over 19 years and to answer questions regarding early disease detection and identification of subjects at risk for severe OA later in life.

This research used previously collected data from earlier visits, quantitative joint space from a novel image analysis program, and some re-analysis of x-rays from earlier study visits. New data collection included using a variety of radiographic scoring methods at year 20 and using a novel measure of joint space on the year 20 x-rays in addition to a sub-sample of baseline and year 5 x-rays. Several different study designs (complete-case analysis, cross-sectional and nested case-control) were used throughout this research in order to address each individual research question depending on the specific risk factors and outcomes being analysed.

3.2 Subjects

Subjects were selected from the Chingford study, a prospective 20-year population-based longitudinal study focused on osteoporosis and osteoarthritis. All women between the ages of 43 and 65 from a single GP practice in Chingford, north London, were contacted to take part between 1988 and 1989. 1003 women out of a possible 1353 (78% response rate) attended the baseline visit with an actual age range of 44 to 67 due to the two year recruitment period (Leyland et al. 2012). Subjects of the Chingford study are a well-described, predominantly Caucasian cohort, who have been found to be representative of women of the general UK population in terms of height, weight, and rates of hysterectomy, but with a lower percentage of current smokers (Hart et al. 1999). The study was approved by the Outer North East London Research Ethics Committee and written consent was obtained from each woman.

3.3 Data collection

3.3.1 Population demographics

Each visit year included a combination of detailed musculoskeletal questionnaires, physical evaluation, blood and urine collection and x-rays, depending on the visit year. Yearly phone interviews were conducted between clinic visit dates in order to gather information on medication, fractures and recent co-morbidities.

Musculoskeletal questionnaire included questions regarding pertinent risk factors for osteoarthritis such as age, current medication, menarche/hysterectomy dates, smoking/alcohol intake, family history of arthritis, location and duration of knee pain, stiffness and swelling, fracture history, and physical activity. Total knee replacements (TKRs) were confirmed using a variety of sources including radiographs, nurses' reports

during physical examinations, self-report by the subject on questionnaires, and confirmation by the subject's general practitioner.

Pain was evaluated by questionnaire at baseline and year 2 clinic visits using the questions 'Have you ever had any knee pain in your knees for at least one month?' (NHANES 1979) and 'Have you had knee pain within the last month?' Clinic visits starting at year 3 added a side specific (left/right) component to the knee pain question 'Have you ever had pain or stiffness in this (knee) joint?' in addition to asking about duration of pain, 'Does it last less than 15 days, 15-30 days or more than a month?' Also included in this version were questions specifically related to knee pain during the previous month; 'Did it occur in the prior month?' and 'How many days did it last in the prior month?' The year 20 visit included the most comprehensive set of knee pain questions which included all the questions given in years 3-15 (NHANES) in addition to both the WOMAC (Bellamy et al. 1986) and Oxford Knee scores (Dawson et al. 1998).

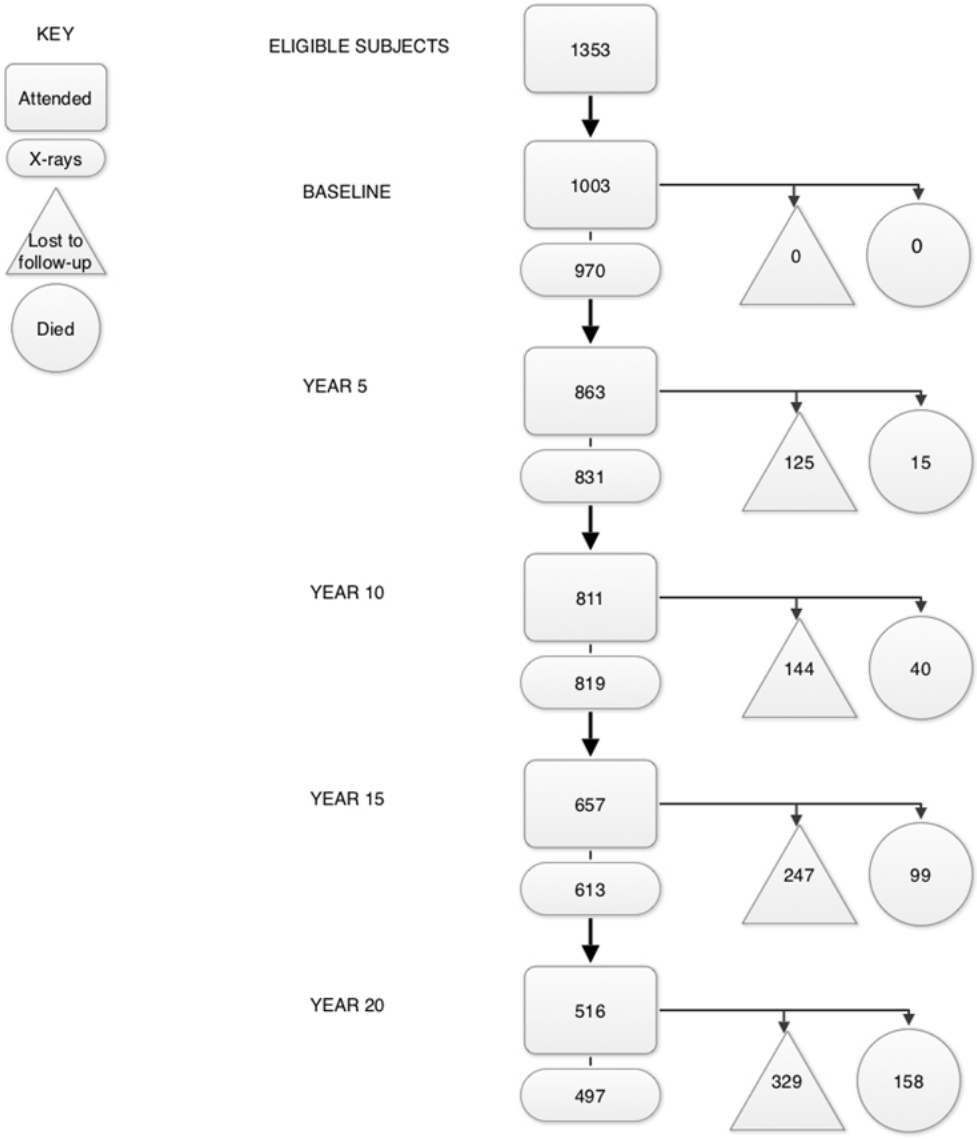
Height and weight were collected by medical staff (from which body mass index (BMI) has been calculated) in addition to blood pressure and serum samples (urine/blood) at each of the clinic visits. BMI groups were calculated using World Health Organisation categories, with normal BMI as less than 25, overweight as 25 to less than 35, and obese as 35 and above.

Some clinic years included more comprehensive examinations with physiotherapy evaluations (grip strength, get up and go test, etc.) in addition to quantitative sensory pain testing (QST). Clinical examination for knee pain included assessment for bony swelling, joint effusion, warmth, tenderness and pain on motion (flexion).

3.3.2 Twenty-year follow-up of the Chingford cohort

Figure 3.1 shows a consort diagram illustrating the attendance, x-rays, mortality and the subjects lost to follow-up over the twenty year study. Mortality (in the circles) increased over time along with the ageing cohort. There was a disparity of up to 6.7% of subjects who attended a study visit, but did not have an x-ray taken for whatever reason.

Figure 3.1 Flow diagram for twenty-year follow-up of the Chingford study



3.3.3 Radiographic risk factors: x-ray protocols and scoring

3.3.3.1 X-ray protocols

Bilateral weight-bearing extended anteroposterior (AP) x-rays were taken at baseline, year 2, year 3, year 5, year 10, year 15 and year 20, with year 20 being the first year of digital x-rays. Standard protocols were established at baseline and continued for all subsequent visits. Both knees for every subject present for the visit were radiographed by experienced radiographers using the same equipment each year. For the baseline to year 15 plain film x-rays, the knee was kept in contact with the cassette and the patella was centered over the lower portion of the femur. The tube-to-film distance was 100cm, with the beam centered 2.5cm below the apex of the patella (Hart et al. 1999).

Digital x-rays were taken at year 20 with protocols as similar as possible to earlier visits, and included a KIDA (Marijnissen et al. 2008) calibration object. Plain film x-rays were scanned on a digital scanner at 600 dpi with a grey scale pixel depth of 16 bits. Pixel size for the digital x-rays was determined using embedded DICOM information.

3.3.3.2 Previously collected data: baseline to year 15

Baseline and year 5 radiographs were read by Prof. Tim Spector (TS) and Dr. Deborah Hart (DJH), with the year 10 and 15 radiographs read by DH alone. Radiographs were read individually (not in sequence) and were blinded to all clinical information (other than study number and visit).

Radiographs were scored using the Kellgren and Lawrence global score ('0= normal; 1= possible osteophyte, no joint space narrowing (JSN); 2= definite osteophyte, possible joint space narrowing; 3= multiple osteophytes, definite JSN, sclerosis and possible deformity of bone ends; 4= large osteophytes, marked JSN, severe sclerosis and definite deformity of bone ends') (Kellgren and Lawrence 1952, Kellgren and Lawrence 1963).

The Chingford atlas (see chapter 2) was also used to score each radiograph. This has scores ranging from 0 to 3 for osteophytes and joint space narrowing, and has a binary score (present/absent) for tibial spiking and subchondral sclerosis (Burnett et al. 1994).

3.3.3.3 New data collection: year 20

3.3.3.3.1 Blinding

Year 20 radiographs were scored by a single reader, Kirsten Leyland (KL) and blinded to all clinical information except study visit and study number. X-rays were read individually and not in a series to keep in practice with reading methods for x-rays from previous study visits. Although this method is not common, it has been used for similar studies (Lachance et al. 2002). Due to the number of years between each x-ray (at least four), it was also determined that progression would be identifiable, which is a primary concern of not reading radiographs in a series (Felson and Nevitt 2009). The regression of grades, which is also a consideration with this type of blinding method, was found to be low (1.7%).

3.3.3.3.2 Radiographic scoring methods

X-rays were scored using the same Kellgren and Lawrence (K/L) grading definition as outlined above in chapter 2. The traditional cut-off for K/L ($\geq$2) was used in analyses unless otherwise stated. This is the cut-off originally indicated by Kellgren and Lawrence (1957) and allows results to be comparable to other epidemiological analyses. Knees were also scored using K&L on a compartmental basis, although this method is not common and is relatively untested. This score was included because although it has only been referenced in a single epidemiological study (McAlindon et al. 1992) and several biomechanical studies (Khan et al. 2008 and Ersoz et al. 2003), it is currently being used as

part of recruitment inclusion criteria for large randomised control trials (NIH clinical trials, 2010).

The Chingford radiographic atlas was used to grade year 20 in order to keep consistency with previous study years although this is not currently a common method for scoring radiographs. Tibial spiking was graded using this atlas, although it was not read for previous x-rays. Although very similar to the Chingford atlas, the OARSI atlas (Altman et al. 2007) was also used to grade the year 20 radiographs due to comparability with other research. Osteophytes (femoral and tibial) and joint space narrowing were graded between 0 and 3, with a binary grade for subchondral sclerosis.

The Line Drawing atlas was used based on the revised version with scores ranging from 0 to 5 for osteophytes and -5 to 5 for joint space width (Wilkinson et al. 2005)). A modified version of this latter score was also used in some analyses, as testing showed that the joint space narrowing grades between -5 and -1 were not as reproducible as when all 'wide' joint space grades were grouped into a single grade (-1) (Wilkinson et al. 2005). Due to the Chingford study using extended protocol x-rays, it was necessary to increase the 'normal' range by one grade to -2 and below.

A manual combined atlas was created for use by the readers, combining all methods of scoring with all available images and descriptions for each method (appendix 1.2). Additional notes and clarification for each grade and/or method were included when any parts were determined to be unclear during the inter-observer training test (see chapter 5).

3.3.3.4 KneeMorf data collection: year 20

KneeMorf (described in detail in chapter 4) is a software program which was developed by the University of Oxford for the purpose of providing an efficient and accurate way to gather quantitative data on a variety of measurements commonly used in

the research of musculoskeletal conditions relating to the knee. Specifically for this research, it was developed to provide a novel way to measure quantitative joint space and identify flexion on older low-contrast radiographs, while providing as much information as newer automated edge detection methods. KneeMorf is also able to record alignment and tibial varus among other factors.

All year 20 x-rays were read by KL using the KneeMorf software package. A nested sample of baseline and year 5 x-rays were also read for quantitative joint space. Minimum joint space, mean joint space and joint space area from both the medial and lateral compartments were all analysed as possible risk factors. Flexion variables were used as covariates during analysis.

3.3.4 Covariates

Adjustment for covariates were done to correct for any imbalances in baseline variables that are related to the outcome (pain or TKR) in addition to the possible adjustment for large differences detected between the groups being tested. Covariates included age, BMI (baseline and year 20), and smoking. While age and BMI are well-established risk factors for osteoarthritis (Garstang and Stitik 2006; Blagojevic et al. 2010), smoking has been more controversial due to conflicting results. A met-analysis in 2011 found that while there was an overall negative association between smoking and OA, in subgroup analyses only case-control studies maintained a significant negative association (cohort and cross-sectional sub-groups were not significant) (Hui 2011). Due to the use of several study designs in this thesis (case-control, cross-sectional and longitudinal), smoking was used as a covariate for all relevant analyses in spite of a possible lack of association with the outcomes.

3.3.5 Outcomes: pain and total knee replacement

3.3.5.1 Year 20 pain

Knee pain was assessed during the year 20 visit as part of the general Chingford musculoskeletal questionnaire, with both a modified NHANES question as well as the full WOMAC questionnaire. The NHANES question came in two parts, with the first asking 'Have you had pain in your right/left knee in the last month?' The second half of the question asked the subject to circle the number of days they had pain each month. The variable for pain used at year 20 was created as positive for pain if subjects answered 'yes' to the first part of the question and selected '15+ days' for the second part of the question.

3.3.5.2 Total knee replacements

Total knee replacements were confirmed using a variety of sources including radiographs, nurses' reports during physical exams, self-report by the subject on questionnaires, and confirmation by the subject's general practitioner. Radiographs were considered the 'gold-standard' evidence for TKR and were used wherever possible.

3.4 Study design and statistical methods

This section will outline the study designs and statistical methods used for each of the results sections; comparison of ROA scoring methods, natural history, cross-sectional and longitudinal analysis. It will also describe several methods and concepts which are used throughout this research; comparing characteristics of subjects in the analysis to subjects lost to follow-up, reproducibility and validity.

3.4.1 Cohort characteristics

Descriptive statistics were used to compare baseline characteristics of subjects included in each analysis to those who were lost to follow-up. Baseline characteristics included age, BMI, history of smoking, pain (at year 5) and baseline K/L grade. Normally

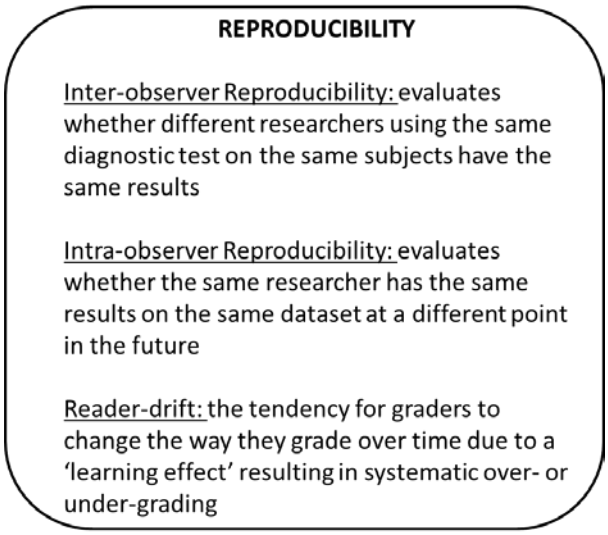
distributed variables were compared by t-test, while non-normal variables were tested using Man-Whitney U tests for continuous data, Pearson Chi-square tests for categorical data, and Fisher exact tests for categorical variables with small expected cell counts.

3.4.2 Reproducibility

Two researchers (KL and NB) graded x-rays using the atlas-based scoring methods for the reproducibility study. Two one hour training sessions were conducted with both readers, with a period of several weeks in-between. Each reader independently scored at least 15 radiographs, which were discussed and compared during the second training session. Official reproducibility was evaluated using 50 radiographs randomly selected from the full cohort (25 from year 15 and 25 from year 20). Each reader read the full set of x-rays, with a single reader (KL) re-reading the set after an interval of one day for short-term intra-observer reproducibility and then again after an interval of ten days for medium-term intra-observer reproducibility.

Linear weighted kappas were calculated in order to evaluate inter- and intra-observer reliability for each atlas-based scoring method (see figure 3.2). The Cohen kappa statistic assesses reproducibility in categorical variables (Kirkwood and Sterne 2003) and compares the 'expected' agreement calculated based on the distribution of the variables to the 'actual' agreement, and gives a range of values between -1 and 1 (primarily between 0 and 1). A negative kappa indicates that 'actual' agreement is less than 'expected' agreement, while a kappa of 0 indicates that agreement is no better than would be expected by chance.

Figure 3.2 Definitions of Reproducibility



A standard interpretation of kappas has been proposed in order to qualify the cut-offs for 'good' or 'poor' agreement. Landis and Koch (1977) gave the following descriptions to describe the strength of the agreement ≤0= poor, .01 -.20= slight, .21-.40= fair, .41-.60= moderate, .61-.80= substantial, and .81-1= almost perfect'. Weighted kappas can be used to reflect a degree of disagreement and are more appropriate to use on ordinal data (such as x-ray grades) than an unweighted kappa (Sim et al. 2005). They take into account the severity of disagreement (i.e. how many categories away from agreement) by counting partial agreement and do not penalise completely for less than perfect agreement. Factors that can affect the kappa include prevalence, bias and non-independence of ratings by the graders (Sim et al. 2005).

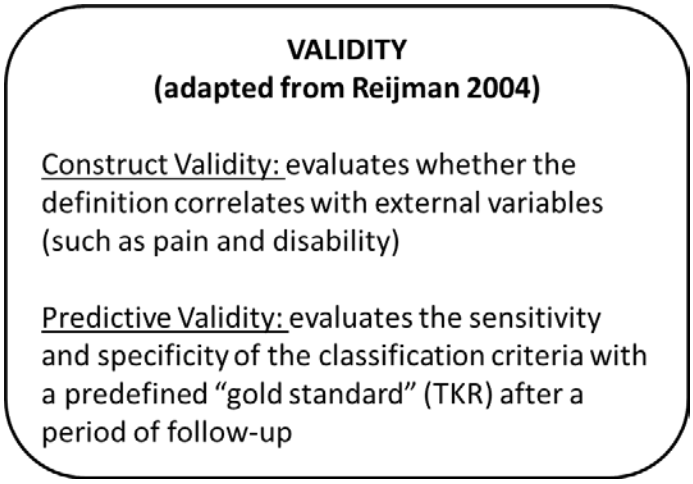
During the year 20 readings, intra-observer reliability was tested at intervals to check for reader drift (see figure 3.1) over the course of the reading period. After reading the first 200 year 20 x-rays, KL re-read the first 20 x-rays and then repeated this process after reading another 200 x-rays. Weighted kappas were used to check for non-systematic change in readings, while cross-tabulations were used to assess if there was any systematic under- or over-grading.

The reproducibility of continuous variables, such as minimum joint space, was tested using intra-class correlations (ICCs) and 95% confidence intervals.

3.4.3 Validity

Validity is one of the measures alongside reproducibility which can help determine the quality of a diagnostic tool. While reproducibility is focused on the consistency of a measurement, validity is concerned with whether the tool or method is actually measuring what is meant to be measured. The definitions for construct validity and predictive validity are outlined in figure 3.3.

Figure 3.3 Definitions of validity



3.4.4 Comparing scoring Methods

Study Design: Cross-sectional analysis using atlas-based ROA data from year 20.

Statistical Methods: All categorical scoring methods were evaluated for each feature and scoring method read for the year 20 x-rays (K/L, Chingford, OARSI and Line Drawing). Descriptive statistical methods were used for the analysis, which included cross-tabulations with frequencies and percentages. Stacked bar charts were used to display cross-tabulation data.

3.4.5 Natural history of radiographic osteoarthritis

Study Design: Complete case analysis of K/L and Chingford atlas data at baseline, year 5, year 10, year 15, and year 20.

Statistical Methods: Prevalence, incidence, progression and worsening (described in more detail below) were calculated using x-ray data from all available visit years. Pearson chi-squared tests were used to compare differences between age groups, BMI groups and pain/pain-free knees. Wilcoxon rank-sum tests were used to assess the significance of trends across age groups.

3.4.5.1 Prevalence

Prevalence was calculated at the knee level (with each subject supplying two knees to the analysis) as well as the compartment level (for both K/L and individual features). It was defined using a K/L grade 2 or above or individual feature grade 1 or above as an indication of disease presence.

3.4.5.2 Incidence

Incidence of K/L was calculated at the knee level and was defined as having a grade 0 or 1 at the first period of observation and a grade 2 or above at the second period of observation. Individual feature incidence was calculated at both the compartmental (*i.e.* 'medial JSN or osteophyte') and knee levels (*i.e.* 'any osteophyte') and were defined as a grade 0 at the first observation period and a grade 1 or above at the second. Annual cumulative incidence was determined by dividing the incidence by the number of years under observation.

3.4.5.3 Progression and worsening

Progression was calculated at the compartmental and knee levels and for K/L was defined as having a grade ≥ 2 at the first period of observation and showing an increase of

at least one grade by the second period of observation. Individual features were defined as having a grade ≥ 1 at the first period of observation and an increase of at least one grade at the second. Worsening was calculated at the compartmental and knee levels and was defined as an increase of one grade from *any* other grade (including grades 0 and 1).

3.4.6 Construct validity of ROA and pain (cross-sectional analysis)

Study Design: Cross-sectional analysis at year 20 of subjects with x-ray (atlas-based ROA scores and quantitative JSW) and pain (NHANES variable)

Statistical Methods: The outcome measure for all analysis is the binary NHANES pain variable (described in chapter 2). Crude associations were tested first between individual ROA features (atlas-based scores and quantitative JSW) and pain using univariable logistic regression. Multivariable logistic models were then used to adjust for covariates; age, BMI, and flexion (for quantitative JSW only). General estimating equation (GEE) models were used for all regression analysis in order to account for co-linearity (clustering) due to the knee level analysis (two knees for each subject) (Burton et al. 1998; Hanley et al. 2003). Construct validity was interpreted using odds ratios and 95% confidence intervals, with ROA features having higher odds ratios demonstrating better construct validity (association with pain).

3.4.7 Predictive validity of ROA with TKR (longitudinal analysis)

Study Design: Nested case-control: 10 controls for each TKR, matched on age and time in the study.

Statistical Methods: The association between baseline ROA (atlas-based and JSW) was assessed against the odds of a future TKR (after year 5). Change (worsening) of risk factors between baseline and year 5 was also evaluated. Conditional logistic regression was used to account for matching of the cases and controls. Results were given in odds ratios and 95% confidence intervals.

3.5 Summary

This chapter describes the unique characteristics of the Chingford study cohort including the nineteen year follow-up, the recruitment from a general population, and the comprehensive x-rays and pain information taken throughout the study, which provides data with which to evaluate the quality of radiographic risk factors in terms of reproducibility, construct validity and predictive validity. It is one of the few cohorts in the world which provides the necessary data to describe the 20 year natural history of radiographic knee osteoarthritis.

This chapter outlines the specific variables that were collected as risk factors, such as the previously read x-ray data, the newly collected data at the year 20 visit, and quantitative joint space measurements assessed at specific time points using the specially developed KneeMorf software program. The methods used to identify the two main outcomes (pain at year 20 and total knee replacement after year 5) are fully described above along with the general collection of variables used to assess the baseline characteristics of the cohort.

The study designs and statistical methods are outlined for each aspect of this research; a complete-case analysis for analysing the natural history of radiographic knee osteoarthritis, a cross-sectional analysis at year 20 to compare ROA methods to each other and then to pain, and a matched nested case-control to analyse baseline (and 4-year change) ROA and the risk for a future knee replacement.

4 KNEEMORF PROGRAM

One of the primary outcomes for testing disease modifying osteoarthritis drugs is minimum joint space measured quantitatively (Abadie et al. 2004). Narrowing joint space over time has also been identified as a possible early indication for both pain and the need for a knee replacement (Bruyere et al. 2005).

Automated methods have been developed to quickly and reproducibly give a variety of joint space measurements including minimum, mean and area on digitised and digital x-rays (Marijnissen et al. 2008; Oka et al. 2009). Most of these methods, however, require high contrast x-rays for their edge-detection algorithms to accurately capture joint space.

Due to the length of the Chingford study, some of the plain-film x-rays are over 20 years old and were only later digitised after being stored for ten to fifteen years. While joint space is still easily identifiable by eye, the lines are not distinct enough to use edge-detection algorithms. A new method using Bézier curves (described in detail below) was developed for this research in order to provide the benefits of edge-detect methods (having a full outline of the entire knee compartment) while allowing the user to identify the compartment edges accurately, quickly and efficiently.

This chapter describes the development of the KneeMorf software, the testing and validation work involved in implementing the Bézier curves, and the long- and short-term reproducibility of the measurements.

4.1 Background

4.1.1 Quantitative joint space width measurements

Joint space narrowing or width is often used as a surrogate outcome marker for knee cartilage loss as a result of osteoarthritis. X-rays are more cost-effective and patient-

friendly than MRIs, and are therefore more often used in determining the extent of the disease process. Plain-film radiographs show the delimitations of the bony edges of the joint and any space between is assumed to represent cartilage. Joint space is scored primarily using either quantitative or categorical methods.

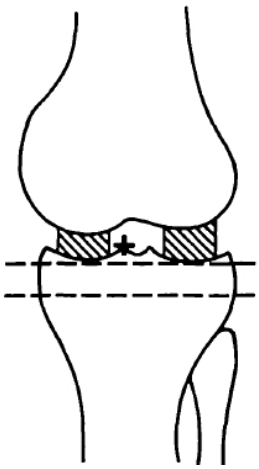
Quantitative methods measure the actual space between the articular surfaces of the bone in millimetres either by hand using callipers on plain-film or by utilising computerised methods on digitised film. Lequesne used specially designed callipers on high quality radiographs to measure the point of maximum narrowing, and this method of measurement is generally preferred to other manual methods such as applying a ruler directly onto the radiograph (1994). Ravaud based his technique on Lequesne's method, but included an additional two measurements; the midpoint of the knee compartment and a standardised point 10mm from the edge of the compartment (1996).

More recent quantitative methods have focused on semi-automated and fully automated computerised measurements in order to minimise error and maximise reproducibility. Automated methods for measuring minimum joint space width (minJSW) have shown high reproducibility when compared to manual measurements. In one study, patients were radiographed twice within a two week time-span and both automated and manual methods were applied to each radiograph. The automated method varied by an average of one pixel between radiographs of the same series, while the manual method differed by 0.10mm (Schmidt et al. 2005).

4.1.2 Computerised methods for measuring joint space width

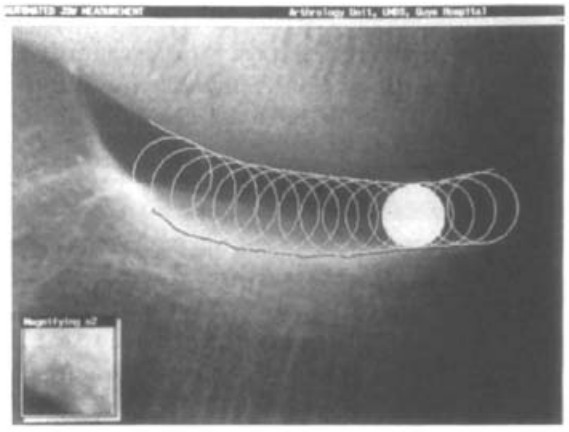
One of the earliest semi-automated methods for calculating joint space was developed by Dacre et al (1988; 1989). They made early use of edge-detection type algorithms, to try to define bone edges within a small defined section of each knee compartment (see figure 4.1). They calculated only joint space area (JSA).

Figure 4.1 Method of measuring JSA (from Dacre et al. 1989)



Lynch et al's work combined automated scoring with precise radiographic protocols (quantitative microfocal radiography), in order to minimise error from rotation and flexion (1993). They used modified edge detection with user-defined parameters within each measure, with the ability to calculate JSA and minJSW (see figure 4.2).

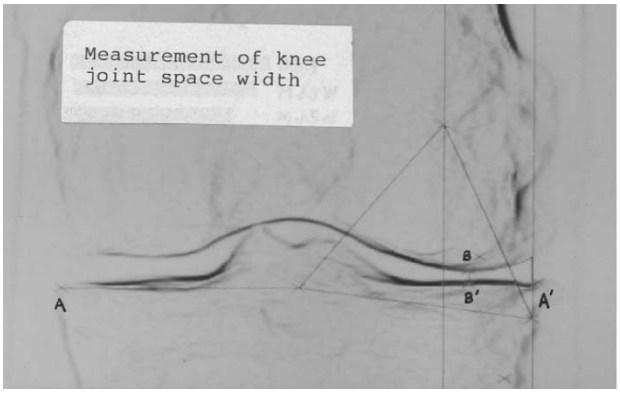
Figure 4.2 Calculating JSW using edge detection (from Lynch 1993)



Conrozier et al developed an automated method for hip minJSW and applied the same method to knee radiographs to calculate minJSW, meanJSW, and JSA (see figure 4.3) (Conrozier et al. 1998; Piperno et al. 1998). Schmidt et al produced a semi-automated

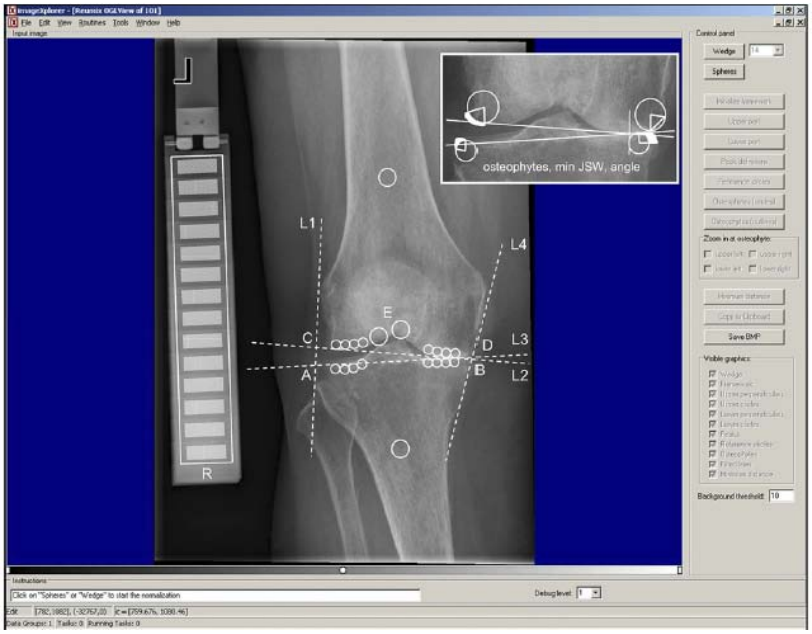
method using established edge-detect methods within a custom Matlab program giving very good reproducibility (2005).

Figure 4.3 Semi-automated edge detection method (Piperno 1998)



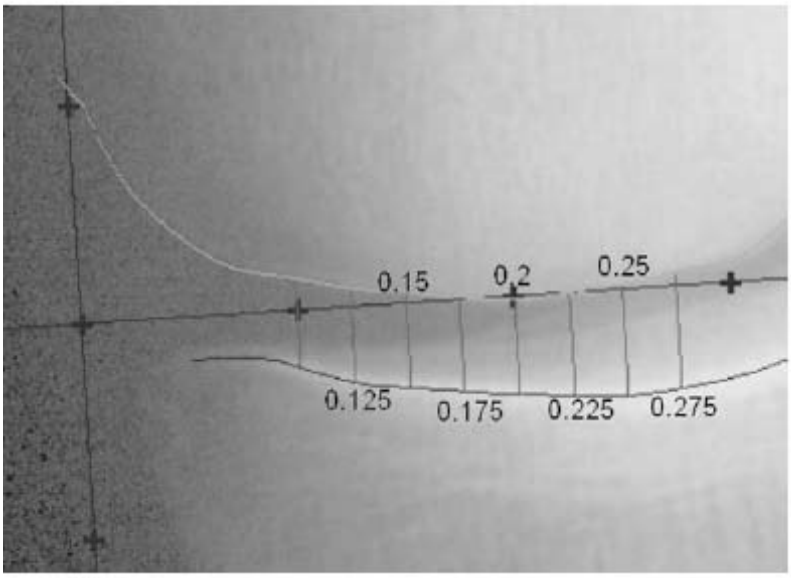
The Knee Image Digital Analysis (KIDA) program was the first program which collected risk-factor data in addition to joint space measures for knee radiographs (Marijnissen et al. 2008). In addition to minJSW and JSA, developers included osteophyte area, joint angle, tibial eminence height and bone density of the bone margin using a proprietary aluminium wedge which was included in x-rays (figure 4.4). Marijnissen et al found the majority of their measures were sensitive for identifying knees with osteoarthritis, and had high levels of reproducibility (Marijnissen et al. 008).

Figure 4.4 KIDA software program (from Marijnissen 2008)



One of the more recently developed methods of measuring joint space width does not measure the 'minimum', but rather uses a coordinate system based on anatomical landmarks to measure joint space at pre-specified locations (Neumann et al. 2009). The benefits of this method include the high reproducibility for a series of radiographs, the ability to make comparisons between patients, and the independence of the measurement location from knee size and magnification (*ibid*). In a small trial the responsiveness of the coordinate system was found to be better than minJSW in knees identified as having narrowing, although it was not significantly different. There was no difference for knees with normal joint space (Benichou et al. 2010). In practice this method is more time-consuming then measuring minimum width and requires perfectly consistent knee positioning (using strict x-ray protocols), all of which may limit its usefulness for future large-scale cohort studies.

Figure 4.5 Coordinate system for measuring JSW (Neumann 2009)

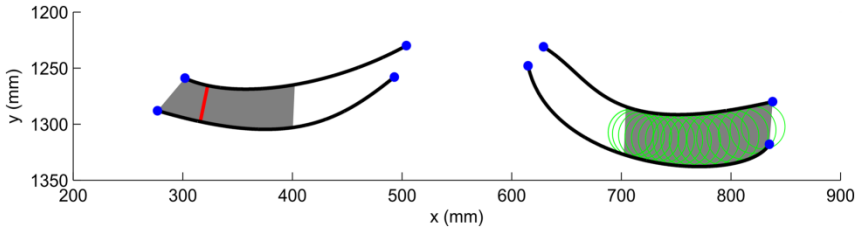


4.1.3 KneeMorf

KneeMorf was developed for the purpose of providing an efficient and accurate way to gather quantitative data on a variety of measurements commonly used in the research of musculoskeletal conditions. Specifically, other automated methods did not provide a way to collect comprehensive data, specialising in either joint space or alignment, but not both.

It was also necessary to develop a method which did not rely on edge-detection and would be equally reproducible and accurate on digitised plain-film x-rays as on digital x-rays. KneeMorf has additional built-in functionality to record semi-quantitative scoring methods (such as Kellgren and Lawrence) commonly used in the research of osteoarthritis. It also features a novel method of recording joint space width, using Bézier curves to give a range of accurate and reproducible measurements (figure 4.6). Additional unique features of KneeMorf include the recording of anatomical features with the potential to identify flexion, which is known to influence joint space measurements (Buckland-Wright et al. 2005)

Figure 4.6 KneeMorf output – Outlines of compartments with measurements



4.2 Software development

4.2.1 Development team

The KneeMorf program was developed by a team of researchers at the University of Oxford. Professor Nigel Arden and Dr. Kassim Javaid were the principal investigators of the project with Dr. Richie Gill as the technological lead. Two programmers were involved over the course of the project, with a contractor providing the primary coding and database development, the user interface and coding of the initial points and measurements. Dr. David Hunter, an image specialist and engineer, modified the original version of the program providing advanced functionality and usability on all aspects of the program including the implementation of novel features such as the Bézier curves. Kirsten Leyland and Nicholas Bottomley developed the original concept for the program including the design of the graphical user interface and database schema. Kirsten Leyland managed the project from inception to the current version and helped develop, evaluate and test all functionality of KneeMorf.

4.2.2 Programming and databases

KneeMorf was developed using open-source software (Python 2.7) and database management systems (SQL 5.1). This allows the program to be as flexible as possible and widely accessible by researchers. Matlab 7.12 was used to calculate final measurements and angles from the raw data (points) exported from KneeMorf.

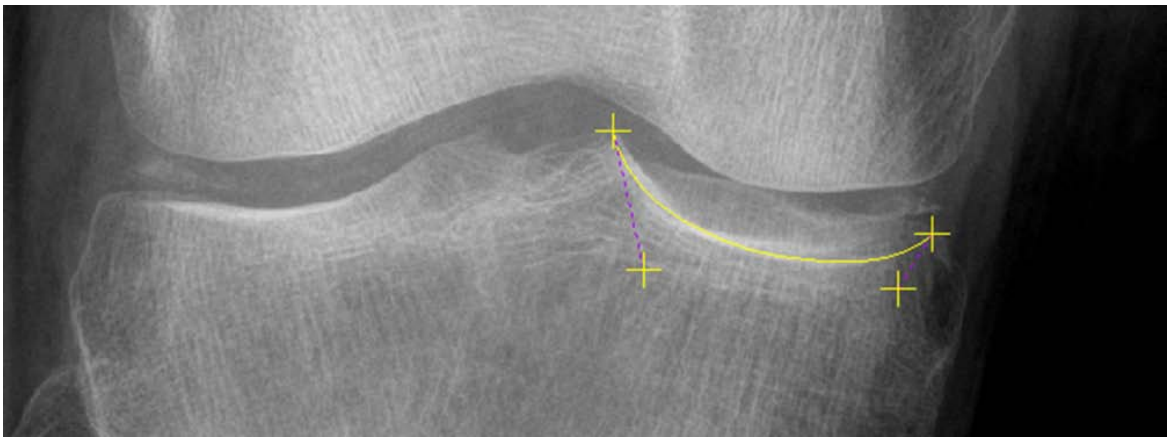
4.3 Bézier curves

4.3.1 Background

As an alternative to edge-detection algorithms and manually determined minJSW, this research explores the use of Bézier curves, a parametric function easily manipulated to represent the curvilinear shape of both a femoral condyle and tibial plateau. Bézier curves provide a similar level of information as automated edge-detect methods, but benefit from the accuracy of the human eye in detecting bone contours. Bézier curves were developed in 1959 by mathematician Paul de Casteljau for use in automobile design, but were popularised by engineer Paul Bézier in 1962 (Riskus et al. 2006). More recently they have been applied to biological systems for use in biometric data (iris and fingerprint images) (Vari et al. 2010), as well as being used to record the shape of complex anatomical structures such as the corpus collosum (Farag et al. 2010).

This research utilised Bézier curves which were able to be defined by between four and six points. Once the two end-points are set, it takes a minimal amount of time (less than 15 seconds) and practice to overlay the bone edge with the Bézier curve with visual accuracy. Each curve can be stored as four discrete points, which require a minimal amount of space to store, but are able to describe a complex biological shape.

Figure 4.7 KneeMorf – Bézier curve placement



4.3.2 Bézier curve placement and measurement definitions

Bézier curves were placed to define the floor of the tibial plateau in the medial and lateral compartments individually (figure 4.7). The first point was placed on the outer edge of the tibial plateau, excluding osteophytes. The point defining the interior of the compartment was placed on the tip of the corresponding tibial spine. Next, two control points were manipulated until the curve corresponded to the middle of the lowest sclerotic line indicating the floor of the tibial plateau.

The placement of the Bézier curves on the medial and lateral femoral condyles was accomplished in a similar manner, with the outer edge determined by trying to include the curve of both the outer edge of the condyle and part way into the femoral notch. The KneeMorf manual in appendix 1.3 has full descriptions of the placement of each point with corresponding images.

4.3.2.1 Minimum joint space width

Minimum JSW was calculated within an area restricted between the two compartmental Bézier curves and the apex of the curve up the tibial spine slope. Two methods were evaluated which determined the bottom of the slope of the tibial spine, past which joint space would not be measured. The first was a point manually determined by

the user (user-slope) and placed where the user determined the floor of the tibial plateau stopped and the slope of the tibial spine began; the second was automatically calculated (auto-slope) using a mathematically derived inflection point on the tibial Bézier curve. Minimum JSW measurements were calculated using both the user-defined slope point as well as the auto-defined slope point to restrict the area of joint space to be measured.

To determine minJSW each curve was calculated into a series of highly discretised points using a customised Matlab function. All distances between each of the resulting points for the femoral Bézier and the tibial Bézier curves were then calculated, with the smallest distance given as the minJSW for the compartment. The axis of measurement was not restricted to any pre-specified angle.

4.3.2.2 Mean joint space width

Mean JSW was calculated within the region established using the automatically-determined tibial slope point as the innermost limit and the tibial plateau edge point as the outer limit. The Bézier curves were used as the superior and inferior limits of the area to be calculated. A series of circles were fit between the curves with the outer edges touching both the femoral and tibial Bézier curves at regular intervals. This was based upon the previous method established by Lynch et al (1993). The mean of the diameters of the circles then gave the meanJSW for each knee compartment.

4.3.2.3 Joint space area

Joint space area was determined for each compartment using the same restricted area for measurement as both minJSW and meanJSW. This measurement was constrained medially by the auto slope point, laterally by the plateau edge, superiorly by the femoral Bézier curve and inferiorly by the tibial Bézier curves. The area was calculated as mm^2 within the pre-defined space.

4.3.2.4 Manual digital calliper (validation measure)

In order to validate the novel measures calculated by the Bézier curves, a 'digital calliper' requiring manual placement to measure minJSW was included in the KneeMorf reproducibility and validity study. The digital calliper was implemented to be as similar to measuring minJSW on a plain film x-ray as possible. A sliding calliper cursor was used, without any restrictions placed on the measurement angle. This measure was used as the gold standard minJSW measure to validate the Bézier minJSW against.

4.4 Validity and reproducibility methods

4.4.1 Reproducibility protocols

Twenty-five digitised plain-film x-rays from year 15 and 25 digital x-rays from year 20 were selected for the study. All joint space measurements available in the KneeMorf program were placed during the testing including digital callipers for manually determining minJSW, and Bézier curves for each of the four bone edges defining the medial and lateral compartments of the knee.

X-rays were selected with a range of disease severity (K/L 0 through 4) and BMI categories. X-rays were subdivided by K/L grade and again by BMI group (<25 kg/m^2 or ≥ 25 kg/m^2) to make sure an equal number of normal/overweight subjects were represented in each K/L grade. Five subjects were randomly selected from each K/L grade as scored at the year 15 visit, with half from each BMI category. Subjects with K/L grade 4 were underrepresented due to a lack of subjects and were substituted with subjects with K/L grade 3. The corresponding x-rays for the 25 subjects chosen from year 15 were then selected from year 20.

Two observers were used to test reproducibility and validity, Kirsten Leyland (KL) and Dr. David Hunter (DH), both of whom were familiar with KneeMorf technology and the definition of the points and curves. Two training sessions were completed with the reading of ten radiographs at the end of each session. A small improvement of agreement was found between the first and second sessions, although the overall agreement remained high from the beginning. Once the training was complete, KL and DH each read 50 radiographs using KneeMorf, blind to study number, disease severity and BMI group. The study year was unable to be blinded due to the noticeable differences between the digitised plain-film and the digital radiographs. KL reread all 50 x-rays after an interval of five days for the short-term intra-observer reproducibility and then for a third time after an interval of five months for the long-term reproducibility. DH reread all radiographs after an interval of five months. There was an informal brief review of training materials with a discussion before each reader read the set after the five month interval.

4.4.2 Statistical methods

4.4.2.1 Reproducibility

Inter- and intra-observer reproducibility was calculated using intra-class correlations (ICC) with 95% confidence intervals (CI) in addition to Bland-Altman plots (Bland and Altman 1986). Short-term intra-observer reproducibility was calculated using KL's first and second readings. Long-term intra-observer calculations were done using KL's first and third readings and DH's first and second readings. Intra-observer reproducibility was calculated with the first readings by both readers and the last readings by both readers.

4.4.2.2 Validity

Validity of the automated minJSW produced by the Bézier curve was tested against minJSW produced by user placement of a digital calliper. To be a valid measure, the automated minJSW would be expected to be highly correlated with the digital calliper minJSW and have a similar or greater level of reproducibility. The measurements were assessed using linear regression in addition to Spearman's correlation coefficient (r).

4.5 Results

4.5.1 Reproducibility

Table 4.1 shows the results for all the joint space measurements taken during the reproducibility study. Inter-observer reproducibility was extremely high with the intra-class correlation coefficients greater than 0.9 for all measurements. When comparing the reproducibility of user-defined tibial slope Bézier minJSW with the automated tibial slope point minimum, the automated measure showed greater short- and long-term reproducibility than the user-determined slope measure. The greatest difference in reproducibility was found for the lateral compartment over the five month interval (0.83 vs 0.97). The meanJSW measures followed a similar pattern with those calculated using the automated slope point being the same or having slightly better reproducibility than the manually determined slope point.

Table 4.1 Intra- and inter- observer short- and long-term Bézier reproducibility

JSW Measurements	Intra-Observer (short)		Intra-Observer (long)		Inter-Observer (first)	
	ICC	*95% CI*	*ICC*	*95% CI*	*ICC*	*95% CI*
minJSW (medial) - user	0.96	(0.94, 0.98)	0.93	(0.89, 0.97)	0.93	(0.90, 0.97)
minJSW (lateral) - user	0.96	(0.94, 0.98)	0.83	(0.74, 0.92)	0.91	(0.87, 0.96)
minJSW (medial) - auto	0.97	(0.95, 0.99)	0.94	(0.90, 0.97)	0.94	(0.91, 0.97)
minJSW (lateral) - auto	0.96	(0.93, 0.98)	0.97	(0.95, 0.99)	0.94	(0.91, 0.97)
meanJSW (medial)	0.96	(0.93, 0.98)	0.91	(0.86, 0.96)	0.89	(0.83, 0.95)
meanJSW (lateral)	0.97	(0.96, 0.99)	0.97	(0.95, 0.99)	0.94	(0.91, 0.97)
Joint Space Area (medial)	0.95	(0.93, 0.98)	0.95	(0.92, 0.98)	0.93	(0.89, 0.97)
Joint Space Area (lateral)	0.97	(0.95, 0.98)	0.97	(0.95, 0.99)	0.94	(0.90, 0.97)

*Intraclass correlations (ICC) and Confidence Intervals (CI)

Table 4.2 shows the reproducibility between the year 15 digitised plain-film low contrast x-rays and the new digital x-rays from year 20. This demonstrates that there is no noticeable difference in reproducibility for x-ray type.

Table 4.2 Inter-observer reproducibility: year 15, year 20, and all x-rays

JSW Measurements	Inter-Observer Year 15 X-rays		Inter-Observer Year 20 X-rays		Inter-Observer (ALL)	
	ICC	*95% CI*	*ICC*	*95% CI*	*ICC*	*95% CI*
Calliper minJSW (medial)	0.95	(0.88, 0.98)	0.92	(0.84, 0.97)	0.94	(0.89, 0.96)
Calliper minJSW (lateral)	0.97	(0.93, 0.99)	0.94	(0.87, 0.97)	0.96	(0.93, 0.98)
minJSW (medial) - user	0.94	(0.87, 0.97)	0.93	(0.83, 0.97)	0.93	(0.90, 0.97)
minJSW (lateral) - user	0.88	(0.76, 0.95)	0.94	(0.88, 0.97)	0.91	(0.87, 0.96)
minJSW (medial) - auto	0.95	(0.89, 0.98)	0.93	(0.83, 0.97)	0.94	(0.91, 0.97)
minJSW (lateral) - auto	0.95	(0.89, 0.98)	0.93	(0.85, 0.97)	0.94	(0.91, 0.97)
meanJSW (medial)	0.89	(0.77, 0.95)	0.88	(0.76, 0.95)	0.89	(0.83, 0.95)
meanJSW (lateral)	0.96	(0.91, 0.98)	0.91	(0.80, 0.96)	0.94	(0.91, 0.97)
Joint Space Area (medial)	0.94	(0.87, 0.97)	0.91	(0.82, 0.96)	0.93	(0.89, 0.97)
Joint Space Area (lateral)	0.97	(0.92, 0.99)	0.89	(0.78, 0.95)	0.94	(0.90, 0.97)

*Intra-class correlations (ICC) and Confidence Intervals (CI)

4.5.2 Validity

The validity of the Bézier curve minJSW was tested against minJSW produced by user placement of the digital calliper. To be a valid measure, the Bézier minJSW would be expected to be highly correlated with the digital calliper minJSW and have a similar or greater level of reproducibility. The measurements were assessed using linear regression in

addition to Spearman's correlation coefficient (CC). The means and standard deviations for the minJSW in the medial compartment were 3.9mm (SD 1.5) and 3.8mm (SD 1.4) for the callipers and Bézier measurements, respectively. In the lateral compartment, the minJSW were 5.1mm (SD 1.5) and 5.1mm (SD 1.4), respectively. The standard deviation of the Bézier minJSW was 0.1mm smaller than the 'gold standard' calliper minJW, indicating a smaller measurement error between grading for the Bézier minJSW.

When the validity of the Bézier -based minJSW was tested against the manual digital callipers, linear regression showed an r^2 of 0.92 in both compartments and a correlation coefficient of 0.89 medially and 0.93 laterally.

Reproducibility (table 4.2) was the same for both measures in the medial compartment, except for the long-term intra-observer analysis which was slightly better (0.96 vs 0.94) for the digital calliper. The reproducibility in the lateral compartment showed slightly better short-term intra-observer (0.96 vs 0.97) and inter-observer (0.94 vs 0.96) but long-term intra-observer was much better using the Bézier minJSW (0.97 vs 0.75).

4.6 Identification of flexion on radiographs

As quantitative joint space width is used as a primary outcome for clinical trials (Abadie et al. 2004), there has been research on ways to minimise the effect that flexion has on the appearance of joint space width. This is a well-known issue addressed by Buckland-Wright et al. 2005, Vignon et al. 2010, and Takahashi et al. 2009. Takahashi showed that mean minimum joint space varied between 2.46mm in an extended view compared to 1.47mm in the same 46 subjects with painful knees (Takahashi et al. 2009).

Extended knee x-rays cause the most problems in regards to flexion due to the variation of each subject. The degree to which a subject can fully straighten their legs is

highly influenced both by pain and fixed-flexion deformities, and is especially evident for x-rays taken at more than one time point if symptoms have changed over that time. Studies have shown that pain relief after extreme flare-ups biased the joint space width which appeared to widen (Mazzucca et al. 2002). This is one of the primary reasons that fixed-flexion and fluoroscopy protocols are now standard where possible.

The primary method to deal with this issue has been to establish the best standardised protocols for positioning subjects during x-ray to minimise the effects of poor flexion. The favoured methods for standardisation have been the metatarsophalangeal (MTP) view and semi-flexed positioning with the aid of fluoroscopy.

A second method of accounting for flexion is to identify the degree of poor positioning found in x-rays which have been previously taken. This method is the only available to long-term cohort studies such as the Chingford study, which is using the same protocols established at the time of the study's inception over twenty years ago. Establishing the presence of flexion requires the identification of morphological features on x-ray which change in some measureable way in relation to the amount of flexion. One of these features which has been previously identified is the inter-margin distance (IMD), which has been indicated as a useful identifier of flexion (Vignon et al. 2010; Takahashi et al. 2009; Le Graverand et al. 2008).

4.6.1 Feature for identifying flexion: inter-margin distance

The inter-margin distance (IMD) is calculated as the maximum distance between the tibial plateau floor (the middle of the sclerotic line) and the upper most rim visible on x-ray (*either* the anterior *or* posterior rim) (Vignon et al. 2010). This distance is calculated both medially and laterally.

For studies using fluoroscopy assisted positioning or MTP views, an IMD distance of greater than 1.0mm is considered an indication of flexion which will have adverse effects on the measurement of joint space (Buckland-Wright et al. 1995).

4.7 Summary

The benefit of using Bézier curves to measure joint space is four-fold; first, a user can accurately define the bone edge of a compartment by placing only four or five points; second, the memory required to record this information and compute results is much less than that needed for edge-detect based software; third, the results are easier to interpret and compare with one another as they are represented by a mathematical function; and finally, edge-detection has not yet reached the same level of accuracy as the human eye when determining bone edge on difficult radiographs.

This work has shown that minimum joint space calculated from Bézier curves is highly reproducible for both short- and long-term inter- and intra-observer readings and is valid when compared to a manual measurement of minimum joint space (using a digital calliper). It also has been shown to have similar reproducibility on older low-contrast digitised plain-film x-rays as on new high quality digital x-rays.

5 REPRODUCIBILITY OF ATLAS-BASED SCORING METHODS

Data quality of a scoring method should be assessed by checking both its reproducibility and its validity. While validity focuses on whether the measurement is giving the *true* answer, reproducibility is concerned with the difficulty in obtaining a *consistent* answer either between readers or between sets of results over time from the same reader (Silman and Macfarlane 1995). If reproducibility of a measurement is not good, it can either indicate that the method itself lacks consistency or that one or both of the readers are inconsistent in the application of the method.

Determining the reproducibility of each atlas-based scoring method (K/L, OARSI, Chingford and Line Drawing) is the first step in evaluating the quality of each method, providing a solid base from which to assess both construct validity and predictive validity of the methods. A comparison of reproducibility between scoring methods will help evaluate whether differences found in prevalence between studies may be due to error between graders as opposed to a true reflection of a difference in disease burden. Assessing reproducibility was also necessary before grading all of the year 20 x-rays needed for testing construct validity of the ROA scoring methods (see chapter 8). This ensured that the atlas grading was consistent and did not change over the course of the grading.

This chapter will address several common issues surrounding reproducibility; comparing reproducibility between graders (inter-observer), comparing reproducibility of a single observer over more than one time (intra-observer), assessing the short and long-term reproducibility, identifying the consistency of grading over a period of time (reader drift), and evaluating whether reproducibility was affected by x-ray type (digitised plain-film vs. digital x-ray).

5.1 Methods

Reproducibility was comprehensively assessed throughout the study. Long- and short-term reproducibility was analysed both between (inter-) and within (intra-) graders for all atlas-based scoring methods (K/L, Chingford, OARSI and Line Drawing). Intra-observer reader drift was evaluated twice throughout the course of reading the year 20 x-rays.

5.1.1 Study design

5.1.1.1 Training

For inter-observer reproducibility, two readers generally familiar with radiographic features of knee osteoarthritis (KL and NB) read a test set of radiographs using a standard set of the radiographic scoring atlases, without any prior consultation, in order to identify natural differences between graders as well as problematic areas in each atlas without external influence.

Ten subjects were randomly selected from the year 15 Chingford x-rays to reflect the full range of K/L grades as the first training set. Each grader independently read the x-rays (right knees only) while keeping a detailed record of specific knees or scores that were found to be problematic. The percentage of agreement for each score was used to give a general indication of where future training should focus.

5.1.1.2 Formal study

Once the training was complete, 50 x-rays were selected from the Chingford study as the formal reproducibility set, 25 from the year 20 visit and the same 25 subjects from the year 15 visit. This was done so that reproducibility could be calculated for both the newer digital x-rays (year 20) as well as the older digitised plain-film x-rays (year 15).

Since the year 20 x-rays had not yet been read for K/L, the K/L grades and BMI measures for year 15 were used to select the reproducibility set. An attempt was made to have an equal representation of the full range of K/L grades (0-4) as well as an even number of subjects representing both the normal (<25 kg/m^2) and overweight/obese (≥25 kg/m^2) BMI categories. Due to an under-representation of subjects with K/L grade 4 at year 15, subjects with K/L grade 3 were substituted.

All radiographs were assigned a random number in order to blind the study number of the subject, but it was not possible to blind the study year, as the difference between digitised and digital x-rays is distinct. K/L grades and BMI categories were not evident once the x-rays were selected and renumbered.

KL read the reproducibility set three times with the first reading also used for the inter-observer analysis. The second reading was completed by KL after an interval of one day, with a third after an interval of ten days. Reproducibility was assessed using the statistical methods previously described.

For inter-observer reproducibility, both readers (KL and NB) independently read all 50 radiographs (right side only).

5.1.1.3 Reader drift

Reader-drift is tested to ensure that readers are grading consistently throughout the time it takes to grade all of the radiographs within a study. Semi-objective scoring methods, such as the radiographic atlases, are prone to several issues including a learning-effect and the tendency to over-read the presence of disease as the study goes on.

To check for reader-drift, KL re-read the first 20 x-rays after grading 250 subjects and a gap of seven months. Ten right knees and ten left knees were selected to be reread. A second reading of the first 20 x-rays was repeated after grading an additional 220 subjects to check for any learning and/or disease over-reading effect.

5.1.2 Statistical methods

The statistical measure, percentage agreement, was used for training sets due to the small number of observations (generally 10 to 15). For all formal reproducibility involving categorical data, a linear weighted kappa statistic (described in chapter 3) was used with standard errors (SE) in addition to percentage agreement. Stata version 12.0 was used for all statistics and graphing.

5.2 Training reproducibility results

The features with the lowest percentage of agreement were lateral K/L, OARSI lateral tibial osteophyte, Line Drawing medial JSN, both Chingford osteophyte measures and medial Chingford JSN, all of which had below 70% agreement. The results were not consistent between features across the scoring methods, as in, the agreement for OARSI scoring of medial JSN was 90%, while Line Drawing medial JSN was 65%. This suggested that the issue was with the specific atlases, as opposed to a general disagreement between graders on how to assess medial JSN. The first testing set was reviewed in detail, with readers coming to a consensus on the interpretation of each atlas. This session resulted in a series of points of clarification for each scoring method to be used in conjunction with the atlas (see appendix 1.2).

Once agreement was reached on the first set of x-rays, a second set of training x-rays was independently read by each grader to check whether the additional training had improved agreement. Only a single score was below 70% agreement during this round; Line Drawing lateral JSN. A majority of the scores (75%) increased in agreement, while the remaining remained steady or decreased slightly.

Unlike the previous training round, lateral JSN was the only feature which was assessed poorly across all scoring methods, indicating a general issue of consistency

between readers in the assessment of lateral joint space narrowing. This measurement is recognised to be difficult and a lot of research focuses on the medial compartment only, excluding the lateral side (Felson et al. 2008). Readers therefore underwent further training together, coming to a consensus on the interpretation and grading of joint space in the lateral compartment. This ultimately came down to agreeing which sclerotic line indicated the 'floor' of the lateral tibial plateau, which is discussed in more detail in the KneeMorf chapter (chapter 4).

5.3 Formal reproducibility results

5.3.1 Formal intra-observer

Table 5.1 demonstrates that the scoring methods generally had good short-term reproducibility, with kappas ranging between 0.43 and 0.95, with the majority (75%) having a kappa of at least 0.70. Several features showed poor reproducibility including tibial spiking (0.43) and the two medial sclerosis scores (0.50). Most osteophyte and joint space grades had very high kappas except for the medial femoral osteophyte in both the OARSI and Line Drawing atlases (Chingford did not have a comparable category). Features with the lowest levels of reproducibility were binary variables such as sclerosis and tibial spiking.

Medium-term readings showed a similarly good level of reproducibility, with kappas ranging from 0.49 to 0.89. As would be expected, the majority of kappas were slightly lower than the short-term reproducibility, with a few exceptions. Both the medial femoral osteophyte scores increased from having 'moderate' reproducibility to the 'substantial' range (Landis and Koch 1977).

Table 5.1. Intra-observer reproducibility of atlas-based methods

Method	Individual Features	Short-term Reproducibility		Medium-term Reproducibility	
		Kappa (w)	Std. Error	Kappa (w)	Std. Error
K&L	whole knee	0.84	0.09	0.79	0.09
	medial compartment	0.79	0.09	0.72	0.09
	lateral compartment	0.71	0.10	0.61	0.09
OARSI	medial femoral osteophyte	0.56	0.10	0.66	0.10
	lateral femoral osteophyte	0.84	0.11	0.89	0.11
	medial tibial osteophyte	0.77	0.11	0.65	0.11
	lateral tibial osteophyte	0.74	0.11	0.72	0.10
	medial jsn	0.79	0.11	0.63	0.10
	lateral jsn	0.95	0.11	0.86	0.11
	medial sclerosis	0.50	0.13	0.53	0.13
	lateral sclerosis	0.69	0.14	0.62	0.14
Line Drawing	medial femoral osteophyte	0.55	0.10	0.71	0.11
	lateral femoral osteophyte	0.84	0.11	0.79	0.11
	medial tibial osteophyte	0.79	0.10	0.70	0.10
	lateral tibial osteophyte	0.78	0.10	0.75	0.10
	medial jsn	0.73	0.08	0.76	0.08
	lateral jsn	0.67	0.08	0.76	0.08
Chingford	medial tibiofemoral osteophyte	0.79	0.10	0.66	0.11
	lateral tibiofemoral osteophyte	0.81	0.10	0.72	0.10
	medial jsn	0.88	0.11	0.78	0.11
	lateral jsn	0.95	0.11	0.86	0.11
	tibial spine spiking	0.43	0.13	0.50	0.13
	medial sclerosis	0.50	0.13	0.53	0.13
	lateral sclerosis	0.56	0.14	0.49	0.14

5.3.2 Formal inter-observer

The results of the atlas-based inter-observer reproducibility (table 5.2) were considerably lower than expected. Kappas ranged between 0.29 and 0.92, with 61% of features having kappas of 0.50 or greater which is considered moderate agreement. Particular issues were seen with medial K/L (kappa 0.35), whole knee K/L (kappa 0.41), medial femoral osteophytes across all atlases (kappa range 0.30-0.42), and joint space in the Line Drawing atlas (kappa 0.40 and 0.42). The percentage of agreement for all

features, however, was relatively high with 82% of features having over 80% agreement between raters.

Table 5.2. Inter-observer reproducibility for atlas-based methods

Method	Individual Features	Percent Agreement	Kappa (w)	SE
K/L	whole knee	81.0%	0.41	0.08
	medial compartment	74.5%	0.35	0.08
	lateral compartment	87.5%	0.54	0.10
OARSI	medial femoral osteophyte	78.0%	0.35	0.08
	lateral femoral osteophyte	86.7%	0.59	0.11
	medial tibial osteophyte	89.5%	0.58	0.10
	lateral tibial osteophyte	87.3%	0.57	0.10
	medial jsn	89.3%	0.68	0.10
	lateral jsn	94.7%	0.62	0.11
	medial sclerosis	72.0%	0.38	0.12
	lateral sclerosis	92.0%	0.56	0.14
Line Drawing	medial femoral osteophyte	76.5%	0.30	0.09
	lateral femoral osteophyte	87.6%	0.53	0.11
	medial tibial osteophyte	88.0%	0.54	0.09
	lateral tibial osteophyte	86.8%	0.52	0.09
	medial jsn	82.4%	0.40	0.07
	lateral jsn	88.4%	0.42	0.07
Chingford	medial tibiofemoral osteophyte	80.0%	0.42	0.09
	lateral tibiofemoral osteophyte	84.0%	0.54	0.10
	medial jsn	88.7%	0.65	0.10
	lateral jsn	96.7%	0.74	0.11
	tibial spine spiking	81.0%	0.33	0.11
	medial sclerosis	68.0%	0.29	0.12
	lateral sclerosis	88.0%	0.33	0.14
Other Features	medial subchondral cysts	90.0%	0.00	0.00
	lateral subchondral cysts	98.0%	0.00	0.00
	medial chondrocalcinosis	98.0%	0.92	0.14
	lateral chondrocalcinosis	90.0%	0.68	0.13

5.4 Reader-drift

The first reader-drift reproducibility showed good results (table 5.3), especially considering the long period of time between readings. A kappa statistic is able to assess any non-systematic changes in grading. Sixty-eight percent of the grades had kappas

greater than 0.7, with the features with lower grades including lateral JSN for all atlases, sclerosis for both atlases as well as tibial spiking. No kappas were in the 'poor' categories (<0.5).

Seventy-one percent of the kappas for the second reader-drift reproducibility test were greater than 0.7 indicating that, in general, the grading remained consistent throughout the course of reading year 20 x-rays. Some of the features with lower reproducibility include joint space narrowing in the lateral compartment for both the OARSI and Line Drawing atlases, as well as sclerosis for both the OARSI and Chingford atlases.

Cross-tabulations were used to check for any systematic over- or under-grading which would not be evident using a kappa statistic. K/L (whole-knee) and all osteophyte and JSN measures from the OARSI atlas were tested to represent the general features being graded for all atlases. While all osteophyte grades had an equal number of over- and under-grading for each reader-drift check, the JSN measures appear to have been under-graded, reading less disease after the first grading. For medial JSN, 25% of grades were lower in the second reading, primarily from a grade 1 to a grade 0. The same was seen for the third reading. Fifteen percent of the grades for lateral JSN were graded as 0 for the first and second readings. This shows that there was a shift to identify less narrowing after the first 20 x-rays that were read, but the fact that the second and third readings show the same trend means that only x-rays read very early in the study will have slightly higher levels of joint space narrowing.

Table 5.3 First and second intra-observer reader drift reproducibility

Method	Individual Features	First Drift Reading		Second Drift Reading	
		Kappa (w)	SE	Kappa (w)	SE
K&L	whole knee	0.86	0.17	0.85	0.15
	medial compartment	0.89	0.16	0.85	0.15
	lateral compartment	0.76	0.17	0.92	0.18
OARSI	medial femoral osteophyte	0.67	0.16	0.74	0.16
	lateral femoral osteophyte	0.83	0.16	0.79	0.17
	medial tibial osteophyte	0.84	0.17	0.84	0.17
	lateral tibial osteophyte	0.91	0.17	0.82	0.17
	medial jsn	0.72	0.15	0.67	0.15
	lateral jsn	0.62	0.19	0.70	0.18
	medial sclerosis	0.34	0.17	0.61	0.21
	lateral sclerosis	0.62	0.21	0.48	0.22
Line Drawing	medial femoral osteophyte	0.81	0.16	0.73	0.16
	lateral femoral osteophyte	0.88	0.17	0.73	0.17
	medial tibial osteophyte	0.86	0.17	0.93	0.17
	lateral tibial osteophyte	0.89	0.16	0.86	0.17
	medial jsn	0.80	0.13	0.50	0.11
	lateral jsn	0.51	0.13	0.49	0.12
Chingford	medial tibiofemoral osteophyte	0.90	0.16	0.85	0.16
	lateral tibiofemoral osteophyte	0.91	0.16	0.83	0.16
	medial jsn	0.87	0.16	0.76	0.16
	lateral jsn	0.57	0.19	0.89	0.19
	tibial spine spiking	0.30	0.22	0.70	0.21
	medial sclerosis	0.34	0.17	0.61	0.21
	lateral sclerosis	0.62	0.21	0.48	0.22
Other Features	medial subchondral cysts	95%*		90%*	
	lateral subchondral cysts	95%*		100%*	
	medial chondrocalcinosis	1.00	0.22	1.00	0.22
	lateral chondrocalcinosis	0.77	0.23	1.00	0.23

*Percentage of agreement used due to lack of categories required by kappa statistic

5.5 Inter-observer analysis with previously read Chingford x-rays

Chingford knee x-rays have been read by three different sets of people. Baseline and year 5 radiographs were read jointly by Tim Spector (TS) and Debbie Hart (DJH) (Hart 1999). Year 10 and year 15 x-rays were read by Debbie Hart alone, and year 20 by Kirsten Leyland (KL). Although all previous plain-film x-rays have been subsequently digitised,

they were all originally read for K/L and the Chingford atlas using the plain-film x-rays. Year 20 is the only year to be read using the digital format with an ability to zoom in on specific areas of the x-ray. This would be expected to cause a difference in readings, specifically that more disease (i.e. osteophytes and joint space narrowing) would be identifiable in year 20.

The 25 year 15 x-rays that were reread for this reproducibility study in their digitised format by KL were compared to the readings done previously by Debbie Hart (DJH) on the plain-film format.

5.5.1 Results

Only the whole knee K/L and the osteophyte and JSN scores from the Chingford atlas had been previously assessed on the plain-film x-rays. In this analysis (table 5.4), K/L showed moderate inter-observer reproducibility (0.54), with medial JSN having the best (0.68) and lateral JSN having the worst (0.22). The low kappas are not reflected by the overall percentage of agreement, which was greater than 85% for all grades. This may be due to the small sample size and the lack of representation in several categories of grades, which has an effect on the kappa statistic. A direct comparison between newer digital and older methods yielded higher grades in direct comparison with an absolute grade increase of 16.0% for K/L, and between 4.0 and 16.0% for individual ROA features.

Table 5.4 Inter-observer reproducibility – previous methods vs. current methods

Method	Individual Features	Percent Agreement	Kappa (w)	SE
K&L	whole knee	85.0%	0.54	0.13
Chingford	medial tibiofemoral osteophyte	85.3%	0.41	0.14
	lateral tibiofemoral osteophyte	85.3%	0.44	0.15
	medial JSN	92.0%	0.68	0.15
	lateral JSN	90.0%	0.22	0.17

5.6 Reproducibility of digitised plain-film compared to digital x-rays

Due to the long-term nature of the Chingford cohort, with x-rays having been collected over a span of 20 years, the methods of acquisition have changed over this time. The most dramatic change was the transition from plain-film to digital x-rays.

While every effort was made to maintain consistent x-ray protocols between the baseline and year 20 visit, the image quality is understandably different between the plain-film x-rays which were subsequently digitised and the digital x-rays acquired using modern techniques. In order to ensure the grading was not dependent on the type of x-ray, the reproducibility of each type of x-ray was checked independently.

5.6.1 Intra-observer comparison

Intra-observer reproducibility (table 5.5) was the same or higher for almost every feature for the year 20 x-rays compared to the year 15 x-rays. The only exceptions were tibial spiking, subchondral sclerosis and chondrocalcinosis. Some of the greatest differences were found for whole knee K/L grades (0.74 vs 0.94), OARSI medial tibial osteophytes (0.67 vs. 0.87) and lateral tibial osteophytes (0.59 vs 0.86).

With the intra-observer reproducibility being consistently higher across the board between digital and plain-film digitised x-rays, it appears that the image clarity seen in digital images improves the reproducibility of grading, as would be expected. The year 15 x-rays, however, still maintained a good level of reproducibility on their own. This demonstrates that the reproducibility of the two different formats of x-ray should not affect analyses between them.

Table 5.5. Intra-observer reproducibility: year 15 digitised vs. year 20 digital x-rays

Method	Individual Features	Year 15 x-rays (n=25)		Year 20 x-rays (n=25)	
		Kappa (w)	SE	Kappa (w)	SE
K&L	whole knee	0.74	0.13	0.94	0.13
	medial compartment	0.71	0.13	0.88	0.13
	lateral compartment	0.62	0.14	0.79	0.13
OARSI	medial femoral osteophyte	0.49	0.14	0.59	0.15
	lateral femoral osteophyte	0.81	0.15	0.86	0.16
	medial tibial osteophyte	0.67	0.16	0.87	0.16
	lateral tibial osteophyte	0.59	0.16	0.86	0.14
	medial jsn	0.74	0.15	0.83	0.15
	lateral jsn	0.87	0.17	1.00	0.16
	medial sclerosis	0.36	0.15	0.45	0.18
	lateral sclerosis	0.62	0.20	0.78	0.20
Line Drawing	medial femoral osteophyte	0.55	0.16	0.50	0.15
	lateral femoral osteophyte	0.76	0.15	0.88	0.16
	medial tibial osteophyte	0.70	0.15	0.86	0.14
	lateral tibial osteophyte	0.68	0.14	0.83	0.14
	medial jsn	0.71	0.12	0.71	0.11
	lateral jsn	0.64	0.11	0.67	0.11
Chingford	medial tibiofemoral osteo	0.69	0.14	0.88	0.15
	lateral tibiofemoral osteophyte	0.66	0.15	0.92	0.14
	medial jsn	0.81	0.15	0.95	0.15
	lateral jsn	1.00	0.17	0.92	0.16
	tibial spine spiking	0.46	0.18	0.39	0.18
	medial sclerosis	0.36	0.15	0.45	0.18
	lateral sclerosis	0.34	0.20	0.78	0.20
Other Features	medial subchondral cysts	100%*		100%*	
	lateral subchondral cysts	100%*		100%*	
	medial chondrocalcinosis	1.00	0.20	0.86	0.20
	lateral chondrocalcinosis	1	0.20	0.75	0.19

*Percent agreement was used due to lack of observations required by the kappa statistic

5.6.2 Inter-observer comparison

Inter-observer reproducibility (table 5.6) was not as consistent in a single direction, with approximately 50% showing an increase and 50% a decrease in the kappas between year 15 and year 20 x-rays. There was no discernible pattern as to which kappas increased or decreased by compartment, feature or scoring method.

This analysis shows mixed results, which is likely due to the issues related to the general inter-observer reproducibility which is addressed further in the discussion. Ultimately, neither set of x-rays had significantly different reproducibility than the other, and both showed acceptable levels of reproducibility for most features and scoring methods.

Table 5.6 Inter-observer reproducibility: year 15 digitised vs. year 20 digital x-rays

Method	Individual Features	Y15 X-rays (n=25)		Y20 X-rays (n=25)	
		Kappa (w)	SE	Kappa (w)	SE
K&L	whole knee	0.43	0.11	0.40	0.12
	medial compartment	0.33	0.10	0.38	0.12
	lateral compartment	0.69	0.14	0.41	0.13
OARSI	medial femoral osteophyte	0.26	0.09	0.42	0.13
	lateral femoral osteophyte	0.56	0.14	0.60	0.16
	medial tibial osteophyte	0.45	0.13	0.69	0.14
	lateral tibial osteophyte	0.55	0.14	0.58	0.14
	medial jsn	0.74	0.14	0.63	0.14
	lateral jsn	0.63	0.15	0.62	0.16
	medial sclerosis	0.29	0.14	0.27	0.16
	lateral sclerosis	0.47	0.17	0.62	0.20
Line Drawing	medial femoral osteophyte	0.25	0.11	0.34	0.13
	lateral femoral osteophyte	0.47	0.15	0.57	0.15
	medial tibial osteophyte	0.43	0.13	0.62	0.13
	lateral tibial osteophyte	0.64	0.15	0.41	0.13
	medial jsn	0.34	0.09	0.45	0.10
	lateral jsn	0.39	0.10	0.45	0.11
Chingford	medial tibiofemoral osteo	0.33	0.11	0.48	0.13
	lateral tibiofemoral osteophyte	0.57	0.13	0.51	0.14
	medial jsn	0.77	0.14	0.54	0.13
	lateral jsn	0.84	0.17	0.68	0.16
	tibial spine spiking	0.27	0.13	0.39	0.18
	medial sclerosis	0.29	0.14	0.13	0.16
	lateral sclerosis	-0.06	0.19	0.50	0.20
Other Features	medial subchondral cysts	96.0%*		84.0%*	
	lateral subchondral cysts	96%*		96.0%*	
	medial chondrocalcinosis	1.00	0.20	0.86	0.20
	lateral chondrocalcinosis	1.00	0.20	0.51	0.17

*Percent agreement used due to lack of number of categories required by the kappa statistic

5.7 Summary of results

Intra-observer results remained high for both short (most kappas >0.70) and medium-term (most kappas >0.60) reproducibility. The features with the lowest kappas (although still in the 'acceptable' range) included medial femoral osteophytes, subchondral sclerosis and tibial spiking.

Inter-observer reproducibility had results ranging between 0.29 (Chingford medial sclerosis) and 0.92 (medial chondrocalcinosis), with the majority of kappas greater than 0.50. The features and/or methods exhibiting the lowest kappas included K/L, osteophyte scores across all methods and JSN in the Line Drawing atlas.

Reader drift reproducibility remained good for both readings, with the majority of kappas greater than 0.70 showing that there were no non-systematic changes in grading. Based on cross-tabulations of the same data, there is some evidence of a shift to identify less narrowing after the first 20 x-rays that were read, although that remains consistent throughout the grading of the rest of the study.

The comparison of the grading of plain-film x-ray previously read by DJH and the digitised x-rays read by KL showed poor to acceptable level of inter-observer reproducibility, with kappas ranging between 0.22 and 0.68. This was due to both a lack of training between the two readers, and the use of very different methods (plain-film on a light-box and a digitised x-ray in imaging software). Despite the low kappa values, the percentage of agreement was greater than 85% for all features and methods.

Both the digitised plain-film x-rays and digital x-rays demonstrated good intra-observer reproducibility, although the digital x-rays had higher kappas for almost all features, possibly due to the better image quality. Inter-observer reproducibility showed mixed results for both types of x-rays, which is consistent with the general problematic results for inter-observer reproducibility.

5.8 Discussion

Assessing reproducibility is one of the most important aspects in determining the quality of a diagnostic measure. This section determined the reproducibility for all of the atlas-based ROA scoring methods which will be used to grade the Chingford year 20 x-rays.

5.8.1 Intra-observer reproducibility

The intra-observer reproducibility for K/L (0.84 short-term and 0.79 medium-term) was better or in line with other reported reproducibility for K/L (0.61-0.85) (Kellgren and Lawrence 1957; Spector et al. 1993; Gunther et al. 1999; Cooper et al. 2000; Neame et al. 2004; Gossec et al. 2008). For the OARSI atlas, the reproducibility for some osteophyte grades seemed slightly on the low side (medial femoral osteophyte = 0.56), but is comparable to the only other research which reported reproducibility (0.57) (Nagaosa et al. 2000). Other OARSI grades were comparable to or better than other reported reproducibility (Lanyon et al. 1998; Gossec et al. 2008).

The reproducibility for the Chingford atlas (>0.88 for JSN, >0.79 for osteophytes) was much better than even the original 'official' reproducibility (>0.48 for JSN, 0.49 for osteophytes) and is similar to the other studies who have used this method (Cicuttini et al. 1996; Gunther et al. 1999; et al. Cooper 2000). In the Line Drawing atlas, scores for medial femoral osteophyte (0.55) were lower than all other reported reproducibility for this feature, although the values for JSN were similar (Nagaosa et al. 2000; Neame et al. 2004; Wilkinson 2004).

The low levels of reproducibility found for medial femoral osteophytes when compared to other research indicates that this may have been due to inconsistencies of the grader (KL) and not a reflection of the atlases. This view is confirmed when the reader-drift is reviewed for the same feature. Reader drift was analysed after the reader had more

experience grading, increasing intra-observer for the OARSI atlas to 0.67 and 0.81 for the Line Drawing atlas, both of which are much higher than the original reproducibility and are more in line with other research.

Overall the intra-observer reproducibility results were good and comparable to reported reproducibility from other studies, indicating that the readings subsequently done for the year 20 x-rays are accurate and reproducible.

5.8.1.1 Inter-observer reproducibility

While intra-observer reproducibility was found to be good, inter-observer tended to be in the poor to fair range. The low kappa for K/L grade (0.41) was lower than other reported reproducibility, which ranged between 0.56 and 0.90 (Kellgren and Lawrence 1957; Spector et al. 1993; Gunther et al. 1999; Cooper et al. 2000; Neame et al. 2004; Gossec et al. 2008).

The reproducibility for the OARSI atlas (>0.62 for JSN and >0.35 for osteophyte) grades were much closer to the reproducibility found for other studies, although the osteophyte grade was still on the low side (Lanyon et al. 1998; Gossec et al. 2008; Nagaosa et al. 2000). The results for the Chingford atlas (>0.65 JSN and >0.42 osteophyte) were also closer to the published ranges (Cicuttini et al. 1996; Gunther et al. 1999; Cooper et al. 2000). The Line Drawing atlas reproducibility (>0.40 JSN and >0.30 osteophyte) was lower than other reported reproducibility, but was in the range of Nagaosa et al who found >0.56 for JSN and 0.46 for osteophytes (2000).

This low reproducibility was worrying particularly in reference to K/L grading, which is historically known to have good levels of inter-observer reproducibility. Because of this, a review was made of the x-rays which showed the highest discrepancy between readers.

Table 5.7 Cross-tabulation of K/L grades read by KL and NB

KL K/L grades	NB K/L grades					
	0	**1**	**2**	**3**	**4**	**total**
0	2	0	2	3	0	7
1	2	0	3	3	0	8
2	0	2	6	9	0	17
3	0	0	1	10	2	13
4	0	0	0	0	5	5
total	4	2	12	25	7	50

*Shaded boxes represent the line of identity

The above cross-tabulation (table 5.7) emphasises the consistent differences between readers, specifically NB reading for greater amounts of disease than KL. The category showing the x-rays with the greatest number of disagreements was x-rays graded as a 0, 1 or 2 by KL and as a grade 3 by NB. This indicates that differences between the two readers were predominantly related to the assessment of joint space, with some disagreement on the presence or absence of osteophytes.

After a review of all 15 radiographs in this group, the different grades for 9/15 of the x-rays were explained by a tendency of NB to grade knees showing one compartment to be *narrower* than the other compartment as being positive for narrowing. If the compartments were evaluated independently, without reference to the contralateral compartment, it would not fit the formal definition provided by the atlas definition of narrowing. The remaining 6/15 radiographs from this group did not show a clear-cut reason for the discrepancy for the disagreement.

5.8.1.2 Expert reader for adjudication for inter-observer results

An expert reader was consulted to adjudicate the results in order to produce a consensus grade for the x-rays with high levels of disagreement between KL and NB. Professor Nigel Arden (NA), a specialist in osteoarthritis research, was asked to review the

x-rays in question after a brief review of the official K/L descriptions and atlas, and to give consensus grades.

The expert reader graded each of the 15 radiographs for K/L in addition to grading for the presence or absence of osteophytes in both compartments. The expert reader agreed with KL for 12/15 of the radiographs, NB for 2/15, and disagreed with both for 1/15 radiographs.

Due to these results, it appeared that NB was more likely to judge the joint as a whole for general indications of disease rather than assess component parts of the joint. For example, he was more likely to grade JSN in cases where there were osteophytes present and generally graded more narrowing overall. This may be due to the fact that he was much more familiar dealing with x-rays of painful knees and knees which had indications for surgery, rather than those in a normal population with less disease.

5.8.2 Comparing atlases

In terms of reproducibility the best complete atlas (for JSN and osteophyte grading only) was the Chingford atlas which had slightly higher levels of reproducibility across all tests (intra-observer, inter-observer, short-term, long-term, reader-drift and digitised vs. digital x-rays). Only just behind was the OARSI atlas, edging out the Line Drawing atlas which was let down by the inter-observer and reader-drift for JSN. This order of atlases also matches the level of description from least to most, which is not likely a coincidence. As the atlas adds more levels of disease description, such as by breaking down osteophyte placement by joint edge rather than grouped by compartment, or by increasing the number of possible grades, the more likely it is that there will be some disagreement between readers. Despite this, the levels of reproducibility did not drop to unacceptable levels in the more descriptive atlases, which is encouraging for the use of both the OARSI and Line Drawing atlases.

This research is the first to compare the reproducibility of a variety of atlas-based knee ROA scoring methods within a single study using the same cohort for comparison. While the reproducibility of an atlas is a very important aspect in determining the quality of the method, there are several other factors which are equally important. The following chapters will address each of these factors; how the grading of each atlas compares to one another and affects the calculation of prevalence (chapter 6), the construct validity of each of the methods with pain (chapter 8) and the predictive validity of atlas-based and quantitative ROA with future TKRs (chapter 9).

6 COMPARISON OF ATLAS-BASED ROA SCORING METHODS

6.1 Introduction

One of the most important aspects of assessing radiographic knee OA using an atlas is that not only is the *location* of the feature recorded, but the *severity* as well. Unfortunately, due to a lack of research and understanding about how the severity of the feature relates to clinical variables, this information is often conflated into a simple binary presence/absence variable which undermines many of the benefits of scoring the size and severity of a radiological feature. This aggregation of information may be hiding meaningful disease information and preventing a better association of ROA with clinically important characteristics such as pain and disability.

This chapter will explore some of the key reasons that atlases are not used to their full potential; the lack of understanding about how grades of each atlas relate to one another, whether having a wider range of categories or recording more locations in grading improves disease description, and how different grade cut-offs affect disease prevalence comparisons between scoring methods.

6.2 Methods

6.2.1 Study design

All Chingford subjects present for the year 20 x-ray visits (n= 497) were eligible for this analysis. Exclusion criteria included subjects with a history of rheumatoid arthritis in any joint recorded at any time-point over the course of the study (n= 21). The knees of any subjects with a total or partial knee replacement on or before year 20 were also excluded from the analysis (8 subjects or 35 knees), although the contralateral knee (if without surgery) was eligible to remain in the analysis.

921 knees (462 subjects) were included in the final analysis. One to two knees were excluded in some analyses due to incomplete data.

The comparability of each atlas's grades was evaluated using cross-tabulation for each feature, although the femoral and tibial osteophytes of the OARSI atlas had to be combined into a single osteophyte variable in order to be directly comparable to the Chingford atlas. The highest graded osteophyte in each compartment (either tibial or femoral) was taken to represent the entire compartment for the modified OARSI osteophyte variable.

Composite grades were created by using a cut-off for each feature and combining them to create a new ROA definition. Composite grades were generated for several different iterations of ROA; for compartments containing at least one feature (JSN *or* osteophytes), for compartments containing both features (JSN *and* osteophytes), for either the medial or lateral compartment containing at least one osteophyte (*any* osteophyte), and for either the medial or lateral compartment containing joint space narrowing (*any* JSN). A cut-off of ≥1 was used for all scoring methods and features except for Line Drawing JSN, where a cut-off of ≥-1 was used. These cut-offs and composite scores were determined based on their use in previous research (Hart et al. 1999; Ingham et al. 2011; Peat et al. 2007; Muraki et al. 2009; Cicuttini et al. 1996).

6.2.2 Statistical methods

The baseline characteristics of subjects (knees) attending the year 20 visit were compared to non-attendees, using Man-Whitney U tests for non-normal for continuous data (age and BMI) and Pearson chi-square tests for categorical variables (smoking, pain and baseline K/L grade). Descriptive statistical methods were used for this analysis including cross-tabulations visually displayed using stacked bar charts. Prevalence was calculated using frequencies and percentages. Stata version 12.0 was used for all analyses.

6.3 Cohort characteristics

921 knees had complete radiographic data at year 20 and were included in the analysis. When the baseline characteristics of knees present at year 20 was compared to those lost to follow-up, statistically significant differences were seen between age, BMI, the percentage of smokers and the percentage of subjects with knee pain at year 5. While statistically significant, the difference in BMI between those with complete and incomplete follow-up does not appear clinically significant (with a difference in medians of 0.8 kg/m^2). Those who attended year 20 were significantly younger, were less likely to smoke and reported less knee pain at year 5. However, importantly for this analysis, there was no significant difference between baseline K/L grades.

Table 6.1. Baseline Characteristics of knees with and without complete follow-up at the year 20 visit

	Entire Cohort	Complete Follow-up	Incomplete Follow-up	p-value**
N	2006	921	1085	
Age (years)*	54.0 (49.0, 60.0)	51.0 (47.0, 56.0)	56.0 (51.0, 61.0)	**>0.001**
BMI (kg/m^2)*	24.8 (22.6, 27.6)	24.4 (22.6, 27.0)	25.2 (22.8, 28.2)	**>0.001**
Current Smokers (%)	46.2	39.7	51.6	**>0.001**
Pain at year 5 %)	20.0	16.3	24.0	**>0.001**
Baseline K/L grade:				
0	84.1	86.1	82.2	
1	5.9	5.4	6.4	
2	6.7	6.0	7.3	0.106
3	3.2	2.4	3.9	
4	0.1	0.0	0.2	

*values for age and BMI (body mass index) are medialns with IQRs

p-values are comparing subjects with complete follow-up versus subjects with incomplete follow-up with significant p-values in **bold

6.4 OARSI versus Chingford atlas

Of all the atlas-based knee ROA assessments addressed here, the Chingford atlas and the OARSI atlas are the most similar. Both atlases grade features individually from

zero through grade three (osteophytes and JSN) and include a few additional binary features (tibial spiking, attrition, etc.). The questions remain whether one atlas is 'better' and should be used over the other, and if they can be reasonably compared to one another for OA prevalence, incidence, and progression in cases where they were used to assess different studies. While 'better' is hard to define, in this context, it would make sense to evaluate reproducibility, the amount of information provided by each atlas, and how easy each atlas is to use correctly with minimal training.

Reproducibility was determined in the previous chapter (chapter 5), with the Chingford atlas having slightly greater reproducibility (for both intra- and inter-observer) overall, although they were roughly comparable to one another. When it comes to the amount of information each atlas gives regarding the disease process, however, the OARSI atlas comes out ahead, specifically in regard to osteophyte location where femoral and tibial osteophytes are graded independently. This is also the likely factor for reproducibility being lower in the OARSI atlas.

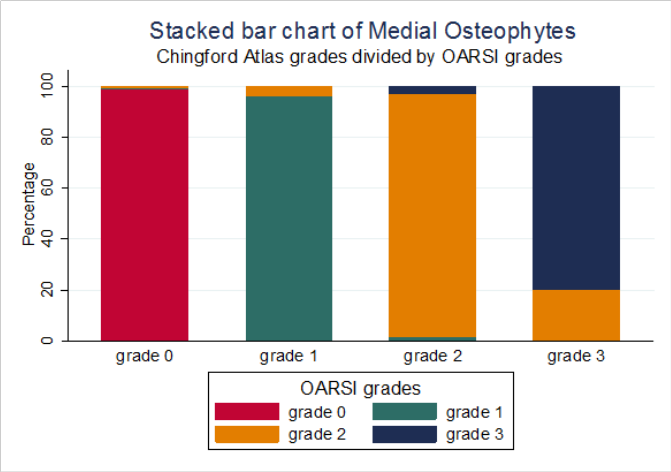
6.4.1 Osteophyte grades

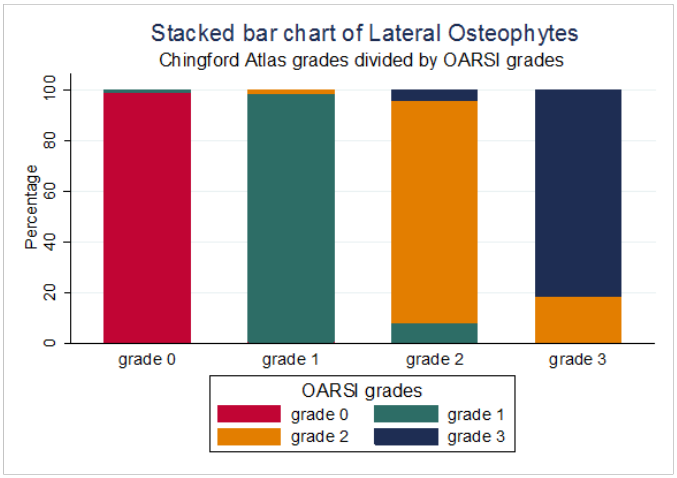
Figure 6.1 shows the comparison of Chingford atlas medial osteophytes versus the modified medial osteophyte variable of the OARSI atlas. For grades 0 through 2, the atlases are similar with only up to a 2.4% difference between the OARSI and Chingford atlas grades. For grade 3 Chingford osteophytes, however, 20.0% of the OARSI grade osteophytes have been recorded as a grade 2. Overall, there was a 96.5% agreement of all OARSI and Chingford atlas medial osteophyte grades.

Lateral osteophytes (figure 6.2) show a similar pattern to the medial osteophyte comparison, although with a slightly lower overall level of agreement (92.8% for all grades). Disagreement between scoring methods occurred more often at the higher grades,

with 12.0% of Chingford grade 2s graded as either a 1 or 3 in OARSI, and 18.4% of Chingford grade 3s graded as an OARSI grade 2.

Figures 6.1 and 6.2 Medial and lateral osteophyte comparison: OARSI and Chingford atlas

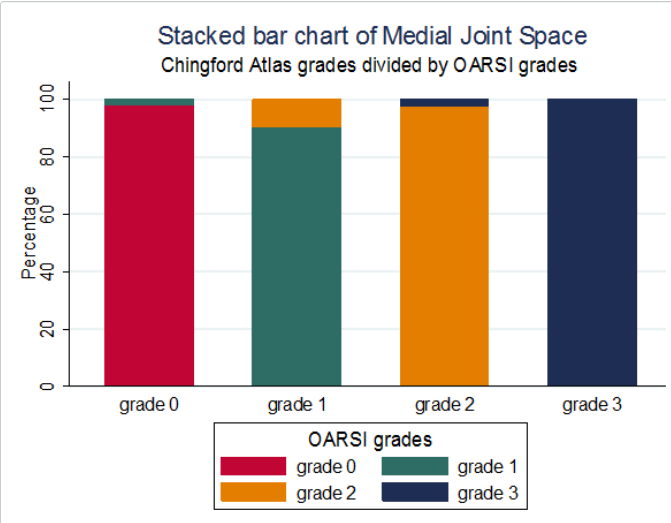


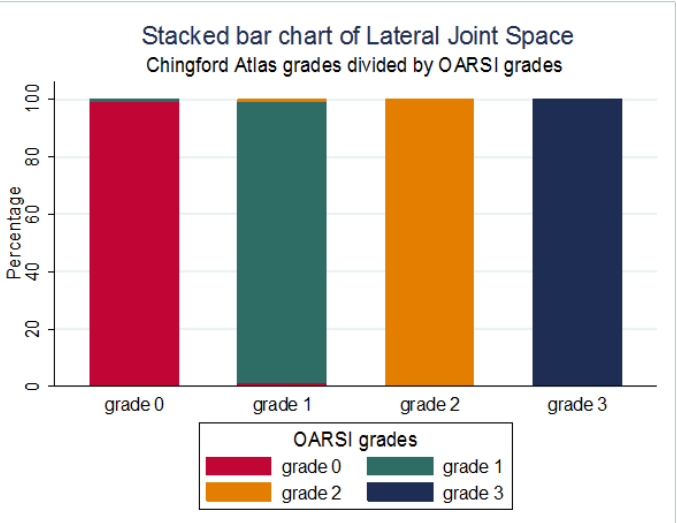


6.4.2 Joint space grades

The medial and lateral joint space grades (figure 6.3 and 6.4) have >97.4% agreement for all grades except for Chingford grade 1 medial JSN, where 10.0% of grade 1s have been scored as a grade 2 OARSI. There were negligible differences between scoring for other binary features of subchondral sclerosis (99.5% agreement).

Figures 6.3 and 6.4 Medial and lateral JSN comparison: OARSI and Chingford atlas





6.5 OARSI versus Line Drawing atlas: are more categories better?

The key differences between the OARSI atlas and the Line Drawing atlas are the increased ranges for both osteophytes (0 to 5) and JSN (-5 to 5), as well as the sex-specific JSN ranges for the Line Drawing atlas. While having expanded ranges of grades provides more information than the narrower ranges of grades of OARSI and Chingford, how those grades relate within a shorter grade range has not been explored.

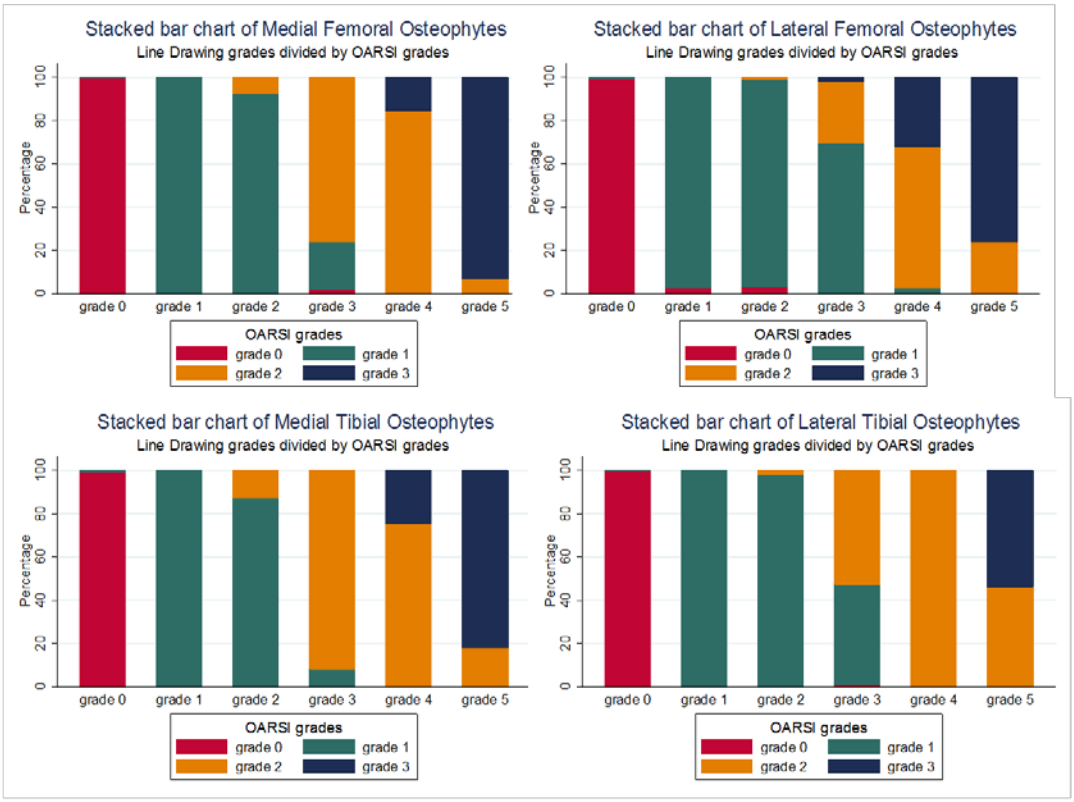
6.5.1 Osteophyte grades

Figures 6.5 and 6.6 show the relationship between medial osteophytes (femoral and tibial) in the OARSI and Line Drawing atlases. They show a similar pattern for both OARSI grade 0 and 3, but much wider for the intermediate grades. For femoral medial osteophytes, a grade 1 OARSI ranged between 0 and 3 Line Drawing, with 31.2% as a grade 1 and 60.1% as a grade 2, and 7.7% as a grade 3 in the Line Drawing atlas. Chingford grade 1 medial tibial osteophytes were similarly spread over the Line Drawing grades 0 through 3, with 31.2% as a grade 1, 60.1% as a grade 2, and 7.7% recorded as a grade 3.

The corresponding Line Drawing grades for an OARSI grade 2 were similarly spread out between Line Drawing grades 2-5. This general pattern of results remained for all osteophyte grades. For medial and lateral JSN, OARSI grade 0 had the widest range of corresponding Line Drawing grades. This was due to the ability of the Line Drawing atlas to capture ‘wide’ joint space in addition to what is considered ‘normal’. Line Drawing grades ranged between -5 and 0 with in an OARSI grade 0 JSN. The line drawing grades for an OARSI grade 1 were relatively narrow with a majority between -1 and 1, while the grade 2 OARSI JSN, as with the osteophyte grades, gave a much wider range with LD grades between 0 and 4. Grade 3 OARSI, which shows distinctive bone-on-bone narrowing, was perhaps surprisingly represented by several Line Drawing grades, 4 and 5.

The lateral osteophytes (figures 6.7 and 6.8) show a slightly different pattern of grade distribution between atlases than the medial osteophytes. The grade 1 OARSI femoral osteophytes are scored as a grade 1 (27.8%), grade 2 (45.8%), and grade 3 (22.2%) on the Line Drawing atlas.

Figures 6.5-6.8 (top left to bottom right) Osteophyte comparison: OARSI and Line Drawing atlas

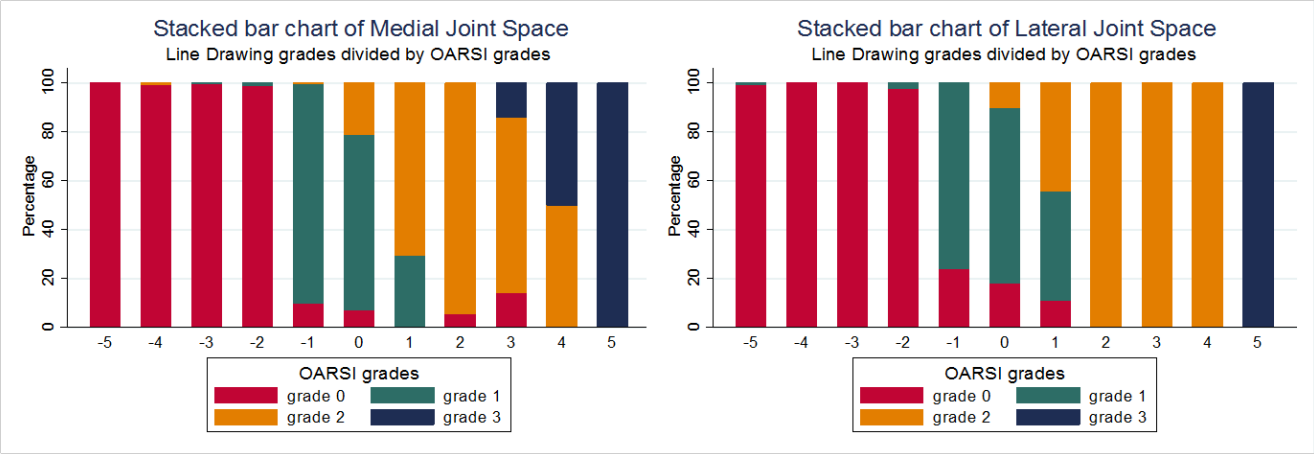


6.5.2 Joint space grades

The figure below shows the cross-tabulation between OARSI and Line Drawing medial (figure 6.9) and lateral (figure 6.10) JSN. The unique design of the Line Drawing atlas capturing a full range of 'normal' joint space is evident by the number of Line Drawing grades which are scored as an OARSI grade 0. Line Drawing grades -5 through -2 compare almost completely to OARSI grade 0 (95.7% medial, 96.4% lateral). OARSI grade 1 is represented by a smaller number of Line Drawing grades, with the majority

(95.8% medial, 91.3% lateral) corresponding to LD grades -1 and 84.1% of medial OARSI grade 2s are represented by Line Drawing grades 0 through 2.

Figures 6.9 and 6.10 Medial JSN and lateral JSN comparison: OARSI and Line Drawing atlas



6.6 Comparison of disease prevalence

One issue in the ROA and pain literature which has not been properly investigated is the effect that using different diagnostic methods for ROA has on disease prevalence and subsequently with ROA's relationship with pain. Although studies have speculated that differences found between the prevalence of ROA in cohorts *may* be due to the method of diagnosis, this limitation has not been addressed.

This section will calculate disease prevalence using four different scoring systems (K/L, OARSI, Chingford and the Line Drawing atlas) as well as looking at composite grades commonly calculated using individual features. These include the definitions discussed in section 6.2.1, for 'JSN *or* osteophytes', 'JSN *and* osteophytes', '*any* osteophyte' and '*any* JSN'.

This analysis eliminates the bias of comparing diagnostic definitions of ROA between different populations, as the ROA data collected for the year 20 Chingford visit is

uniquely suited to answer this research question due to the multiple ways ROA was assessed for each subject in the study.

6.6.1 Kellgren and Lawrence prevalence

Table 6.2 shows the prevalence of radiographic knee OA when using whole knee K/L grade, medial compartment K/L and compartmental K/L grades. The percentage of ROA is shown when using three different cut-offs; ≥1, ≥2, and ≥3. The highest prevalence was found using K/L ≥1 cut-off, with 89.7 percent of subjects having ROA, with only a slightly lower prevalence (74.4%) using the traditional K/L cut-off of ≥2. The more restrictive definition of ≥3, which is often equated to 'severe' OA, gave a more modest prevalence of 32.1%.

Grading each compartment independently using the K/L criteria is not common and therefore prevalence for this has not yet been reported in any study. The medial compartment showed higher prevalence than the lateral for each of the cut-offs, although only slightly higher for both the ≥1 and ≥2 cut-off, with the largest difference found in subjects with K/L ≥3 (25.4% vs. 8.7%).

Table 6.2 Prevalence of K/L grades (percentage)

K/L	N	1+	2+	3+
Whole Knee	921	89.7	74.4	32.1
Medial	921	83.5	64.7	25.4
Lateral	921	77.4	59.8	8.7

6.6.2 Individual feature prevalence

The OARSI and Chingford atlases showed similar prevalence across all comparable categories, including the modified OARSI compartmental osteophyte and the Chingford compartmental osteophyte scores for both cut-offs (≥1 and ≥2) (table 6.3). For methods which score the femoral and tibial osteophytes separately (OARSI, Line

Drawing), the tibial osteophyte seems to be driving the prevalence for the overall compartmental osteophyte prevalence, i.e. tibial osteophytes are almost always present when femoral osteophytes are present. The Line Drawing atlas ≥1 osteophyte definition was directly comparable to the same OARSI and Line Drawing osteophyte definitions, however the ≥2 osteophyte cut-off gave almost double the prevalence for both medial and lateral osteophytes (20.4% OARSI vs. 40.9% LD medial osteophyte; 17.2% OARSI vs. 48.2% LD lateral osteophyte). Line Drawing joint space categories were not directly comparable to any OARSI or Chingford JSN categories, however ≥-1 for medial and lateral JSN was the closest match to the ≥1 cut-off for both Chingford and OARSI (34.6% OARSI vs. 37.0% LD medial JSN; 13.6% OARSI vs. 16.2% LD lateral JSN).

Table 6.3 Prevalence of individual feature scoring methods (percentage)

Method	N	-1	0	1+	2+	3+
Chingford						
Medial Osteophyte	920			64.2	18.8	
Lateral Osteophyte	921			59.5	17.5	
Medial JSN	920			33.3	6.0	
Lateral JSN	921			13.1	2.2	
OARSI						
Medial Fem Osteophyte	919			31.2	11.3	
Medial Tib Osteophyte	920			63.4	16.8	
Medial Osteophyte	920			64.7	20.4	
Lateral Fem Osteophyte	921			23.5	7.8	
Lateral Tib Osteophyte	921			57.3	14.3	
Lateral Osteophyte	921			59.7	17.2	
Medial JSN	920			34.6	8.7	
Lateral JSN	921			13.6	2.3	
Line Drawing						
Medial Fem Osteophyte	919			31.1	24.9	12.0
Medial Tib Osteophyte	920			63.2	34.6	15.1
Medial Osteophyte	920			64.5	40.9	19.8
Lateral Fem Osteophyte	921			23.2	18.8	11.3
Lateral Tib Osteophyte	921			57.2	45.3	18.6
Lateral Osteophyte	921			59.6	48.3	22.7
Medial JSN	920	37.0	16.7	7.5	4.9	2.8
Lateral JSN	921	16.2	6.6	2.4	1.4	1.0

6.6.3 Composite grades

All of the composite definitions created using 'osteophyte *or* JSN' for each atlas give extremely comparable prevalence (table 6.4). This definition also roughly compares the prevalence reported for compartmental K/L with a ≥2 cut-off for each compartment (64.7% medial K/L and 59.8% lateral K/L). 'Any osteophyte' was the best composite score to compare with the traditional K/L≥2 using the whole knee definition. As would be expected by using a more restrictive definition for ROA as defined by both JSN and an osteophyte ('JSN *and* Ost'), prevalence is much lower for all methods, although they are still comparable to each another. The percentages reported for the 'JSN *and* Ost' composite definition of ROA compares to the prevalence reported for the ≥3 cut-off for compartmental K/L (25.4% medial K/L and 8.7% lateral K/L).

Table 6.4 Prevalence for ROA using composite feature scores (percentage)

Method	JSN *or* Ost*		JSN *and* Ost*		Any Ost	Any JSN*
	Medial	Lateral	Medial	Lateral		
Chingford	72.3	63.8	25.2	8.8	74.2	40.9
OARSI	73.3	64.3	26.0	9.0	74.5	42.5
Line Drawing (-1+)	74.3	65.6	27.1	10.2	74.3	45.6

Medial (n=920), Lateral (n=921)

6.7 Summary

As would be expected from how similar each atlas was to use in practice, the Chingford and OARSI atlases showed overall high levels of agreement for all grades (≥92.8% osteophytes, ≥95.7% JSN). The majority of disagreement for osteophyte grades occurred between grades 2 and 3, and grades 1 and 2 for JSN scores.

The comparison between the OARSI and Line Drawing atlases showed a marked difference between the scoring methods for both osteophytes and JSN grades. Grade 0 had the highest agreement between methods, followed by an OARSI grade 3 which was

consistently graded as a LD 4 or 5. OARSI grades 1 and 2 were both represented by a wide range of LD grades (1 through 3 for OARSI grade 1s and 2 through 5 for OARSI grade 2s). For the comparison between joint space grades in the OARSI and LD atlas, OARSI grade 0 corresponded to a wide range of LD grades (-5 through -2). The majority of OARSI grade 1s (≥91.3%) were graded as LD grades -1 or 0, while ≥84.1% of grade 2s are represented by LD grades 0 through 2.

The prevalence for whole knee K/L grade was 89.7% for the ≥1cut-off, 74.4% for the ≥2 cut-off, and 32.1% for the ≥3cut-off. Medial K/L prevalence was up to 9% lower than whole knee for all cut-offs, while lateral K/L was 6% lower for both the ≥1 and ≥2 cut-offs and 17% lower for the ≥3 cut-off. The prevalence calculated for the Chingford and OARSI atlases were directly comparable for both osteophytes and JSN using both cut-off points (≥1 and ≥2). For the scoring methods which grade the femoral and tibial osteophytes independently, the tibial osteophytes seem to be driving the compartmental osteophyte prevalence, with the tibial osteophytes almost always present when the femoral osteophytes are present. The femoral osteophytes did not generally appear in isolation.

The prevalence calculated for the composite ROA definition 'osteophyte *or* JSN' gives roughly the same prevalence for the Chingford, OARSI and LD atlases (range 72.3-74.3% medial, 63.8-65.6% lateral) and is generally similar, although higher, than the ≥2 cut-off compartmental K/L prevalence (64.7% medial, 59.8% lateral) which would be the most comparable K/L definition. The ROA composite definition 'osteophyte *and* JSN' again gives similar prevalence with a tight range of values between the three scoring methods with individual features (range 25.2-27.1% medial, 8.8-10.2% lateral) and is very similar to the ≥3 cut-off compartmental K/L prevalence (25.4% medial, 8.7% lateral) with the medial prevalence falling inside the range for the composite definition and the lateral prevalence just under the range of prevalence calculated.

6.8 Discussion

The Chingford atlas and OARSI atlas (with modified osteophyte grades) were found to have very high agreement (≥92.8%) both for individual features and compartments. This makes sense due to the fact that both use the same grading scale (0 through 3), although with different atlas images. The highest agreement occurred for grades 0 and 1 osteophytes and JSN which was likely due to grade 1 being almost a presence/absence grade. Higher osteophyte grades (2 and 3) had greater differences in agreement which were not found in the JSN grades. The major differences between methods are the images used to illustrate osteophyte grades. The Chingford atlas uses grading images for osteophytes which switch between tibial osteophytes and femoral osteophytes of very different shapes to illustrate the increasing severity in each compartment, making the grades much more difficult to interpret accurately than the OARSI atlas.

There are major differences between the OARSI (and therefore Chingford) atlas and the Line Drawing atlas, with the most dominant differences being the increase in grading scale for osteophytes (from 4 to 6) and a range of negative joint space grades for capturing JSN in the 'normal' range. The increase in grades addresses criticisms of other atlases where it is difficult to capture progression of features due to the large jump in severity between each grade (Felson et al. 2008; Guermazi et al. 2009). This would also affect the sensitivity in capturing variations in severity in cross-sectional analyses. This is especially important as increasing grade has been shown to be associated with an increasing risk of pain (Lethbridge-Cejku et al. 2005; Muraki et al. 2009; Neogi et al. 2009) and therefore the ability to distinguish between different levels of severity is key in helping improve the understanding of the ROA and pain relationship.

The other unique aspect of the Line Drawing atlas is the use of negative joint space grades. Categorical JSN grades with a negative range have been explored previously, but without the benefit of associated images (Schouten et al. 1992). For the OARSI and Chingford atlases, graders have to make a subjective judgement as to when a knee has 'normal' joint space and when this changes to 'abnormal'. This range of normal space can vary by age, sex, and x-ray protocol and therefore the atlas image for 'normal' joint space cannot reflect these variations (Dacre et al. 1991; Buckland-Wright et al. 1995). It is then up to the grader to make an educated decision based on the characteristics of the cohort they are studying as to when joint space becomes abnormal. The Line Drawing atlas removes several of these issues; firstly by having sex-specific JSN grades, and secondly by having a range of grades capable of capturing a wide range of possible joint space regardless of the population. In a way, this is a categorical equivalent to a continuous joint space measure. A third benefit is the ability to capture progressive widening of joint space. This is a known phenomenon, particularly in the lateral compartment, which often occurs at the same time as narrowing in the medial compartment (Felson et al. 2008; Duryea et al. 2010).

The prevalence of K/L≥2 in the Chingford cohort at year 20 (74.4% of knees) is high, but is comparable to prevalence in population-based cohorts with a similarly elderly population (mean age 72.5 (SD 5.5)). In a large-scale cohort based in Japan, 70.2% of women (mean age 74.0 (SD 6.4) had K/L≥2 (Muraki et al. 2009), while in a UK-based cohort of men and women (mean age 65.2 (SD 8.6)), 68.0% had K/L≥2 (Duncan et al. 2007). Several US-based cohorts have reported lower prevalence, one identifying 34.0% of women in the Framingham cohort (mean age 73) (Felson et al. 1987) and only 6.1% in the national health and nutrition examination with a much lower age range (35-74) (Anderson et al. 1988). A Dutch cohort found a prevalence somewhere in the middle, with 40.2% of

women aged 70 to 79 (van Saase et al. 1989). There appears to be a gap in the reported prevalence between older studies (Felson et al. 1987, van Saase et al. 989) and the most recent population studies (Duncan et al. 2007, Muraki et al. 2009) independent of mean age and location, suggesting that new methodology (x-ray acquisition and assessment) is possibly influencing the resulting prevalence. This is explored further in this current research in chapter 7.

In the previous chapter (chapter 6), the merits of each atlas were evaluated in respect to their reproducibility. With all things being (relatively) equal in those respects, the benefits of each method comes down to the detail and level of information given by each scoring method and whether it accurately captures the structural features of knee OA (location, type, size) in a meaningful way.

In comparing the atlases, it was found that the OARSI and Chingford atlas are the most comparable grade for grade and would calculate the same disease prevalence regardless of the cut-offs chosen. The Line Drawing atlas is only comparable to these atlases using osteophyte scores grouped into the lowest score. The Line Drawing JSN grades are very different from the more traditional atlases, and are only comparable if a careful choice of the grade where 'normal' joint space changes to 'narrowing'. In this case a cut-off of -1 gave the closest result for prevalence in the medial and lateral compartments, although at least 3-4% higher. Prevalence created using different K/L grades as cut-offs, compared reasonably well to the prevalence calculated using composite grades, with 'any osteophyte' comparing directly to K/L≥2, and 'JSN *and* Ost' comparing to the compartmental K/L≥3 grades. This section reviews the comparability of these methods to each other, and whether there is a difference in the amount of disease detected when using the respective scoring methods.

7 NATURAL HISTORY OF RADIOGRAPHIC KNEE OA

Natural History is a systematic account of a chronological disease process based on observation rather than experimentation. It describes the disease from exposure through progression and final outcome, providing a framework for identifying methods for preventing and controlling disease progression.

Research into the natural history of radiographic knee osteoarthritis has primarily focused on symptomatic patients (Thorstensson et al. 2008; Pavelka et al. 2000; Massardo et al. 1989), and the progression of disease in older community-based cohorts (Felson et al. 1995; Felson et al. 1997). Population-based studies have been limited by relatively short follow-up times that vary between 3 and 12 years (Lachance et al. 2002; Schouten et al. 1992). These studies also lack repeated measures between baseline and follow-up, which may lead to grouping slow and fast progressing knees into the same group

This section will analyse the 19-year prevalence, incidence and progression/worsening of radiographic knee osteoarthritis in the Chingford cohort. The analysis includes atlas-based scoring methods and composite definitions to establish ROA (K/L and the Chingford atlas), and the novel use of incremental 5-year data between baseline and follow-up to better understand the pattern of disease incidence and progression over time. The 14-year natural history (baseline through year 15) of this cohort has been previously published (Leyland 2012) and the full article is available in **appendix 1.1**.

The strengths of this study is that it is the longest natural history of radiographic knee osteoarthritis to date, it has repeated measures of follow-up, it has a large proportion of disease free subjects at baseline, and it evaluates both whole-knee as well as individual feature progression.

7.1 Methods

7.1.1 Subjects

All subjects present for x-rays for each clinic visit between baseline and year 20 and with no evidence of rheumatoid arthritis (RA) were eligible for inclusion in this analysis (n=834 knees). Subjects with TKRs were included in the analysis for K/L prevalence, incidence and progression/worsening as a proxy for K/L ≥2, so as to prevent underestimation of the overall disease burden. TKRs were excluded from the individual feature and compartmental analyses since it was not possible to extrapolate which compartment, feature or location was the driving force for the structural presence of OA, and therefore the results may be slightly underestimated in these sections. 834 knees were available for compartmental and individual feature analysis for baseline through year 10, 825 knees for year 15 and 805 for year 20.

K/L grades and the Chingford atlas were used to evaluate ROA to maintain consistency within the cohort, because all previous x-rays from baseline through year 15 had been analysed using these scoring methods. Age groups were determined by grouping subjects into 5-year age categories; 45-49, 50-54, 55-59 and 60-64. BMI groups were determined using the WHO categories with normal subjects having a BMI less than 25.0 kg/m^2 and overweight subjects with a BMI 25.0 kg/m^2 or higher. Prevalence, incidence and worsening were all stratified by age group, BMI group and pain. Incidence was not analysed due to the small number of knees with disease at baseline.

The methods used to calculate prevalence, progression and worsening are described in detail in chapter 3.

7.1.2 Statistical methods

Chi-squared tests were used to assess differences in age, BMI and pain between groups. Fisher's exact tests were used in cases where there was a small cell count. Linear trends were tested using the Cuzick extension of the Wilcoxon rank-sum test, which is a non-parametric test for trend across ordered groups (Kirkwood and Sterne 2003).

7.1.3 Novel data collection: year 20 data

Although the year 20 data is presented in the following analysis alongside the previously collected Chingford x-ray data (baseline through year 15), there are several major caveats to interpretation of the 20-year data. Different methods were used to take the x-rays and to assess K/L for the year 20 visit (see section chapter 3). Digitised x-rays were taken only at year 20, and which had greater clarity and definition then the plain-film x-rays used for the previous visits. The x-rays were also assessed using a computer program (see chapter 4) with the capability to zoom and focus on small aspects of the x-ray providing much greater detail of the joint than would be possible looking at a plain-film x-ray on a light box.

An analysis comparing DJH's scoring on 25 plain-film x-rays using traditional methods (light-box) to the same x-rays in a digital format scored by KL is described in chapter 5. The digitised plain-film x-rays were originally selected for the reproducibility analysis, and therefore represent a wide variety of ages, BMI and severity of ROA. These new readings using the digital methods were then compared to the grades originally given by DJH using traditional methods during the normal grading of the x-rays for the study year. Reproducibility between graders (and methods) was found to be in the poor to acceptable range (weighted kappa 0.22-0.68), but the percentage of agreement was good (≥85.0%) for all features.

Thirty-two percent of K/L grades were higher when scored using the new digital methods versus the traditional plain-film grading techniques, while 16.0% were lower. Twenty-four percent of medial osteophytes were graded as higher using digital methods, with 16.0% graded lower as a 0 rather than a grade 1. Lateral osteophytes had 20.0% higher grades and 16.0% lower grades using the new methods. Medial JSN had 20.0% higher grades and only one (4.0%) x-ray graded as a 0 instead of a grade 1. Lateral JSN showed a very similar pattern to medial JSN, with 12.0% graded higher and 4.0% as lower.

Although there will be an obvious jump between the baseline through year 15 data and the year 20 data due to the different scoring methods, the year 20 data was included into this analysis. The original intention of this work was to produce a 20-year natural history, and short of rereading 4000 knee x-rays (which was outside the scope of this study), combining the data was the best option. This methodological change can be viewed in the same light as studies which dramatically change x-ray protocols from extended to semi-flexion over the course of a study and due to the type of rare longitudinal data, present the data regardless (Thorstensson 2008). Both methods are correct, but are known to give slightly different results. Due to this difference, however, the results below have been given between baseline and year 15, and between baseline and year 20 in order to clearly highlight the differences between the data.

7.2 Prevalence

7.2.1 Kellgren and Lawrence

Out of all knees (n=834), ROA prevalence (K/L $\geq$ 2) was 9.0% at baseline, 17.3% at year 5, 27.7% at year 10, 36.5% at year 15 and 74.5% at year 20 (data included in figure 7.1). Using the incremental (repeated measure) data, prevalence increased by 2.1%

between year 1 and year 5 (4-years), 1.7% between year 5 and 10, 2.2% between 10 and 15, and 7.6% between year 15 and 20. When OA was determined using a more severe definition of radiographic changes (K/L ≥ 3), prevalence was 2.6% at year 1, 4.9% at year 5, 7.8% at year 10, 17.0% at year 15 and 34.4% at year 20.

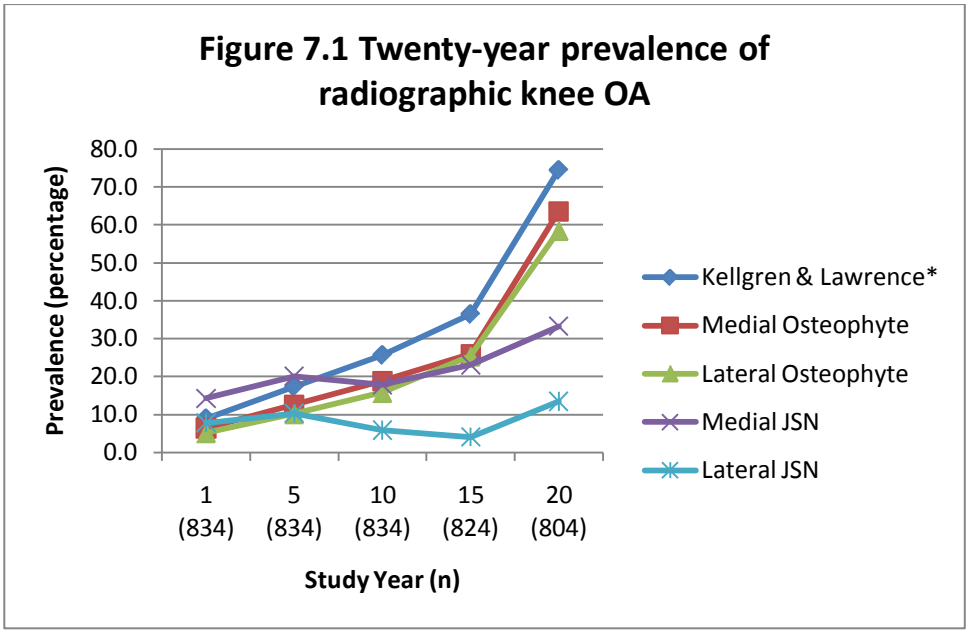
7.2.2 Individual radiographic features

Table 7.1 shows the prevalence of medial osteophytes, lateral osteophytes, medial joint space and lateral joint space at each clinic visit. Medial compartment features show a slightly higher prevalence at each time point, with prevalence of lateral osteophytes between 0.5% and 5.0% lower than medial osteophytes. Medial JSN has a much higher prevalence than lateral JSN at all time-points, which has a prevalence up to 19.8% lower than medial JSN.

Table 7.1. Prevalence of individual ROA features by compartment

	Y1	Y5	Y10	Y15	Y20
N	834	834	834	824	805
Medial Osteophyte	6.4	12.6	18.8	25.9	63.5
Lateral Osteophyte	5.2	10.2	15.8	25.4	58.5
Medial JSN	14.3	20.1	17.8	23.1	33.3
Lateral JSN	7.8	10.3	5.9	4.1	13.5

Figure 7.1 illustrates the temporal trends in prevalence for K/L as well as the individual compartment features. The upward trends are clear for K/L and osteophytes in both compartments for the first 15 years, with a sharp jump in prevalence up to year 20. The temporal trend for joint space narrowing is not as clear in either compartment. Medial JSN flattens out through year 10 and 15 before increasing in year 20, while lateral JSN shows a slight downward trend until year 15, when it increases again at year 20.



*K/L was calculated including TKRs with n of 834 for all years

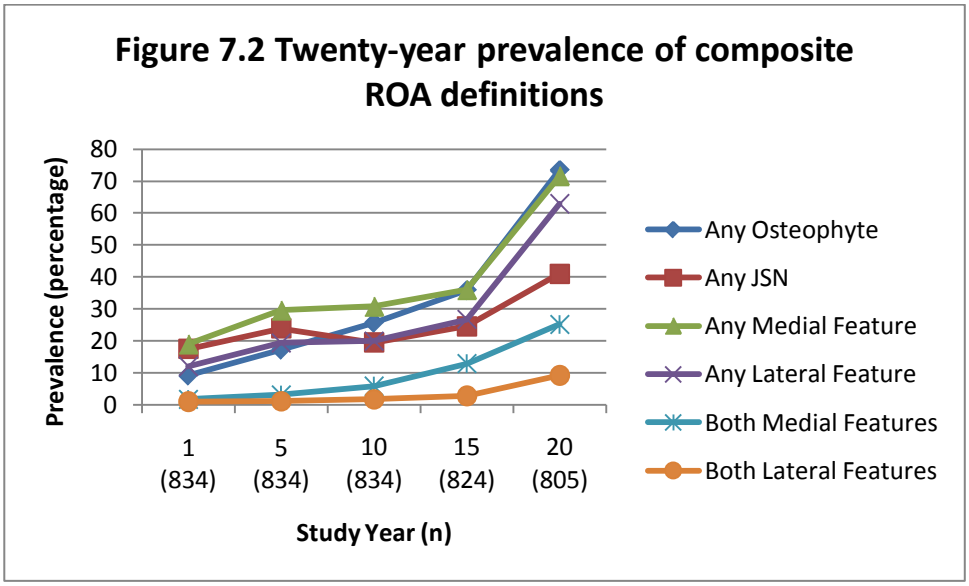
7.2.3 Composite ROA grades

Table 7.2 shows the prevalence of composite ROA definitions created using the individual feature scores of the Chingford atlas. The prevalence of 'any osteophyte' increases for each visit over the 20 years starting from 9.2% at baseline and ending at 73.5% by year 20. 'Any JSN' remains relatively stable in comparison, ranging between 17.5% and 24.6% for the first 15 years and then increasing to 41.0% by year 20. Prevalence for any medial feature is approximately 10.0% higher than any lateral feature at each time point. The definition requiring both osteophytes and narrowing in the lateral compartment has the lowest prevalence of all the composite ROA definitions.

Table 7.2 Prevalence of composite ROA definitions: baseline through year 20

	Y1	Y5	Y10	Y15	Y20
N	834	834	834	824	805
Any Osteophyte	9.2	17.3	25.7	35.9	73.5
Any JSN	17.5	23.9	19.5	24.6	41.0
Any Medial Feature	19.0	29.6	30.8	36.0	71.6
Any Lateral Feature	12.0	19.4	20.0	26.7	62.9
Both Medial Features	1.7	3.1	5.8	12.9	25.2
Both Lateral Features	1.0	1.1	1.7	2.8	9.2

Figure 7.2 demonstrates the prevalence of the composite ROA definitions over the 19-year study period. As with the individual feature prevalence, there is a noticeable jump in prevalence at year 20 due to the change in grading methods. The prevalence for the definitions of 'any osteophyte', 'any medial feature' and 'any lateral feature' all increase through each study year through year 20. The definitions for 'any JSN', 'both medial features' and 'both lateral features' either increase only modestly or remain stable until year 20.



7.2.4 ROA prevalence stratified by age

The prevalence of K/L (≥ 2) was significantly different between the four age groups for each visit year ($p \leq 0.001$) except for year 20 ($p=0.159$) (**table 7.3**). There was a significant trend for increasing prevalence from the youngest to oldest age groups for baseline to year 15 ($p \leq 0.001$) but not year 20 ($p=0.135$).

Prevalence of medial osteophytes (≥1) were significantly different for all age groups ($p \leq 0.023$), except for year 20 ($p=0.334$). There was also a trend for increasing prevalence between the youngest and oldest groups for all visit years ($p \leq 0.008$), except for visit year 20 ($p=0.315$). The prevalence of lateral osteophytes was significantly different between age groups ($p \leq 0.003$), with an increasing trend from youngest to oldest groups ($p \leq 0.007$) for all visit years. For medial and lateral joint space narrowing, only year 20 showed any significant difference in prevalence between age groups ($p \leq 0.008$), with an increasing trend with increasing age group ($p \leq 0.022$). Prevalence for lateral joint space at year 5 did not show any significant difference between age groups ($p=0.089$), but an increasing trend from youngest to oldest age group was found to be significant ($p=0.023$).

Table 7.3 Prevalence (%) by visit year stratified by age groups

Feature/Method	Visit year	Age groups			
		45-49	**50-54**	**55-59**	**60-64**
N		*300*	*250*	*158*	*126*
K/L	1	5.0	6.8	11.4	19.8
	5	10.7	13.2	21.5	35.7
	10	17.0	22.0	31.7	46.0
	15	28.0	34.4	38.0	58.7
	20	73.3	73.2	72.2	82.5
Medial osteophytes	1	4.3	4.8	8.9	11.1
	5	9.0	9.2	16.5	23.0
	10	12.0	16.0	25.3	32.5
	15	19.0	25.6	31.7	41.3
	20	63.7	64.8	61.4	71.4
Lateral osteophytes	1	2.0	4.4	5.7	13.5
	5	5.3	8.0	12.0	23.8
	10	10.3	14.4	19.0	27.8
	15	18.7	27.2	24.7	44.4
	20	56.3	59.6	55.7	74.6
Medial JSN	1	10.3	16.0	18.4	15.1
	5	19.3	20.4	22.2	19.1
	10	15.0	20.8	19.0	16.7
	15	20.3	24.8	26.6	27.8
	20	28.3	41.6	36.7	39.7
Lateral JSN	1	8.0	7.2	7.0	9.5
	5	13.7	8.8	9.5	6.4
	10	5.4	5.6	7.0	6.4
	15	4.0	4.8	8.9	4.8
	20	10.3	17.6	24.7	19.1

7.2.5 ROA prevalence stratified by BMI

There was a significant increase for prevalence of K/L in overweight subjects (baseline BMI) for all years compared to normal subjects ($p \leq 0.001$) (table 7.4). Overweight subjects had a higher prevalence of both medial ($p \leq 0.001$) and lateral osteophytes ($p \leq 0.02$) for all visit years. The prevalence of medial JSN was significantly higher in overweight subjects for visit 5, 15 and 20 ($p \leq 0.05$), and a significant difference for prevalence of lateral JSN was only found for visit 1 and visit 20 ($p \leq 0.02$).

When baseline BMI was analysed as a continuous variable, the risk of having prevalent K/L ≥ 2 at each time point was 15% to 20% higher for each unit increase of BMI (Y1: OR 1.15 (95%CI 1.09, 1.23); Y5: OR 1.15 (95%CI 1.10, 1.21); Y10: OR1.17 (95%CI 1.12, 1.22); Y15: OR 1.16 (95%CI 1.11, 1.21); Y20: OR 1.20 (95%CI 1.14, 1.27)). The risk of having prevalent osteophytes increased by between 13-18% medially and between 19-29% laterally for each unit increase of BMI, and was significant for all visits. The risk of prevalent joint space narrowing increased by between 12-13% medially (for visits 10, 15 and 20) and between 10-13% laterally (for visits 15 and 20).

7.2.6 ROA prevalence stratified by pain

When the prevalence of K/L ≥ 2 was stratified by NHANES knee pain (evaluated at each visit, excluding baseline due to a lack of this information), a significant difference was seen between the pain and pain-free knees at each visit year ($p \leq 0.003$).

The prevalence of both medial and lateral osteophytes was significantly increased in knees with pain for all visit years ($p \leq 0.01$). There was a difference between medial narrowing in knees with and without pain at visits 10 through 20 ($p \leq 0.008$), and for lateral narrowing at visits 10 and 20 ($p \leq 0.002$).

Table 7.4 Prevalence (%) by visit year stratified by BMI groups and pain

Feature/Method	Visit year	BMI group		Pain	
		Normal	Overweight	No Pain	Pain
N	1	464	370		
	5			600	122
	10			580	248
	15			549	280
	20			676	127
K/L	1	5.8	13.0		
	5	12.3	23.5	13.5	34.4
	10	18.3	34.9	20.0	37.1
	15	29.5	45.1	28.8	52.1
	20	69.0	81.4	71.8	84.3
Medial osteophyte	1	3.7	9.7		
	5	8.6	17.6	9.8	23.0
	10	12.9	26.2	14.1	28.2
	15	19.8	35.4	19.5	41.4
	20	58.6	72.4	62.0	74.0
Lateral osteophyte	1	3.5	7.3		
	5	6.7	14.6	7.5	23.8
	10	10.6	22.4	11.0	25.4
	15	19.6	34.6	19.3	40.4
	20	54.1	67.3	55.3	80.3
Medial JSN	1	12.7	16.2		
	5	17.0	24.1	20.8	19.7
	10	17.2	18.4	15.3	23.8
	15	21.1	27.6	21.3	29.6
	20	32.8	39.2	32.3	50.4
Lateral JSN	1	5.8	10.3		
	5	8.8	12.2	10.7	9.0
	10	4.5	7.6	4.2	10.1
	15	4.3	6.5	4.4	7.1
	20	12.5	21.6	14.1	28.4

7.3 Incidence

7.3.1 Kellgren and Lawrence

At baseline, there were 759 knees (out of 834) that were not classified as having disease defined as K/L≥2. The annual cumulative incidence increased steadily between

each five-year period, from 2.3% between baseline and year 5, 2.2% between year 5 and 10, 3.2% between years 10 and 15, and 12.1% between year 15 and 20. 30.4% percent of knees developed incident ROA between baseline and year 15, while 72.1% developed K/L≥2 between baseline and year 20. The annual cumulative incidence between baseline and year 15 was 2.2% and 3.8% between baseline and year 20.

7.3.2 Individual radiographic features

Table 7.5 shows the incidence of individual ROA features between each x-ray visit. Incidence of medial JSN and osteophytes are similar for each visit until year 20 incidence, where osteophytes occur 31.0% more than JSN. This likely reflects the difference found in grading year 20 x-rays compared with year 15 x-rays discussed previously in section 1.1.3. Overall incidence remains relatively low during each time-point (between 2.0 and 13.1%), until year 20 where medial osteophytes have an incidence of 52.5%.

Table 7.5. Five-year incidence of individual ROA features

	N	Year 5	N	Year 10	N	Year 15	N	Year 20
Medial Osteophytes	781	6.8	729	8.9	677	12.3	611	52.5
Lateral Osteophytes	791	5.4	749	6.9	702	13.1	615	46.5
Medial JSN	715	7.6	666	8.0	686	12.7	634	21.5
Lateral JSN	769	3.9	748	2.0	783	2.2	790	11.4

Table 7.6 shows the incidence between baseline and year 15, with annual cumulative incidence of 1.6% for medial osteophytes, 1.5% for lateral osteophytes, 1.2% for medial JSN and 0.2% for lateral JSN. The 20-year incidence (baseline to year 20 only) gives annual cumulative incidence of 3.2% for medial osteophytes, 2.9% for lateral osteophytes, 1.5% for medial JSN and 0.6% for lateral JSN.

Table 7.6 Fifteen and nineteen-year incidence of individual ROA features

	N	Y1-Y15	Y1-Y20
Medial Osteophytes	**781**	21.8	60.3
Lateral Osteophytes	**791**	21.6	55.0
Medial JSN	**715**	16.2	28.7
Lateral JSN	**769**	2.6	10.5

7.3.3 Composite ROA grades

The highest incidence for ROA using composite definitions (table 7.7) at all time-points is for 'any medial' feature up until year 20 where 'any osteophytes' has the highest incidence with 75.2%. The definition requiring the presence of both osteophytes and JSN in a single compartment remains low for all years, with the highest incidence of this definition occurring at year 20, with 16.2% for 'both medial' and 9.6% for 'both lateral'.

Table 7.7 Five-year incidence of composite ROA definitions

	N	Year 5	N	Year 10	N	Year 15	N	Year 20
Any Osteophytes	757	9.0	690	11.0	620	16.1	528	75.2
Any JSN	688	8.9	635	7.9	671	13.4	621	29.8
Any Medial	676	13.9	587	13.5	577	16.3	527	59.6
Any Lateral	734	9.4	672	8.6	667	13.6	604	52.8
Both Medial	820	1.6	808	4.0	786	9.7	718	16.2
Both Lateral	826	0.5	825	0.7	818	2.8	801	9.6

Based on the incidence presented in table 7.8, annual cumulative incidence between baseline and year 15 is 2.1% for 'any osteophyte', 1.2% for 'any JSN', 2.0% for 'any medial', 1.7% for 'any lateral', 0.9% for 'both medial', and 0.3% for 'both lateral'. The annual cumulative incidence between baseline and year 20 grades is 3.7% for 'any osteophyte' 1.8% for 'any JSN', 3.6% for 'any medial', 3.2% for 'any lateral', 1.4% 'both medial' and 0.6% for 'both lateral'.

Table 7.8 Fifteen and nineteen-year Incidence of composite ROA definitions

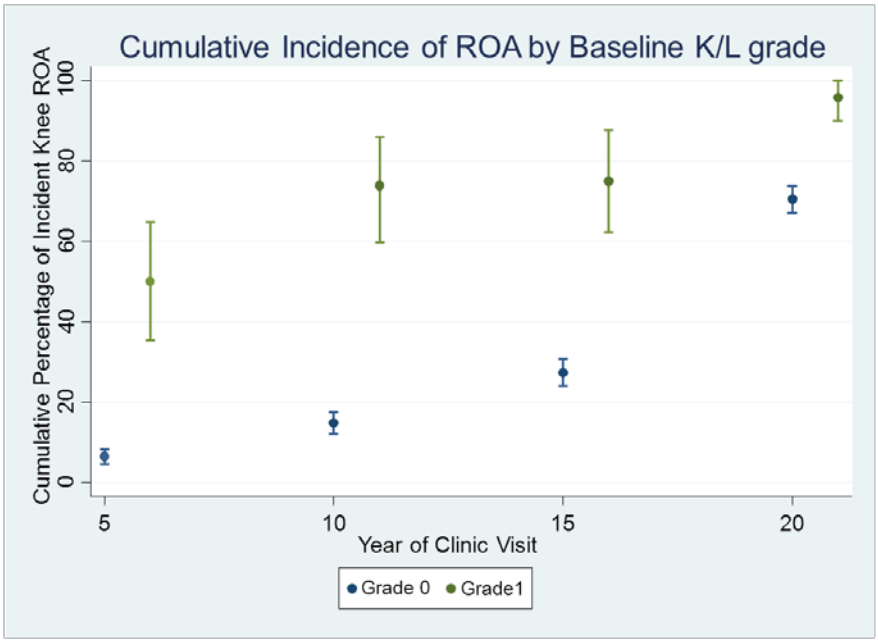
	N	Y1-Y15	Y1-Y20
Any Osteophytes	757	29.9	72.0
Any JSN	688	17.2	34.4
Any Medial	676	28.3	68.5
Any Lateral	734	23.3	61.6
Both Medial	820	12.8	27.3
Both Lateral	826	3.6	12.0

7.3.4 Incidence of K/L stratified by covariates

7.3.4.1 Incidence of K/L stratified by baseline grade

While a K&L grade 1 is not considered diagnostic of ROA, when stratifying by an initial baseline grade of 0 or 1 (figure 7.3), there is a significant difference in cumulative incidence between the groups at each visit ($p<0.001$). The odds of a baseline grade 1 developing incident ROA by year 15 is 4.8 (95% CI 2.8, 8.1) times the odds for a baseline grade 0, while the odds increase to 5.2 (95% CI 2.1, 13.3) by year 20.

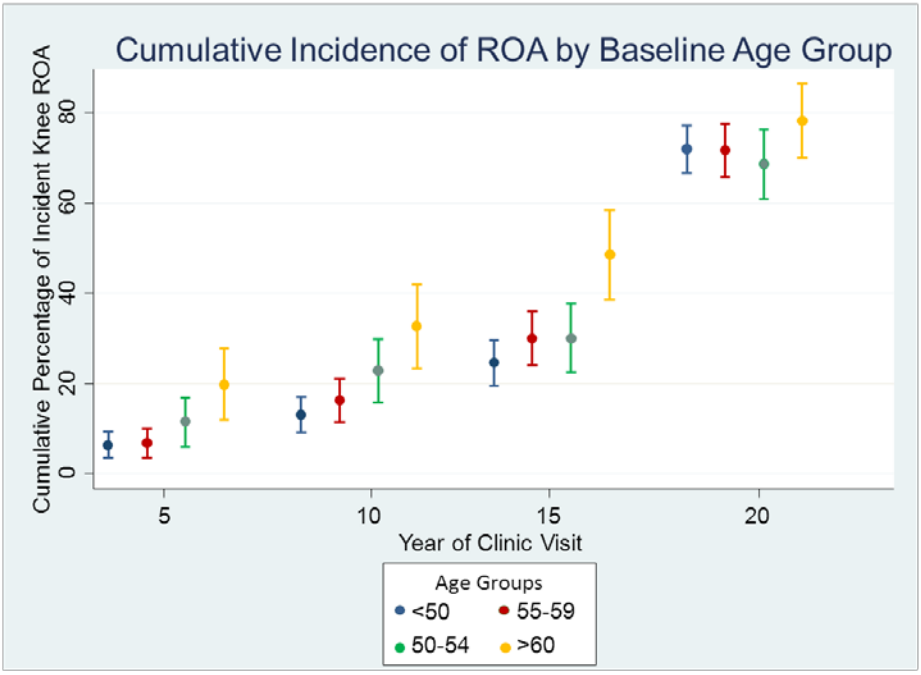
Figure 7.3 Cumulative incidence of ROA (K&L≥ 2), baseline grade 0s versus 1s



7.3.4.2 K/L Incidence stratified by age

When 5-year cumulative incidence of ROA was examined within each visit year and stratified by age group, a linear trend ($p<0.02$) is evident for years 5 and 10 with the oldest age group having the highest percentage of incident ROA, but this trend is not significant for incident ROA by age group at year 15 or year 20 (figure 7.4). Between baseline and year 15, 24.6% of subjects with a baseline age under 50 developed incident ROA, 30.0% of subjects aged 50 to 54, 30.0% of 55 to 60 year olds and 48.5% of subjects with a baseline age greater than 60. Between baseline and year 20, 71.9% of subjects with a baseline age less than 50 had developed incident ROA, 71.9% of subjects aged 50 to 54, 68.6% of 55 to 60 year olds and 78.2% of subjects who had a baseline age greater than 60. There was a significant difference found between the youngest and oldest age groups ($p<0.001$), in addition to between the two middle age groups ($p<0.02$).

Figure 7.4 Cumulative incidence of ROA (%) stratified by age groups (n= 759 knees)

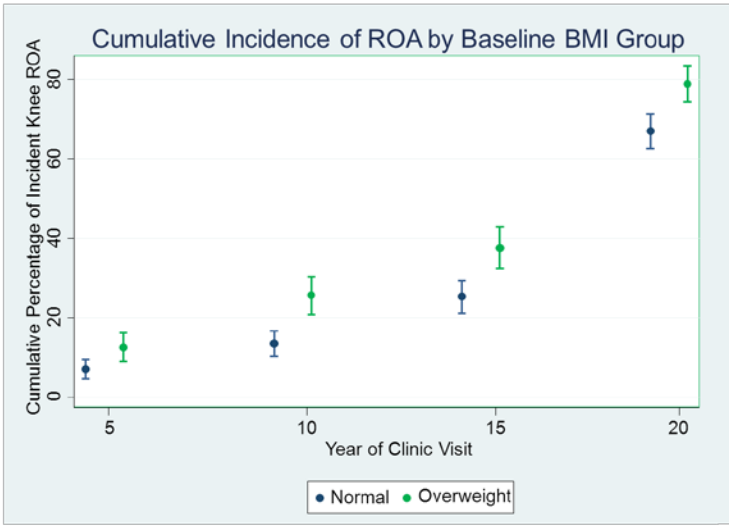


7.3.4.3 Incidence of K/L stratified by BMI

Incidence between subjects with a normal baseline BMI and subjects who were overweight was significantly different at year 5, year 10 and year 20 ($p<0.02$). There was no difference between groups for year 15 ($p=0.429$). When cumulative incidence was examined between normal and overweight subjects (figure 7.5), it showed a significant difference between overweight and normal subjects at each year ($p<0.009$).

For each unit increase of baseline BMI (as a continuous measure), knees had a significantly increased risk of incident K/L ≥ 2 of 14% by year 5, 16% by year 10, 15% by year 15 and 18% by year 20.

Figure 7.5 Cumulative incidence of ROA (%) stratified by BMI groups (n= 759 knees)



7.3.5 Cumulative incidence of ROA individual features stratified by covariates

7.3.5.1 Incidence of individual features stratified by age

The cumulative incidence of medial and lateral osteophytes (table 7.9) was significantly different between age groups at all visit years ($p\leq0.05$) except for medial

osteophytes at year 20 (p=0.329). A trend of increasing incidence between the youngest and oldest age groups was also found for all years except for medial osteophytes at year 20. Cumulative incidence of narrowing was the same for age groups at all years except for medial narrowing at year 20 (p=0.049), and lateral narrowing at year 5 (p=0.001) and year 20 (p≤0.001).

Table7.9 Cumulative incidence by visit year stratified by age

Feature/Method	Visit year	Age groups			
		45-49	**50-54**	**55-59**	**60-64**
N		*287*	*238*	*144*	*112*
Medial osteophyte	1				
	5	5.23	4.62	8.33	13.39
	10	9.41	12.61	18.75	25
	15	16.03	22.69	25.69	34.82
	20	62.02	63.45	57.64	68.75
N		*294*	*239*	*149*	*109*
Lateral osteophyte	1				
	5	3.74	3.77	6.71	11.93
	10	8.5	10.46	14.09	17.43
	15	17.01	24.27	20.13	35.78
	20	55.44	57.74	53.02	70.64
N		*269*	*210*	*129*	*107*
Medial JSN	1				
	5	10.41	5.71	6.98	4.67
	10	9.29	12.38	9.3	7.48
	15	14.87	15.71	20.16	22.43
	20	25.28	36.19	32.56	35.51
N		*276*	*232*	*147*	*114*
Lateral JSN	1				
	5	7.25	1.72	4.08	0
	10	3.28	3.02	4.08	1.75
	15	2.54	2.59	7.48	4.39
	20	7.25	15.52	22.45	15.79

7.3.5.2 Incidence of individual features stratified by baseline BMI

Cumulative incidence of both medial and lateral osteophytes was significantly higher for overweight compared to normal subjects for all visit years (p≤0.04) (table 7.10).

Overweight subjects had significantly higher cumulative incidence of medial JSN at years 5, 15 and 20 ($p \leq 0.05$), and higher incidence of lateral JSN only at year 20 ($p=0.007$).

For each unit increase of baseline BMI (as a continuous measure), the risk of incident medial osteophytes increased by 27% at year 5, 14% at year 10, 10% at year 15 and 13% at year 20. The risk of having incident lateral osteophytes was 30% higher at year 5, 37% at year 10, 24% at year 15 and 17% at year 20. The risk of incident medial JSN increased by 24% at year 10, 13% by year 15, and 17% by year 20 for each unit increase of baseline BMI. There was no significant increase for year 5. The risk of incident lateral JSN was only significantly increased at year 15 (15%) and year 20 (11%).

7.3.5.3 Incidence of individual features stratified by pain

The cumulative incidence of medial and lateral osteophytes (table 7.10)were significantly higher in subjects with knee pain than without for all visit years ($p \leq 0.02$), except for medial osteophytes at year 15 ($p=0.076$). The incidence of medial JSN was significantly higher in painful knees at year 20 only ($p \leq 0.001$), and at year 15 and 20 for lateral JSN ($p \leq 0.001$).

Table 7.10 Cumulative incidence by visit year stratified by BMI group and pain

Feature/Method	Visit year	BMI		Pain	
		Normal	Overweight	No Pain	Pain
Medial osteophyte	1				
	5	5.15	8.98		
	10	9.84	20.36	12.5	22.86
	15	17	29.94	20.82	26.73
	20	57.05	70.06	59.13	70.8
Lateral osteophyte	1				
	5	3.57	7.87		
	10	7.37	16.62	9.64	18.1
	15	16.74	29.74	19.79	27.75
	20	52.46	64.72	52.28	70
Medial JSN	1				
	5	5.43	10.32		
	10	9.63	10.32	10.55	8.65
	15	14.81	20.32	16.21	20.2
	20	28.4	35.16	24.48	45.49
Lateral JSN	1				
	5	4.12	3.61		
	10	2.98	3.32	2.7	3.64
	15	3.43	4.22	2.24	7.49
	20	10.98	17.77	11.11	19.44

7.4 Progression and worsening

7.4.1 Kellgren and Lawrence

Out of knees with a grade 2 or above at baseline (n=75), 58.7% progressed between baseline and year 15, and 66.7% progressed between baseline and year 20. 11.5% of knees (n=834) showed worsening (increase of at least one grade) between baseline and year 5, 21.8% between year 5 and 10, 23.1% between 10 and 15 and 62.6% between year 15 and 20. Analysing only baseline and year 15 data, 39.8% of knees showed worsening by at least one grade.

A cross-tabulation of individual K/L grades (0-4) and TKR's between baseline and year 20 (table 7.11) demonstrates that 10.6% of all knees (n=834) remained a grade 0

throughout the study period, while 85.0% of knees worsened by at least one grade. Out of the subjects with a K&L of 1 or above at baseline (n=123), 23.6% remained at the same grade and 69.9% worsened (including to TKR) by year 20. Knees with a baseline grade 1 (n=48) had a higher percentage of worsening than any other baseline grade (95.8%). Grade 0's (n=711) were the next most likely to worsen, with 87.6% increasing by at least one grade over 20 years. Only 1.0% of knees were scored as having regressed to a lower grade by year 20.

Table 7.11 Cross-tabulation of baseline and year 20 K/L grades*
(n= 834)

Baseline	N	Year 20					
		0	**1**	**2**	**3**	**4**	**TKR**
0	711	88 (12.4)	122 (17.2)	285 (40.1)	190 (26.7)	13 (1.9)	13 (1.8)
1	48	0 (0.0)	2 (4.2)	25 (52.1)	14 (29.2)	2 (4.2)	5 (10.4)
2	53	0 (0.0)	0 (0.0)	17 (32.1)	24 (45.3)	5 (9.4)	7 (13.2)
3	22	0 (0.0)	1 (4.6)	7 (31.8)	10 (45.5)	0 (0.0)	4 (18.2)

*Results are displayed as N (%), with percentages calculated by row

7.4.2 Individual ROA features

62.3% of subjects with medial osteophytes at baseline (grade 1 or above, n=53) progressed to a higher grade over the next 19 years. 72.1% of subjects with lateral osteophytes (n= 31/43), 15.1% with medial JSN (n=18/119), and 13.8% with lateral JSN (n=9/65) also progressed over the duration of the study. Tables 7.12 – 7.15 show the cross-tabulations between baseline and year 20 ROA features showing both incidence and progression.

Table 7.12 Cross-tabulation of baseline and year 20 medial osteophytes

Baseline	N	Year 20				
		0	**1**	**2**	**3**	**TKR**
0	781	292 (37.4)	349 (44.7)	97 (12.4)	25 (3.2)	18 (2.3)
1	47	2 (4.3)	17 (36.2)	16 (34.0)	5 (10.6)	7 (14.9)
2	6	0 (0.0)	0 (0.0)	1 (16.7)	1 (16.7)	4 (66.7)

Table 7.13 Cross-tabulation of baseline and year 20 lateral osteophytes

Baseline	N	Year 20				
		0	**1**	**2**	**3**	**TKR**
0	791	334 (42.2)	322 (40.7)	89 (11.3)	24 (3.0)	22 (2.8)
1	38	0 (0.0)	10 (26.3)	17 (44.7)	6 (15.8)	5 (13.2)
2	4	0 (0.0)	0 (0.0)	2 (50.0)	1 (25.0)	1 (25.0)
3	1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)

Table 7.14 Cross-tabulation of baseline and year 20 medial JSN

Baseline	N	Year 20				
		0	**1**	**2**	**3**	**TKR**
0	715	491 (68.7)	164 (22.9)	27 (3.8)	14 (2.0)	19 (2.7)
1	116	46 (39.7)	54 (46.6)	7 (6.0)	1 (0.9)	8 (6.9)
2	1	0 (0.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
3	2	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (100.0)

Table 7.15 Cross-tabulation of baseline and year 20 lateral JSN

Baseline	N	Year 20				
		0	**1**	**2**	**3**	**TKR**
0	769	662 (86.1)	69 (9.0)	7 (0.9)	5 (0.7)	26 (3.4)
1	65	34 (52.3)	22 (33.9)	6 (9.2)	0 (0.0)	3 (4.6)

7.4.3 Worsening stratified by age

When stratified by age, worsening of K/L grade was significantly different between age groups between baseline and year 5 ($p<0.001$) and between year 15 and 20 only ($p=0.04$). The difference between age groups was significant when worsening was analysed between baseline and year 15 ($p<0.001$), but not between baseline and year 20 ($p=0.304$).

There was a significant increase in knees with worsening medial osteophytes between the youngest and oldest age groups at year 5 ($p \leq 0.001$) (table 7.16). There were also trends identified between age groups for the year 10 and 20 visits, although for year 20 there is a decrease in the percentage of worsening medial osteophytes as subjects increase in age (54.7% to 45.2% [p=0.028]). For lateral osteophytes, worsening was significantly different between age groups only for visit 5 and visit 15 ($p \leq 0.02$), and trends were identified for increasing percentage of worsening with increasing age ($p \leq 0.02$). No difference was found for worsening of medial JSN for any visit year, but both baseline and year 20 showed significant differences for lateral narrowing between age groups ($p \leq 0.02$).

Table7.16 Worsening by visit year stratified by age

Feature/Method	Visit year	Age groups			
		45-49	**50-54**	**55-59**	**60-64**
N		*300*	*250*	*158*	*126*
Medial osteophyte	1				
	5	5.0	4.8	7.6	14.3
	10	6.3	10.4	11.4	13.5
	15	10.3	13.6	12.7	15.9
	20	54.7	52.0	45.6	45.2
Lateral osteophyte	1				
	5	3.7	3.6	7.0	11.9
	10	7.0	7.2	8.2	6.4
	15	12.3	19.2	14.6	23.0
	20	46.3	44.8	43.7	43.7
Medial JSN	1				
	5	9.3	5.2	8.2	4.0
	10	6.3	10.8	6.3	4.8
	15	13.0	13.6	14.6	17.5
	20	17.7	26.0	22.8	26.2
Lateral JSN	1				
	5	6.7	2.4	5.1	0.0
	10	1.3	3.6	3.2	1.6
	15	2.7	3.2	7.0	3.2
	20	9.3	14.0	19.6	15.1

7.4.4 Worsening stratified by baseline BMI

The difference in worsening K/L grades between normal and overweight subjects was only significant between baseline and year 5, and year 5 and year 10 ($p<0.04$). As with the age groups, baseline BMI groups were only significantly different between baseline and year 15 ($p<0.001$) and not between baseline and year 20 ($p=0.207$).

Both medial and lateral osteophytes demonstrated an increase in worsening in overweight subjects for all years ($p\leq0.02$), except for year 20. There was a significant increase in the worsening of medial JSN for years 10 and 15 ($p\leq0.008$), and for lateral JSN at year 20 ($p\leq0.009$).

7.4.5 Worsening stratified by pain

No difference was found in worsening K/L between subjects with and without knee pain ($p=0.681$). Subjects with knee pain had significantly more worsening of medial osteophytes at year 10, and of lateral osteophytes at year 20. Worsening of joint space narrowing was significantly increased for painful knees at all visits ($p\leq0.03$), except for year 10 medial JSN ($p=0.987$).

Table 7.17 Worsening by visit year stratified by BMI and pain

Feature/Method	Visit year	BMI		Pain	
		Normal	Overweight	No Pain	Pain
N	1	464	370		
	5			600	122
	10			580	248
	15			549	280
	20			676	127
Medial osteophyte	1				
	5	4.96	9.19		
	10	6.9	12.97	8.33	14.75
	15	10.34	15.41	11.55	14.92
	20	48.71	53.24	49.36	53.93
Lateral osteophyte	1				
	5	3.45	8.11		
	10	4.31	10.81	6.5	7.38
	15	11.64	22.43	14.83	20.16
	20	42.89	47.57	42.44	50.71
Medial JSN	1				
	5	4.96	9.73		
	10	7.33	7.57	7.33	7.38
	15	11.21	17.84	12.24	18.95
	20	20.47	24.86	16.39	34.29
Lateral JSN	1				
	5	4.31	3.78		
	10	1.94	2.98	1.67	4.96
	15	3.02	4.61	2.42	6.88
	20	10.78	17.03	10.56	18.93

7.5 Summary

Prevalence of K/L≥2 was 36.5% by year 15 and 74.5% by year 20. Individual ROA features had a prevalence of 63.5% for medial osteophytes, 58.5% for lateral osteophytes, 33.3% for medial JSN, and 13.5% for lateral JSN by year 20. The composite ROA definition of ‘any medial’ feature had the highest prevalence at each year except for year 20 (19.0-71.6%), while the definition for ‘any osteophytes’ had the greatest increase of prevalence over 19 years ranging from 9.2% to 73.5%. K/L prevalence is significantly different between pain and pain-free subjects, BMI groups and between age groups for the

majority of visit years. The risk of having prevalent K/L ≥ 2 was 15 to 20% higher at each visit for each unit increase in baseline BMI. For both medial and lateral osteophytes, there was a significant increase in prevalence with increasing age, obesity (in overweight category) and knee pain. The prevalence of JSN was significantly associated with increasing age group only at year 20, with BMI for roughly half the visits (for both compartments) and with pain toward the later visit years (10 through 20). The risk of prevalent osteophytes increased from 13 to 29% each visit for each unit increase of baseline BMI, while the risk prevalent JSN increased from 10 to 13% for later visit years.

The annual cumulative incidence over 19 years was 3.8% for K/L, 3.2% for medial osteophytes, 2.9% for lateral osteophytes, 1.5% for medial JSN and 0.6% for lateral JSN. For the composite ROA grades, annual cumulative incidence was 3.7% for 'any osteophyte', 1.8% for 'any JSN', 3.6% for 'any medial' feature, 3.2% for 'any lateral' feature, 1.4% for 'both medial' features and 0.6% for 'both lateral' features. There was a significant difference in the 19 year annual cumulative incidence of K/L≥2 in subjects with a baseline grade of 1 versus 0, and between both age and BMI groups. The cumulative incidence of medial and lateral osteophytes was significantly higher for older age groups, in knees of overweight subjects and in painful knees. When baseline BMI was analysed as a continuous variable, the risk of incident osteophytes was much higher at earlier visits (27-30% at year 5) compared to later visits (13-17% at year 20). The incidence of JSN was only associated with age group at the year 5 and year 20 visits. Incidence of medial narrowing was significantly higher in overweight subjects for visits 5, 15 and 20, although this association was only seen at year 20 for lateral narrowing. Cumulative incidence for medial narrowing was higher in painful knees only at year 20, with lateral narrowing showing a significant difference at year 15 and 20.

Thirty-six percent of knees with K/L≥2 at baseline progressed between baseline and year 15, with 67.9% progressing between baseline and year 20. Eighty-five percent of knees worsened by at least one K/L grade over the 19 years of the study, with knees with a baseline grade of 1 having the highest percentage of worsening (95.8%) compared to any other baseline grade. Out of the knees which had ROA features present at baseline (grade≥1), 62.3% of medial osteophytes, 72.1% of lateral osteophytes, 15.1% of medial JSN and 13.8% of lateral JSN increased by at least one grade over 19 years. Subjects with knee pain had significantly more worsening of medial osteophytes at year 10, and of lateral osteophytes at year 20. Worsening of joint space narrowing was significantly increased for painful knees at all visits ($p \leq 0.03$), except for year 10 medial JSN ($p=0.987$).

7.6 Discussion

The novel findings of this research were that annual cumulative incidence, progression and worsening of K/L≥2 were 3.8%, 3.6% and 4.5% respectively over the 19 years of follow-up; medial osteophytes had the highest prevalence of any individual ROA feature by year 20 (63.5%); lateral osteophytes had high levels of prevalence, incidence and progression between baseline and year 20 (58.5%, 55.0% and 72.1% respectively); and the risk of a baseline K/L grade 1 was 5.2 times (95% CI 2.1, 13.3) that of a grade 0 to develop incident ROA (K/L≥2).

The increasing prevalence of K/L≥2 over each visit year is related to the increasing age of the study population as well as the change in method for the year 20 data. The prevalence of ROA is well known to be correlated to increasing age and continues to increase through elderly years (Felson et al. 1987). The jump between the year 15 and year 20 prevalence is also a reflection of the change in methodology between grading the two sets of x-rays. Year 15 x-rays were graded with traditional techniques, as plain-film x-rays

using a light-box. The year 20 digital x-rays were graded using the KneeMorf program which has both zooming and contrast adjustment capabilities. Therefore, there was a much greater chance to identify disease at year 20 than at year 15, even if it was present at year 15 resulting in misclassification. This was also borne out by the results in the reproducibility section (chapter 5) comparing year 15 x-rays read by DJH to those read by KL on digitised x-rays and KneeMorf. A direct comparison between newer digital and older methods yielded higher grades in direct comparison with an absolute grade increase of 16.0% for K/L, and between 4.0 and 16.0% for individual ROA features.

Furthermore, the reported prevalence for K/L≥2 at year 20 (74.5%) was directly comparable to two recent population-based cohort studies. A large-scale Japanese cohort found that 70.2% of women had K/L≥2 (Muraki et al. 2009), and a UK-based cohort of both men and women found a prevalence of 68.0% (Duncan et al. 2007), both of which were populations with similar mean ages to the Chingford study at year 20.

Annual rates of knee progression (3.6%) and worsening (4.5%) between baseline and year 20 were on the lower end for progression than those observed in other community-based cohorts which varied between 3.5% and 8.0% (Thorstensson et al. 2008; Felson et al. 1995; Felson et al. 1997), although was on par for annual worsening 4.4% (Cooper et al. 2000). Rates in established symptomatic cohorts were similar, varying between and 4.0% to 8.8% for progression (Pavelka et al. 2000; Thorstensson et al. 2008) and from 3.3% to 7.7% for worsening (Massardo et al. 1989; Spector et al. 1992). The slightly lower rates found in Chingford for progression are likely a consequence of both the relatively young age of the cohort at the start of the study, and the much longer length of the study in comparison to other research.

There were no comparable population studies against which to evaluate compartmental incidence and worsening of individual radiographic features. Despite these

features being measured in several longitudinal studies (Cooper et al. 2000, Felson et al. 1997), the natural history of individual features or compartments were not reported. Due to the emphasis on the medial compartment as the driving force for incidence and progression, it was surprising to find the high prevalence, incidence and worsening of osteophytes and JSN in the lateral compartment. Using the composite definition which included 'JSN *or* osteophytes', the prevalence of any feature in the lateral compartment was 62.9% at year 20 compared to 71.6% in the medial compartment at year 20. The prevalence of the 'JSN or osteophyte' definition was only approximately 10% lower in the lateral compartment compared to the medial compartment at each visit year.

The comparison of incident ROA between subjects with baseline grades 0 and 1 extends the timeline of an earlier nested case-control study within this cohort (Hart et al. 1999) and other recent research (Lachance et al. 2002) which emphasise the importance of grade 1s being treated distinctly from grade 0s. The higher risk of a grade 1 going on to develop incident ROA (5.2 times the chance of a grade 0) suggests that they are an important indicator of longitudinal radiographic incidence. Lachance et al. looked at incidence and progression of mild K&L grades over 3 years and found that a grade 1 was 6.4 times more likely to progress to a grade ≥2 than a grade 0 (Lachance et al. 2002).

K/L has been the primary method for evaluating radiographic knee OA in similar studies (Pavelka et al. 2000; Thorstensson et al. 2008; Felson et al. 1995; Felson et al. 1997; Cooper et al. 2000, etc.) but it is commonly criticised (as discussed in chapter 2). The issue specifically related to this analysis is the use of K/L to evaluate progression and worsening, despite several major limitations. K/L assumes a non-validated natural disease progression which is extremely osteophyte-centric, as well as having several different 'official' and modified versions which are all in use (Schiphof et al. 2011). Despite these negatives, K/L scores have a high level of reproducibility (Thorstensson et al. 2008;

Cooper et al. 2000) a strong correlation with pain and increasing K&L grades (Neogi et al. 2009, Cho et al. 2010), and is present in the majority of subjects who present with knee pain (Duncan et al. 2007). Although increasing K/L grades may be criticised for not being 'linear or equidistant' (Ding et al. 2012), the evidence of increasing grades having greater association with pain (Duncan et al. 2007; Muraki et al. 2009; Neogi et al. 2009) implies that while the grades are not perfect, they are capturing something important about the disease process and the resulting severity. And more importantly, they are identifying something that may be clinically relevant.

This analysis establishes the longest natural history study of radiographic knee OA to date, and provides novel information about incidence and worsening of whole knee and compartmental features in the intervals over the 19 year study. These results provide the necessary basis for understanding the structural component of knee osteoarthritis, allowing for further research to relate these features to clinical outcomes such as pain and risk of knee replacement.

8 CONSTRUCT VALIDITY OF ROA WITH PAIN (CROSS-SECTIONAL)

This section will evaluate the construct validity of radiographic knee OA, as assessed by a cross-sectional analysis of ROA (atlas-based and quantitative JSW) and knee pain. Year 20 data was used due to several factors; this visit provided the most comprehensive data for pain assessment, and the year 20 x-rays were the only high-quality digital x-rays available for the cohort.

Previous studies evaluating this relationship between ROA and pain are common, although very few studies have assessed community-based subjects and they often target subjects either with knee pain or with pre-defined risk factors for ROA. One of the most important aspects of this relationship, which has not been explored, is whether the relationship with pain improves when ROA is broken down by compartment, feature (i.e. osteophytes or JSN), and the severity of the feature.

This chapter will establish which diagnostic scoring method of ROA has the best construct validity with pain, if there is an individual feature which has better construct validity than whole-knee grades (K/L) and if evaluating the severity (individual grades) improves the relationship with pain.

8.1 Methods

8.1.1 Study design

All subjects who attended the year 20 x-ray visit were eligible for inclusion in the analysis. 994 knees were radiographed at year 20 and were graded for K/L. Thirty-five of the knees had a TKR by year 20, 104 were missing either quantitative joint space or flexion data and 21 subjects (42 knees) had evidence of RA in any joint at any time during the study. A total of 174 knees were excluded from the analyses due to one or more of

these issues, leaving 820 knees for this cross-sectional analysis. The methods used to score the x-rays have been previously described in detail (chapters 3 and 4), and the outcome measure for pain was the NHANES question.

8.1.2 Statistical methods

Baseline characteristics of subjects with all pertinent data at year 20 (ROA, quantitative JSW and pain) were compared to subjects who were lost to follow-up, with Man-Whitney U tests for non-normal continuous data and Pearson's chi-squared test for categorical variables. GEE logistic regression modeling was used for all analyses assessing ROA and pain, in order to account for within person clustering due to the knee level analysis. Both univariable and multivariable models adjusted for age, BMI (kg/m^2), history of smoking and flexion (for quantitative JSW only) were produced for each radiographic variable, giving results in odds ratios and 95% confidence intervals. Construct validity was evaluated by comparing odds ratios and confidence intervals of the multivariable model for each feature.

8.2 Results

8.2.1 Cohort characteristics

Table 8.1 shows the comparison of demographic characteristics between the knees used in the analysis (820) and those that were lost to follow-up between baseline and year 20 (1186). There was a significant difference for all characteristics tested between subjects who had complete data compared to those who were lost to follow-up. The subjects lost to follow-up were significantly older, had a higher BMI (although only slightly), were more likely to have a history of smoking, had knee pain at year 5, and had a higher percentage of K/L $\geq$2 at baseline.

Table 8.1 Comparison of baseline demographics for knees with and without year 20 follow-up

Characteristic	Knees in Analysis	Knees lost to follow-up	p-value
N (knees)	820	1186	
Age (median (IQR))	51.0 (47.0, 56.0)	56.0 (50.0, 61.0)	***<0.001***
BMI (median (IQR))	24.4 (22.5, 27.0)	25.2 (22.8, 28.1)	***<0.001***
Smoking (% ever)	39.9	50.5	***<0.001***
Pain at year 5 (%)	16.3	23.2	***0.001***
K/L 2+ (%)	8.6	11.6	***0.012***

8.2.2 Descriptive statistics for year 20 radiographic data

8.2.2.1 Categorical atlas-based ROA descriptive data

The prevalence (overall) of radiographic features shown here for year 20 (table 8.2), demonstrate the same trends discussed in previous chapters (chapters 6 and 7). There are higher levels of medial involvement than lateral, both for osteophytes and JSN. Osteophytes overall have a much higher prevalence than narrowing in both compartments. Prevalence of pain was relatively low overall, with 16.3% of knees having pain for more than 15 days in the last month (NHANES pain question).

The prevalence for almost all ROA features was significantly higher in the older age group compared to the younger group (table 8.2), except for whole-knee K/L, medial K/L and medial osteophytes. The prevalence for all ROA features was higher in overweight subjects ($\geq$25 kg/m^2), except for medial narrowing in the OARSI and Line Drawing atlases.

Table 8.2 Descriptive statistics for year 20 atlas data

Scoring Method	Feature	Overall (%)	Percentage by Age		p-value*	Percentage by BMI		p-value‡
			<72	*≥72*		*Normal*	*Overweight*	
	N	**820**	**410**	**410**		***249***	***565***	
K/L	Whole Knee (≥2)	73.7	71.7	75.6	0.205	63.1	78.1	*0.000*
	Medial (≥2)	63.9	62.4	65.4	0.383	55.0	67.6	*0.001*
	Lateral (≥2)	59.2	54.9	63.4	*0.013*	45.0	65.3	*0.000*
Chingford	Medial Osteo (≥1)	63.5	62.2	64.9	0.425	54.6	67.3	*0.001*
	Lateral Osteo (≥1)	58.9	54.4	63.4	*0.009*	44.5	65.1	*0.000*
	Medial JSN (≥1)	32.8	28.1	37.6	*0.004*	28.1	35.2	*0.047*
	Lateral JSN (≥1)	13.1	8.3	17.8	*0.000*	7.2	15.8	*0.001*
OARSI	Medial Fem Osteo (≥1)	30.0	28.1	32.0	0.223	18.5	34.9	*0.000*
	Medial Tib Osteo (≥1)	62.6	61.2	63.9	0.427	55.0	65.7	*0.004*
	Lateral Fem Osteo (≥1)	22.7	18.8	26.6	*0.008*	13.7	26.6	*0.000*
	Lateral Tib Osteo (≥1)	56.8	53.4	60.2	*0.048*	44.6	62.1	*0.000*
	Medial JSN (≥1)	34.2	28.5	39.8	*0.001*	29.7	36.5	0.062
	Lateral JSN (≥1)	13.5	8.5	18.5	*0.000*	7.2	16.5	*0.000*
Line Drawing	Medial Fem Osteo (≥1)	29.8	27.8	31.7	0.222	18.1	34.7	*0.000*
	Medial Tib Osteo (≥1)	62.3	61.0	63.7	0.428	54.6	65.5	*0.003*
	Lateral Fem Osteo (≥1)	22.6	19.3	25.9	*0.024*	13.3	26.6	*0.000*
	Lateral Tib Osteo (≥1)	56.7	53.2	60.2	*0.041*	44.2	62.1	*0.000*
	Medial JSN (≥-1)	36.5	31.5	41.5	*0.003*	33.7	38.1	0.239
	Lateral JSN (≥-1)	16.1	9.8	22.4	*0.000*	10.0	18.9	*0.002*
Composite (Chingford)	Medial Osteo or JSN	71.8	68.8	74.9	*0.052*	65.9	74.3	*0.013*
	Latera Osteo or JSN	63.2	58.5	67.8	*0.006*	47.0	70.3	*0.000*
	Medial Osteo and JSN	24.5	21.5	27.6	*0.042*	16.8	28.1	*0.001*
	Lateral Osteo and JSN	8.8	4.2	13.4	*0.000*	4.8	10.6	*0.007*
NHANES Pain	pain at year 20	16.3	14.7	17.9	0.217	9.5	19.5	*0.001*

Italicised p-values are significant (p<0.05); *comparison of age groups; ‡comparison of BMI groups

8.2.2.2 Quantitative JSW descriptive data

Table 8.3 shows the descriptive statistics for the quantitative JSW. The means for both minJSW and meanJSW are in a small range between 4.3 and 6.1 millimetres for the four measurements. The ranges, however show that meanJSW in the lateral compartment has much larger maximum values. The lateral JSA measurements has the widest range of all the joint space widths, ranging between 4.9 and 61.5 millimetres, with the medial compartment showing a similarly wide range. This is consistent with the way this

measurement was calculated compared to minJSW and meanJSW. The inter-margin distances for determining flexion, has small means/medians for both compartments, despite having wide ranges. When stratifying by age or BMI group, the only significant difference was found for medial inter-margin distance, which was significantly larger in normal subjects.

Table 8.3 Descriptive statistics for year 20 quantitative JSW data (in millimetres)

Feature	Mean (SD)	Range	Mean by Age		Mean by BMI	
			<72	*≥72*	*Normal*	*Overweight*
N	***820***	***820***	***410***	***410***	***249***	***565***
Medial minJSW	4.3 (1.1)	0.0 - 8.1	4.3 (1.0)	4.3 (1.1)	4.3 (1.0)	4.3 (1.1)
Lateral minJSW	5.2 (1.2)	0.7 - 9.5	5.2 (1.2)	5.1 (1.2)	5.1 (1.1)	5.2 (1.2)
Medial meanJSW	5.2 (1.2)	1.3 - 9.7	5.2 (1.1)	5.2 (1.1)	5.3 (0.9)	5.2 (1.1)
Lateral meanJSW	6.1 (1.2)	1.1 - 11.8	6.1 (1.1)	6.0 (1.3)	6.0 (1.1)	6.1 (1.3)
Medial JSA	31.2 (6.6)	5.2 - 61.0	31.4 (6.2)	31.0 (6.9)	31.3 (5.9)	31.1 (6.8)
Lateral JSA	35.5 (7.6)	4.9 - 61.5	35.7 (7.1)	35.4 (8.0)	34.9 (7.3)	35.8 (7.7)
Medial IMD	5.0 (1.9)	0.0 - 13.9	4.9 (2.1)	5.1 (1.8)	5.3(2.0)**	4.9 (1.9)**
Lateral IMD‡	1.3 (0.0, 2.8)	0.0 - 25.8	1.7 (2.8)	1.9 (2.9)	2.1 (3.5)	1.6 (2.3)

‡Given in median (interquartile range) due to skewed nature of data

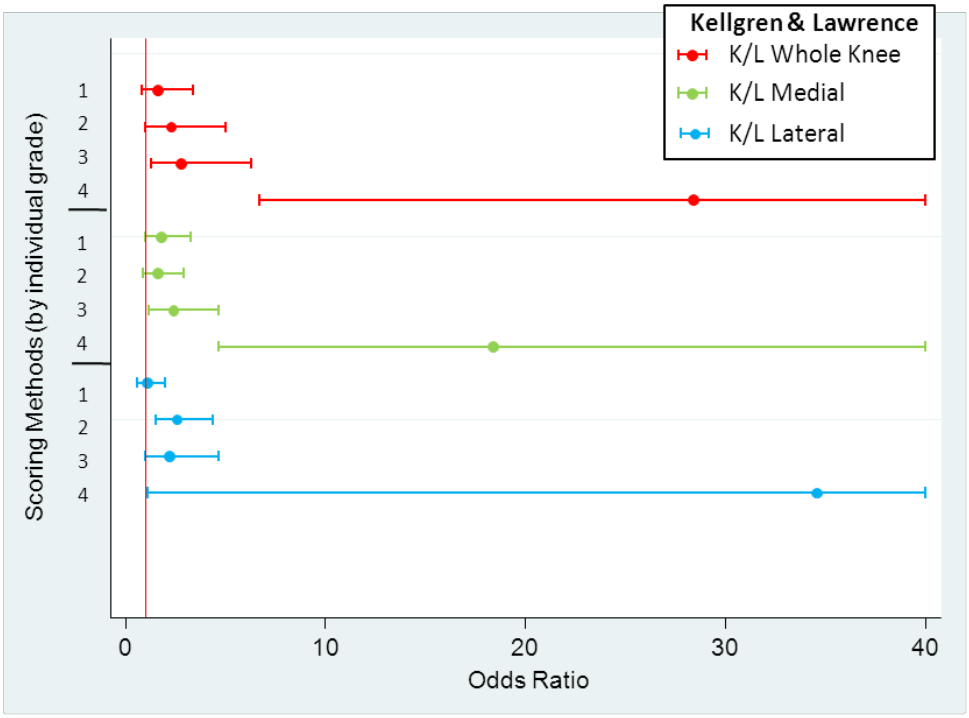
**Statistically sigificant difference

8.2.3 Construct validity of atlas-based radiographic scoring methods

8.2.3.1 Kellgren and Lawrence

Figure 8.1 shows the odds ratios of whole knee K/L, and both compartmental K/L grades. Only grade 3 and 4 were significantly associated with pain for whole-knee K/L (grade 3: OR 2.8 [95% CI 1.3, 6.3], grade 4: OR 28.4 [95% CI 6.7, 121.0]). Lateral K/L grade 4 had the highest overall association with pain (OR 34.6[95% CI 1.1, 1093.8) although the relationship was only marginally significant. The only definition where a grade 2 was significant was lateral K/L (OR 2.6 [95% CI 1.5, 4.4]).

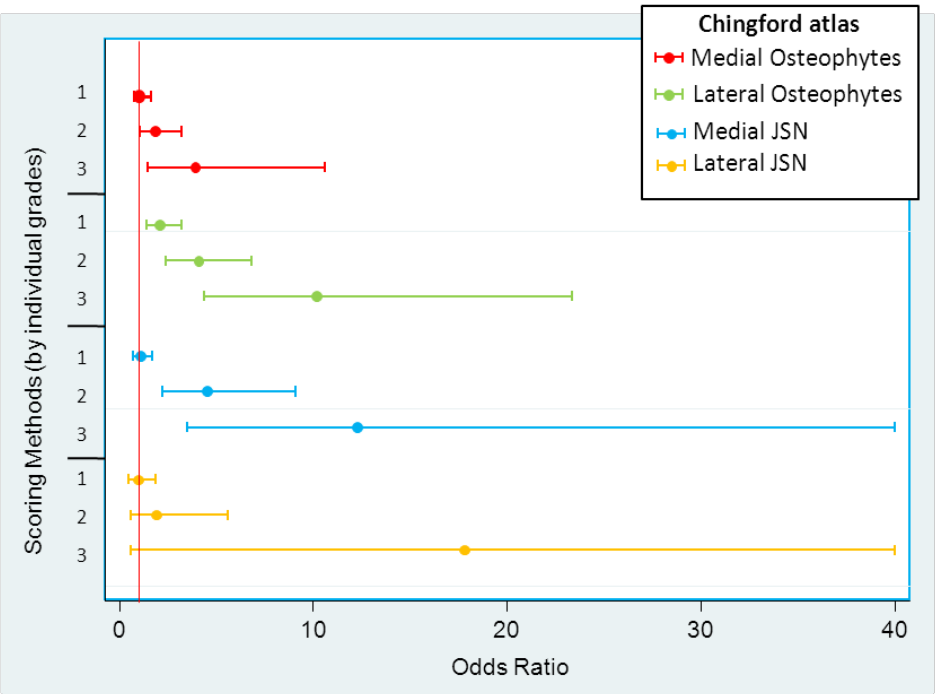
Figure 8.1 Construct validity of Kellgren and Lawrence scoring method



8.2.3.2 Chingford atlas

Out of the Chingford atlas scores (figure 8.2), lateral osteophytes and medial JSN grades 2 and 3 had the highest significant associations with pain (OR ≥4.1). Grade 3 medial JSN had the best association with pain (OR 12.3 [95% CI 3.5, 43.6]) out of all features and grades. Lateral osteophytes were the only feature which were significantly associated with pain for all individual grades (grade 1: OR 2.1 [95% CI 1.4, 3.2], grade 2: OR 4.1 [95% CI 2.4, 6.8], grade 3: OR 10.2 [95% CI 4.4, 23.4]).

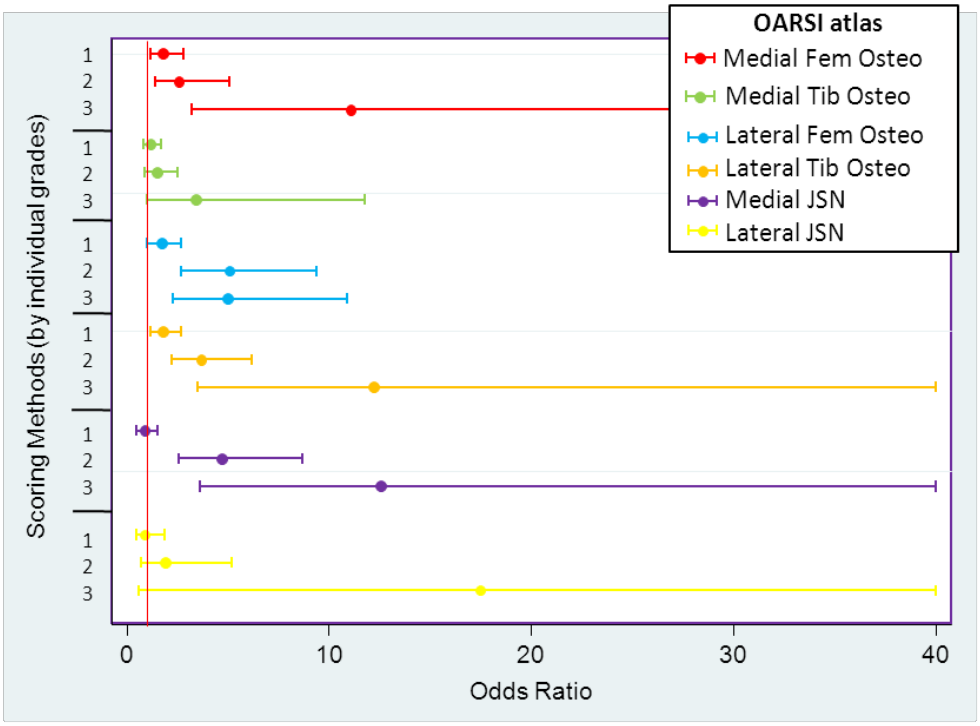
Figure 8.2 Construct validity of the Chingford atlas



8.2.3.3 OARSI atlas

Testing the construct validity of the OARSI atlas (figure 8.3) demonstrated that medial narrowing, lateral tibial osteophytes and medial femoral osteophytes all had good associations with pain for their highest individual grade (OR 12.6 [95% CI 3.6, 8.7], OR 12.2 [95% CI 3.5, 42.3], OR 11.1 [95% CI 3.2, 38.0], respectively). Lateral tibial osteophytes were the only feature to have a significant association with pain for all three grades (grade 1: OR 1.8 [95% CI 1.2, 2.7], grade 2: OR 3.7 [95% CI 2.2, 6.2], grade 3: OR 12.2 [95% CI 3.5, 42.3]). Medial tibial osteophytes and lateral JSN were not significantly associated with pain for any grades.

Figure 8.3 Construct validity of the OARSI atlas



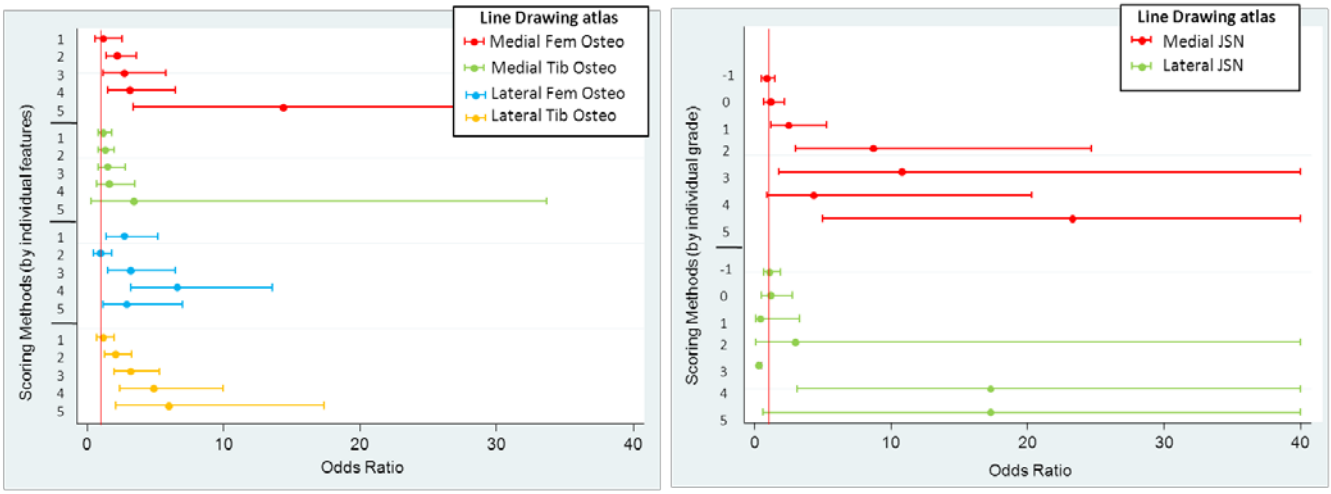
8.2.3.4 Line Drawing atlas

Of all the compartmental osteophyte grades in the Line Drawing atlas (figure 8.4), medial femoral osteophyte grade 5s had the best association with pain (OR 14.4 [95% CI 3.4, 61.7]). Similar to the OARSI atlas, lateral tibial osteophytes had the most number of grades which had a significant association with pain (grade 3: OR 3.2 [95% CI 2.0, 5.3], grade 4: OR 4.9 [95% CI 2.4, 10.1], grade 3: OR 6.0 [95% CI 2.1, 17.4]). There was no significant association between medial tibial osteophytes and pain. Lateral femoral osteophytes had an unusual association with pain, where each increasing grade did not have an increasing association. The grade 4 lateral femoral osteophyte had the best association with pain (OR 6.6 [95% CI 3.2, 13.6]).

The medial compartment (figure 8.5) had the best association between JSN and pain (grade 1: OR 2.5 [95% CI 1.2, 5.3], grade 2: OR 8.7 [95% CI 3.0, 24.7], grade 3: OR

10.8 [95% CI 1.8, 64.3], grade 5: OR 23.3 [95% CI 5.0, 108.6]). Grade 4, was the only lateral JSN grade significantly associated with pain (OR 17.3 [95% CI 3.1, 95.5]).

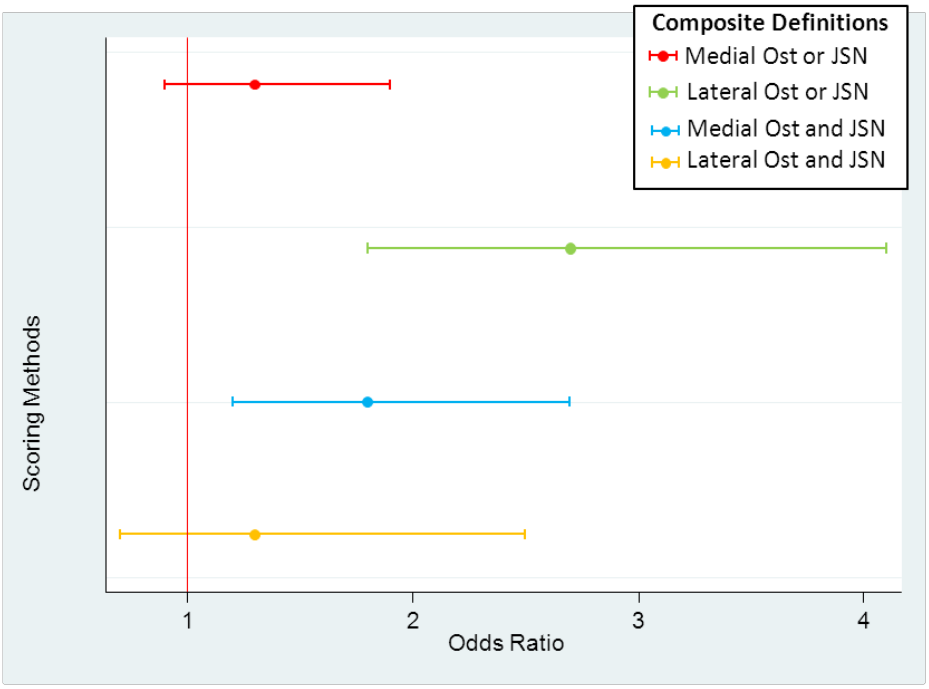
Figure 8.4 and 8.5 Construct validity of the Line Drawing atlas: osteophytes and JSN



8.2.3.5 Composite ROA definitions

The composite ROA definitions (figure 8.6) with the best construct validity was having either joint space narrowing or an osteophyte in the lateral compartment (figure 8.6) had the best construct validity of any of the composite ROA definitions (OR 2.7 [95% CI 1.8, 4.1]) followed by having both JSN and an osteophyte in the medial compartment (OR 1.8 [95% CI 1.2, 2.7]).

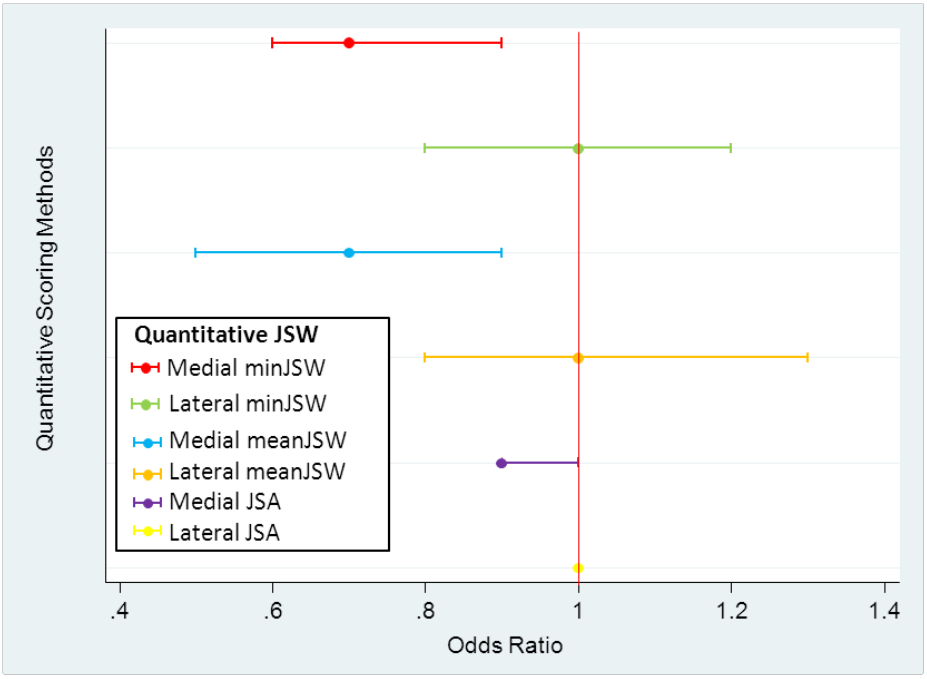
Figure 8.6 Construct validity of the composite ROA definitions



8.2.4 Quantitative JSW measurements

For the quantitative JSW measurements, having higher (wider) minJSW or meanJSW was significantly protective against having knee pain (medial minJSW: OR 0.7 [95% CI 0.6, 0.9], medial meanJSW: OR 0.7 [95% CI 0.5, 0.9]). This translates to a 30% decrease in the odds of having pain with every 1 mm increase in minJSW. None of the lateral quantitative measurements or medial JSA were significantly associated with pain in either direction.

Figure 8.7. Construct validity of quantitative JSW measurements



8.3 Summary

The OARSI atlas had the best overall construct validity with pain, taking into account the number of grades of each feature which had a statistically significant association with pain and the magnitude of each of the associations.

Grade 4 lateral K/L had the highest association with pain of all individual grades (OR 34.6 [95% CI 1.1, 1093.8]). The individual feature and grade with the highest association was grade 5 JSN in the Line Drawing atlas (OR 23.3 [95% CI 5.0, 108.7]), which represents bone-on-bone narrowing in that atlas. Neither of the compartmental K/L grades had a better association with pain than the original whole-knee grade.

The feature with the most number of individual grades associated with pain was the lateral tibial osteophyte. This was found for all atlases (lateral osteophyte in Chingford atlas), with the highest odds ratios associated with the OARSI atlas grade 3 lateral tibial osteophyte (OR 12.2 [95% CI 3.5, 42.3]). Having a large femoral osteophyte (OARSI

grade 3 or Line Drawing grade 5) in the medial compartment also had very good construct validity with pain (OR 11.1 and OR 14.4, respectively). There was no significant association found for medial tibial osteophytes for any grades.

Two of the composite ROA definitions showed comparatively modest association with pain; 'lateral JSN *or* osteophyte' had an odds ratio of 2.7 (95%CI 1.8, 4.1) while 'medial JSN *and* osteophyte' had an odds ratio of 1.8 (1.2, 2.7).

Two of the medial quantitative joint space measurements showed a significant cross-sectional association with pain; medial minimum JSW (OR 0.7 [95% CI 0.6, 0.9]), and medial mean JSW (OR 0.7 [95%CI 0.5, 0.9]). This result showed that subjects with overall wider medial joint space were less likely to have knee pain.

8.4 Discussion

Although there are many studies which analyse the cross-sectional relationship between radiographic osteoarthritis and pain, few break down the analysis by feature and compartment, and there is no previous research assessing how the severity of each feature to relate to the presence of pain. For the atlas-based features, the results demonstrated that any lateral tibial osteophyte, large medial femoral osteophytes and severe medial narrowing all had good construct validity with pain regardless of the atlas method. For quantitative JSW, having higher medial minJSW and meanJSW was significantly protective against pain.

It would be expected with the general emphasis in ROA research on medial disease that using a compartmental specific K/L grade would increase the association with pain compared to the older whole-knee K/L grade. These results did not find any benefit of using individual compartmental K/L grades over the traditional whole-knee score. This result can be explained by the other findings in this section showing the importance of

medial narrowing *and* lateral osteophytes. Restricting these features into one compartment reduced the overall construct validity with pain.

The OARSI atlas was found to have the best overall association with pain. It provides a good balance of having distinctive grades for both compartmental osteophytes (rather than grouping as the Chingford atlas), and a narrower range of individual grades (0-3) which likely gives each grade more statistical power (rather than the Line Drawing atlas with expanded grades).

This analysis showed that the presence of any size lateral tibial osteophyte is significantly associated with pain. This result was only visible because of the use of the OARSI and Line Drawing atlas, and the fact that lateral tibial osteophytes were not grouped together with femoral osteophytes which would hide the association. The results show that the tibial osteophyte is driving the cross-sectional relationship with pain in the lateral compartment and the femoral osteophyte is driving this relationship in the medial compartment.

Although the mechanisms for why lateral tibial osteophytes have an increased association with pain have not yet been researched, a study by Nagaosa et al. found that they were significantly associated with patellofemoral attrition and narrowing, chondrocalcinosis, and medial JSN (2002). Their data also showed that lateral tibial osteophytes were different from the majority of those at other sites which extended outward from the joint. The majority of lateral tibial osteophytes grew upwards appearing as 'traction spurs', which have been described as being able to reduce valgus motion increasing joint stability (Nagaosa et al. 2002; Pottenger et al. 1990).

Osteophytes have been shown to have a good cross-sectional relationship with pain. The presence of osteophytes has been shown to have a better association with pain than narrowing in any x-ray view (AP, lateral, skyline), and increasing osteophyte size was

also found to be significantly associated with pain (Cicuttini et al. 1996). This relationship was confirmed on MRI, where the number of osteophytes present was associated with pain, although not the size of the osteophyte (Kornaat et al. 2006).

This research did not find any increase in construct validity for K/L grading by compartment instead of the traditional whole-knee method. Szebenyi et al. found similar results where knees with structural changes in both compartments were more likely to be painful than knees with only single compartment involvement (2006). Felson et al. also found that a combination of ROA features (moderate definite osteophyte and 'moderate JSN or osteophyte') had good relationship with pain, although did not explore this finding by compartment (1997).

The degree of structural severity being associated with pain has previously been demonstrated for K/L grades, although not for individual features (osteophytes and JSN). Cho et al. looked at the relationship of individual K/L grades to the WOMAC pain subscale and found a significant increase in the pain score between grade 2 and grade 3 women and men (2010). Laxafoss et al. demonstrated increasing odds of pain between K/L grades 1 through 4 (OR 1.1 to 6.4) where the positive predictive validity (PPV) improved for each increasing grade as well (PPV 42.6 to 80.0).

The mean of the quantitative medial minJSW measurement for this study was 4.3mm (SD 1.1). This is partway between the means reported for a study on ROA normal knees without pain (4.8mm (SD 0.7), and a study on obese, symptomatic women with evidence of ROA (3.6mm (SD 1.1) (Beattie et al. 2008; Mazzuca 2006). This is due to the non-restrictive inclusion criteria for this analysis. This type of analysis looking at the cross-sectional relationship of the continuous range of quantitative JSW measures and pain has not previously been done. The results show that wide joint space is protective against

the presence of pain, which directly relates to the categorical atlas-based JSN where narrowing predicts the presence of pain.

9 PREDICTIVE VALIDITY OF ROA WITH FUTURE TKR

9.1 Introduction

While joint space narrowing and osteophytes are well-known features of knee osteoarthritis and are considered key disease indications on x-ray, their relationship with future total knee joint replacements (TKRs) has not been well explored. Total knee replacements are the best hard outcome to establish end-stage osteoarthritis as they require the presence of both structural evidence of OA in addition to severe pain. Knee replacement surgery is important both at the patient level, but at the public health level as well due to the economic burden of surgery. The need for surgery is 'widely recognised as the single most relevant outcome in knee osteoarthritis' (Bruyere et al. 2005).

Research has focused primarily on the cross-sectional relationship between joint space and pain or the predictive power of joint space at baseline to predict future pain (Bedson and Croft 2008; Emrani et al. 2008). Very little research has explored whether the change in joint space can predict the need for a total knee replacement, and there is no research to date looking into the predictive power of osteophytes, which have been shown to have good construct validity with pain (Spector et al. 1993; Cucuttini et al. 1996; Kornaat et al. 2006).

This section will look at both baseline measures of radiographic OA including quantitative joint space measures, categorical joint space and osteophytes, as well as the worsening of these same features over a period of four years and how predictive these variables are in regards to the need for a future total knee replacement over the subsequent fifteen years. This will be the first research to look at the predictive validity of osteophytes and to explore these joint space variables as predictors for TKRs over such a long period of time.

9.2 Methods

9.2.1 Study design

For this analysis, it was decided to use a nested case control study design in order to increase the power and precision of the study due to the use of a rare outcome (TKR). A case-control design also limited any bias due to loss to follow-up, and limited the amount of extra data needed to be collected for quantitative joint space at baseline and year 5. Cases were selected from subjects present for the baseline and year 5 x-rays, and had a TKR after the year 5 x-ray visit, and had no evidence of RA over the 20 years of the study. Ten controls were selected for each case from subjects who had both baseline and year 5 x-rays and did not have any indications for a total or partial knee replacement at any time up until the year 20 x-ray visit.

Controls were matched to cases by baseline age, with the majority (85.0%) matched to the exact age of the case at baseline. 86.7% of those not matched exactly (n=60), were matched within two years of the baseline age of the case. Controls were also matched by the length of time they were in the study. They had to be actively involved in the study, and not listed as missing, dropped out, or deceased by the time their matched case had their TKR (e.g. for a case with a TKR at visit 10, controls must have study information at least until the year 10 visit but not necessarily beyond).

Controls provided the same side knee to the analysis as the case knee (e.g. 1 bilateral TKR [2 case knees] was matched with 10 subjects [20 control knees]; 1 right unilateral TKR [1 case knee] was matched with 10 subjects [10 right control knees]). This allowed for the use of a knee level analysis without having to adjust for within person clustering.

9.2.2 Statistical methods

Change in the quantitative joint space measurements was calculated by subtracting the year 5 variable from the corresponding baseline variable. A positive number indicated narrowing over the four year period, with a negative number indicating widening. These measures were analysed as continuous measurements unless specifically described as having a set cut-off creating a binary variable. Change for categorical features was calculated similarly with the grade at year 5 subtracted from the baseline grade. All negative numbers (which would indicate widening joint space or osteophytes getting smaller) were relabelled as noughts to group them within the 'no change' group.

The baseline demographics and radiographic predictors of cases and controls were compared using t-tests for continuous and chi squared tests for categorical variables. Conditional logistic regression was used to account for matching the cases and controls by age at baseline, with odds ratios (ORs) and 95% confidence intervals (CIs) given in the results. A univariable model is presented where only the predictor of interest was included in the model as well as a multivariable model where baseline BMI has been adjusted for. Neither pain nor baseline K/L were included in the multivariable analysis. Pain was excluded because it is likely on the causal pathway between radiographic risk factors and the outcome (TKR), and adding this into the model would cause overadjustment bias (Schisterman 2009). Baseline K/L was not included in the model because it is a composite measure which is made up of part of the risk factors being testing (osteophytes and joint space narrowing).

9.3 Descriptive statistics

9.3.1 Baseline demographics

Baseline demographics of the cases, controls and the subjects missing from the analysis can be seen in table 9.1 There was a statistically significant difference between BMI, knee pain at year 5 and baseline K/L grade between the case and control group.

Table 9.1 Baseline demographics of cases and controls

Characteristic	Cases (n=40)	Controls (n=383)	p-value
Age, mean (SD) years	53.9 (5.3)	53.8 (5.2)	0.94
BMI, mean (SD) kg/m^2	27.0 (2.9)	25.3(3.6)	***0.05***
Knee Pain at Year 5	54.3%*	16.7%*	***>0.001***
Baseline K/L grade:			
Grade 0	55.0%	86.7%	***>0.001***
Grade 1	15.0%	6.3%	
Grade 2	17.5%	4.4%	
Grade 3	12.5%	2.6%	

9.3.2 Descriptive statistics of quantitative risk factors (joint space)

Table 9.2 compares quantitative measurements collected using the KneeMorf software between cases and controls. Baseline and change measures were compared for minimum JSW, mean JSW, and joint space area for both the medial and lateral compartments. None of the baseline measurements were found to be significantly different between cases and controls, while a single change variable, change of mean JSW in the lateral compartment showed a modest difference (p=0.048) between groups. This result shows that there was more joint space widening in the lateral compartment over four years in cases compared to the control group.

Table 9.2 Descriptive statistics of quantitative JSW (cases vs. controls) in mm (±SD)

Predictor	Cases (n=40)	Controls (n=383)	p-value
Baseline:			
Medial minJSW	4.2 (1.3)	4.4 (0.8)	0.257
Lateral minJSW	4.8 (1.6)	5.1 (1.3)	0.102
Medial meanJSW	5.2 (1.2)	5.3 (0.8)	0.277
Lateral meanJSW	6.0 (1.4)	6.1 (1.1)	0.634
Medial JSA	30.6 (8.2)	31.6 (5.4)	0.268
Lateral JSA	34.3 (8.6)	35.3 (7.4)	0.431
Change:			
Medial minJSW	0.07 (0.84)	0.03 (0.67)	0.736
Lateral minJSW	-0.51 (1.05)	-0.17 (1.19)	0.081
Medial meanJSW	0.91 (0.81)	0.00 (0.63)	0.415
Lateral meanJSW	-0.42 (1.27)	-0.09 (0.98)	***0.048***
Medial JSA	-0.12 (5.25)	-0.12 (4.39)	0.992
Lateral JSA	-2.84 (6.99)	-0.84 (6.58)	0.070
Proportional med. minJSW	0.02 (0.26)	-0.00 (0.16)	0.500
Proportional lat. minJSW	-0.13 (0.36)	-0.08 (0.33)	0.363

9.3.3 Descriptive statistics of categorical risk factors

Table 9.3 compares the percentage of ROA features present in cases to controls both for baseline and whether any feature worsened over 4 years (any increase of 1 grade or more). All baseline ROA features were significantly higher in cases than controls except for lateral JSN. The percentage of subjects of subjects with worsening grades over four years was also significantly higher in cases than controls.

Table 9.3 Descriptive statistics of ROA features(cases vs. controls)

Characteristic	Cases (n=40)	Controls (n=383)	p-value
Baseline:			
Medial Osteophyte	27.5%	5.7%	**>0.001**
Lateral Osteophyte	20.0%	3.45	**>0.001**
Medial JSN	30.0%	13.8%	**0.007**
Lateral JSN	10.0%	7.35	0.541
K/L	30.0%	7.1%	**>0.001**
Change:			
Medial Osteophyte	20.0%	5.5%	***0.001***
Lateral Osteophyte	22.5%	5.55	***>0.001***
Medial JSN	20.05	7.3%	***0.006***
Lateral JSN	7.5%	2.15	***0.041***
K/L	30.0%	11.2%	***0.001***

9.4 Baseline risk factors as predictors of future TKR

9.4.1 Demographic predictors

The results of the conditional logistic regression analysis assessing the risk of common demographic predictors of OA (both structural and symptomatic) including BMI, pain and history of smoking is shown in table 9.4. Age was not included, as it was used as a matching variable for the controls. These results show that baseline BMI and pain at year 5 significantly predict having a future knee replacement. Smoking was not significant both with and without adjustment for BMI in the multivariable model. An increase of one unit of BMI (kg/m^2), increases the odds of having a TKR by ten percent. As would be expected, pain was the strongest demographic predictor of future TKR, even after adjustment for baseline BMI, with painful knees having 4.8 times the risk of having a future TKR compared to knees without pain.

Table 9.4 Conditional logistic regression: demographic predictors

Predictors	Univariable			Multivariable (adj for BMI only)		
	OR	95% CI	p-value	OR	95% CI	p-value
Baseline BMI	1.1	(1.0, 1.2)	***0.005***			
Pain at year 5	5.5	(2.6, 11.5)	***<0.001***	4.8	(2.3, 10.2)	***<0.001***

9.4.2 Quantitative JSW predictors

Table 9.5 shows the baseline joint space variables calculated using KneeMorf as predictors for TKR. Minimum JSW, mean JSW and joint space area were not found to be predictive of a TKR in either compartment, except for baseline minimum JSW in the lateral compartment only after adjustment for BMI. This is possibly a type I error (a rejection of a true null hypothesis) due to multiple testing.

Table 9.5 Conditional logistic regression: baseline JSW predictors

Predictors	Univariable			Multivariable (adj for BMI only)		
	OR	95% CI	p-value	OR	95% CI	p-value
medial minJSW	0.8	(0.5, 1.2)	0.233	0.8	(0.5, 1.1)	0.166
lateral minJSW	0.8	(0.6, 1.0)	0.110	0.8	(0.6, 1.0)	***0.035***
medial meanJSW	0.9	(0.7, 1.3)	0.657	0.8	(0.5, 1.2)	0.223
lateral meanJSW	0.9	(0.7, 1.3)	0.657	0.9	(0.6, 1.2)	0.348
medial JSA	1.0	(0.9, 1.0)	0.260	1.0	(0.9, 1.0)	0.212
lateral JSA	1.0	(0.9, 1.0)	0.456	1.0	(0.9, 1.0)	0.214

9.4.3 Baseline categorical ROA predictors

Analysis of the predictive validity of baseline categorical features showed that the presence of any feature significantly predicted future TKR, except for lateral joint space (table 9.6). The odds ratios remained significant, although slightly attenuated in the multivariable model once adjusted for BMI. The presence of lateral osteophytes at baseline was the strongest predictor of future TKR, with an increased risk of 5.5 times that of a

knee without lateral osteophytes at baseline. Medial osteophytes (OR 4.9) and K/L grade (OR 4.7) at baseline were also both strong predictors of TKR.

Table 9.6 Conditional logistic regression: baseline categorical ROA predictors

Predictors	Univariable			Multivariable (adj for BMI only)		
	OR	**95% CI**	**pavalue**	**OR**	**95% CI**	**pvalue**
medial JSN	2.9	(1.3, 6.2)	***0.007***	2.6	(1.2, 5.7)	***0.017***
lateral JSN	1.4	(0.5, 4.3)	0.540	1.5	(0.5, 4.6)	0.481
medial osteophytes	6.0	(2.7, 13.5)	***<0.001***	4.9	(2.1, 11.5)	***<0.001***
lateral osteophytes	7.3	(2.8, 19.1)	***<0.001***	5.5	(2.0, 15.0)	***<0.001***
Kellgren and Lawrence	5.8	(2.6, 13.0)	***<0.001***	4.7	(2.0, 11.0)	***<0.001***

9.5 Four-year change predictors of TKR

9.5.1 Quantitative change predictors

Table 9.7 shows the results of the analysis of the four-year joint space change variables including minimum, mean, area and proportional change in both the medial and lateral compartments. The only variable which showed any significance was mean joint space width change in the lateral compartment. This was modestly significant, as the confidence intervals were close to 1.0, and this result disappeared in the multivariable model.

Table 9.7 Conditional logistic regression: quantitative JSW change predictors

Predictors	Univariable			Multivariable (adj for BMI only)		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Joint Space Change:						
medial minJSW	1.1	(0.7, 1.8)	0.731	1.0	(0.7, 1.6)	0.878
lateral minJSW	0.8	(0.6, 1.0)	0.078	0.8	(0.6, 1.0)	0.100
medial meanJSW	0.8	(0.5, 1.4)	0.409	0.8	(0.5, 1.4)	0.446
lateral meanJSW	0.7	(0.5, 1.0)	***0.043***	0.8	(0.6, 1.0)	0.089
medial JSA	1.0	(0.9, 1.1)	0.982	1.0	(0.9, 1.1)	0.866
lateral JSA	1.0	(0.9, 1.0)	0.064	1.0	(0.9, 1.0)	0.134
proportional medminJSW	2.0	(0.3, 13.7)	0.492	1.6	(0.2, 10.8)	0.616
proportional latminJSW	0.7	(0.3, 1.6)	0.363	0.7	(0.3, 1.5)	0.324

9.5.2 Categorical change predictors

All four-year categorical change variables significantly predicted future TKRs (table 9.8). Four-year worsening of lateral joint space was the strongest predictor, where knees with an increase of at least one joint space grade were 4.2 times more likely to have a future TKR than knees without any evidence of worsening. Worsening of lateral osteophytes (OR 3.8) and medial osteophytes (OR 3.4) were the next strongest predictors.

Table 9.8 Conditional logistic regression: categorical ROA change predictors

Predictors	Univariable			Multivariable (adj for BMI only)		
	OR	95% CI	p-value	OR	95% CI	p-value
Medial minJSN	3.3	(1.4, 8.0)	***0.009***	3.3	(1.3, 8.2)	***0.011***
Lateral minJSN	4.1	(1.0, 17.2)	**0.052**	4.2	(1.0, 17.7)	**0.052**
Medial Osteophytes	4.3	(1.7, 10.4)	***0.001***	3.4	(1.4, 8.5)	***0.009***
Lateral Osteophytes	4.7	(2.0, 10.9)	***<0.001***	3.8	(1.6, 9.2)	***0.003***
Kellgren and Lawrence	3.5	(1.6, 7.4)	***0.001***	2.9	(1.3, 6.3)	***0.007***

9.6 Summary

There were statistically significant differences found between baseline BMI, pain at year 5, and baseline K/L grade between the cases and controls, although no differences were seen between the controls and missing data.

There was a 0.2mm difference in mean between medial minJSW for cases and controls, and a 0.1mm difference in mean for medial meanJSW. The largest difference in means was seen for change in lateral meanJSW, where cases had a 0.33mm larger mean than the controls. In contrast, all categorical variables showed significant differences for baseline variables when cases and controls were compared except for lateral JSN. All four-year change variables were significantly different between cases and controls.

Baseline BMI and pain at year 5 both significantly predicted having a future TKR. Lateral minJSW was the only quantitative predictor of TKR, although was only marginally

significant. All baseline categorical variables were significant except for lateral JSN, with lateral osteophytes having the highest predictive validity. Subjects with lateral osteophytes at baseline were 5.5 times more likely to go on to have a TKR than subjects without lateral osteophytes.

When analysing change between baseline and year 5, no quantitative joint space measures were found to be a significant predictor of TKR. All four-year categorical atlas-based ROA features significantly predicted TKRs. Change in lateral joint space was the strongest predictor (OR 4.2) followed by osteophytes in both the medial and lateral compartments (OR 3.4 medial, OR 3.8 lateral).

9.7 Discussion

For the evaluation of four-year change in the categorical variables, narrowing in the lateral compartment had the highest odd ratio (OR 4.2), although due to the extremely small sample size this risk factor was only borderline significant (p=0.052). This finding is very interesting since lateral joint space is often ignored or dropped from analyses due to a lack of data. 27.3% of knees with incident lateral joint space narrowing by year 5 (n=11) went on to have a knee replacement by year 20 (n=3). Only a very small number of control knees showed any evidence of lateral narrowing (2.1%). So while lateral narrowing appears to be rare, when it is present it appears to be an important disease indicator.

A change in lateral osteophyte had the next highest predictive validity (OR 3.8), but had a much stronger statistical significance (p=0.007). This finding correlates well with the result in the previous chapter, where the presence of any size of lateral tibial osteophyte had good construct validity with pain (OR 2.1-10.2). Although the osteophyte grade is made up of both incident and progressive osteophytes in order to have enough

statistical power, out of all the knees which were positive for osteophyte 4-year change (n=30), all knees with incident osteophytes went on to have a TKR (n=9).

This finding contrasts a study looking at the predictive validity of ROA features, where joint space narrowing was a better predictor for joint failure (complete loss of joint space) than osteophytes. Wolfe et al. found that osteophytes were a valuable indicator for progression only until moderate JSN was also present in the joint (2001). The difference of this study, however, is that it is not using a marker of pain and is evaluating progression to structural failure in isolation of any accompanying symptoms. There are no comparable studies for predictive validity of osteophytes to the need for a future TKR.

The lack of significance for four-year change in quantitative JSW variables as predictors for TKR in this study was surprising based on the results of several other studies. Bruyere et al. found that by dichotomising their continuous joint space change variables using cut-offs between 0.5 and 0.8mm, that eight-year change in minJSW was predictive of a TKR. (2005). Two MRI studies have also shown that loss of cartilage volume is predictive of a future TKR (Cicuttini et al. 2004; Raynauld et al. 2011).

The reason for this lack of significance is likely due to the number of healthy knees at the baseline of the study. 83.9% of knees (cases and controls) had a baseline K/L grade of 0. The three studies which found an association with TKR all used subjects with radiographic and/or symptomatic osteoarthritis present at baseline (Bruyere et al. 2005; Raynauld et al. 2011; Cicuttini et al. 2004). This indicates that quantitative joint space measures may be identifying change within the 'normal' range and are grouping change within the normal range and abnormal range together, thereby obscuring any associations that may be present. This requires a different study design and inclusion criteria to further explore these results.

10 STRENGTHS AND LIMITATIONS

This section will outline the strengths and limitations of this thesis, reflecting specifically on the study population and study design, radiographic protocols, measurement of risk factors and the selection of outcome measures.

10.1 Population – the Chingford study

There were many strengths in using the Chingford study to assess long-term natural history of ROA and to test the construct and predictive validity of ROA and pain. The subjects had a relatively young age (median 54.0) at the start of the study allowing for research into incident disease and the progression of mild disease, rather than into established disease, which is more common. The follow-up of twenty-years is one of the longest for OA research, and the retention rate is extremely high, with almost half the women returning for the year 20 visit. The population was selected from a single GP practice in an area with a relatively homogenous population (Hart et al. 1999), eliminating the need to adjust for geographical and social factors. The inclusion criteria were not restrictive, based only on being female and age (45-65), unlike a lot of osteoarthritis cohorts which recruit on the presence of ROA (Hornberg et al.1976), pain (Thorstensson et al. 2008, Massardo et al. 1989) or a combination of both (Pavelka et al. 2000).

Subjects lost to follow-up are a major limitation of all long-term cohort studies, such as Chingford. There is a potential for study bias due to deaths, subjects dropping out because of disability and illness and generally having a healthier cohort attending the follow-up visits. While the comparison of baseline characteristics of those included in analysis versus those lost to follow-up are similar enough not to imply a severe bias (i.e. slightly older, more likely to smoke and slightly more knee pain), the unknown possibility remains that subjects lost to follow-up, for whatever reason, would have had significantly

worse ROA than those included in this analysis. There is no way to know the potential effect of this type of bias on any study of this design.

Another limitation of the Chingford study is that it is comprised solely of women who are predominantly Caucasian. Because of this, the results of this study are not necessarily applicable to the general population. While the results should be able to be loosely applied to men, and women of other ancestry, there are known differences in prevalence, incidence and progression of knee osteoarthritis between these groups (Lachance et al. 2001, McAlindon et al. 1992, Cho et al. 2010, Odding et al. 1998). Felson et al. followed an elderly cohort (mean age 71) and found that incident radiographic disease was 1.7 times higher in women than men at the 8 year follow-up (1995). Differences in prevalence, incidence and pain due to OA has also been found when comparing groups of different ancestry, with subjects of both African and Chinese descent having more radiographic evidence of knee OA and pain than subjects of European ancestry (Zhang et al. 2001, Jordan et al. 2007)

10.2 Risk factors – radiographic features

10.2.1 X-ray views and positioning

The Chingford study is unusual in the extent of radiography available throughout the course of the study. Knees have been radiographed at each 5-yearly visit over twenty years following a similar protocol and using the same x-ray machine for the first 15 years. The year 20 visit, as previously discussed, was the first to have digital x-rays. Due to the age of the study, at the time the original protocols were established, the best x-rays for research was considered to be fully-extended, bilateral and weight-bearing.

Even though this view has changed over the past decade, and currently semi-flexed and fluoroscopy-assisted x-rays are preferred (Brandt et al. 2002, Buckland-Wright et al. 1999, Mazzuca et al. 1997), the original fully-extended protocols were maintained during the year 20 visit. This allowed for the ability to accurately evaluate change, both for incidence and progression over the course of the study. The use of extended x-rays are now known to underestimate joint space (Buckland-Wright et al. 1995), therefore no attempt was made in this research to reduce this limitation by assessing the tibial inter-margin distance as a marker of flexion. It was included as a covariate during multivariable regression of quantitative joint space.

Censoring was utilised due to the lack of an exact date of radiographic incidence/progression/worsening for each knee. Scores were only evaluated on the date of each radiograph, so that knees that progressed immediately after a visit would be recorded as not having worsened until the next clinic visit 4-5 years later. This may have contributed to an overestimation of the amount of time it took subjects to have incident or progressive ROA.

The Chingford study also has skyline x-rays (patellofemoral (PF) views) at five-year intervals starting at the year 5 visit. Unfortunately these x-rays are not digitised and have not yet been made available for research. The lack of this view is a limitation for this type of research, as the PF compartment is known to be an important component of 'whole organ' knee OA, and the presence of PF ROA is highly associated with pain and disability (McAlindon et al. 1992, Duncan et al. 2006; Peat et al. 2007). The lack of additional radiographic views in this study is a major limitation which should be addressed in future research.

10.2.2 Blinding x-rays for grading

A potential limitation of reading radiographs with the reader blinded to order is that the grades may decrease over time, because they are not being read in a method that allows for the evaluation of change. The percentage of 'regressive' grades in this study (1.7%) was much lower than that in studies using similar blinding methods (5.5-7.5%) (Lachance et al. 2002; Hart et al. 2001).

Reading in a series, blinded to time, is meant to reduce the bias of over-reading for disease and for limiting the bias for readers who have information about risk factors or treatment (Felson and Nevitt 2009). The blinding method used for Chingford is even more restrictive and is completely reliant on objective grading (with either an atlas or software program), and therefore could be underrepresenting disease compared to other longitudinal studies.

10.2.2.1 Effect of blinding on categorical measures

Botha-Scheepers et al. focused specifically on the sensitivity-to-change issue when looking at blinding (2005). They used categorical measures of JSN, and assessed radiographic change after 2 years in 20 subjects. They found that change in JSN and osteophyte progression was smaller for paired x-rays than for x-rays read in a known series. The actual number of radiographs which showed any change was very small in relation to the number of radiographs assessed, and the clinically important difference when grouping by *any* progression did not appear to be significant (Botha-Scheepers and Watt 2005).

Auleley et al. compared the grading and blinding methods in 104 x-rays taken three years apart for determining change in hip ROA (2000). They read K/L, JSN and minJSW unpaired, paired (random-order) and chronologically. They found that categorical ROA

grades (K/L and JSN) showed the most change using unpaired, than by paired x-rays blinded to time and finally, with a known time sequence (Auleley et al. 2000)

10.2.3 Quantitative JSW - KneeMorf program

Many of the digital methods used to assess radiographic joint space use an automated edge-detection algorithm which generally has high levels of reproducibility (Dacre et al. 1989, Lynch et al. 1993, Buckland-Wright et al. 1995, Piperno et al. 1998, Marijnissen et al. 2008, Oka et al. 2009). This study found similarly high levels of reproducibility despite the method having a large manual component. This study also did not find a significant difference in the reproducibility of lateral compartment compared with the medial compartment minJSW which has been identified in previous studies (Lynch et al. 1993).

MinJSW is used to assess knee joint space and as a proxy for cartilage damage, but the true relationship between minJSW on knee radiographs and actual cartilage volume is not yet fully understood. Many factors unrelated to cartilage are known to have an effect on radiographic joint space such as meniscal extrusion, using extended knee AP radiographs (versus fixed-flexion or fluoroscopy) and the quality of the radiographic image (Gale et al. 1999; Buckland-Wright et al. 1995). This work shows that joint space measurements calculated from Bezier curves as highly reproducible as automated edge-detect measurements, and are valid when compared to digital manual methods.

10.3 Outcome Measures

10.3.1 Pain

The pain assessment used for this research, although commonly-used and validated, is ultimately a crude evaluation of pain presence. The NHANES pain question

gives a simple binary result of present or absent, which may be capturing pain due to any cause. As it is a self-evaluating response, it has all the biases involved for any subjective diagnosis. For the purposes of this analysis, where ROA was the primary risk factor under examination, the use of a simplified pain variable was necessary to avoid adding an additional layer of uncertainty. The benefits of using a more comprehensive pain assessment method, such as WOMAC, a variable that identifies both sub-scales of pain and function as well as levels of severity, would have unnecessarily complicated the results for this stage of the research.

10.3.2 Total knee replacements

There are many factors which influence whether a subject has a joint replacement, in addition to radiographic evidence of OA and symptoms. Cultural views, access to healthcare and willingness for surgery have all been shown to influence whether someone has a TKR (Raynauld et al. 2011).

Although this is one of the best outcome measures for joint failure, the variability in subjects who have surgery that does not directly relate to the disease (i.e. personal choice, access to care), is a limitation of its use as an outcome measure. This has prompted work in having a surrogate outcome combining radiographic and symptomatic aspects of the disease which would produce a 'indication of TKR' outcome measure which would eliminate some of the non-controllable factors (Raynauld et al. 2011).

.

11 CONCLUSION

11.1 Principal findings

This research contributes several major findings to the body of knowledge surrounding the relationship between radiographic knee osteoarthritis and clinical outcome measures (pain and TKR).

It emphasises the usefulness of radiographs as tools for evaluating structural joint changes, both in terms of construct and predictive validity, and that x-rays should remain in the arsenal of both clinicians and researchers for the identification of subjects at risk for disease.

In comparing the reproducibility of atlas-based scoring methods, this research shows that within the same cohort all atlases showed relatively good reproducibility, despite grading different features and using varying grading scales. Methods which tend to group features by compartment have slightly better reproducibility (K/L and Chingford), than those which aim for a greater number of features by separating tibial and femoral osteophytes (OARSI and Line Drawing). This overall good reproducibility suggests that each of these methods is a possible candidate for use in validation testing against pain (construct validity) or the longitudinal relationship with knee replacements (predictive validity).

For this research, a software program was required which would be able to evaluate quantitative joint space on older low-contrast x-rays. The KneeMorf program was developed with specialised methods to measure joint space in the same way as fully-automated software, but employing manually positioned Bezier curves. This is the first time such techniques have been used in this way and they were found to be highly valid and reproducible. This method offers a novel way to evaluate older x-rays using the most up-to-date quantitative scores.

Due to the wide range of radiographic scoring methods used for research in different populations, it was not known whether they could be reasonably compared. This study assessed both how the individual grades of each atlas related to one another, and whether prevalence differed when it was calculated using the respective atlases. These findings show that only the Chingford and OARSI atlas have high agreement across all grades and features. Prevalence from the composite scores created from the individual feature atlases all compared well to each other, but had higher prevalence than the corresponding compartmental K/L grade. Some of the differences in prevalence found here may help explain the wide range in reported results seen in various studies, both in terms of prevalence and the relationship between ROA and clinical outcomes (Bedson 2008). These findings may provide guidance as to which atlases or grades can be reasonably compared when assessing the results of different studies.

An understanding of the natural history of disease is the first step needed for determining the risk factors for incidence and progression. The Chingford study offered a unique opportunity to look at the natural history of radiographic knee osteoarthritis over the course of 20 years, which sheds new light on the long-term course of ROA. This is the first study to establish the natural history of individual radiographic features over 19 years. Novel findings include that the features with the highest prevalence, incidence and progression between baseline and year 20 are medial osteophytes, closely followed by lateral osteophytes. A K/L grade 1 at baseline was 5.2 times more likely to develop incident ROA (K/L≥2) than a grade 0 to over the nineteen year study.

The OARSI atlas was found to have the best overall construct validity (association with pain) out of all the atlases. Severe medial JSN grades had the highest magnitude of association with pain (i.e. Line Drawing atlas grade 5), although significant association with pain was found for lateral tibial osteophytes of any severity. Having wider medial

joint space measured quantitatively (minJSW and meanJSW) was significantly protective against the presence of knee pain.

This thesis demonstrates the continued usefulness of x-rays in the study of knee osteoarthritis; for understanding the long-term disease process, for highlighting structural features with increased associations with pain, and identifying early structural risk factors for progression to joint failure (TKR).

11.2 Future work

The results of this thesis open up several new avenues of research. The importance of lateral osteophytes was highlighted throughout this research, but the biological reasons for their increased association with pain are unknown. Their biomechanical properties and interaction with the surrounding soft-tissue should be explored using other imaging methods, such as MRI. Also, further analyses should be done using this dataset to identify their interaction with other structural features over time.

While significant associations between ROA and pain were found, it was acknowledged that the pain measurement used was simplistic. Further analysis on the same dataset using a more comprehensive measure of pain (such as WOMAC) should be done to explore the relationship between the severity of each feature to the severity of pain.

A limitation of this work was the lack of patellofemoral compartment view. This compartment has a well-known association with pain, and would complete the picture for 'whole-knee' disease.

The quantitative JSW measurements read using KneeMorf, showed mixed levels of association with pain. These measurements need to be further assessed in a more pre-defined population to improve the predictive validity.

12 REFERENCES

Abadie, E., Ethgen, D., Avouac, B., Bouvenot, G., Branco, J., Bruyere, O., et al. (2004). Recommendations for the use of new methods to assess the efficacy of disease-modifying drugs in the treatment of osteoarthritis. Osteoarthritis Cartilage, 12(4), 263-268.

Ahlback, S. (1968). Osteoarthrosis of the knee. A radiographic investigation. Acta Radiol Diagn (Stockh), Suppl 277:277-272.

Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., et al. (1986). Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum, 29(8), 1039-1049.

Altman, R. D., & Gold, G. E. (2007). Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage, 15 Suppl A, A1-56.

Altman, R. D., Hochberg, M., Murphy, W. A., Jr., Wolfe, F., & Lequesne, M. (1995). Atlas of individual radiographic features in osteoarthritis. Osteoarthritis Cartilage, 3 Suppl A, 3-70.

Anderson, J. J., & Felson, D. T. (1988). 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, 128(1), 179-189.

Auleley, G. R., Giraudeau, B., Dougados, M., & Ravaud, P. (2000). Radiographic assessment of hip osteoarthritis progression: impact of reading procedures for longitudinal studies. Ann Rheum Dis, 59(6), 422-427.

Beattie, K. A., Duryea, J., Pui, M., O'Neill, J., Boulos, P., Webber, C. E., et al. (2008). Minimum joint space width and tibial cartilage morphology in the knees of healthy individuals: a cross-sectional study. BMC Musculoskelet Disord, 9, 119.

Bedson, J., & Croft, P. R. (2008). The discordance between clinical and radiographic knee osteoarthritis: A systematic search and summary of the literature. BMC Musculoskelet Disord, 9, 116.

Bellamy, N., & Buchanan, W. W. (1986). A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clin Rheumatol, 5(2), 231-241.

Benichou, O. D., Hunter, D. J., Nelson, D. R., Guermazi, A., Eckstein, F., Kwoh, K., et al. (2010). One-year change in radiographic joint space width in patients with unilateral joint space narrowing: data from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken), 62(7), 924-931.

Blagojevic, M., Jinks, C., Jeffery, A., & Jordan, K. P. (2010). Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. Osteoarthritis Cartilage, 18(1), 24-33.

Bland, J. A., DG.;. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, i, 307-310.

Botha-Scheepers, S., Watt, I., Breedveld, F. C., & Kloppenburg, M. (2005). Reading radiographs in pairs or in chronological order influences radiological progression in osteoarthritis. Rheumatology (Oxford), 44(11), 1452-1455.

Brandt, K. D., Fife, R. S., Braunstein, E. M., & Katz, B. (1991). Radiographic grading of the severity of knee osteoarthritis: relation of the Kellgren and Lawrence grade to a grade based on joint space narrowing, and correlation with arthroscopic evidence of articular cartilage degeneration. Arthritis Rheum, 34(11), 1381-1386.

Brandt, K. D., Heilman, D. K., Slemenda, C., Katz, B. P., Mazzuca, S., Braunstein, E. M., et al. (2000). A comparison of lower extremity muscle strength, obesity, and depression scores in elderly subjects with knee pain with and without radiographic evidence of knee osteoarthritis. J Rheumatol, 27(8), 1937-1946.

Brandt, K. D., Mazzuca, S. A., Conrozier, T., Dacre, J. E., Peterfy, C. G., Provvedini, D., et al. (2002). Which is the best radiographic protocol for a clinical trial of a structure modifying drug in patients with knee osteoarthritis? J Rheumatol, 29(6), 1308-1320.

Brandt, K. D., Radin, E. L., Dieppe, P. A., & van de Putte, L. (2006). Yet more evidence that osteoarthritis is not a cartilage disease. Ann Rheum Dis, 65(10), 1261-1264.

Bruyere, O., Richy, F., & Reginster, J. Y. (2005). 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, 64(12), 1727-1730.

Buckland-Wright, C. (1995). Protocols for precise radio-anatomical positioning of the tibiofemoral and patellofemoral compartments of the knee. Osteoarthritis Cartilage, 3 Suppl A, 71-80.

Buckland-Wright, C. (1999). Radiographic assessment of osteoarthritis: comparison between existing methodologies. Osteoarthritis Cartilage, 7(4), 430-433.

Buckland-Wright, J. C., Bird, C. F., Ritter-Hrncirik, C. A., Cline, G. A., Tonkin, C., Hangartner, T. N., et al. (2003). X-ray technologists' reproducibility from automated measurements of the medial tibiofemoral joint space width in knee osteoarthritis for a multicenter, multinational clinical trial. J Rheumatol, 30(2), 329-338.

Buckland-Wright, J. C., Macfarlane, D. G., Williams, S. A., & Ward, R. J. (1995). Accuracy and precision of joint space width measurements in standard and macroradiographs of osteoarthritic knees. Ann Rheum Dis, 54(11), 872-880.

Buckland-Wright, J. C., Macfarlane, D. G., Lynch, J. A., Jasani, M. K., & Bradshaw, C. R. (1995B). Joint space width measures cartilage thickness in osteoarthritis of the knee: high resolution plain film and double contrast macroradiographic investigation. Ann Rheum Dis., 54(4), 263-268.

Buckland-Wright, J. C., Ward, R. J., Peterfy, C., Mojcik, C. F., & Leff, R. L. (2004). Reproducibility of the semiflexed (metatarsophalangeal) radiographic knee position and automated measurements of medial tibiofemoral joint space width in a multicenter clinical trial of knee osteoarthritis. J Rheumatol, 31(8), 1588-1597.

Burnett, S., Hart, D. J., Cooper, C., & Spector, T. (1994). A Radiographic Atlas of Osteoarthritis: Pfizer Ltd. .

Cho, H. J., Chang, C. B., Yoo, J. H., Kim, S. J., & Kim, T. K. (2010). Gender differences in the correlation between symptom and radiographic severity in patients with knee osteoarthritis. Clin Orthop Relat Res, 468(7), 1749-1758.

Cicuttini, F. M., Baker, J., Hart, D. J., & Spector, T. D. (1996). Association of pain with radiological changes in different compartments and views of the knee joint. Osteoarthritis Cartilage, 4(2), 143-147.

Cicuttini, F. M., Baker, J., Hart, D. J., & Spector, T. D. (1996). Association of pain with radiological changes in different compartments and views of the knee joint. Osteoarthritis Cartilage, 4(2), 143-147.

Claessens, A. A., Schouten, J. S., van den Ouweland, F. A., & Valkenburg, H. A. (1990). Do clinical findings associate with radiographic osteoarthritis of the knee? Ann Rheum Dis, 49(10), 771-774.

Coggon, D., Reading, I., Croft, P., McLaren, M., Barrett, D., & Cooper, C. (2001). Knee osteoarthritis and obesity. Int J Obes Relat Metab Disord, 25(5), 622-627.

Conrozier, T., Jousseaume, C. A., Mathieu, P., Tron, A. M., Caton, J., Bejui, J., et al. (1998). Quantitative measurement of joint space narrowing progression in hip osteoarthritis: a longitudinal retrospective study of patients treated by total hip arthroplasty. Br J Rheumatol, 37(9), 961-968.

Conrozier, T., & Vignon, E. (1996). Quantitative radiography in osteoarthritis: computerized measurement of radiographic knee and hip joint space. Baillieres Clin Rheumatol, 10(3), 429-433.

Cooper, C., Snow, S., McAlindon, T. E., Kellingray, S., Stuart, B., Coggon, D., et al. (2000). Risk factors for the incidence and progression of radiographic knee osteoarthritis. Arthritis Rheum, 43(5), 995-1000.

Creamer, P. (2000). Osteoarthritis pain and its treatment. Curr Opin Rheumatol, 12(5), 450-455.

Culliford, D. J., Maskell, J., Beard, D. J., Murray, D. W., Price, A. J., & Arden, N. K. (2010). Temporal trends in hip and knee replacement in the United Kingdom: 1991 to 2006. J Bone Joint Surg Br, 92(1), 130-135.

Dacre, J. E., & Huskisson, E. C. (1989). The automatic assessment of knee radiographs in osteoarthritis using digital image analysis. Br J Rheumatol, 28(6), 506-510.

Davis, M. A., Ettinger, W. H., Neuhaus, J. M., Barclay, J. D., & Segal, M. R. (1992). Correlates of knee pain among US adults with and without radiographic knee osteoarthritis. J Rheumatol, 19(12), 1943-1949.

Dieppe, P., Cushnaghan, J., Tucker, M., Browning, S., & Shepstone, L. (2000). The Bristol 'OA500 study': progression and impact of the disease after 8 years. Osteoarthritis Cartilage, 8(2), 63-68.

Dillon, C. F., Rasch, E. K., Gu, Q., & Hirsch, R. (2006). Prevalence of knee osteoarthritis in the United States: arthritis data from the Third National Health and Nutrition Examination Survey 1991-94. J Rheumatol, 33(11), 2271-2279.

Ding, C., Garnero, P., Cicuttini, F., Scott, F., Cooley, H., & Jones, G. (2005). Knee cartilage defects: association with early radiographic osteoarthritis, decreased cartilage volume, increased joint surface area and type II collagen breakdown. Osteoarthritis Cartilage, 13(3), 198-205.

Ding, C., Zhang, Y., & Hunter, D. (2013). Use of imaging techniques to predict progression in osteoarthritis. Curr Opin Rheumatol, 25(1), 127-135.

Dore, D., Quinn, S., Ding, C., Winzenberg, T., Zhai, G., Cicuttini, F., et al. (2006). Natural history and clinical significance of MRI-detected bone marrow lesions at the knee: a prospective study in community dwelling older adults. Arthritis Res Ther, 12(6), R223.

Duncan, R., Peat, G., Thomas, E., Hay, E., McCall, I., & Croft, P. (2007). Symptoms and radiographic osteoarthritis: not as discordant as they are made out to be? Ann Rheum Dis, 66(1), 86-91.

Duncan, R. C., Hay, E. M., Saklatvala, J., & Croft, P. R. (2006). Prevalence of radiographic osteoarthritis--it all depends on your point of view. Rheumatology (Oxford), 45(6), 757-760.

Eckstein, F., Cotofana, S., Wirth, W., Nevitt, M., John, M. R., Dreher, D., et al. (2011). Greater rates of cartilage loss in painful knees than in pain-free knees after adjustment for radiographic disease stage: data from the osteoarthritis initiative. Arthritis Rheum, 63(8), 2257-2267.

Emrani, P. S., Katz, J. N., Kessler, C. L., Reichmann, W. M., Wright, E. A., McAlindon, T. E., et al. (2008). Joint space narrowing and Kellgren-Lawrence progression in knee osteoarthritis: an analytic literature synthesis. Osteoarthritis Cartilage, 16(8), 873-882.

Englund, M., & Lohmander, L. S. (2004). Risk factors for symptomatic knee osteoarthritis fifteen to twenty-two years after meniscectomy. Arthritis Rheum., 50(9), 2811-2819.

Ersoz, M., & Ergun, S. (2003). Relationship between knee range of motion and Kellgren-Lawrence radiographic scores in knee osteoarthritis. Am J Phys Med Rehabil, 82(2), 110-115.

Farag, A., Elhabian, S., Abdelrahman, M., Graham, J., Chen, D., & Casanova, M. F. (2010). Shape modeling of the corpus callosum. Conf Proc IEEE Eng Med Biol Soc, 2010, 4288-4291.

Felson, D. T., McAlindon, T. E., Anderson, J. J., Naimark, A., Weissman, B. W., Aliabadi, P., et al. (1997). Defining radiographic osteoarthritis for the whole knee. Osteoarthritis Cartilage, 5(4), 241-250.

Felson, D. T., Naimark, A., Anderson, J., Kazis, L., Castelli, W., & Meenan, R. F. (1987). The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis Rheum, 30(8), 914-918.

Felson, D. T., & Nevitt, M. C. (2009). Blinding images to sequence in osteoarthritis: evidence from other diseases. Osteoarthritis Cartilage, 17(3), 281-283.

Felson, D. T., Nevitt, M. C., Yang, M., Clancy, M., Niu, J., Torner, J. C., et al. (2008). A new approach yields high rates of radiographic progression in knee osteoarthritis. J Rheumatol, 35(10), 2047-2054.

Felson, D. T., Niu, J., Guermazi, A., Roemer, F., Aliabadi, P., Clancy, M., et al. (2007). Correlation of the development of knee pain with enlarging bone marrow lesions on magnetic resonance imaging. Arthritis Rheum, 56(9), 2986-2992.

Felson, D. T., Zhang, Y., Hannan, M. T., Naimark, A., Weissman, B., Aliabadi, P., et al. (1997). Risk factors for incident radiographic knee osteoarthritis in the elderly: the Framingham Study. Arthritis Rheum, 40(4), 728-733.

Felson, D. T., Zhang, Y., Hannan, M. T., Naimark, A., Weissman, B. N., Aliabadi, P., et al. (1995). The incidence and natural history of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis Rheum, 38(10), 1500-1505.

Gale, D. R., Chaisson, C. E., Totterman, S. M., Schwartz, R. K., Gale, M. E., & Felson, D. (1999). Meniscal subluxation: association with osteoarthritis and joint space narrowing. Osteoarthritis Cartilage, 7(6), 526-532.

Garstang, S. V., & Stitik, T. P. (2006). Osteoarthritis: epidemiology, risk factors, and pathophysiology. Am J Phys Med Rehabil, 85(11 Suppl), S2-11; quiz S12-14.

Gossec, L., Jordan, J. M., Lam, M. A., Fang, F., Renner, J. B., Davis, A., et al. (2009). Comparative evaluation of three semi-quantitative radiographic grading techniques for hip osteoarthritis in terms of validity and reproducibility in 1404 radiographs: report of the OARSI-OMERACT Task Force. Osteoarthritis Cartilage, 17(2), 182-187.

Goulston, L. M., Kiran, A., Javaid, M. K., Soni, A., White, K. M., Hart, D. J., et al. (2011). Does obesity predict knee pain over fourteen years in women, independently of radiographic changes? Arthritis Care Res (Hoboken), 63(10), 1398-1406.

Guermazi, A., Burstein, D., Conaghan, P., Eckstein, F., Hellio Le Graverand-Gastineau, M. P., Keen, H., et al. (2008). Imaging in osteoarthritis. Rheum Dis Clin North Am, 34(3), 645-687.

Guermazi, A., Hunter, D. J., & Roemer, F. W. (2009). Plain radiography and magnetic resonance imaging diagnostics in osteoarthritis: validated staging and scoring. J Bone Joint Surg Am, 91 Suppl 1, 54-62.

Gunther, K. P., & Sun, Y. (1999). Reliability of radiographic assessment in hip and knee osteoarthritis. Osteoarthritis Cartilage, 7(2), 239-246.

Hannan, M. T., Felson, D. T., & Pincus, T. (2000). Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol, 27(6), 1513-1517.

Hart, D. J., Spector, T. D., Brown, P., Wilson, P., Doyle, D. V., & Silman, A. J. (1991). Clinical signs of early osteoarthritis: reproducibility and relation to x ray changes in 541 women in the general population. Ann Rheum Dis, 50(7), 467-470.

Hart, D. J., Mootoosamy, I., Doyle, D. V., & Spector, T. D. (1994). The relationship between osteoarthritis and osteoporosis in the general population: the Chingford Study. Ann Rheum Dis., 53(3), 158-162.

Hart, D. J., Doyle, D. V., & Spector, T. D. (1995). Association between metabolic factors and knee osteoarthritis in women: the Chingford Study. J Rheumatol., 22(6), 1118-1123.

Hart, D. J., & Spector, T. D. (1995). The classification and assessment of osteoarthritis. Baillieres Clin Rheumatol, 9(2), 407-432.

Hart, D. J., Doyle, D. V., & Spector, T. D. (1999). Incidence and risk factors for radiographic knee osteoarthritis in middle-aged women: the Chingford Study. Arthritis Rheum, 42(1), 17-24.

Hart, D. J., Doyle, D. V., & Spector, T. D. (1999). Incidence and risk factors for radiographic knee osteoarthritis in middle-aged women: the Chingford Study. Arthritis Rheum, 42(1), 17-24.

Hart, D. J., & Spector, T. D. (2003). Kellgren & Lawrence grade 1 osteophytes in the knee--doubtful or definite? Osteoarthritis Cartilage, 11(2), 149-150.

Hawker, G. A., Davis, A. M., French, M. R., Cibere, J., Jordan, J. M., March, L., et al. (2008). Development and preliminary psychometric testing of a new OA pain measure--an OARSI/OMERACT initiative. Osteoarthritis Cartilage, 16(4), 409-414.

Hawker, G. A., Gignac, M. A., Badley, E., Davis, A. M., French, M. R., Li, Y., et al. (2011). A longitudinal study to explain the pain-depression link in older adults with osteoarthritis. Arthritis Care Res (Hoboken). 63(10), 1382-1390.

Hayes, C. W., Jamadar, D. A., Welch, G. W., Jannausch, M. L., Lachance, L. L., Capul, D. C., et al. (2005). Osteoarthritis of the knee: comparison of MR imaging findings with radiographic severity measurements and pain in middle-aged women. Radiology, 237(3), 998-1007.

Hernandez-Molina, G., Guermazi, A., Niu, J., Gale, D., Goggins, J., Amin, S., et al. (2008). Central bone marrow lesions in symptomatic knee osteoarthritis and their relationship to anterior cruciate ligament tears and cartilage loss. Arthritis Rheum, 58(1), 130-136.

Hernborg, J. S., & Nilsson, B. E. (1977). The natural course of untreated osteoarthritis of the knee. Clin Orthop Relat Res(123), 130-137.

Hochberg, M. C. (2010). Structure-modifying effects of chondroitin sulfate in knee osteoarthritis: an updated meta-analysis of randomized placebo-controlled trials of 2-year duration. Osteoarthritis Cartilage, 18 Suppl 1, S28-31.

Hough, A. J. (2007). Pathology of Osteoarthritis. In R. W. A. Moskowitz, R. D.; Hochberg, M. C.; Buckwalter, J. A.; Goldberg, V. M. (Ed.), Osteoarthritis: Diagnosis and Medical/Surgical Management (4 ed., pp. 51-72). Philadelphia: Wolters Kluwer.

Hui, M., Doherty, M., & Zhang, W. (2011). Does smoking protect against osteoarthritis? Meta-analysis of observational studies. Ann Rheum Dis., 70(7), 1231-1237. doi: 1210.1136/ard.2010.142323. Epub 142011 Apr 142327.

Ingham, S. L., Zhang, W., Doherty, S. A., McWilliams, D. F., Muir, K. R., & Doherty, M. (2011). Incident knee pain in the Nottingham community: a 12-year retrospective cohort study. Osteoarthritis Cartilage, 19(7), 847-852.

Jordan, J. M., Helmick, C. G., Renner, J. B., Luta, G., Dragomir, A. D., Woodard, J., et al. (2007). Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. J Rheumatol, 34(1), 172-180.

Keen, R. W., Hart, D. J., Lanchbury, J. S., & Spector, T. D. (1997). Association of early osteoarthritis of the knee with a Taq I polymorphism of the vitamin D receptor gene. Arthritis Rheum, 40(8), 1444-1449.

Kellgren, J. H., & Lawrence, J. S. (1957). Radiological assessment of osteo-arthrosis. Ann Rheum Dis, 16(4), 494-502.

Kellgren, J. H., & Lawrence, J. S. (1963). Atlas of standard radiographs: the epidemiology of chronic rheumatism. Vol. 2: Oxford: Blackwell.

Khan, F. A., Koff, M. F., Noiseux, N. O., Bernhardt, K. A., O'Byrne, M. M., Larson, D. R., et al. (2008). Effect of local alignment on compartmental patterns of knee osteoarthritis. J Bone Joint Surg Am, 90(9), 1961-1969.

Kim, S. (2008). Changes in surgical loads and economic burden of hip and knee replacements in the US: 1997-2004. Arthritis Rheum, 59(4), 481-488.

Kornaat, P. R., Bloem, J. L., Ceulemans, R. Y., Riyazi, N., Rosendaal, F. R., Nelissen, R. G., et al. (2006). Osteoarthritis of the knee: association between clinical features and MR imaging findings. Radiology, 239(3), 811-817.

Lachance, L., Sowers, M., Jamadar, D., Jannausch, M., Hochberg, M., & Crutchfield, M. (2001). The experience of pain and emergent osteoarthritis of the knee. Osteoarthritis Cartilage, 9(6), 527-532.

Lachance, L., Sowers, M. F., Jamadar, D., & Hochberg, M. (2002). The natural history of emergent osteoarthritis of the knee in women. Osteoarthritis Cartilage, 10(11), 849-854.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174.

Lane, N. E., Brandt, K., Hawker, G., Peeva, E., Schreyer, E., Tsuji, W., et al. (2011). OARSI-FDA initiative: defining the disease state of osteoarthritis. Osteoarthritis Cartilage, 19(5), 478-482.

Lanyon, P., O'Reilly, S., Jones, A., & Doherty, M. (1998). Radiographic assessment of symptomatic knee osteoarthritis in the community: definitions and normal joint space. Ann Rheum Dis, 57(10), 595-601.

Lau, E. C., Cooper, C., Lam, D., Chan, V. N., Tsang, K. K., & Sham, A. (2000). Factors associated with osteoarthritis of the hip and knee in Hong Kong Chinese: obesity, joint injury, and occupational activities. Am J Epidemiol., 152(9), 855-862.

LaValley, M. P., McAlindon, T. E., Chaisson, C. E., Levy, D., & Felson, D. T. (2001). The validity of different definitions of radiographic worsening for longitudinal studies of knee osteoarthritis. J Clin Epidemiol, 54(1), 30-39.

Lawrence, J. S. (1977). Rheumatism in Populations. London, UK: William Heinemann Medical Books Ltd.

Laxafoss, E., Jacobsen, S., Gosvig, K. K., & Sonne-Holm, S. (2010). Case definitions of knee osteoarthritis in 4,151 unselected subjects: relevance for epidemiological studies: the Copenhagen Osteoarthritis Study. Skeletal Radiol, 39(9), 859-866.

Le Graverand, M. P., Vignon, E. P., Brandt, K. D., Mazzuca, S. A., Piperno, M., Buck, R., et al. (2008). Head-to-head comparison of the Lyon Schuss and fixed flexion radiographic techniques. Long-term reproducibility in normal knees and sensitivity to change in osteoarthritic knees. Ann Rheum Dis, 67(11), 1562-1566.

Lequesne, M., Brandt, K., Bellamy, N., Moskowitz, R., Menkes, C. J., Pelletier, J. P., et al. (1994). Guidelines for testing slow acting drugs in osteoarthritis. J Rheumatol Suppl, 41, 65-71; discussion 72-63.

Lethbridge-Cejku, M., Scott, W. W., Jr., Reichle, R., Ettinger, W. H., Zonderman, A., Costa, P., et al. (1995). Association of radiographic features of osteoarthritis of the knee with knee pain: data from the Baltimore Longitudinal Study of Aging. Arthritis Care Res, 8(3), 182-188.

Leyland, K. M., Hart, D. J., Javaid, M. K., Judge, A., Kiran, A., Soni, A., et al. (2012). The natural history of radiographic knee osteoarthritis: a fourteen-year population-based cohort study. Arthritis Rheum, 64(7), 2243-2251.

Lynch, J. A., Buckland-Wright, J. C., & Macfarlane, D. G. (1993). Precision of joint space width measurement in knee osteoarthritis from digital image analysis of high definition macroradiographs. Osteoarthritis Cartilage, 1(4), 209-218.

Marijnissen, A. C., Vincken, K. L., Vos, P. A., Saris, D. B., Viergever, M. A., Bijlsma, J. W., et al. (2008). Knee Images Digital Analysis (KIDA): a novel method to quantify individual radiographic features of knee osteoarthritis in detail. Osteoarthritis Cartilage, 16(2), 234-243.

Massardo, L., Watt, I., Cushnaghan, J., & Dieppe, P. (1989). Osteoarthritis of the knee joint: an eight year prospective study. Ann Rheum Dis, 48(11), 893-897.

Maurer, K. (1979). Basic data on arthritis knee, hip, and sacroiliac joints in adults ages 25-74 years. Vital Health Stat 11(213), 1-31.

Mazzuca, S. A., Brandt, K. D., & Katz, B. P. (1997). Is conventional radiography suitable for evaluation of a disease-modifying drug in patients with knee osteoarthritis? Osteoarthritis Cartilage, 5(4), 217-226.

Mazzuca, S. A., Brandt, K. D., Katz, B. P., Lane, K. A., & Buckwalter, K. A. (2006). Comparison of quantitative and semiquantitative indicators of joint space narrowing in subjects with knee osteoarthritis. Ann Rheum Dis, 65(1), 64-68.

McAlindon, T. E., Snow, S., Cooper, C., & Dieppe, P. A. (1992). Radiographic patterns of osteoarthritis of the knee joint in the community: the importance of the patellofemoral joint. Ann Rheum Dis, 51(7), 844-849.

McAlindon, T. E., Felson, D. T., Zhang, Y., Hannan, M. T., Aliabadi, P., Weissman, B., et al. (1996). Relation of dietary intake and serum levels of vitamin D to progression of osteoarthritis of the knee among participants in the Framingham Study. Ann Intern Med., 125(5), 353-359.

McAlindon, T. E., Wilson, P. W., Aliabadi, P., Weissman, B., & Felson, D. T. (1999). Level of physical activity and the risk of radiographic and symptomatic knee osteoarthritis in the elderly: the Framingham study. Am J Med., 106(2), 151-157.

Muraki, S., Oka, H., Akune, T., Mabuchi, A., En-yo, Y., Yoshida, M., et al. (2009). Prevalence of radiographic knee osteoarthritis and its association with knee pain in the elderly of Japanese population-based cohorts: the ROAD study. Osteoarthritis Cartilage, 17(9), 1137-1143.

Murphy, L., Schwartz, T. A., Helmick, C. G., Renner, J. B., Tudor, G., Koch, G., et al. (2008). Lifetime risk of symptomatic knee osteoarthritis. Arthritis Rheum, 59(9), 1207-1213.

Nagaosa, Y., Lanyon, P., & Doherty, M. (2002). Characterisation of size and direction of osteophyte in knee osteoarthritis: a radiographic study. Ann Rheum Dis, 61(4), 319-324.

Nagaosa, Y., Mateus, M., Hassan, B., Lanyon, P., & Doherty, M. (2000). Development of a logically devised line drawing atlas for grading of knee osteoarthritis. Ann Rheum Dis, 59(8), 587-595.

Neame, R., Zhang, W., Deighton, C., Doherty, M., Doherty, S., Lanyon, P., et al. (2004). Distribution of radiographic osteoarthritis between the right and left hands, hips, and knees. Arthritis Rheum, 50(5), 1487-1494.

Neame, R. L., Muir, K., Doherty, S., & Doherty, M. (2004B). Genetic risk of knee osteoarthritis: a sibling study. Ann Rheum Dis., 63(9), 1022-1027.

Neogi, T., Felson, D., Niu, J., Nevitt, M., Lewis, C. E., Aliabadi, P., et al. (2009). Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. BMJ, 339, b2844.

Neogi, T., Felson, D., Niu, J., Nevitt, M., Lewis, C. E., Aliabadi, P., et al. (2009). Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. BMJ, 339, b2844.

Neumann, G., Hunter, D., Nevitt, M., Chibnik, L. B., Kwoh, K., Chen, H., et al. (2009). Location specific radiographic joint space width for osteoarthritis progression. Osteoarthritis Cartilage, 17(6), 761-765.

Odding, E., Valkenburg, H. A., Algra, D., Vandenouweland, F. A., Grobbee, D. E., & Hofman, A. (1998). Associations of radiological osteoarthritis of the hip and knee with locomotor disability in the Rotterdam Study. Ann Rheum Dis, 57(4), 203-208.

Oka, H., Muraki, S., Akune, T., Nakamura, K., Kawaguchi, H., & Yoshimura, N. Normal and threshold values of radiographic parameters for knee osteoarthritis using a computer-assisted measuring system (KOACAD): the ROAD study. J Orthop Sci, 15(6), 781-789.

O'Reilly, S. C., Muir, K. R., & Doherty, M. (1996). Screening for pain in knee osteoarthritis: which question? Ann Rheum Dis, 55(12), 931-933.

Pavelka, K., Gatterova, J., & Altman, R. D. (2000). Radiographic progression of knee osteoarthritis in a Czech cohort. Clin Exp Rheumatol, 18(4), 473-477.

Peat, G., Thomas, E., Duncan, R., Wood, L., Wilkie, R., Hill, J., et al. (2007). Estimating the probability of radiographic osteoarthritis in the older patient with knee pain. Arthritis Rheum, 57(5), 794-802.

Peterfy, C. G., Guermazi, A., Zaim, S., Tirman, P. F., Miaux, Y., White, D., et al. (2004). Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage., 12(3), 177-190.

Petersson, I. F., Boegard, T., Saxne, T., Silman, A. J., & Svensson, B. (1997). Radiographic osteoarthritis of the knee classified by the Ahlback and Kellgren & Lawrence systems for the tibiofemoral joint in people aged 35-54 years with chronic knee pain. Ann Rheum Dis, 56(8), 493-496.

Piperno, M., Hellio Le Graverand, M. P., Conrozier, T., Bochu, M., Mathieu, P., & Vignon, E. (1998). Quantitative evaluation of joint space width in femorotibial osteoarthritis: comparison of three radiographic views. Osteoarthritis Cartilage, 6(4), 252-259.

Pottenger, L. A., Phillips, F. M., & Draganich, L. F. (1990). The effect of marginal osteophytes on reduction of varus-valgus instability in osteoarthritic knees. Arthritis Rheum, 33(6), 853-858.

Ravaud, P., Giraudeau, B., Auleley, G. R., Chastang, C., Poiraudeau, S., Ayral, X., et al. (1996). Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. J Rheumatol, 23(10), 1756-1764.

Raynauld, J. P., Martel-Pelletier, J., Haraoui, B., Choquette, D., Dorais, M., Wildi, L. M., et al. (2011). 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, 70(8), 1382-1388.

Reijman, M., Hazes, J. M., Bierma-Zeinstra, S. M., Koes, B. W., Christgau, S., Christiansen, C., et al. (2004). A new marker for osteoarthritis: cross-sectional and longitudinal approach. Arthritis Rheum, 50(8), 2471-2478.

Riskus, A. (2006). Approximation of a cubic Bezier curve by circular arcs and vice versa. Information Technology and Control, 35(4), 371-378.

Roos, E. M. (2005). Joint injury causes knee osteoarthritis in young adults. Curr Opin Rheumatol, 17(2), 195-200.

Samanta, A., Jones, A., Regan, M., Wilson, S., & Doherty, M. (1993). Is osteoarthritis in women affected by hormonal changes or smoking? Br J Rheumatol., 32(5), 366-370.

Schiphof, D., Boers, M., & Bierma-Zeinstra, S. M. (2008). Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. Ann Rheum Dis, 67(7), 1034-1036.

Schiphof, D., de Klerk, B. M., Kerkhof, H. J., Hofman, A., Koes, B. W., Boers, M., et al. (2011). Impact of different descriptions of the Kellgren and Lawrence classification criteria on the diagnosis of knee osteoarthritis. Ann Rheum Dis, 70(8), 1422-1427.

Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology, 20(4), 488-495.

Schmidt, J. E., Amrami, K. K., Manduca, A., & Kaufman, K. R. (2005). Semi-automated digital image analysis of joint space width in knee radiographs. Skeletal Radiol, 34(10), 639-643.

Schouten, J. S., van den Ouweland, F. A., & Valkenburg, H. A. (1992). A 12 year follow up study in the general population on prognostic factors of cartilage loss in osteoarthritis of the knee. Ann Rheum Dis, 51(8), 932-937.

Sharma, L. (2001). Local factors in osteoarthritis. Curr Opin Rheumatol, 13(5), 441-446.

Slemenda, C., Heilman, D. K., Brandt, K. D., Katz, B. P., Mazzuca, S. A., Braunstein, E. M., & Byrd, D. (1998). Reduced quadriceps strength relative to body weight: a risk factor for knee osteoarthritis in women? Arthritis Rheum., 41(11), 1951-1959.

Sofat, N., Ejindu, V., & Kiely, P. (2011). What makes osteoarthritis painful? The evidence for local and central pain processing. Rheumatology (Oxford), 50(12), 2157-2165.

Soni, A., Kiran, A., Hart, D. J., Leyland, K. M., Goulston, L., Cooper, C., et al. (2012). Prevalence of reported knee pain over twelve years in a community-based cohort. Arthritis Rheum, 64(4), 1145-1152.

Spector, T. D., Cooper, C., Cushnaghan, J., Hart, D. J., & Dieppe, P. A. (1992). A radiographic atlas of knee osteoarthritis: London: Springer Verlag.

Spector, T. D., Dacre, J. E., Harris, P. A., & Huskisson, E. C. (1992). Radiological progression of osteoarthritis: an 11 year follow up study of the knee. Ann Rheum Dis, 51(10), 1107-1110.

Spector, T. D., Hart, D. J., Byrne, J., Harris, P. A., Dacre, J. E., & Doyle, D. V. (1993). Definition of osteoarthritis of the knee for epidemiological studies. Ann Rheum Dis, 52(11), 790-794.

Spector, T. D., Cicuttini, F., Baker, J., Loughlin, J., & Hart, D. (1996). Genetic influences on osteoarthritis in women: a twin study. BMJ., 312(7036), 940-943.

Symmons, D., Mathers, C., Pfleger, B. (2002). Global burden of osteoarthritis in the year 2000. World Health Report, 5.

Szebenyi, B., Hollander, A. P., Dieppe, P., Quilty, B., Duddy, J., Clarke, S., et al. (2006). Associations between pain, function, and radiographic features in osteoarthritis of the knee. Arthritis Rheum, 54(1), 230-235.

Takahashi, T., Yamanaka, N., Ikeuchi, M., & Yamamoto, H. (2009). Reproducibility of joint space width and the intermargin distance measurements in patients with medial osteoarthritis of the knee in various degrees of flexion. Skeletal Radiol, 38(1), 37-42.

Thorstensson, C. A., Andersson, M. L., Jonsson, H., Saxne, T., & Petersson, I. F. (2008). The natural course of knee osteoarthritis in middle-aged individuals with knee pain - A 12 year follow-up using clinical and radiographic criteria. Ann Rheum Dis.

van Saase, J. L., van Romunde, L. K., Cats, A., Vandenbroucke, J. P., & Valkenburg, H. A. (1989). Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. Ann Rheum Dis, 48(4), 271-280.

Vani, P. J., R. (2010). A novel mathematical algorithm to represent and reconstruct some Bio-Metric Patterns. IJCA, 11(7), 26-33.

Vignon, E., Brandt, K. D., Mercier, C., Hochberg, M., Hunter, D., Mazzuca, S., et al. (2010). Alignment of the medial tibial plateau affects the rate of joint space narrowing in the osteoarthritic knee. Osteoarthritis Cartilage, 18(11), 1436-1440.

Wilkinson, C. E., Carr, A. J., & Doherty, M. (2005). Does increasing the grades of the knee osteoarthritis line drawing atlas alter its clinimetric properties? Ann Rheum Dis, 64(10), 1467-1473.

Williams, D. A., Farrell, M. J., Cunningham, J., Gracely, R. H., Ambrose, K., Cupps, T., et al. (2004). Knee pain and radiographic osteoarthritis interact in the prediction of levels of self-reported disability. Arthritis Rheum, 51(4), 558-561.

Wolfe, F., & Lane, N. E. (2002). The longterm outcome of osteoarthritis: rates and predictors of joint space narrowing in symptomatic patients with knee osteoarthritis. J Rheumatol, 29(1), 139-146.

Wolfe, F., & Lane, N. E. (2002). The longterm outcome of osteoarthritis: rates and predictors of joint space narrowing in symptomatic patients with knee osteoarthritis. J Rheumatol, 29(1), 139-146.

Woolf, A. D., & Pfleger, B. (2003). Burden of major musculoskeletal conditions. Bull World Health Organ, 81(9), 646-656.

Yoshimura, N., Muraki, S., Oka, H., Tanaka, S., Kawaguchi, H., Nakamura, K., & Akune, T. (2012). Accumulation of metabolic risk factors such as overweight, hypertension, dyslipidaemia, and impaired glucose tolerance raises the risk of occurrence and progression of knee osteoarthritis: a 3-year follow-up of the ROAD study. Osteoarthritis Cartilage., 20(11), 1217-1226.

Zhang, Y., Xu, L., Nevitt, M. C., Aliabadi, P., Yu, W., Qin, M., et al. (2001). 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, 44(9), 2065-2071.

Zhang, W., Robertson, J., Doherty, S., Liu, J. J., Maciewicz, R. A., Muir, K. R., & Doherty, M. (2008). Index to ring finger length ratio and the risk of osteoarthritis. Arthritis Rheum., 58(1), 137-144.

1 APPENDIX

1.1 Publications arising from the thesis

1.1.1 Journal articles

ARTHRITIS & RHEUMATISM
Vol. 64, No. 7, July 2012, pp 2243–2251
DOI 10.1002/art.34415

The Natural History of Radiographic Knee Osteoarthritis

A Fourteen-Year Population-Based Cohort Study

K. M. Leyland,[1] D. J. Hart,[2] M. K. Javaid,[1] A. Judge,[1] A. Kiran,[1] A. Soni,[1] L. M. Goulston,[3] C. Cooper,[4] T. D. Spector,[2] and N. K. Arden[5]

Objective. **To establish the natural history of radiographic knee osteoarthritis (OA) over 14 years in a community-based cohort.**

Methods. **We examined women from the Chingford Women's Study, a community-based cohort followed up for more than 14 years. We selected women for whom bilateral radiographs of the knees (with the legs in full extension) were obtained at approximately 5-year intervals. Radiographs were scored for OA in a blinded manner, using Kellgren/Lawrence (K/L) grades. Descriptive statistics and odds ratios (ORs) were used to compare the incidence, worsening, and progression of radiographic knee OA.**

Results. **A complete radiography series was available for 561 of the original 1,003 subjects enrolled in the study. The median age of these subjects at baseline was 53 years (interquartile range 48–58 years). At baseline, 13.7% of the subjects had radiographic knee OA (K/L grade ≥2) in at least one knee, and the prevalence increased to 47.8% by year 15. The annual cumulative incidence of radiographic knee OA was 2.3% between baseline and year 15. The annual rates of disease progression and worsening between baseline and year 15 were 2.8% and 3.0%, respectively. Subjects with a K/L grade of 1 at baseline were more likely to experience worsening by year 15 compared with subjects with a baseline grade of 0 (OR 4.5, 95% confidence interval 2.7–7.4).**

Conclusion. **This is the longest natural history study of radiographic knee OA to date. The results showed relatively low rates for the incidence and progression of radiographic knee OA; more than half of all subjects had no radiographic evidence of knee OA over a 15-year period of time. Subjects with a baseline K/L grade of 1 were more likely than subjects with other baseline K/L grades to experience worsening of knee OA.**

Knee osteoarthritis (OA) is one of the leading health burdens; in 2004 alone, the cost for knee replacement in the US was $14.6 billion (1). This dollar amount does not address the additional expenses associated with pain management, loss of work due to disability, and various treatment options such as physiotherapy and revision surgery. The economic burden of OA is increasing; 54% more knee replacements were performed in 2004 compared with 4 years earlier, and this number is estimated to increase to 1.4 million by 2015 (1). The trend has been further substantiated in a long-term study based in the UK, where the rate of knee replacements tripled between 1991 and 2006 (2). Because of the increasing health burden due to the aging population and a projected 45% lifetime risk of symptomatic knee OA developing, there is an urgent need to understand the natural course of knee OA in order to target preventative therapies and reduce known risk factors for both the incidence and progression of knee OA (3).

Plain film radiography is the diagnostic imaging technique used most commonly to evaluate knee OA.

Supported by the NIHR Musculoskeletal Biomedical Research Unit, University of Oxford, and by Arthritis Research UK.

[1]K. M. Leyland, MSc, M. K. Javaid, MBBS, PhD, MRCP, A. Judge, PhD, A. Kiran, PhD, A. Soni, BMBCh, MRCP: University of Oxford, Oxford, UK; [2]D. J. Hart, PhD, T. D. Spector, MD, MSc, FRCP: King's College London, London, UK; [3]L. M. Goulston, BSc (Hons): University of Southampton, Southampton, UK; [4]C. Cooper, MA, DM, FRCP, FFPH, FMedSci: University of Oxford, Oxford, UK, and University of Southampton, Southampton, UK; [5]N. K. Arden, MBBS, FRCP, MSc, MD: University of Oxford, Oxford, UK, and University of Southampton and Southampton University Hospitals, Southampton, UK.

Address correspondence to N. K. Arden, MBBS, FRCP, MSc, MD, Botnar Research Centre, Institute of Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, UK. E-mail: nka@mrc.soton.ac.uk.

Submitted for publication February 4, 2011; accepted in revised form January 26, 2012.

Although other imaging modalities such as magnetic resonance imaging (MRI) are being assessed within the research community, their advantage over radiographic assessment in clinical practice remains uncertain (4).

Research into the natural history of radiographic OA (ROA) of the knee has focused primarily on the incidence and progression in symptomatic subjects (5–8) and the progression of disease in older cohorts (9–11). Community-based studies have been limited by followup times that varied from 3 years to 12 years (9–14), and few data are available for followup times extending beyond this period. Previous studies have also focused only on baseline and followup radiography data. In those studies, the reported annual incidence rates ranged from 2% to 4% (9,11,12), the rates of ROA were significantly higher in women than in men, and the rates of incident ROA were twice as high as the rates of incident symptomatic OA (9).

The aim of this study was to assess the long-term prevalence, incidence, and progression of mild and moderate ROA of the knee in a well-described population-based cohort of middle-aged women; to compare incident unilateral and bilateral disease and progressive unilateral ROA using data for 5-year intervals; and to assess the changes in individual Kellgren/Lawrence (K/L) grades (15) over 14 years. This research will establish a long-term natural history of ROA of the knee.

SUBJECTS AND METHODS

Subjects. The participants were selected from the Chingford Women's Study, a prospective population-based longitudinal study of osteoporosis and OA. All women derived from the register of a large general practice in Chingford, North London, UK, who were between the ages of 45 years and 64 years were contacted in 1988–1989 and asked to participate. Of the 1,353 women contacted, 1,003 (78% response rate) attended the baseline visit; due to the 2-year recruitment period, the actual age range of these women was 44–67 years.

Clinic visits included administration of detailed questionnaires regarding musculoskeletal symptoms, physical evaluations, and knee radiography; pertinent risk factors for OA such as physical activity, smoking, and age were ascertained using a nurse-administered questionnaire. Height and weight measurements were obtained by staff members, and these data were used to calculate the body mass index (BMI) of the subjects. Groupings according to BMI were based on the World Health Organization categories, with a normal BMI defined as <25 kg/m^2, overweight defined as 25 kg/m^2 to <35 kg/m^2, and obese defined as ≥ 35 kg/m^2. Pain was evaluated using data from the baseline clinic visit assessing the presence/absence of current pain in each knee. Much of this information was evaluated repeatedly over the course of the prospective study. Information about comorbidities such as rheumatoid arthritis (RA) and fractures was collected, and total knee replacements (TKRs) were confirmed by general practice records in addition to self-report. Subjects identified as having RA (n = 6), including those who underwent TKRs due to RA (as determined from self-report), were excluded from the final analysis.

Subjects in the Chingford Study are a well-described predominantly white cohort who have been shown to be representative of women in the general UK population in terms of height, weight, and rates of hysterectomy but with a lower percentage of current smokers (16). By the year 15 clinic visit, 98 women had died, 76 had moved away, 22 were unable to be contacted, and 149 declined to attend. The study was approved by the Outer North East London Research Ethics Committee, and written consent was obtained from each woman.

Radiography protocols. Weight-bearing anteroposterior radiographs of the knees with the legs in full extension were obtained at baseline (year 1), year 5, year 10, and year 15. Both knees of each subject who was present were radiographed by experienced radiographers, using the same equipment each year. Standardized protocols were established at baseline and used for all subsequent visits. According to these protocols, views were standardized, with the back of the knees kept in contact with the cassette and the patella centralized over the lower end of the femur. A tube-to-film distance of 100 cm was used, with the beam centered 2.5 cm below the apex of the patella (12).

Radiographic grading. Radiographs were scored using a K/L global score (0 = normal; 1 = possible osteophyte, no joint space narrowing [JSN]; 2 = definite osteophyte, possible JSN; 3 = multiple osteophytes, definite JSN, sclerosis, and possible deformity of bone ends; 4 = large osteophytes, marked JSN, severe sclerosis, and definite deformity of bone ends) (15,17). TKRs and partial knee replacements were identified using a combination of self-report and general practice records and were further confirmed by reviewing the original radiographs. Subjects with knee replacements were included in the final analysis and were coded separately. Radiographs were read individually by year (unpaired), with blinding regarding subject identity and symptoms. The baseline and year 5 radiographs were read by the same 2 observers (TS and DH), and a single reader (DH) read the year 10 and year 15 radiographs. As previously reported, reproducibility of the radiographic grading system was confirmed using films from 50 women (100 knees), with observers reading the films 3 weeks apart. Kappa values for intraobserver reproducibility were 0.88 (95% confidence interval [95% CI] 0.87–0.89) and 0.79 (95% CI 0.78–0.80). Interobserver reproducibility was high, with a kappa value of 0.80 (95% CI 0.79–0.81) (18).

Statistical analysis. The present analysis was conducted using a subset of the Chingford cohort, which included only subjects for whom complete knee radiographs obtained at baseline, year 5, year 10, and year 15 were available as well as data for all pertinent baseline characteristics. Due to the inherent limitations of complete case analysis, a post hoc available-case analysis was performed when possible to check for dropout bias. Subjects with knee replacements were included in the analysis and placed in the groups "K/L grade 2 or above" and "ROA" unless they were explicitly listed sepa-

rately. The baseline age, BMI, and K/L grade of subjects lost to followup were compared with those of subjects selected for this study. Because none of the continuous variables had normal distributions, Mann-Whitney U tests were used. For categorical data, Pearson's chi-square test was used except when the expected cell counts were ≤5, in which case Fisher's exact test was used.

Prevalence was calculated at both the subject level (using the "worse knee" of each subject) and the knee level (with each subject supplying 2 knees to the analysis) and was defined using a K/L grade of ≥2 as the indication of disease presence and K/L grades of 0 and 1 as the lack of disease. The "worse knee" of each subject was determined by the knee with the higher K/L grade and was used as the index knee in the analysis.

Incidence was calculated at both the subject level (worse knee) and the knee level and was defined by having a K/L grade of 0 or 1 at the first period of observation and a grade of ≥2 at the second period of observation. The annual cumulative incidence was calculated by dividing the incidence by the number of years under observation. Incident unilateral and bilateral disease was defined as having a K/L grade of 0 or 1 in both knees at the first observation and having a grade of ≥2 in one or both knees at the next observation point, respectively.

Progression was calculated at the knee level and was defined as having a K/L grade of ≥2 at the first period of observation and showing an increase of at least one K/L grade by the second period of observation. At the subject level, progression was defined as unilateral disease at the first period of observation and bilateral disease at the second period of observation.

Worsening was calculated at the knee level and was defined as an increase of one K/L grade from any other grade (including grades 0 and 1). The group with worsening essentially includes incident cases, subjects with disease progression, as well as subjects with mild progression who moved from a K/L grade of 0 to a K/L grade of 1.

The development of incident ROA at each time point (year 5, year 10, and year 15) from a baseline K/L grade of 1 was compared with baseline K/L grades of 0 by calculating percentages (with 95% CIs) for each. This was stratified by quartiles of age (<50 years, 50–54 years, 55–60 years, and >60 years). Differences between groups were assessed by chi-square tests.

Odds ratios (ORs) were used to compare incident ROA at year 15 among subjects with a baseline K/L grade of 0 with that among subjects with a baseline K/L grade of 1 and to assess the odds of subjects with each baseline K/L grade progressing to TKR by year 15. These ORs were calculated by generalized estimating equation logistic regression models in order to account for clustering due to each subject contributing 2 knees to the analysis. The baseline characteristics (age, BMI, pain, and smoking status) of subjects in whom unilateral or bilateral disease developed were compared using logistic regression models. Finally, cross-tabulation was used to assess individual K/L grades at baseline and year 15. Statistical analysis was carried out using Stata version 10 (19) and SPSS version 17.0 (20).

RESULTS

The baseline median age of the subjects with complete followup was 53 years (interquartile range [IQR] 48–58 years). Of the original 1,003 women who were seen at baseline, 970 women underwent radiography at baseline, 831 had radiographs obtained at year 5, 819 had radiographs obtained at year 10, and 613 had radiographs obtained at year 15. Five hundred sixty-one women underwent radiography at all 4 visits and had complete demographic data recorded at baseline. Four hundred forty-two women had incomplete followup data and were excluded from the complete case analysis. Subjects lost to followup were slightly older than those with complete followup ($P < 0.0001$) and were more likely to be current smokers. In addition, the percentage of subjects with knee pain was slightly higher in the group lost to followup compared with the group with complete followup (33.5% versus 28.5%) (Table 1).

The prevalence of ROA (worse knee having a K/L grade of ≥2) was 13.7% at baseline, 23.9% at year 5, 36.4% at year 10, and 47.8% at year 15. Among all knees (n = 1,122), the prevalence of ROA was 9.5% at baseline, 17.5% at year 5, 27.5% at year 10, and 38.6% at year 15. Interval rates of the annual cumulative incidence at the subject level (worse knee) were 3.0% between baseline and year 5, 3.4% between year 5 and year 10, and 3.9% between year 10 and year 15, with 39.5% of subjects developing incident ROA in at least one knee between baseline and year 15. At the knee level, the annual cumulative incidence also increased steadily between each 5-year period, with increases of 2.3% between baseline and year 5, 2.6% between year 5 and year 10, and 3.3% between year 10 and year 15, and incident ROA developed in 32.5% of knees between baseline and year 15.

The annual cumulative incidence between baseline and year 15 was 2.3% at the knee level and 2.8% per year for the worse-knee subject-level analysis. Among the 106 knees with a K/L grade of ≥2 at baseline, 38.7% had disease progression between baseline and year 15. Approximately 12% of all knees showed worsening (increase of at least one K/L grade) between baseline and year 5, 23.4% showed worsening between year 5 and year 10, and 23.8% showed worsening between year 10 and year 15. When only baseline and year 15 data were analyzed, 41.5% of knees showed worsening by at least one K/L grade. A sensitivity analysis using all available knees at each time point did not show any significant differences compared with the complete case analysis (data not shown).

Table 1. Baseline characteristics of the entire cohort, subjects with complete followup, and subjects lost to followup*

Characteristic	Entire cohort (n = 1,003)	Complete followup (n = 561)	Lost to followup (n = 442)
Age, median (IQR) years	54.0 (49.0–60.0)	53.0 (48.0–58.0)	56.0 (50.0–61.0)†
BMI, median (IQR) kg/m^2	24.8 (22.6–27.6)	24.7 (22.7–27.3)	25.1 (22.6–28.2)
BMI <25 kg/m^2	51.4	53.7	48.4
Current smoker	22.8	19.3	27.4‡
K/L grade in worse knee			
Grade 0	79.2	80.9	76.8
Grade 1	6.1	5.4	7.1
Grade 2	9.6	9.8	9.3
Grade 3	4.9	3.9	6.1
Grade 4	0.1	0.0	0.2
TKR of worse knee	0.2	0.0	0.5
Knee pain	30.7	28.5	33.5§

* Except where indicated otherwise, values are the percent. A body mass index (BMI) of <25 kg/m^2 was considered normal. IQR = interquartile range; K/L = Kellgren/Lawrence; TKR = total knee replacement.
† $P < 0.0001$ versus subjects with complete followup.
‡ $P < 0.007$ versus subjects with complete followup.
§ $P < 0.052$ versus subjects with complete followup.

When the 5-year cumulative incidence of ROA was examined at each time point and age was stratified into quartiles (<50 years, 50–54 years, 55–60 years, and >60 years), a linear trend ($P < 0.002$) was evident, with the oldest age group having the highest percentage of incident ROA (Figure 1). By year 15, incident ROA had developed in 26.0% of subjects who were younger than age 50 years at baseline, 34.1% of subjects who were 50–54 years of age at baseline, 31.7% of those ages 55–60 years at baseline, and 42.2% of subjects who were older than age 60 years at baseline. The difference between the youngest and oldest age groups was significant ($P < 0.01$), although the difference between the 2 middle-aged groups was not ($P = 0.584$). When age was stratified into 2 age bands (<55 years and ≥55 years), the difference between incident ROA in the 2 groups was statistically significant ($P = 0.017$). When the 5-year cumulative incidence was analyzed according to BMI category, the percentage of knees with incident ROA at year 5 was roughly similar between the groups. By year 10 and year 15, however, the cumulative incidence among obese subjects was almost 20% higher than that among subjects in both the normal and overweight categories (Figure 2). No difference in the incidence of ROA between subjects who were premenopausal and those who were postmenopausal at baseline was observed ($P = 0.193$).

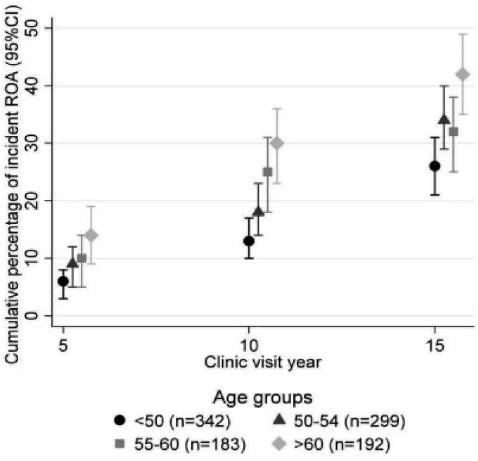


Figure 1. Point estimates and 95% confidence intervals (95% CIs) for the cumulative percentage of women with incident radiographic osteoarthritis (ROA) at each visit, stratified by baseline age group (n = 1,016 knees).

Cross-tabulation of individual K/L grades and TKRs at baseline and year 15 (Table 2) demonstrated that 51.3% of 1,122 knees had a K/L grade of 0 throughout the study period, while 41.5% of knees worsened by at least one grade. Among the subjects with a K/L grade of ≥1 at baseline (n = 167), 37.1% remained at the same grade, and 51.5% worsened (including progression to TKR) by year 15. Knees with a baseline K/L grade of 1 (n = 61) had a higher percentage of progression (73.8%) compared with knees with any other K/L grade at baseline. Knees with a baseline K/L grade of 2 (n = 76) were the next most likely to undergo progression, with 47.7% increasing by at least one K/L grade over 15 years; 1.7% of knees were scored as having

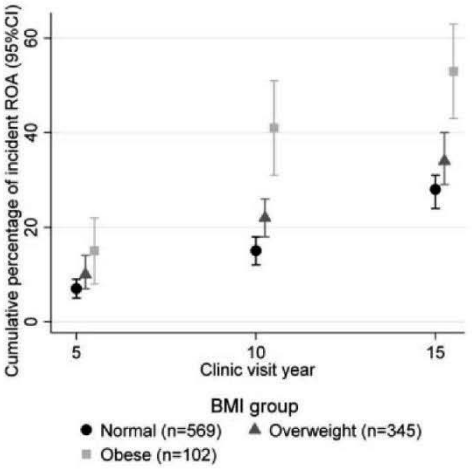


Figure 2. Point estimates and 95% confidence intervals (95% CIs) for the cumulative percentage of women with incident radiographic osteoarthritis (ROA) at each visit, stratified by baseline body mass index (BMI) category (n = 1,016 knees).

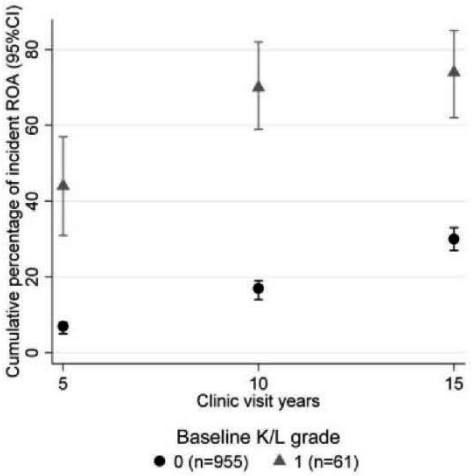


Figure 3. Point estimates and 95% confidence intervals (95% CIs) for the cumulative percentage of women with incident radiographic osteoarthritis (ROA) at each visit, stratified by baseline Kellgren/Lawrence (K/L) grades of 0 or 1 (n = 1,016 knees).

regressed to a lower K/L grade by year 15. Ten (1.1%) of 955 knees with a baseline K/L grade of 0 progressed to TKR by year 15, compared with 3 (4.9%) of 61 knees with a K/L grade of 1 at baseline, 4 (5.3%) of 76 knees with a K/L grade of 2 at baseline, and 2 (6.7%) of 30 knees with a K/L grade of 3 at baseline. Among subjects with baseline K/L grades of 0, 1, 2, and 3, the respective percentages of knees with pain at baseline were 21.5%, 19.7%, 39.5%, and 26.7%.

When rates of worsening for all knees (n = 1,122) were analyzed, knees in the group with the highest number of TKRs by year 15 had a baseline K/L grade of 0. The odds of having a TKR by year 15 were similar for knees with baseline K/L grades of 1–3 and those with a baseline grade of 0 (for grade 1, OR 4.7 [95% CI 1.0–22.2]; for grade 2, OR 5.9 [95% CI 1.9–18.2]; for grade 3, OR 4.6 [95% CI 0.3–65.3]). Among subjects who underwent a TKR by year 15, 52.4% reported having pain at the baseline visit.

Although a K/L grade of 1 is not considered diagnostic of ROA, when the data were stratified by an initial baseline K/L grade of 0 or 1 (Figure 3), the difference between groups in the cumulative incidence was significant at each visit ($P < 0.001$). The odds of a subject with a baseline K/L grade of 1 developing incident ROA by year 15 was 4.5-fold (95% CI 2.7–7.4) that of the odds for a subject with a baseline K/L grade of 0. When the data were stratified by age groups, these differences remained significant ($P < 0.05$) at all visits

Table 2. Cross-tabulation of baseline and year 15 K/L grades in 1,122 knees*

	K/L grade at year 15					
Baseline K/L grade	0	1	2	3	4	TKR at year 15
0 (n = 955)	575 (60.2)	95 (10.0)	157 (16.4)	116 (12.2)	2 (0.2)	10 (1.1)
1 (n = 61)	12 (19.7)	4 (6.6)	24 (39.3)	18 (29.5)	0 (0.0)	3 (4.9)
2 (n = 76)	0 (0.0)	1 (1.3)	39 (51.3)	32 (42.1)	0 (0.0)	4 (5.3)
3 (n = 30)	1 (3.3)	1 (3.3)	4 (13.3)	19 (63.3)	3 (10.0)	2 (6.7)

* Values are the number (%), with percentages calculated by row. K/L = Kellgren/Lawrence; TKR = total knee replacement.

Table 3. Cross-tabulation of radiographic knee OA status at baseline versus year 5, year 5 versus year 10, and year 10 versus year 15*

Radiographic knee OA	None	Unilateral	Bilateral
Baseline versus year 5			
None (n = 484)	426 (88.0)	42 (8.7)	16 (3.3)
Unilateral (n = 48)	1 (2.1)	30 (62.5)	17 (35.4)
Bilateral (n = 29)	0 (0.0)	0 (0.0)	29 (100.0)
Year 5 versus year 10			
None (n = 427)	355 (83.1)	51 (11.9)	21 (4.9)
Unilateral (n = 72)	2 (2.8)	44 (61.1)	26 (36.1)
Bilateral (n = 62)	0 (0.0)	5 (8.1)	57 (91.9)
Year 10 versus year 15			
None (n = 357)	287 (80.4)	43 (12.0)	27 (7.7)
Unilateral (n = 100)	6 (6.0)	56 (56.0)	38 (38.0)
Bilateral (n = 104)	0 (0.0)	4 (3.8)	100 (96.2)

* Values are the number (%), with percentages calculated by row. OA = osteoarthritis.

and in all age groups except at year 5 in the group of subjects >60 years of age (OR 2.8, 95% CI 0.9–9.0 [P = 0.085]).

The prevalences of unilateral and bilateral ROA at baseline were 8.6% and 5.2%, respectively (Table 3), and at year 15 were 18.4% and 29.4%, respectively. Among the 484 subjects without ROA at baseline, 293 (60.5%) remained free of disease at year 15, compared with 14 (29.2%) of 48 subjects with unilateral disease at baseline and 28 (96.6%) of 29 subjects with bilateral disease at baseline. Among the subjects in whom either incident unilateral or bilateral disease developed within any 5-year followup period (n = 200), bilateral disease developed in 32.0%, and unilateral disease developed in 68.0%. Among the subjects in whom unilateral disease developed between baseline and year 10 (n = 143), 56.6% had progressed to bilateral disease by year 15.

When the cumulative incidence was stratified by age quartiles for baseline and year 15 data, a significant difference was observed between the youngest (age <55 years) and oldest (age >60 years) age groups (P = 0.003), although adjacent age groups were not significantly different from one another (P = 0.642). Baseline characteristics were compared between subjects who remained free of disease over the 15-year study period and subjects who experienced progression to unilateral or bilateral disease.

All subjects who experienced disease progression were more likely to have a higher BMI, and those in whom bilateral disease developed (from no disease or unilateral disease) were more likely to be older. Subjects in whom unilateral disease developed and those who experienced progression from unilateral to bilateral disease were more likely to have pain at baseline compared with subjects in whom ROA did not develop.

DISCUSSION

The novel findings of this research were as follows: the annual rates of disease incidence, progression, and worsening between baseline and year 15 were 2.3%, 2.8%, and 3.0%, respectively; there are 3 potential symmetry-based phenotypes for knee ROA (incident unilateral, incident bilateral, and progressive unilateral to bilateral disease); and although the risk of TKR was associated with an increasing baseline K/L grade, the majority of knees that underwent a total replacement by year 15 had a baseline K/L grade of 0.

More than half of the subjects (52.2%) remained free of radiographic knee OA over the course of the study. At the year 15 visit, 38.6% of knees had prevalent ROA, compared with 9.5% of knees at baseline. Annual rates of knee progression (2.8%) and worsening (3.0%) between baseline and year 15 were slightly lower than those observed in other community-based cohorts, which were 3.5–8.0% for progression (8,9,11) and 4.4% for worsening (11). Rates in established symptomatic cohorts were similar, varying between 3.3% and 7.7% for worsening (6,21) and from 4.0% to 8.8% for progression (7,8). The slightly lower rates observed in the Chingford Study are likely a consequence of both the relatively young age of the cohort at the start of the study and the length of the study.

This study demonstrated an annual cumulative incidence of radiographic knee OA of 2.3% between baseline and year 15. When the data were broken down by 5-year intervals, the annual rate between baseline and year 5 was in the lower range (2.3%), with a high of 3.3% between year 10 and year 15. This is likely a result of the increasing age of the sample, with a median age of 53 years (IQR 48–58 years) at baseline and 68 years (IQR 63.5–72.5 years) by year 15. When the cumulative incidence was stratified by age quartiles, a significant difference was observed between the youngest (<55 years) and oldest (>60 years) age groups, although adjacent age groups were not significantly different from one another. Analyses conducted in primarily symptomatic cohorts have shown much higher rates of annual cumulative incidence (up to 4.0%) (8). Cohort studies and case–control studies that included both symptomatic and asymptomatic subjects were more comparable with these results, with percentages ranging between 2.0% and 2.5% (9–11). As would be expected, however, the cumulative incidence among obese subjects was almost 20% higher by year 15 than that among subjects in both the normal and overweight categories.

Assessing a cross-tabulation of individual K/L

grades over 15 years, rather than using the K/L grade as a binary variable, demonstrated that knees with specific K/L grades of 1 and 3 are more likely to progress to a higher grade or to remain stable, respectively, even over a long period of time. Fewer than half of all knees (41.5%) worsened by at least one K/L grade over the 14 years of the study. The majority of subjects (68.4%) who underwent a knee replacement by year 15 did not have evidence of conventional ROA (K/L grade ≥2) at baseline, and more than half had current knee pain. This suggests that radiographs are not necessarily the optimal tool for predicting TKR as the long-term outcome in younger subjects (median age 53.0 years at baseline).

The comparison of incident ROA between subjects with baseline K/L grades of 0 and 1 extends the time line of an earlier nested case–control study within this cohort (22) and other recent research (23) that emphasizes the importance of subjects with a K/L grade of 1 being treated distinctly from those with a K/L grade of 0. The higher risk of a subject with a baseline K/L grade of 1 progressing to incident ROA (4.5-fold the odds associated with a K/L grade of 0) suggests that grades of 1 are an important indicator of longitudinal incidence. Lachance et al examined the incidence and progression of mild K/L grades over 3 years and reported that women with a K/L grade of 1 were 6.4-fold more likely than those with a K/L grade of 0 to progress to a grade of ≥2 (13).

Among the subjects in whom incident ROA developed, approximately one-third developed bilateral ROA between each clinic visit, while the remaining two-thirds developed unilateral ROA. More than one-third of the knees with unilateral disease progressed to bilateral disease between each clinic visit, while the rest remained stable. The baseline characteristics of these groups (age, BMI, and pain) were different from those of subjects who remained free of ROA over the study period. These data are possibly describing 3 distinct subsets of ROA, in which some subjects have slow progression from no disease to unilateral and then bilateral ROA, while others have more rapid progression to bilateral disease within a 5-year period. This could reflect a difference between environmental factors (i.e., functional effect of having contralateral knee OA) and genetic factors (i.e., genetic predisposition to ROA) (24), although further work is required to validate these findings.

Limitations of this study include the effect of radiographic views, scoring methods, radiographic blinding, inclusion criteria, and loss to followup (which is the most common limitation of studies of this length). The standard view used for radiography of knees at the start of the Chingford Study was anteroposterior, fully extended, and weight-bearing. Although the prevailing opinion is to use standard semiflexed views (25,26) and the more rigorous fluoroscopy-assisted positioning (27) due to underestimation of joint space narrowing in fully extended views (28), long-term studies often continue using the same radiographic protocol as that used at the baseline visit in order to more accurately evaluate change. The patellofemoral compartment is known to be an important component of "whole organ" knee OA, and the presence of patellofemoral ROA is highly associated with pain and disability (29,30). The lack of additional views that allow imaging of the patellofemoral compartment, such as skyline and/or lateral, is a limitation of this study that should be addressed in future natural history studies of ROA.

The K/L scoring system has been the primary method used to evaluate radiographic knee OA in similar studies (5,7–11,13,14) but is commonly criticized for several known limitations. The K/L system assumes a nonvalidated natural disease progression that is extremely osteophyte-centric, and also has several different "official" and modified versions, all of which are in use (15,17,31). Despite these negative attributes, K/L scores have a high level of reproducibility (8,11,32), there is a strong correlation between pain and increasing K/L grades (33,34), and K/L grade–defined ROA is present in the majority of subjects who present with knee pain (35). Although the relationship between K/L grades and pain is by no means perfect, other imaging modalities, such as MRI, have not yet demonstrated a better specificity than plain film radiography (36).

A potential limitation of reading radiographs with the reader blinded to order is that grades may decrease over time, because they are not being read in a method that allows for the evaluation of change. The percentage of "regressive" grades in this study (1.7%) was much lower than that in studies using similar blinding methods (5.5–7.5%) (13,22). Interval censoring was used due to the lack of exact dates for radiographic incidence, progression, and worsening for each knee. Scores were evaluated only on the date that each radiograph was obtained, so that knees in which disease progressed immediately after a visit would be recorded as not having worsened until the next clinic visit 4–5 years later. This may have contributed to an overestimation of time interval until the development of incident or progressive ROA.

Due to the original study design, the results of this study are also restricted to the natural history of

ROA in white women. Although it is possible that the results could be loosely applied generally, there are known differences in prevalence, incidence, and progression between sexes and between subjects of different ancestry (23,29,34,37). The high projected lifetime risk of symptomatic knee OA (45%) emphasizes the importance of using symptoms when defining OA (3). The lack of information regarding pain in this analysis limits its clinical application and should be addressed in future work.

Subjects lost to followup represent a major limitation of all long-term cohort studies, such as the Chingford Study. There is a potential for study bias due to deaths, subjects withdrawing because of disability and illness, and generally having a healthier cohort attending the followup visits. Although the baseline characteristics of subjects included in the analysis and those of subjects lost to followup were similar enough not to imply a severe bias (i.e., slightly older, more likely to smoke, and slightly more knee pain), the possibility remains that subjects lost to followup, for whatever reason, would have had significantly worse ROA than those included in this analysis. There is no way to know the potential effect of this type of bias on any study with this design. The use of complete case analysis can add further bias due to the need to exclude subjects who were not present for intermediate visits; however, only a small percentage of subjects who were present at both baseline and year 15 missed any other visits. Among the subjects for whom both baseline and year 15 data were available, 29 were excluded due to missing year 5 data, 8 were excluded due to missing year 10 data, and 3 were excluded due to missing data for both the year 5 and year 10 visits.

This analysis represents the longest natural history study of radiographic knee OA to date and is intended to provide novel data regarding the trend of individual K/L grades during 5-year intervals over 14 years. No other natural history study of this type has included intermediate radiography scores (i.e., scores obtained between the baseline visit and the followup visit). The inclusion criteria were deliberately nonrestrictive in order to gain a more accurate picture of ROA in a normal population, including analysis of the progression of mild ROA in a relatively young cohort. The Chingford Study has an extremely high response rate for a study of this length, with >50% of the original 1,003 women attending all clinic visits involving knee radiography over the 14 years of followup.

In conclusion, this study showed that the annual rates of disease incidence, progression, and worsening between baseline and year 15 were 2.3%, 2.8%, and 3.0%, respectively; that more than half of the subjects (52.2%) remained free of radiographic knee OA over 14 years; that 3 potential phenotypes exist for knee ROA based on symmetry; that knees with a baseline K/L grade of 1 had a 4.5-fold greater risk of developing incident ROA compared with knees with a baseline K/L grade of 0; and that the majority of knees that had undergone total replacement by the time of the followup visit did not have ROA at baseline.

ACKNOWLEDGMENTS

We would like to thank all of the participants in the Chingford Women's Study and Maxine Daniels and Dr. Alan Hakim for their time and dedication.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Arden had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Leyland, Hart, Javaid, Spector, Arden.
Acquisition of data. Leyland, Hart, Javaid, Goulston, Spector, Arden.
Analysis and interpretation of data. Leyland, Hart, Javaid, Judge, Kiran, Soni, Goulston, Cooper, Arden.

REFERENCES

1. Kim S. Changes in surgical loads and economic burden of hip and knee replacements in the US: 1997–2004. Arthritis Rheum 2008; 59:481–8.
2. Culliford DJ, Maskell J, Beard DJ, Murray DW, Price AJ, Arden NK. Temporal trends in hip and knee replacement in the United Kingdom: 1991 to 2006. J Bone Joint Surg Br 2010;92:130–5.
3. Murphy L, Schwartz TA, Helmick CG, Renner JB, Tudor G, Koch G, et al. Lifetime risk of symptomatic knee osteoarthritis. Arthritis Rheum 2008;59:1207–13.
4. Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol 2000;27:1513–7.
5. Spector TD, Dacre JE, Harris PA, Huskisson EC. Radiological progression of osteoarthritis: an 11 year follow up study of the knee. Ann Rheum Dis 1992;51:1107–10.
6. Massardo L, Watt I, Cushnaghan J, Dieppe P. Osteoarthritis of the knee joint: an eight year prospective study. Ann Rheum Dis 1989;48:893–7.
7. Pavelka K, Gatterova J, Altman RD. Radiographic progression of knee osteoarthritis in a Czech cohort. Clin Exp Rheumatol 2000;18:473–7.
8. Thorstensson CA, Andersson ML, Jonsson H, Saxne T, Petersson IF. The natural course of knee osteoarthritis in middle-aged subjects with knee pain: 12-year follow-up using clinical and radiographic criteria. Ann Rheum Dis 2009;68:1890–3.
9. Felson DT, Zhang Y, Hannan MT, Naimark A, Weissman BN, Aliabadi P, et al. The incidence and natural history of knee osteoarthritis in the elderly: the Framingham Osteoarthritis Study. Arthritis Rheum 1995;38:1500–5.
10. Felson DT, Zhang Y, Hannan MT, Naimark A, Weissman B, Aliabadi P, et al. Risk factors for incident radiographic knee

osteoarthritis in the elderly: the Framingham Study. Arthritis Rheum 1997:40:728–33.
11. Cooper C, Snow S, McAlindon TE, Kellingray S, Stuart B, Coggon D, et al. Risk factors for the incidence and progression of radiographic knee osteoarthritis. Arthritis Rheum 2000;43: 995–1000.
12. Hart DJ, Doyle DV, Spector TD. Incidence and risk factors for radiographic knee osteoarthritis in middle-aged women: the Chingford Study. Arthritis Rheum 1999;42:17–24.
13. Lachance L, Sowers MF, Jamadar D, Hochberg M. The natural history of emergent osteoarthritis of the knee in women. Osteoarthritis Cartilage 2002;10:849–54.
14. 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;51:932–7.
15. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957;16:494–502.
16. Hart DJ, Mootoosamy I, Doyle DV, Spector TD. The relationship between osteoarthritis and osteoporosis in the general population: the Chingford Study. Ann Rheum Dis 1994;53:158–63.
17. Kellgren JH, Lawrence JS. Atlas of standard radiographs. Vol. II. The epidemiology of chronic rheumatism. Oxford: Blackwell Scientific; 1963.
18. 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;52:790–4.
19. Stata statistical software: release 10. College Station (TX): StataCorp; 2007.
20. Statistical package for the social sciences:version 17.0.1. Chicago: SPSS; 2008.
21. Spector TD, Cooper C, Cushnaghan J, Hart DJ, Dieppe PA. A radiographic atlas of knee osteoarthritis. London: Springer Verlag; 1992.
22. Hart DJ, Spector TD. Kellgren & Lawrence grade 1 osteophytes in the knee: doubtful or definite? Osteoarthritis Cartilage 2003;11: 149–50.
23. 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;9:527–32.
24. Driban JB, Sitler MR, Barbe MF, Balasubramanian E. Is osteoarthritis a heterogeneous disease that can be stratified into subsets? Clin Rheumatol 2010;29:123–31.
25. 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;29:1308–20.
26. Buckland-Wright C. Radiographic assessment of osteoarthritis: comparison between existing methodologies. Osteoarthritis Cartilage 1999;7:430–3.
27. Mazzuca SA, Brandt KD, Katz BP. Is conventional radiography suitable for evaluation of a disease-modifying drug in patients with knee osteoarthritis? Osteoarthritis Cartilage 1997;5:217–26.
28. Buckland-Wright C. Protocols for precise radio-anatomical positioning of the tibiofemoral and patellofemoral compartments of the knee. Osteoarthritis Cartilage 1995;3 Suppl A:71–80.
29. 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: 844–9.
30. Duncan RC, Hay EM, Saklatvala J, Croft PR. Prevalence of radiographic osteoarthritis: it all depends on your point of view. Rheumatology (Oxford) 2006;45:757–60.
31. Lawrence JS. Rheumatism in populations. London: William Heinemann Medical Books; 1977.
32. Neame RL, Muir K, Doherty S, Doherty M. Genetic risk of knee osteoarthritis: a sibling study. Ann Rheum Dis 2004;63:1022–7.
33. 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:b2844.
34. Cho HJ, Chang CB, Yoo JH, Kim SJ, Kim TK. Gender differences in the correlation between symptom and radiographic severity in patients with knee osteoarthritis. Clin Orthop Relat Res 2010;468: 1749–58.
35. Duncan R, Peat G, Thomas E, Hay E, McCall I, Croft P. Symptoms and radiographic osteoarthritis: not as discordant as they are made out to be? Ann Rheum Dis 2007;66:86–91.
36. Kornaat PR, Bloem JL, Ceulemans RY, Riyazi N, Rosendaal FR, Nelissen RG, et al. Osteoarthritis of the knee: association between clinical features and MR imaging findings. Radiology 2006;239: 811–7.
37. 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;57:203–8.

1.1.2 Conference papers

1.1.2.1 British Society for Rheumatology 2010 – oral abstract

Friday 23 April 2010, 13:45-15:15

BSR CONCURRENT ORAL PRESENTATION OF ABSTRACTS

Concurrent Oral 11 – Osteoarthritis

OP73. MECHANICAL LOAD DRIVES INFLAMMATORY GENE EXPRESSION AND DISEASE IN MURINE OA

Annika Burleigh, Jeremy Saklatvala and Tonia L. Vincent
Cell Signalling, Kennedy Institute of Rheumatology, London, UK

Background: The role of mechanical factors in OA pathogenesis is undisputed, but how these factors drive the processes that lead to joint disease is unknown. What is clear is that OA is not simply a disease of cartilage attrition as a result of abnormal or repetitive wearing of the surfaces, but that it requires activation of pathways that lead to expression of degradative enzymes, which break down the matrix. We have established a murine surgical model of OA, that causes destabilization of the medial meniscus (DMM) and which produces robust cartilage degeneration within 8–12 weeks of surgery. The aim of this study was to examine the early expression of inflammatory genes in the joints of OA mice and to assess the influence of joint loading on gene expression and the development of OA.
Methods: DMM surgery was performed on 10 week old male C57Bl6 mice by cutting the right menisco-tibial ligament. Some mice underwent sham surgery where the capsule of the joint was opened but the menicso-tibial ligament was left intact. The joints of these mice were either left for up to 12 weeks and sectioned for histological scoring or had RNA extracted at early time points post surgery. Microarray analysis was performed and regulated genes were selected and validated quantitatively by RT-PCR using Taqman high density microfluidic cards. A subtotal reduction in weight bearing through the ipsilateral hind limb was induced by cutting the sciatic nerve at the time of DMM surgery.
Results: Compared with sham operated mice, DMM surgery strongly induced a number of genes within 6 h of surgery. Of these, the chemokine CCL2, TNF-stimulated gene 6 (TSG-6), IL-6, serum amyloid A (SAA) and arginase were the most highly regulated. Some metalloproteinases such as ADAMTS4, ADAMTS5 and MMP3 were regulated from 3 days, albeit less strongly. Mice that had undergone sciatic neurectomy exhibited abnormal gait; some weight was born through the limb, but the leg was maintained in full extension and walking was achieved by flexion at the hip. When DMM surgery was performed at the same time as sciatic neurectomy, the joints showed no evidence of OA even 12 post surgery. Analysis of early gene expression in these animals revealed a striking abrogation of greater than 80% of the inflammatory genes studied.
Conclusions: These data show that there is an early inflammatory response in the joint in response to knee destabilization injury/surgery and importantly, indicate that inflammatory gene expression is dependent upon mechanical loading through the joint. We hypothesize that upon joint destabilization the perceived mechanical loads in the articular cartilage reach a threshold at which inflammatory genes are expressed. An intact meniscus (which helps dissipate load) or partial immobilization can reduce the perceived loads to below this threshold thus protecting the joint from degradation.
Disclosure statement: All authors have declared no conflicts of interest.

OP74. CHONDROCYTE IL-1β EXPRESSION IS A CLINICALLY SIGNIFICANT MARKER OF PROGRESSION AND RADIOGRAPHIC SEVERITY IN OSTEOARTHRITIS OF THE HIP

Evin Sowden[1], David Kynaston[1], Madhurima Rai[3], Elliot Yates[3], Winston Kim[2] and Anthony Freemont[3]
[1]*Rheumatology, Salford Royal NHS Foundation Trust, Salford, UK;* [2]*Orthopaedics, Salford Royal NHS Foundation Trust, Salford, UK;* [3]*School of Clinical and Laboratory Sciences, University of Manchester, Manchester, UK*

Background: Osteoarthritis is a leading cause of hip arthritis and disability in the industrialized world. Symptoms of pain, stiffness and functional limitation are typically associated with progressive degeneration of joint cartilage and characteristic radiographic changes in affected individuals. IL-1β is a powerful catabolic cytokine naturally expressed by chondrocytes in osteoarthritic joints. The purpose of our study was to ascertain whether chondrocyte IL-1β expression is a clinically significant marker of the rate of progression and severity of hip osteoarthritis.
Methods: Patients about to undergo elective hip arthroplasty for primary osteoarthritis were assessed for symptom duration, clinical severity with the Western Ontario and McMaster Osteoarthritis (WOMAC) Index, radiographic stage and pattern using routine preoperative standard antero-posterior pelvic roentgenograms as well as potential clinical and demographic confounders. A total of 23 femoral head specimens were collected from participants and 10 of these were dissected and cartilage stained for IL-1β expressed as the percentage of positively staining chondrocytes in both superficial and deeper zones. Our preliminary results were analysed by simple and multiple linear regression. The study had Research Ethics Committee approval and was conducted in accordance with the principles of the Declaration of Helsinki.
Results: Our results show that chondrocyte IL-1β expression in superficial zone cartilage is a statistically significant and independent marker of symptom duration (univariate $P = 0.006$, multivariate $P = 0.018$), a surrogate measure known to correlate with radiographic progression. No association was found with symptom severity scores (univariate $P = 0.853$). In multivariate analysis, a statistically significant association with pre-operative femoro-acetabular minimum joint space width was detected (multivariate $P = 0.022$).
Conclusions: Our findings suggest that chondrocyte IL-1β expression is a clinically significant marker of progression and radiographic severity in osteoarthritis of the hip joint. Despite the study's methodological limitations and concern that our findings may only be applicable to a sub-group of individuals with a more aggressive disease phenotype, they would support the continued development toward effective anti-IL-1β blockade in osteoarthritis and suggest that the use of disease progression may be a more appropriate clinical trial outcome than short-term measures of symptom relief.
Disclosure statement: All authors have declared no conflicts of interest.

OP75. MILD RADIOGRAPHIC FEATURES PREDICT STRUCTURAL PROGRESSION OF KNEE OSTEOARTHRITIS

Kirsten M. White[1], Anushka Soni[1], Amit Kiran[1], Lyndsey M. Goulston[2], Deborah J. Hart[3], Tim D. Spector[3], Kassim Javaid[1] and Nigel K. Arden[1,2]
[1]*NIHR Musculoskeletal BRU, University of Oxford, Oxford, UK;* [2]*MRC Epidemiology Resource Centre, University of Southampton, Southampton, UK;* [3]*Department of Twin Research and Genetic Epidemiology, King's College London, London, UK*

Background: The natural history of radiographic knee OA in a normal population is not fully understood. Previous studies had restrictive inclusion criteria, limited time-frames or radiographs only at the start and end of the study. This research addresses those issues by analysing trends and patterns of Kellgren and Lawrence grades (KandL) at four time-points over a span of 15 years.
Methods: 1003 women, aged 43–65 years, from a single general practice in Chingford, UK were recruited in 1988–89 for a longitudinal population-based cohort studying OA and osteoporosis. Knee radiographs were performed at baseline, year 5, 10 and 15 and read blinded for KandL grades using standard methods. 527 women (1052 knees), present for all radiographs and complete baseline data, were included in the analysis. Methods included assessing the frequency distribution of empirical permutations and cross-tabulations of KandL grades between years.
Results: The mean age of subjects included in the analysis was 53.4 years (S.D. 5.8) and the mean BMI was 25.3(S.D. 3.7) with 46.8% classified by WHO guidelines as overweight (>25.0). The prevalence of radiographic knee OA (defined as KandL grade 2+) at baseline was 8.2% (37.5% by year 15). The cumulative incidence rate out of knees at risk at baseline ($n = 966$) was an average of 2.2% (95% CI 2.0–2.5) per year over 15 years. These results were not significantly changed

when tested in the complete cohort. The Table demonstrates the movement of KandL grades, by cross-tabulation, between baseline and year 15 with percentages calculated by row. The majority of knees remained stable over the 15 years; 60.6% of grade 0's, 50.0% of grade 2s and 65.4% of grade 3s. The glaring exception was grade 1, where only 5.3% remained stable. 75.4% of baseline grade 1s progressed (an increase by at least one grade from any baseline grade) to a higher grade. A rate-ratio showed that a KandL grade of 1 at baseline was 3.2 (95% CI 2.3–4.4) times more likely to progress to a higher grade by year 15 than a baseline grade of 0. Empirical permutations of KandL grades at four time points were assessed for each knee (0011, 0123, etc.). Eighty different patterns were identified, with 77% of all knees accounted for by only 10 different KandL patterns. Of the top 10 most common, only one did not have a KandL grade 0 at baseline. This pattern had grade 2s at all four time-points.

Conclusions: This study demonstrates that the majority of subjects in this population cohort remained stable over a 15 year period. The group with the most rapid progression were those with a grade 1 K and L at baseline.

Disclosure statement: All authors have declared no conflicts of interest.

K and L Movement: baseline and year 15 ($n=1052$)

K and L Grades	N	0	1	2	3	4 and 5
0	909	551(60.6%)	91(10.0%)	142(15.6%)	111(12.2%)	14(1.5%)
1	57	11(19.3%)	3(5.3%)	22(38.6%)	18(31.6%)	3(5.3%)
2	60	0(0%)	1(1.7%)	30(50.0%)	25(41.7%)	4(6.7%)
3	26	0(0%)	1(3.8%)	4(15.4%)	17(65.4%)	4(15.4%)

1. Baseline KandL grades in the left hand column and year 15 in the top row 2. Grade 5s are TKRs and Uni Knee replacements

OP76. INCIDENCE AND PROGRESSION OF RADIOGRAPHIC KNEE OSTEOARTHRITIS IN A SYMPTOMATIC POPULATION

Rachel Duncan, George Peat, Elaine Thomas, Elaine Hay and Peter Croft
Arthritis Research Campaign National Primary Care Centre, Keele University, Keele, UK

Background: Establishing the incidence and progression of radiographic osteoarthritis (ROA) of the knee is important in determining cause and impact. Many previous population studies have failed to include changes in the patellofemoral (PF) compartment in their estimates. Our objectives were to describe the incidence and progression of ROA in both the patellofemoral and tibiofemoral (TF) compartments and explore the interaction between them.

Methods: Population-based study of adults aged ≥ 50 years with knee pain who had knee X-rays in 2002-2003 and again in 2005–2006 (mean interval 36.7 months): the CAS(K) study. Three radiographic views of the knee joint were obtained: a posteroanterior semiflexed/metatarsophalangeal view, a skyline view and a lateral view. The PF and TF compartments were scored separately for radiographic severity.

An incident case of ROA was defined as the appearance of mild or moderate/severe ROA at 3-year follow up in a knee that had no definite ROA at baseline. Progression of ROA (in a knee with definite ROA in at least one compartment at baseline) was defined by either: (1) an increase in radiographic severity from mild to moderate/severe ROA within that compartment and/or (2) the development of ROA in the other compartment. The influence of baseline ROA on progressive disease was examined using logistic models and results presented as odds ratios (95% CIs).

Results: Of 777 participants with full radiographic data at baseline, 481 (62%) attended 3-year follow-up clinic: 414 provided full X-ray data and were at risk of either incident ($n=144$) or progressive ($n=270$) disease. 47 out of 144 participants (32.6%) developed incident disease and 94 out of 270 (34.8%) had radiographic progression at 3 years. Approximately three-quarters of incident disease involved the PF compartment ($n=37$) and about half involved the TF compartment ($n=23$). Isolated PF ROA was the commonest pattern of incident disease (51%). At 3 years, progression was a little more likely to have occurred in the TF compartment than in the PF compartment (20% *vs* 17%). Severity of baseline radiographic TF ROA significantly increased the likelihood of developing progressive PF ROA [odds ratio for mild TF ROA: 3.6 (95% CI 1.2, 10.4); moderate/severe TF ROA: 5.1 (1.9, 13.8), while baseline PF ROA severity was not significantly associated with TF ROA progression.

Conclusions: We report a high incidence and progression of ROA in the knee at 3 years. The incidence of PF ROA is particularly high and we hypothesize that ROA of the knee may commence in the PF compartment in many individuals. In studies examining the incidence and natural history of knee ROA, radiological examination of the PF compartment in fundamental.

Disclosure statement: All authors have declared no conflicts of interest.

OP77. INTER-RATER RELIABILITY AND VALIDITY OF ULTRASOUND IMAGING FOR FEATURES OF OSTEOARTHRITIS IN THE KNEE IN THE COMMUNITY

Ajay Abraham[1,3], Iain Goff[1,3], Mark S. Pearce[3], Roger M. Francis[2,3] and Fraser Birrell[1,3]
[1]*Northumbria Healthcare NHS Trust, Newcastle, UK;* [2]*The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK;* [3]*Newcastle University, Newcastle, UK*

Background: Radiographs are the main outcome measure in epidemiological studies of osteoarthritis (OA). However, their use has significant limitations: exposure to ionizing radiation and inability to view soft tissue structures. There is a need for a more valid and reliable outcome measure. Ultrasound (US) imaging has unique advantages: involves no radiation, easy to use, visualizes soft tissue structures like cartilage and synovium and measures inflammation.

Objectives: To measure the inter-rater reliability and validity of US imaging in the detection of knee OA.

Methods: 18 participants from a community cohort (Northumberland over 85), aged 88 years or older, had both knees scanned by two trained musculoskeletal sonographers, within 6 weeks. A Mylab 5 US machine with a 10-18 MHz linear transducer was used. The US protocols were derived from EULAR recommendations and reported on the size of effusions and presence of osteophytes, with the knee in 20 degrees of flexion. Femoral condyle cartilage thickness was measured medially, laterally and in the notch, with the knee in full flexion. The inter-rater reliability was determined by estimating Kappa (κ) and Intraclass correlation coefficients (ICC), as appropriate.

The participants had weight bearing AP radiographs of both knees. The films were scored by a single trained observer using Kellgren and Lawrence criteria for minimal joint space and osteophytes. A measure of construct validity was determined by estimating κ between cartilage thickness on US and minimal joint space on radiographs (comparing κ between each in quartiles) and for osteophytes.

Results: Reliability—κ for femoral osteophyte presence was 0.77 (right), 0.65 (left) and 0.88 for tibial osteophytes. ICCs for effusion size were 0.70 (right) and 0.85 (left). Moderate to substantial agreement was found for cartilage thickness: ICCs from 0.42 to 0.68.

Validity—for osteophytes, κ was moderate to excellent at 0.45 (right) and 0.86 (left). The κ agreement between quartiles of cartilage thickness and minimal joint space was poor medially [κ -0.06 (right) and 0.16 (left)] and fair laterally[κ 0.27 (right) and 0.28 (left)].

Conclusions: This study has established reliability for ultrasound assessment of OA and measured the process of acquizition and reading of the images between US observers, including all main potential sources of variation. Intra-rater reliability is likely to be as good, if not better. Joint space width is influenced by femoral/tibial cartilage plus meniscal position. This could explain the limited agreement between minimal joint space on radiographs and femoral cartilage thickness on US. This is the first study to look at inter-rater reliability and validity of US features of OA in the community. Future studies should look at the construct validity of ultrasound by comparing against MRI and symptoms of OA. Also, the predictive validity of ultrasound should be measured in longitudinal community cohort studies.

Disclosure statement: F.B., Roche - Unrestricted research grant, Servier - Unrestricted research grant, NIHR Biomedical Research Centre for Ageing, Newcastle - Support. All other authors have declared no conflicts of interest.

OP78. SHORT-TERM EFFECTS OF TRANSCUTANEOUS ELECTRICAL NERVE STIMULATION (TENS) AND EXERCISE ON KNEE OSTEOARTHRITIS (OA)

Shea Palmer[1], Melissa Domaille[2], Fiona Cramp[1] and Mark I. Johnson[3]
[1]*School of Health and Social Care, University of the West of England, Bristol, UK;* [2]*Rheumatology Centre, University Hospitals Bristol NHS Foundation Trust, Bristol, UK;* [3]*Faculty of Health, Leeds Metropolitan University, Leeds, UK*

Background: Systematic reviews suggest that TENS (1) and exercise (2) reduce OA knee pain. TENS is advocated as an adjunct to other treatments (3), although combined effects with exercise have yet to be clearly demonstrated. This study aimed to determine whether a

1.1.2.2 OARSI 2011 poster abstract– Bézier curves

tibial subregions. WORMS score ranges from 0–6 where 6 represents cartilage loss to bone in 75% of region. Analysis was performed for the compartment showing bone-on-bone appearance ("index") on radiograph and also for the other TF compartment of the same knee. Hoffa-synovitis and effusion-synovitis were assessed for the whole knee. Changes in scores at follow-up were noted for each feature. For cartilage and BML, within-grade changes were also recorded.

Results: 67 knees from 63 subjects were included (51% women, 84% White, mean age 65.1±8.6 years, mean BMI 30.2±5.2 kg/m2). At baseline, in the index TF compartment, all knees showed severe cartilage loss (max WORMS score from 5 subregions was 5 in 1 knee and 6 in 66 knees), 54 knees (80%) showed moderate to large BMLs (max WORMS score 2 or 3), and 62 knees (94%) had severe meniscal lesions (i.e. displaced tear or maceration). In the other TF compartment, 12 knees (18%) had severe cartilage loss, but 47 (71%) had no BML and 57 (97%) had no meniscal damage. 39 knees (58%) had moderate to severe effusion-synovitis, 56 knees (86%) had mild or moderate Hoffa-synovitis. Longitudinally, 22 index compartments (35%) showed an increase in the sum of cartilage scores from all subregions, and 2 (3%) showed increase in the maximum cartilage score. In the other TF compartment, 22% showed an increase in the sum score for cartilage damage, while 15% showed increase in maximum score. For BMLs in the index TF compartment, 19 knees (31%) showed an increase in maximum score and 11 (18%) showed a decrease. Fluctuation of BMLs was also seen in the other TF compartment, but to a lesser extent. Meniscal status mostly remained the same in the index (98%) and other TF (95%) compartments. Effusion-synovitis worsened in 15 knees (27%) and improved in 2 knees (4%). Hoffa-synovitis worsened in 6 knees (11%) and improved in 2 knees (4%).

Conclusion: In KL4 knees, MRI detected progression of cartilage loss, effusion-synovitis, and Hoffa-synovitis, and fluctuation in size of BMLs. Meniscal damage remained stable. Our findings support the idea that disease progression still occurs in KL4 knees. KL4 knees can be a potential target for assessing therapeutic interventions and should not necessarily be excluded from studies evaluating therapeutic response.

390

GEOMETRY OF THE ARTICULAR CARTILAGE OF THE TIBIAL PLATEAU IS RELATED TO ANTERIOR CRUCIATE LIGAMENT INJURY RISK

B. Beynnon[1], H.C. Smith[1], A. Scheinman[1], D. Sturnick[1], P. Vacek[1], L. Holtermann[1], M. Gardner-Morse[1], T. Tourville[1], J. Slauterbeck[1], I. Bernstein[1], S. Shultz[2], J. Hashemi[3], R. Johnson[1]. [1]*The Univ. of Vermont, Burlington, VT, USA;* [2]*Univ. of North Carolina at Greensboro, Greensboro, NC, USA;* [3]*Texas Tech Univ., Lubbock, TX, USA*

Purpose: Injuries to the Anterior Cruciate Ligament (ACL) of the knee are common and can lead to post traumatic osteoarthritis (PTOA). Recent studies have shown that tibial plateau geometry may play an important role in controlling transmission of intersegmental forces across the the knee during weight-bearing activity. Factors that have been shown to influence the risk of ACL injury explored thus far have focused on subchondral bone geometry, and include the depth of the concave surface of the medial tibial plateau and the posterior-inferior directed slopes of the medial and lateral plateaus of the tibia. The goal of our study was to build on prior studies of bony geometry by studying the influence of the articular cartilage geometry of the tibial plateau on the risk of suffering ACL injury.

Methods: The study used a matched case-control design. Knee MRI images of 20 ACL injured cases and 20 uninjured controls matched by age and sports team were obtained in order to control for exposure. The DICOM images were uploaded into a viewer program (Osirix, Pixmeo, version 3.6.1., open-source). Using the Cintiq digitizing tablet (Wacom, 2010), the cartilaginous articular surface of the medial tibial plateau was segmented in a standardized and reproducible coordinate system aligned with the tibia. The maximum depth of concavity in the tibial plateau was defined as the point with the greatest depth of concavity within the central 20% of the total surface area. For each of the 40 knees that were segmented, the data points defining the sagittal profiles that contained this value were subsequently used in the statistical analysis. A hierarchical mixed model was used to fit fourth order polynomials to the sagittal profile data. Interaction terms were included in the model as fixed effects to permit the regression coefficients to vary between cases and controls. Variation in the coefficients between individuals and deviations between the estimated and observed data points within individuals were modeled as random effects. Model parameters were estimated by maximum likelihood and the difference in the fit of models with and without the interaction terms was assessed by the likelihood ratio test.

Results: Polynomial fit lines for medial tibial geometry (Figure 1) were significantly different between cases and controls ($p < 0.001$). Polynomial fit equations are as follows:

Case: $-0.5682 + 0.041495x + 0.007804x^2 - 0.000045x^3 - 0.00001143x^4$

Control: $-0.6073 + 0.03446x + 0.007512x^2 - 0.00013x^3 - 0.00001x^4$

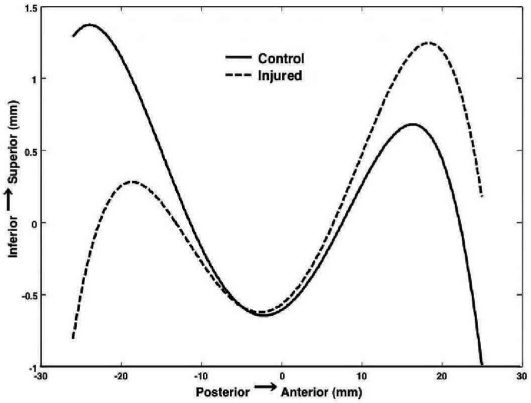


Fig. 1. Polynomial fits of sagittal articular cartilage profiles.

Conclusions: There was a significant difference in tibial articular surface morphometry measured using the fourth order polynomial models between ACL injured case subjects and uninjured matched controls. Uninjured controls appeared to have a tibial articular cartilage profile that conformed to the femoral condyle, while this was not the case for the injured subjects. The increased conformity in uninjured controls was characterized by a substantial increase in the depth of concavity that may act to control the joint biomechanics, particularly during impulsive loading conditions when the knee transitions from non-weightbearing to weightbearing conditions such as during an ACL injury. This divergence of shape of the articular cartilage may further our understanding of how the forces transmitted across the knee influence risk of ACL injury, how individual knee joints respond to loading, and subsequent risk of development of PTOA.

391

BEZIER CURVES FOR MEASURING JOINT SPACE ON KNEE RADIOGRAPHS – REPRODUCIBILITY AND VALIDITY

K.M. Leyland[1], D. Hunter[1], A. Judge[1], N. Bottomley[1], D. Hart[2], R. Gill[1], M.K. Javaid[1], N.K. Arden[1,3]. [1]*NIHR Musculoskeletal BRU, Univ. of Oxford, Oxford, United Kingdom;* [2]*Dept. of Twin Res. and Genetic Epidemiology, King's Coll. London, London, United Kingdom;* [3]*MRC Epidemiology Resource Ctr., Univ. of Southampton, Southampton, United Kingdom*

Purpose: Radiographic joint space width (JSW) is a key feature for evaluating severity and progression of knee osteoarthritis. Computerized methods for evaluating JSW are ideal for large studies, but fully-automated programs which detect bone edge are not as accurate on older digitized plain-film radiographs while manual methods are time consuming. We explored using manually placed Bézier curves to automatically find several measures of JSW, and compared minimum JSW with a manual digital measure.

Methods: 25 digitized plain-film and 25 digital knee radiographs from the Chingford cohort were selected with a range of disease severity (K&L 0–4). Minimum JSW (minJSW) as measured by digital calipers placed by the user was the 'gold-standard' measure. Mean and minimum JSW measurements were calculated based on the Bézier curve, with a user selected point to constrain the area of analysis (outer slopes of the tibial spines) as well as an automated constraint point selected by the program based on curvature. Two observers (KL and DH) completed two training sessions before independently reading the radiographs in a random order, with KL re-reading all radiographs after several days. Intra and inter-observer reproducibility was tested using intraclass correlations (ICC)

with 95% confidence intervals (CI). Validity was evaluated by comparing Bézier minJSWs against the manual caliper minJSW and assessing them using R2s from linear regression in addition to Spearman's correlation coefficient (r). Data was collected using a proprietary software program developed by University of Oxford, minimums and means were calculated from Bézier data using Matlab 7.10, and statistical tests were completed in Stata version 11.0.

Results: The means (in millimeters) and standard deviations for the minJSW in the medial compartment were 3.9 (SD 1.5), 3.8 (SD 1.4) and 3.9 (SD 1.4) for the manual calipers, user Bézier and auto Bézier measurements, respectively. In the lateral compartment minJSWs were 5.1 (SD 1.5), 5.1 (SD 1.4) and 5.2 (SD 1.4) for the same measurements. Means (in millimeters) and standard deviations for mean JSW in the medial compartment were 5.0 (SD 1.1) and 5.0 (1.1) for the user Bézier and auto Bézier measurements, respectively. In the lateral compartment means for mean JSW were 5.7 (SD 1.5) and 5.8 (SD 1.5) for the same measures. The table shows the intra- and inter-observer reproducibility, with all joint space measures showing high levels of reproducibility. When the validity of the user constrained Bézier curve was tested against the manual measure of minJSW, linear regression showed an R2 of 0.92 in both compartments and an r of 0.89 medially and 0.93 laterally. When the automated Bézier curve was compared with the manual measure, R2s of 0.92 and 0.93 and r of 0.92 and 0.93 were found in the medial and lateral compartments, respectively.

Conclusions: The measurements from this proposed method are extremely reproducible, highly correlated with the gold-standard, and simple to use. The curves provide a wealth of additional information about joint morphology by accurately reflecting the curvature of the bone edges, are easily adaptable to future measurements (e.g. area and maximum), and show great promise in measuring JSW in cohort studies.

Table: Intra- and inter-observer reproducibility

JSW measurements	ICC (95% CI)	
	Intra-observer	Inter-observer
mJSW (medial) – manual caliper	0.97 (0.96, 0.99)	0.94 (0.90, 0.97)
mJSW (lateral) – manual caliper	0.97 (0.95, 0.98)	0.96 (0.93, 0.98)
mJSW (medial) – user Bezier	0.96 (0.94, 0.98)	0.93 (0.90, 0.97)
mJSW (lateral) – user Bezier	0.96 (0.94, 0.98)	0.91 (0.87, 0.96)
mJSW (medial) – auto Bezier	0.97 (0.95, 0.99)	0.94 (0.91, 0.97)
mJSW (lateral) – auto Bezier	0.96 (0.93, 0.98)	0.94 (0.91, 0.97)
Mean JSW (medial) – user Bezier	0.93 (0.90, 0.97)	0.89 (0.83, 0.95)
Mean JSW (lateral) – user Bezier	0.97 (0.96, 0.99)	0.93 (0.91, 0.97)
Mean JSW (medial) – auto Bezier	0.96 (0.93, 0.98)	0.89 (0.83, 0.95)
Mean JSW (lateral) – auto Bezier	0.97 (0.96, 0.99)	0.94 (0.91, 0.97)

392
INTERMUSCULAR FAT VOLUME IN THE THIGH RELATES TO KNEE STRENGTH AND PHYSICAL PERFORMANCE AMONG WOMEN AT RISK FOR OR WITH KNEE OSTEOARTHRITIS

M.R. Maly, N.J. MacIntyre, K.A. Beattie. *McMaster Univ., Hamilton, ON, Canada*

Purpose: While measures of thigh muscle and fat volumes from magnetic resonance images show robust psychometric properties, the role of these measures in understanding knee function in the presence of osteoarthritis remains unclear. The purpose of this study was to determine whether quadriceps muscle (QM) and intermuscular thigh fat (IMF) volumes explain variance in knee strength and physical performance of women in the incident and progression cohorts of the Osteoarthritis Initiative (OAI).

Methods: From women over age 50 enrolled in the OAI, baseline data from 93 right knees in the incident or progression cohorts were randomly selected. Knee strength was the peak isometric extensor force of the right knee produced against a fixed force transducer, expressed per kilogram of body mass to control for body size (N/kg). Physical performance was the time (s) required to stand from a seated position 5 times without the use of hands. Longer time corresponded with reduced performance. QM and IMF volumes were determined from T1-weighted axial magnetic resonance imaging scans (cm^3) consisting of 15 contiguous slices (5 mm slice thickness) of the right mid-thigh. A single reader segmented each anonymized scan on a slice-by-slice basis using SliceOmatic 4.3 (Tomovision, Magog, QC, Canada) to generate QM and IMF volumes. Because age and local knee factors affect knee strength and physical performance, these were included in the analyses. Local knee factors included (i) OA status, or assignment to the incident or progression cohort (ii) right knee frontal plane alignment, measured with a goniometer in standing, and (iii) pain intensity reported during the strength test. In addition to descriptive statistics, 2 hierarchical multiple regressions were used to test whether QM and IMF would explain variance in each of knee strength and physical performance. For both regressions, block 1 included age; block 2 included local knee factors; and block 3 included QM, IMF (probability of F to enter: p less than 0.05; to remove: p greater than 0.10).

Results: The mean (SD) age of the 93 participants was 62.7 (7.4) years and knee strength was 4.12 (1.20) N/kg. The 49 incident knees [4.42 (1.3) N/kg] were stronger than 44 progression knees [3.78 (1.01) N/kg, p = 0.009]. No differences were found between incident and progression in physical performance, QM or IMF. Physical performance of the 93 participants was 12.0 (3.6) s, QM volume was 252.01 (49.21) cm^3 and IMF volume was 103.42 (32.61) cm^3. A model explaining 20.1% of variance in knee strength included alignment, OA status, pain intensity and IMF (Table 1). Also, 7.3% of variance in physical performance was explained by age and IMF. QM volume was unrelated to knee strength and physical performance.

Conclusions: After controlling for the effect of age and local knee factors, IMF explained a small amount of variance in knee strength and physical performance among women at risk for, and with knee osteoarthritis. Larger volumes of IMF corresponded with poorer knee function. Interestingly, QM was unrelated to knee function, perhaps because muscle volume does not reflect the magnitude of intramuscular fatty infiltration, or the muscle mechanics (e.g., cross-sectional area, architecture, muscle activation) known to affect force output from muscle.

Table 1. Hierarchical multiple regressions of knee strength and physical performance

Model	R	Adjusted R^2	Standardized beta coefficient	p
Knee strength				
1. Alignment			0.205	0.039
2. OA status	0.486	0.201	−0.186	0.058
3. Pain intensity			−0.190	0.055
4. IMF			−0.221	0.025
Physical performance				
1. Age	0.306	0.073	0.219	0.033
2. IMF			0.220	0.033

393
SPATIO-TEMPORAL ANALYSIS OF THE SIGNIFICANT CHANGES IN CARTILAGE MORPHOLOGY: DATA FROM THE OSTEOARTHRITIS INITIATIVE

J.G. Tamez-Pena[1,2], P. Gonzalez[2], E. Schreyer[2], S. Tottermann[2]. [1]*Tec de Monterrey, Monterrey, Nuevo Leon, Mexico;* [2]*Qmetrics Technologies, Rochester, NY, USA*

Purpose: The purpose of this analysis is to study the spatio-temporal presentation of the changes in cartilage morphology using significant change detection from standardized thickness maps and to report the rate of progression of significant changes in cartilage thickness.

Methods: OAI MRI data sets releases: 0.3C.2 (Baseline), 1.C.2 (12 month) and 3.C.1 (24 month) were used in this study. The most diseased knee per subject and with the three complete MRI observations were selected and analyzed (138 subjects, 3 Timepoints). The baseline-image analysis data extracted from the OAI site was used to stratify knees into three groups: Non-denuded: Knees without full cartilage thickness defects (n = 52). Low-Denuded: Subjects with small full thickness defects (n = 43). Top-Denuded: Subjects with large full thickness defects (n = 43). All the 414 DESS MRI images were independently segmented using a fully automated multi-atlas segmentation algorithm that created atlas-referenced thickness maps of the tibia and the femur cartilage. The segmentation quality was visually inspected. The Pilot OAI scan-rescan data was used to estimate the paired-measurement noise of the method in creating atlas-referenced thickness maps. The 12-month and 24-month cartilage thickness maps were compared to the baseline observation and standardized thickness change maps were created by subtracting the two

1.1.2.3 OARSI 2011 poster abstract – Reproducibility

407

A COMPARISON OF FOUR RADIOGRAPHIC SCORING METHODS FOR KNEE OSTEOARTHRITIS – SHORT AND MEDIUM TERM REPRODUCIBILITY

K. Leyland[1], N. Bottomley[1], A. Judge[1], T. Spector[2], D. Hart[2], M.K. Javaid[1,3], N.K. Arden[1,3]. *[1]NIHR Musculoskeletal BRU, Univ. of Oxford, Oxford, United Kingdom; [2]Dept. of Twin Res. and Genetic Epidemiology, King's Coll. London, London, United Kingdom; [3]MRC Epidemiology Resource Ctr., Univ. of Southampton, Southampton, United Kingdom*

Purpose: Severity of radiographic knee osteoarthritis is currently evaluated using a variety of methods. The comparable short and medium term reproducibility of many of these methods has not been evaluated within a single cohort. This study aims to provide a comparison of reproducibility between four commonly used radiographic scoring methods.

Methods: We used 25 digitized plain-film and 25 digital AP fully-extended weight-bearing knee radiographs from the Chingford cohort with a range of K&L grades, subject age, and BMI. Four scoring methods were used to grade all radiographs: Kellgren and Lawrence (K&L); OARSI Atlas; Line Drawing Atlas; and the Chingford Study Atlas. Radiographic features (osteophytes, joint space, etc.) were evaluated individually as specified by each method in each tibio-femoral compartment. For training, a single grader (KL) read 10 radiographs twice, with discrepancies and grading issues addressed by a panel of more experienced graders. All 50 blinded radiographs were then read three times after an interval of one and 10 days. Linear weighted kappas with standard errors were calculated for all scoring methods in order to assess short-term and medium-term reproducibility.

Results: The tables gives weighted kappas for the short and medium-term reproducibility for all four scoring methods. Most features had good short-term reproducibility, with the majority having kappas above 0.7. Most osteophyte and joint space narrowing (JSN) grades had very high kappas except for medial femoral osteophytes in both the OARSI and Line Drawing atlases. The Chingford atlas did not have a comparable category as osteophytes are grouped by compartment only. The features with the lowest kappas included those with binary categories such as sclerosis and tibial spiking. Readings after 10 days showed a similar level of reproducibility with kappas ranging from 0.49 to 0.89.

Table 1. Short and medium term reproducibility – K&L and OARSI

Method	Individual features	Reproducibility (Kappa and standard error)	
		Short term	Medium term
K&L	Whole Knee	0.84 (0.09)	0.79 (0.09)
	Medial compartment	0.79 (0.09)	0.72 (0.09)
	Lateral compartment	0.71 (0.10)	0.61 (0.09)
OARSI	Medial femoral osteophyte	0.56 (0.10)	0.66 (0.10)
	Lateral femoral osteophyte	0.84 (0.11)	0.89 (0.11)
	Medial tibial osteophyte	0.77 (0.11)	0.65 (0.11)
	Lateral tibial osteophyte	0.74 (0.11)	0.72 (0.10)
	Medial JSN	0.79 (0.11)	0.63 (0.10)
	Lateral JSN	0.95 (0.11)	0.86 (0.11)
	Medial sclerosis	0.50 (0.13)	0.53 (0.13)
	Lateral sclerosis	0.69 (0.14)	0.62 (0.14)

Table 2. Short and medium term reproducibility – Line Drawing and Chingford

Method	Individual features	Reproducibility (Kappa and standard error)	
		Short term	Medium term
Line Drawing	Medial femoral osteophyte	0.55 (0.10)	0.71 (0.11)
	Lateral femoral osteophyte	0.84 (0.11)	0.79 (0.11)
	Medial tibial osteophyte	0.79 (0.10)	0.70 (0.10)
	Lateral tibial osteophyte	0.78 (0.10)	0.75 (0.10)
	Medial JSN	0.73 (0.08)	0.76 (0.08)
	Lateral JSN	0.67 (0.08)	0.76 (0.08)
Chingford	Medial tibiofemoral osteophyte	0.79 (0.10)	0.66 (0.11)
	Lateral tibiofemoral osteophyte	0.81 (0.10)	0.72 (0.10)
	Medial JSN	0.88 (0.11)	0.78 (0.11)
	Lateral JSN	0.95 (0.11)	0.86 (0.11)
	Tibial spine spiking	0.43 (0.13)	0.50 (0.13)
	Medial sclerosis	0.50 (0.13)	0.53 (0.13)
	Lateral sclerosis	0.56 (0.14)	0.49 (0.14)

Conclusions: All four scoring methods generally show high levels of both short-term and medium-term reproducibility. Scoring methods that group more than one compartment (such as K&L) or more than one bone (such as the Chingford atlas) into a single grade were expected to show higher levels of reproducibility, however all methods in this study show similar levels of reproducibility. These results demonstrate that using methods which break down scores into individual features and compartments (OARSI, Chingford and Line Drawing) and have larger scales (Line Drawing) does not necessarily reduce the reliability of the measurements.

408

MULTI-ECHO PATTERN BASED CARTILAGE SEGMENTATION FOR OA EARLY DIAGNOSIS

X. Cui[1], S. Li[1], N. June[1], H. Kim[1], K-S. Kwack[2], B-H. Min[3]. *[1]Sch. of Information and Communication Engineering, INHA Univ., INCHEON, Korea, Republic of; [2]Dept. of Radiology, Ajou Univ. Sch. of Med., SUWON, Korea, Republic of; [3]Dept. of Orthopaedic Surgery, Ajou Univ. Sch. of Med., SUWON, Korea, Republic of*

Purpose: It is very important to visualize the articular cartilage in diagnosing OA (osteoarthritis). In general, the cartilage evaluation is performed by signal intensity, signal contrast and T2 values based on a single echo MR images or the multi-echo MR images. However, the cartilage segmentation results only based on T2 value may not be suitable for diagnosis of OA (Fig.1 (a)). This paper proposes a new cartilage segmentation method using spectral feature matching in the multi-echo MR Images for early diagnosis of OA. The VTK (Visualization ToolKit) based 3D visualization tool developed for displaying the detected cartilage volume.

Methods: In our study using the multi-echo MR images, the different tissues show different multi-echo patterns called as spectral features. The spectral features include more information than the single-echo MR image because they accumulate information over echo time. This paper develops a new spectral feature matching algorithm, Normalized Spectral Angle Mapper (NSAM) for segmenting the tissues in the multi-echo MR images. In the hyperspectral imaging field, the Euclidean distance (ED) and Spectral angle mapper (SAM) method are proposed. However, the ED method just considers the magnitude of vector and the SAM consider the angle between two vectors only. In this study, the proposed NSAM method considers the similarity of the angle and the magnitude simultaneously. The bone areas segmented by active contour model (ACM) based semi-automatically method, and then create the 2D bone cartilage interface (BCI) using 3D gradient information. Finally, combine the NSAM and BCI to get the cartilage information.

Results: Our new experimental method demonstrates that the contour of articular cartilage and the boundary between articular cartilage and meniscus more effectively (Fig. 1). 3D volume images (Fig. 2) are reconstructed 3D images from the multi-echo pattern based segmentation algorithm. Fig. 2 shows the VTK based 3D visualization tool for displaying the detected cartilage information. The 3D visualization tool can shows both 3D morphologic information and T2 values of articular cartilage at the same time, which could be used for the early diagnosis of OA.

Conclusions: The proposed cartilage segmentation method is based on the NSAM and 2D BCI. Therefore, the cartilage can be segmented efficiently because the multi-echo patterns include the much information and the 2D BCI give additional location information. Nevertheless, the segmented cartilage still contains the misclassification because the noises like the arrow in the Fig. 2. Future studies will consider the noise removing method for accuracy segmentation method.

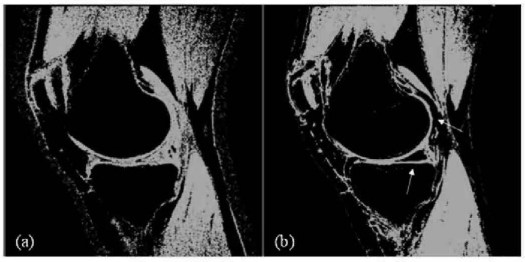


Fig. 1. Comparison between (a) T2 value based and (b) multi-echo based segmentation.

1.1.2.4 British Society of Rheumatology 2013 - poster abstract (accepted)

TITLE: JOINT SPACE NARROWING OVER FIVE-YEARS PREDICTS FUTURE NEED FOR KNEE REPLACEMENTS UP TO FIFTEEN YEARS LATER

AUTHORS (FIRST NAME, LAST NAME): Kirsten M. Leyland1, Amit Kiran1, Andrew Judge1,3, Dave Hunter1, Debbie Hart2, M.K.Javaid1,3, N.K.Arden1,3, Cyrus Cooper1,3

INSTITUTIONS: 1. NIHR Musculoskeletal BRU, University of Oxford, Oxford, United Kingdom.

2. Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom.

3. MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom.

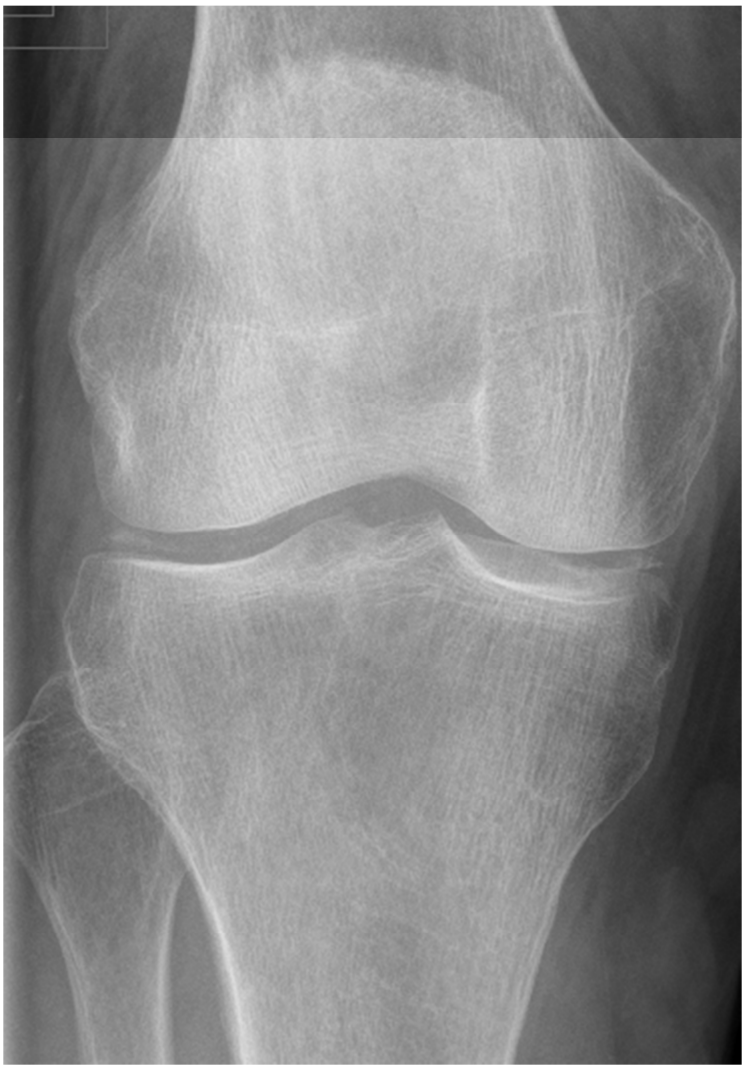
BACKGROUND: The primary outcome measure recommended by regulatory agencies for structure modifying drugs in osteoarthritis (OA) is a reduction in joint space narrowing (JSN) over 3 years. While JSN is an established feature of knee OA on radiographs, its relationship with knee replacement is not well understood. This study evaluated the relationship between 5-year change of joint space and the future risk of a TKR over 15year follow-up.

METHODS: A nested case control study was performed using the Chingford Women's Cohort, a UK population-based study with a twenty-year follow-up. At the knee level, 40 TKRs (cases) were observed between year 5 and year 20 visits (of which 10 were bilateral). Cases and controls (1:10 ratio) were matched by baseline age (within 3 years), side of TKR and time in the study. Subjects were required to have baseline and year 5 radiographs, and must not have had a TKR over this period. Digitised radiographs were read for minimum quantitative joint space width (minJSW) (using the KneeMorf software program) as well as categorical joint space narrowing (catJSN) using a standard atlas (grade 0 [normal] through grade 3 [bone-on-bone]). Reproducibility for minJSW was calculated using intra-class correlations (ICC) on 50 x-rays read by two readers. Reproducibility has previously been reported to be good for catJSN and Kellgren and Lawrence (K/L) grade (ICC >0.7). Five-year change was calculated as an increase of one or more grades for catJSN or K/L and as any change of joint space for JSW (both widening and narrowing). Multivariable Conditional logistic regression was used to analyse the risk of TKR in relation to change in JSN and K/L grade.

RESULTS: The ICC's for inter-observer reproducibility for medial minJSW were 0.96 (95% CI 0.93, 0.98) and 0.81 (95% CI 0.69, 0.89) for minJSW in the lateral compartment. Cases (n=40 knees) had higher mean BMI (27.0 versus 25.3, p=0.05) and more knee pain at year 5 (54.3% versus 16.7%, p<0.001) compared to controls (n=383). The mean minJSW at baseline was 4.2 mm (1.2 SD) for cases and 4.4 mm (0.8 SD) for controls. Over 5 years, 20% of cases and 7% of controls showed medial catJSN narrowing, while 52% of cases and 53% of controls showed any medial minJSW narrowing. 5-year change in catJSN was significantly associated with TKR in both the medial (OR 3.32 [95% CI (1.34, 8.19)]) and lateral compartment (OR 4.29 [95% CI (1.01, 18.28)]); a commensurate increase in risk was observed with change in K/L grade (OR 2.98 [95% CI (1.37, 6.49)]). 5-year change in quantitative minJSW was not significant in either the medial (OR 1.04 [95% CI (0.65, 1.65)]) or lateral compartment (OR 0.79 [95% CI (0.60, 1.05)]).

CONCLUSION: These results show that 5-year change in categorical joint space and overall K/L grade predict future knee replacements up to 15 years later. Quantitative joint space change was not predictive, perhaps due to the large number of subjects whose values remained within the normal range over this period of follow-up. The performance characteristics of these measures require further evaluation.

CHINGFORD STUDY: KNEE RADIOGRAPHIC SCORING ATLAS and MANUAL



KELLGREN & LAWRENCE (K/L)

WHOLE KNEE K/L

Reference: Images from Kellgren and Lawrence 1957, Descriptions from Kellgren et al. 1963

Additional Information:

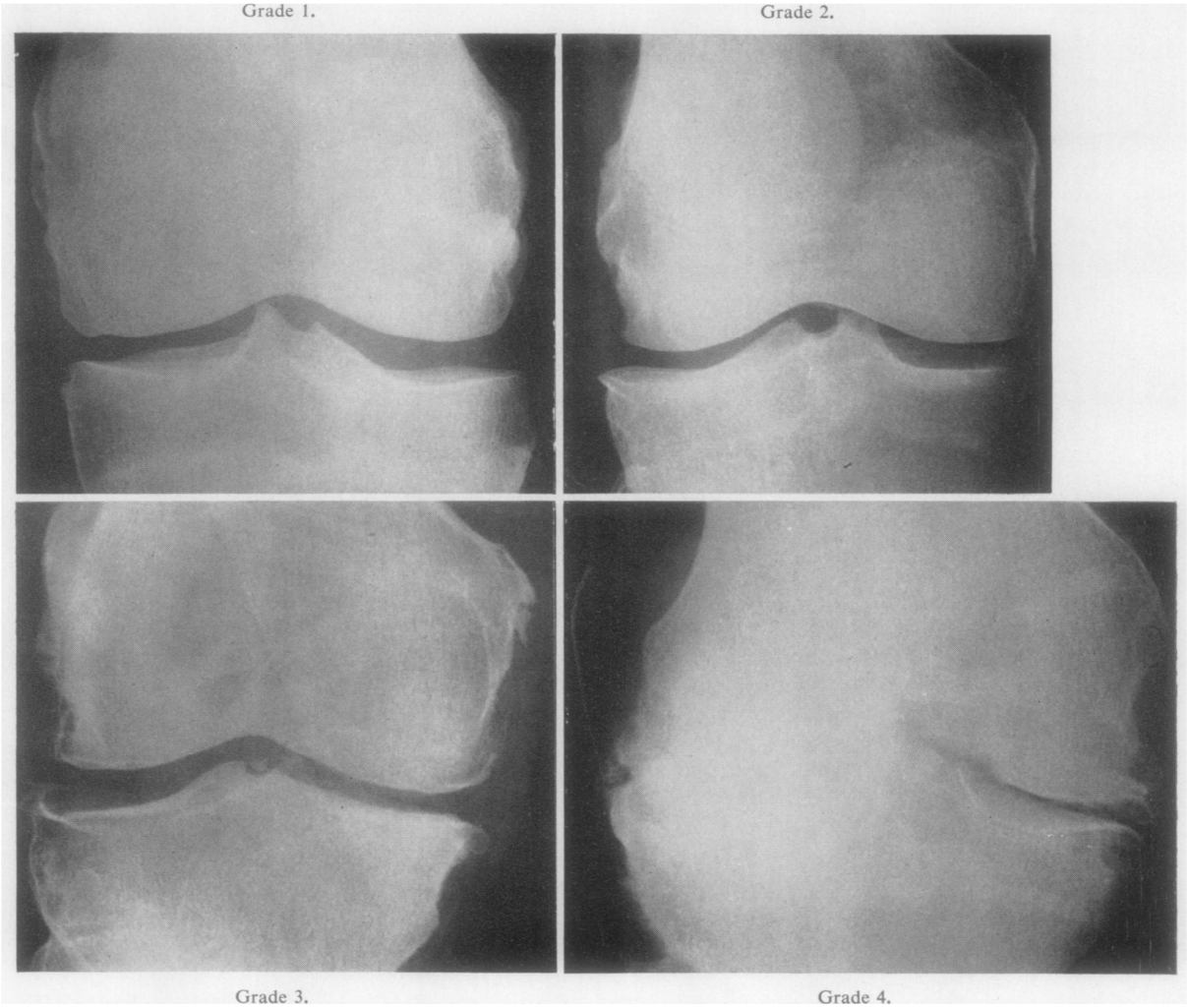
1. Grade 1 = Possible Osteophyte, ***not*** a small definite osteophyte as in some versions of K&L. A grade 1 should be marked when there is evidence of a disruption of cortical bone and/or abnormal morphology in common osteophytic areas, without the definition of an actual osteophyte. A grade 1 K&L should not equal a grade 1 in any other scoring method for osteophyte as an individual feature.
2. Grade 2 = Definite osteophyte present with a possible JSN. This should be scored as a grade 2 even if joint space appears normal.
3. Grade 3 = At least one definite osteophyte (even if small) plus definite JSN

COMPARTMENTAL K/L

Additional Information:

1. Using same definitions as whole knee K&L, but features must all be present in the ***same*** compartment, i.e. definite osteophyte plus possible narrowing in the medial compartment equals a Grade 2

(Kellgren and Lawrence 1957)



(Kellgren et al. 1963)

Grade 0 – Normal

Grade 1 – Doubtful narrowing of joint space and possible osteophytic lipping

Grade 2 – Definite osteophytes and possible narrowing of joint space

Grade 3 – Moderate, multiple osteophytes, definite narrowing of joint space, some sclerosis and possible deformity of bone ends

Grade 4 – Large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends

OARSI ATLAS

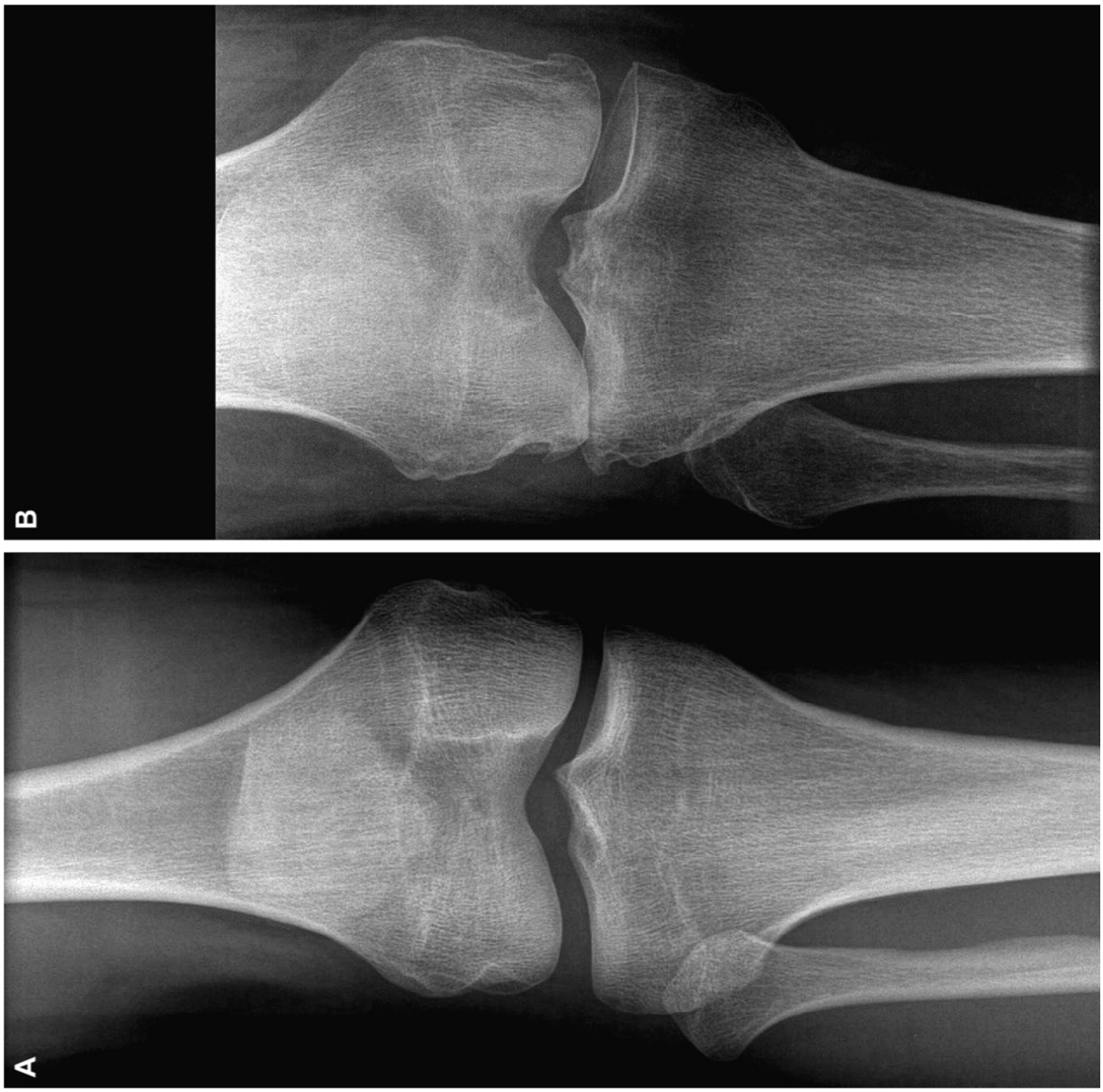
Reference: Images from Altman et al. 2007 (Revised atlas)

Additional Information:

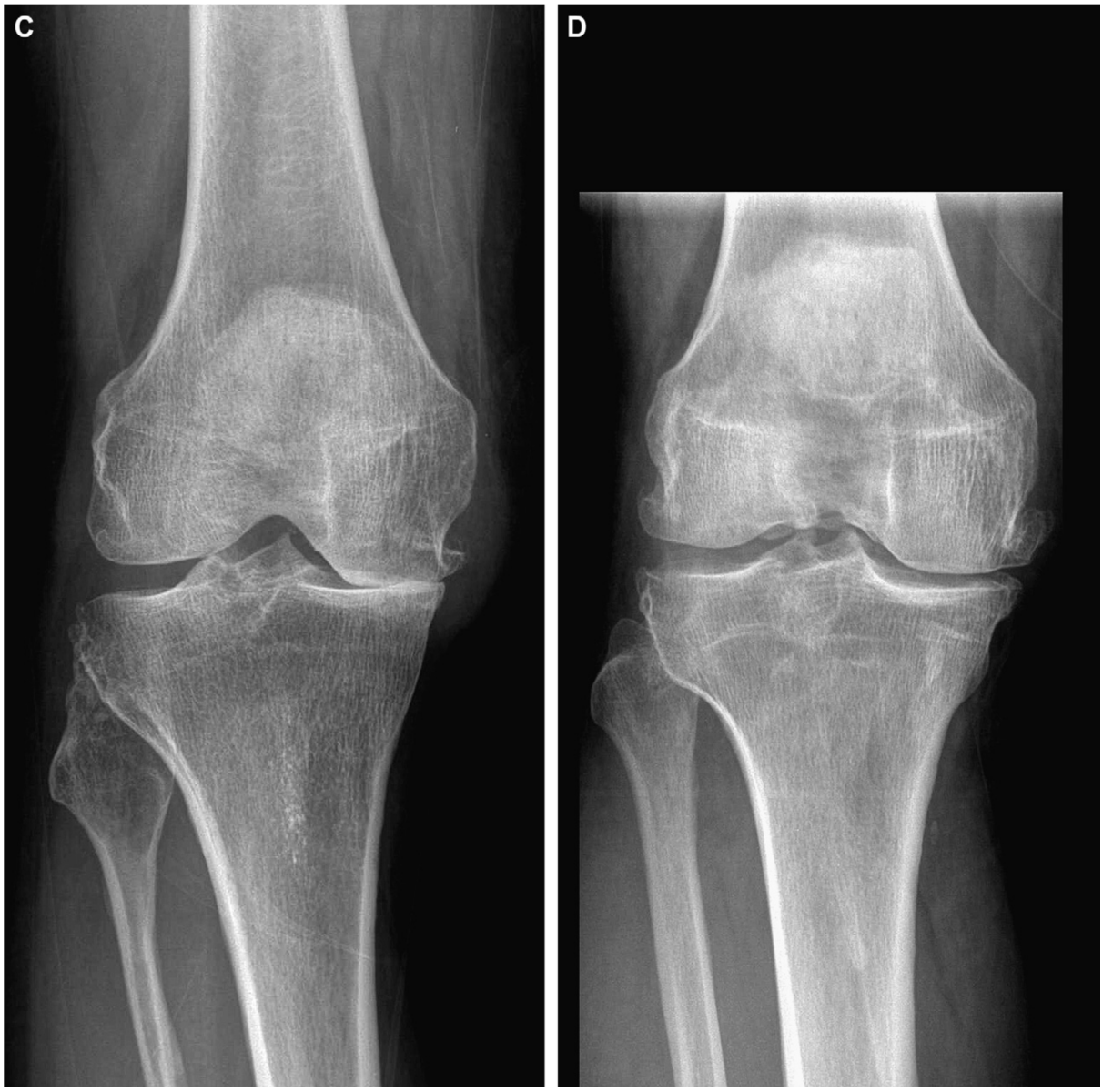
1. Osteophyte Grade 1 = Definite osteophyte (not possible osteophyte as used in K&L)
2. JSN Grade 1 = possible/minimum narrowing
3. JSN Grade 2 = moderate/definite narrowing
4. JSN Grade 3 = bone on bone

(Altman 2007)

Medial Femoral Osteophytes

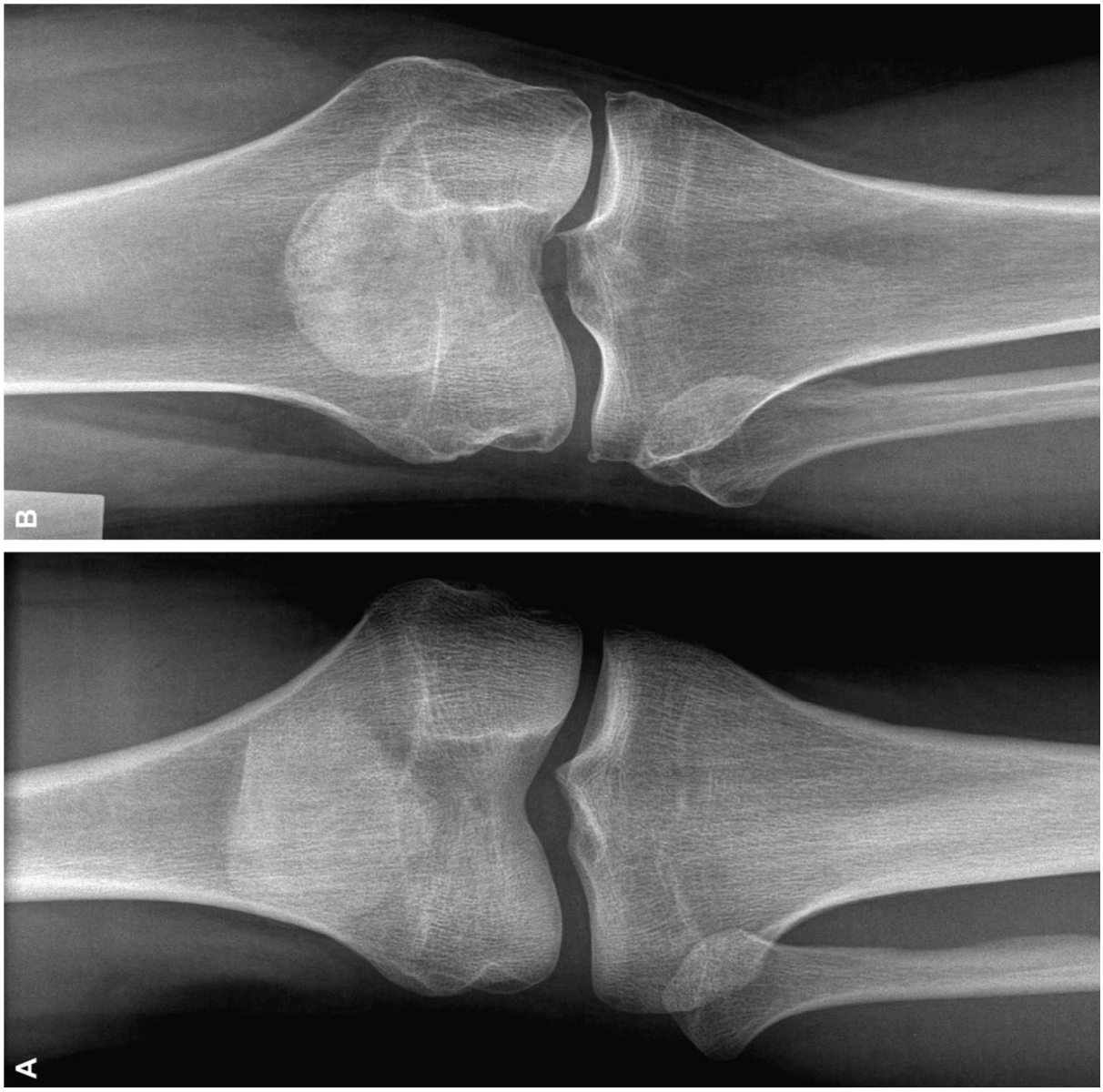


A) Grade 0, Normal
B) Grade 1 Medial femoral osteophyte

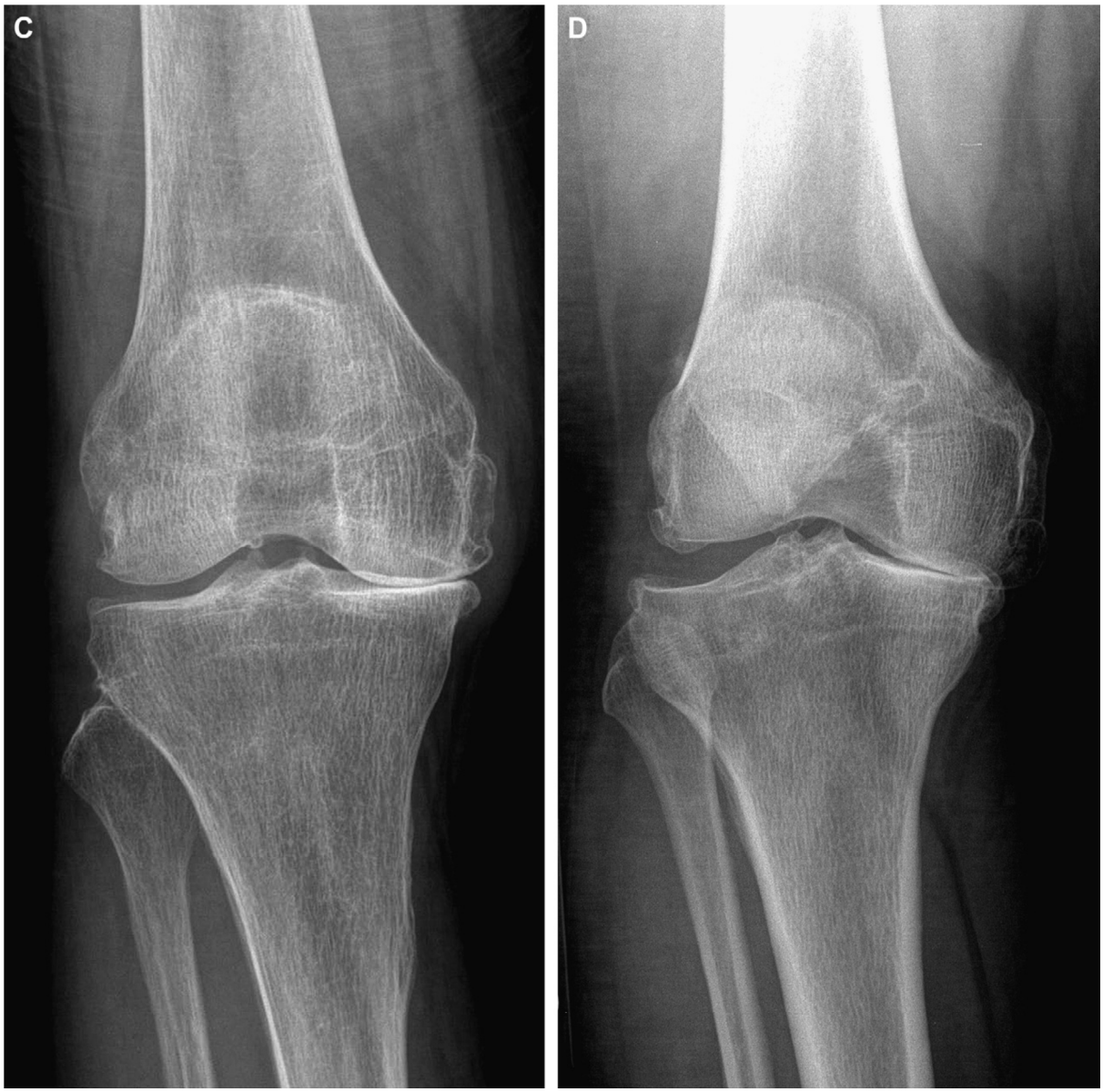


C) Grade 2 Medial femoral osteophyte
D) Grade 3 Medial femoral osteophyte

Medial Tibial Osteophytes

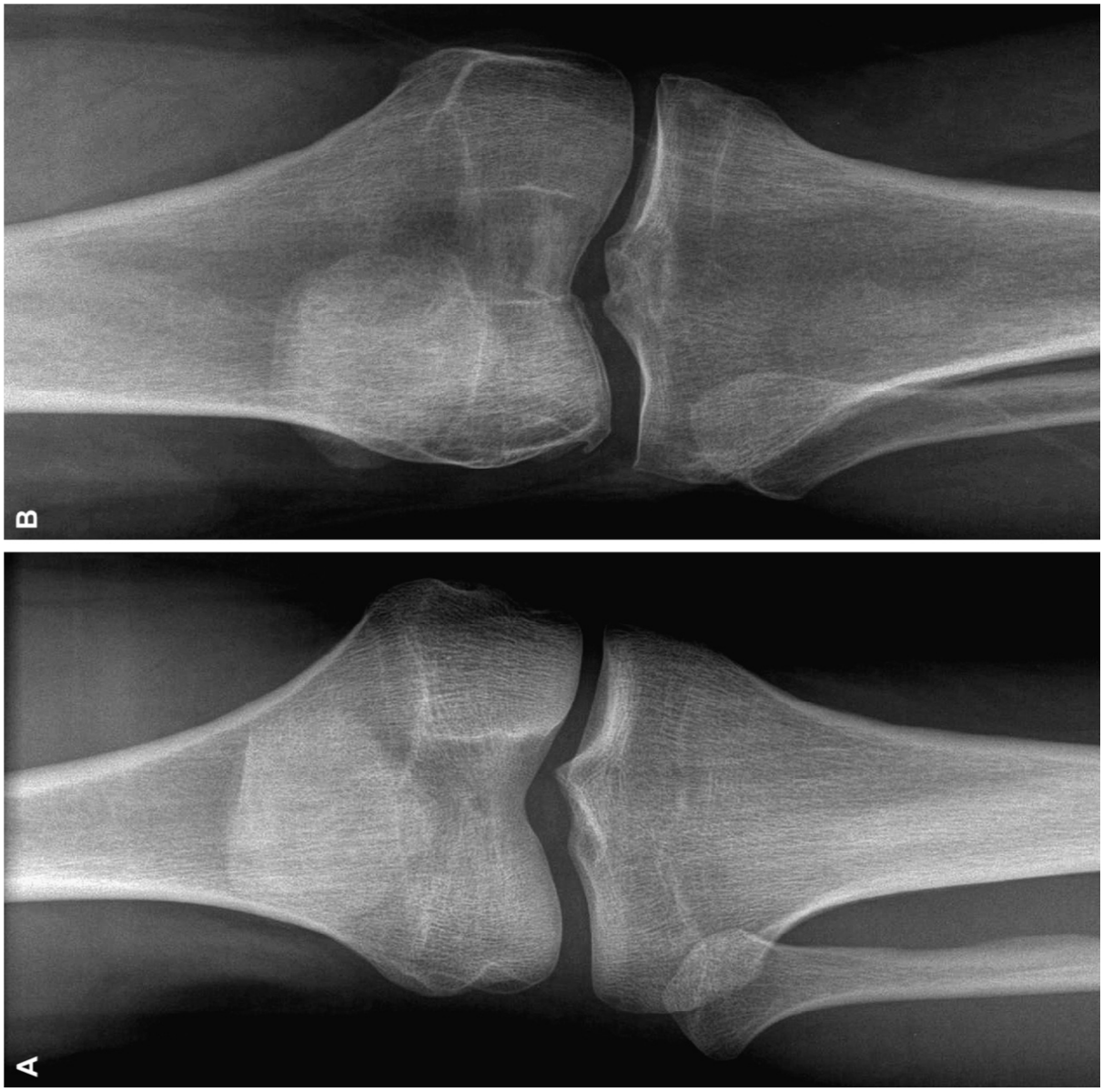


A) Grade 0, Normal
B) Grade 1 Medial tibial osteophyte

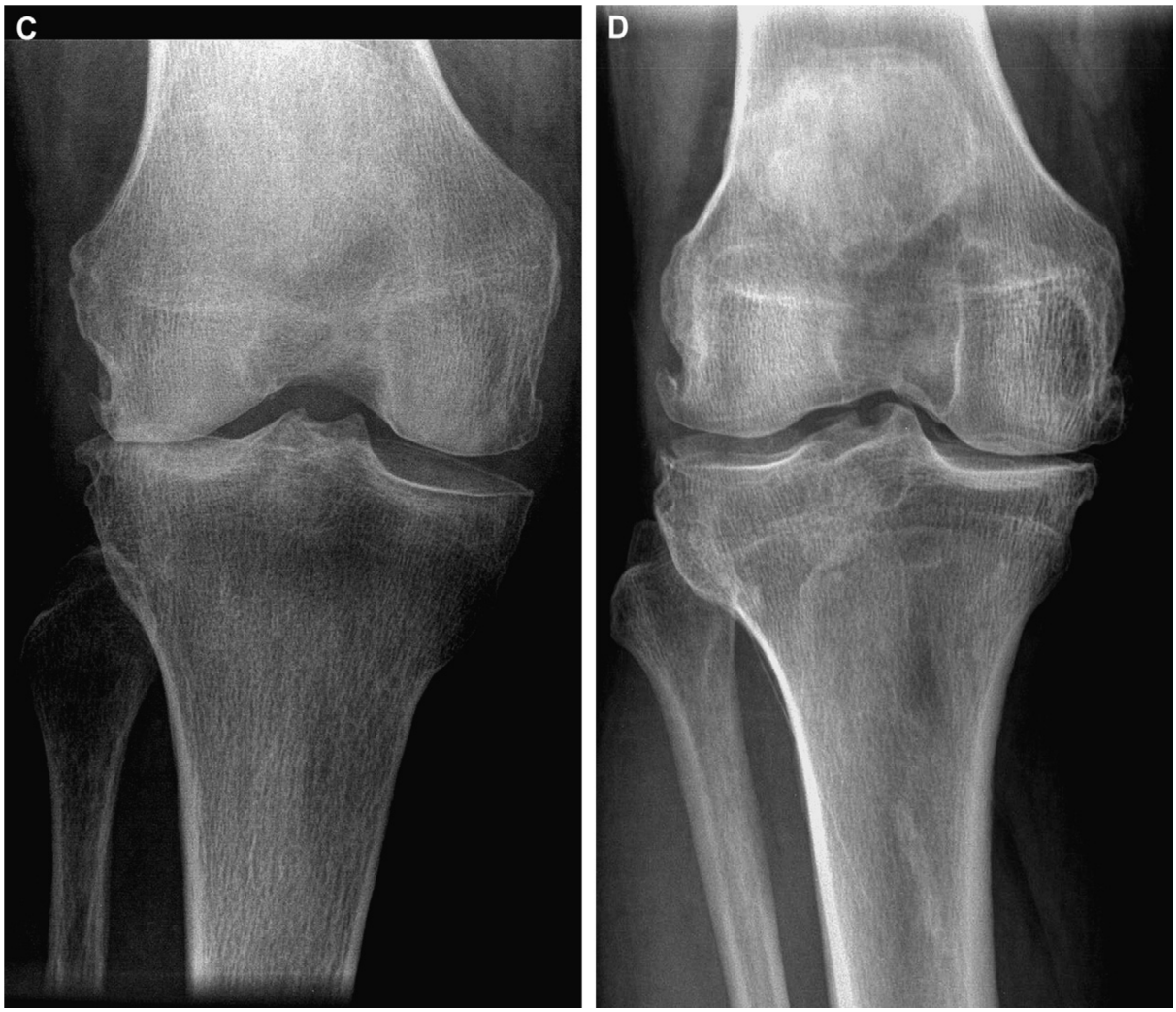


C) Grade 2 Medial tibial osteophyte
D) Grade 3 Medial tibial osteophyte

Lateral Femoral Osteophytes

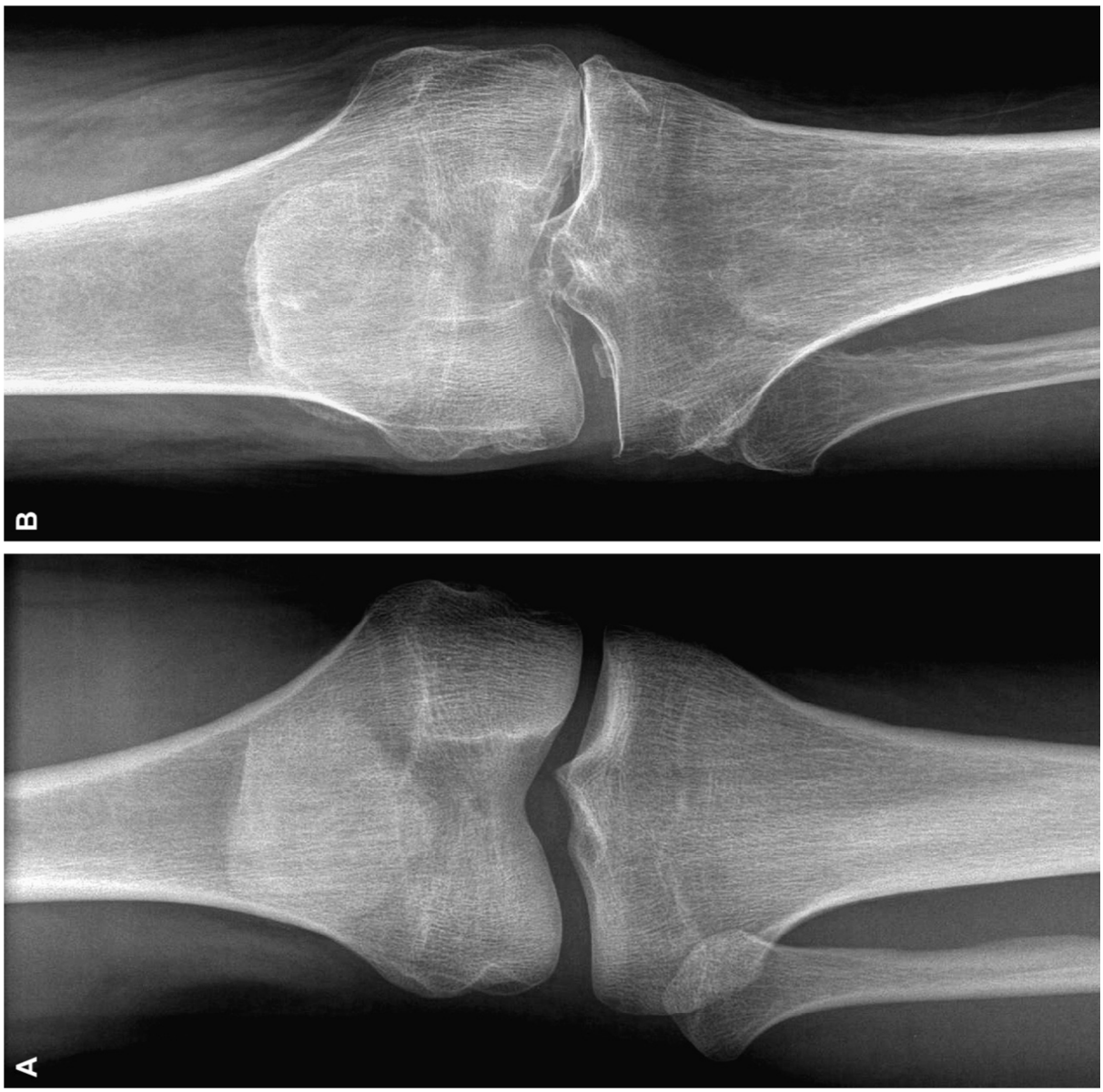


A) Grade 0 Normal
B) Grade 1 Lateral femoral osteophyte

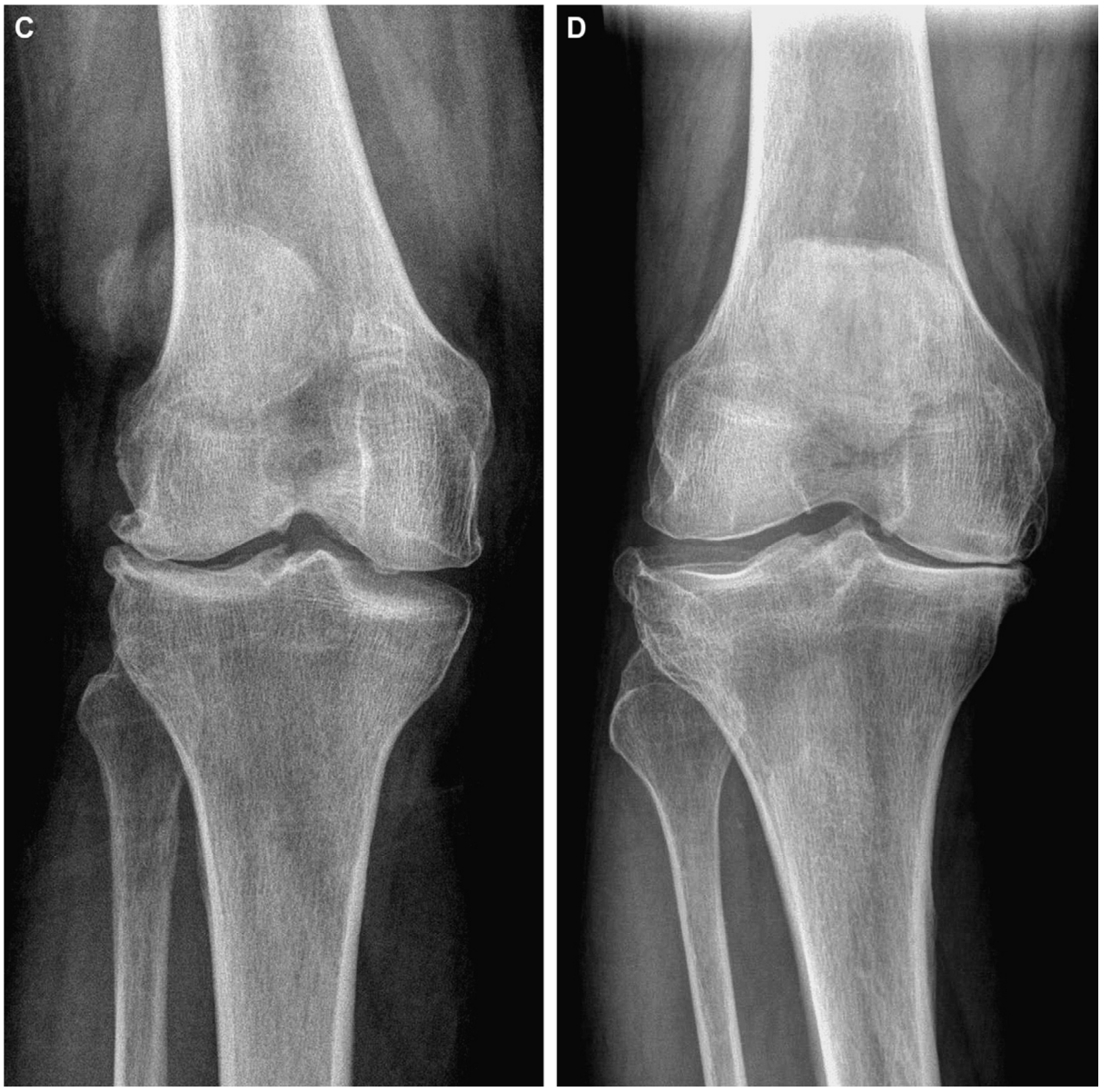


C) Grade 2 – Lateral femoral osteophyte
D) Grade 3 – Lateral femoral osteophyte

Lateral Tibial Osteophytes

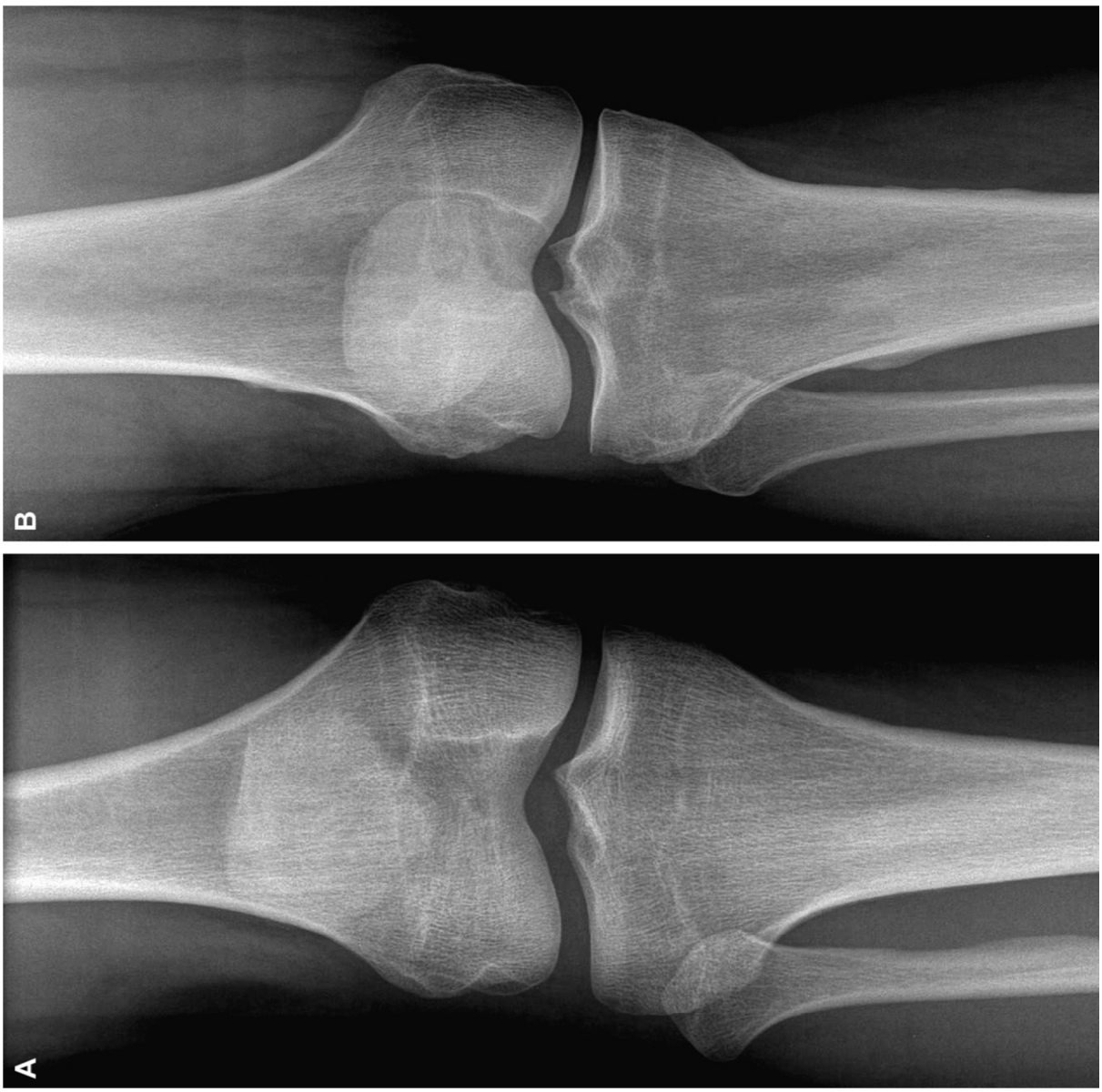


A) Grade 0 Normal
B) Grade 1 Lateral tibial osteophyte

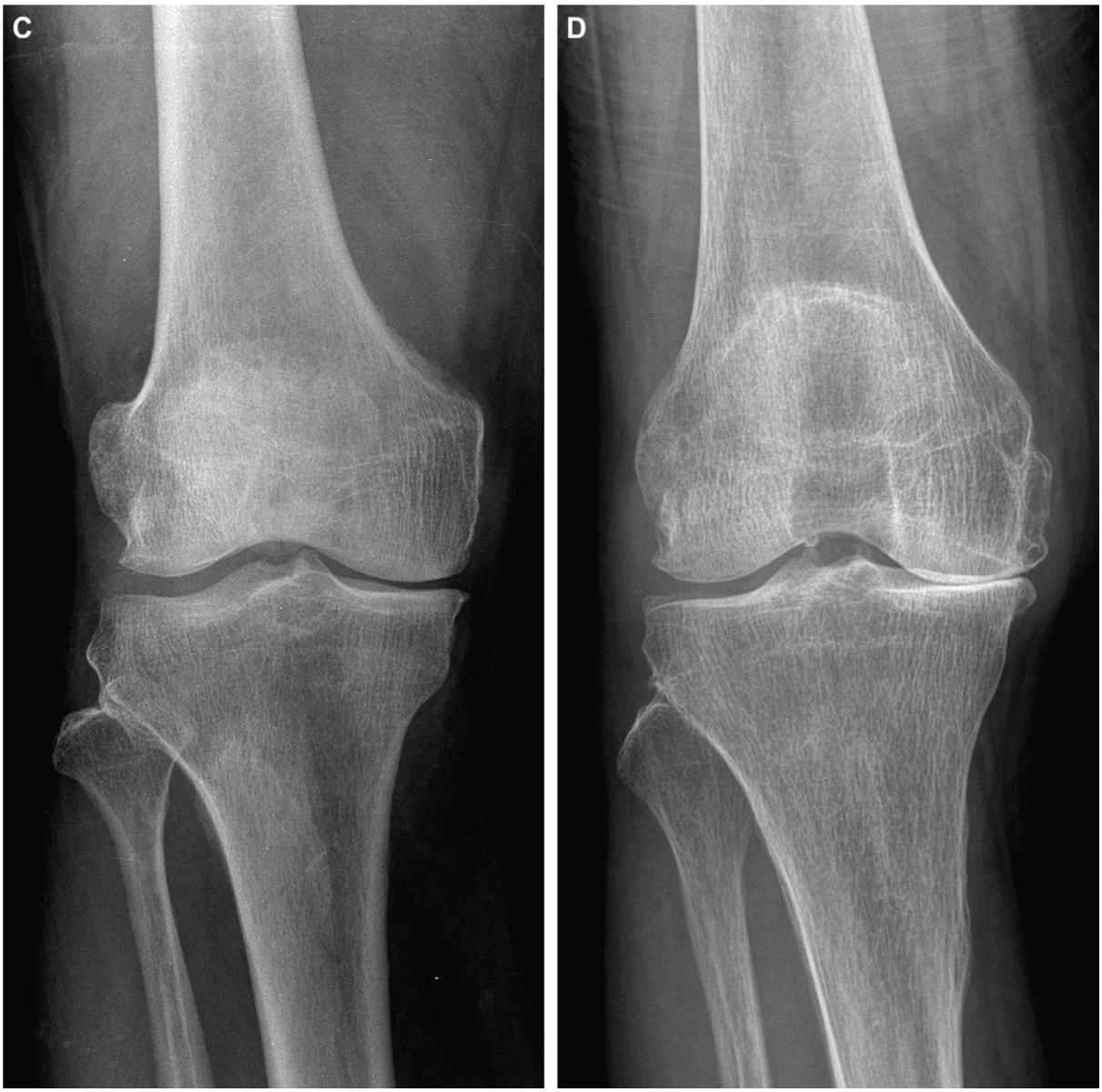


C) Grade 2 Lateral tibial osteophyte
D) Grade 3 Lateral tibial osteophyte

Medial Tibiofemoral Joint Space Narrowing

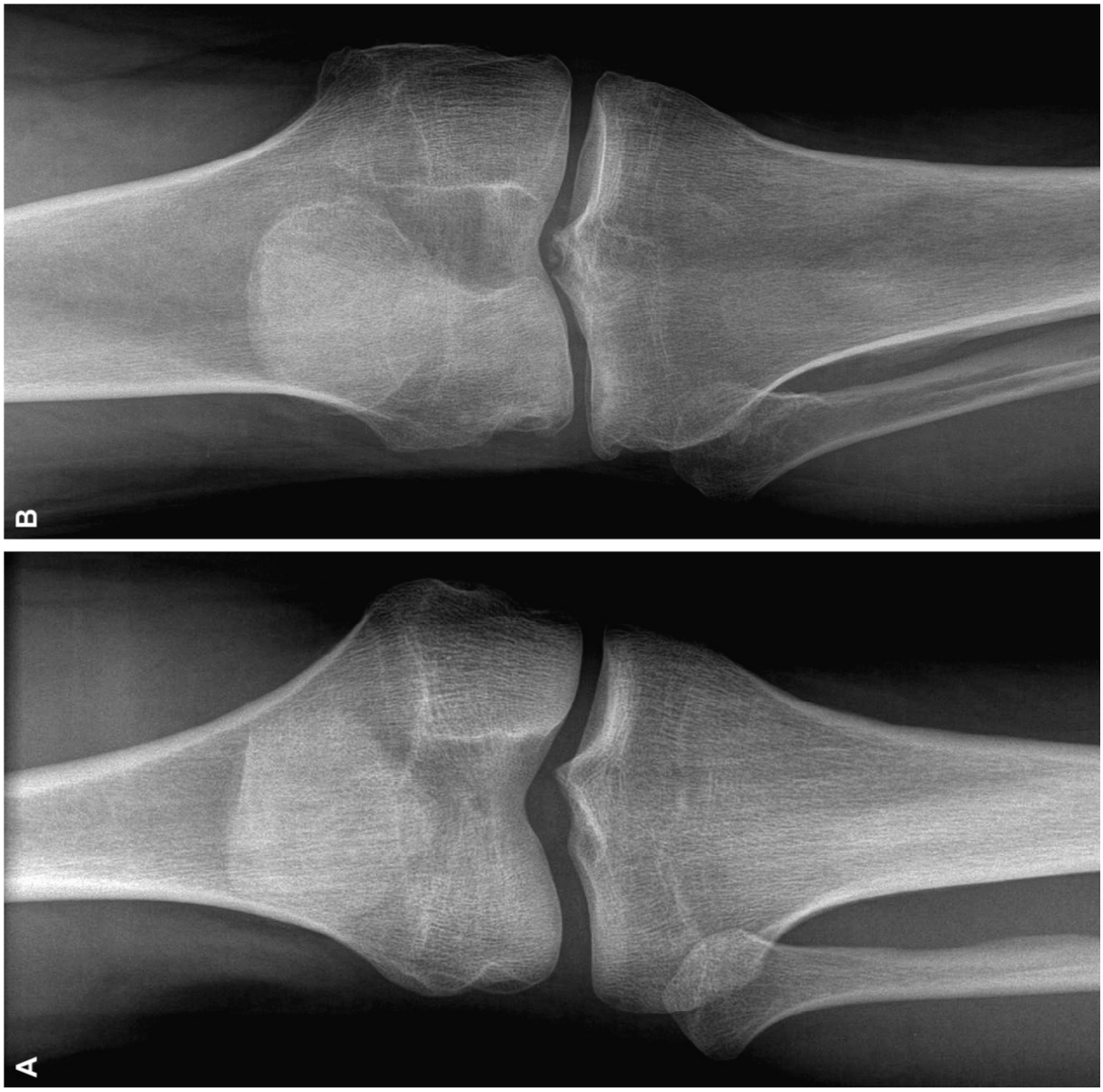


A) Grade 0 Normal
B) Grade 1 Medial tibiofemoral narrowing

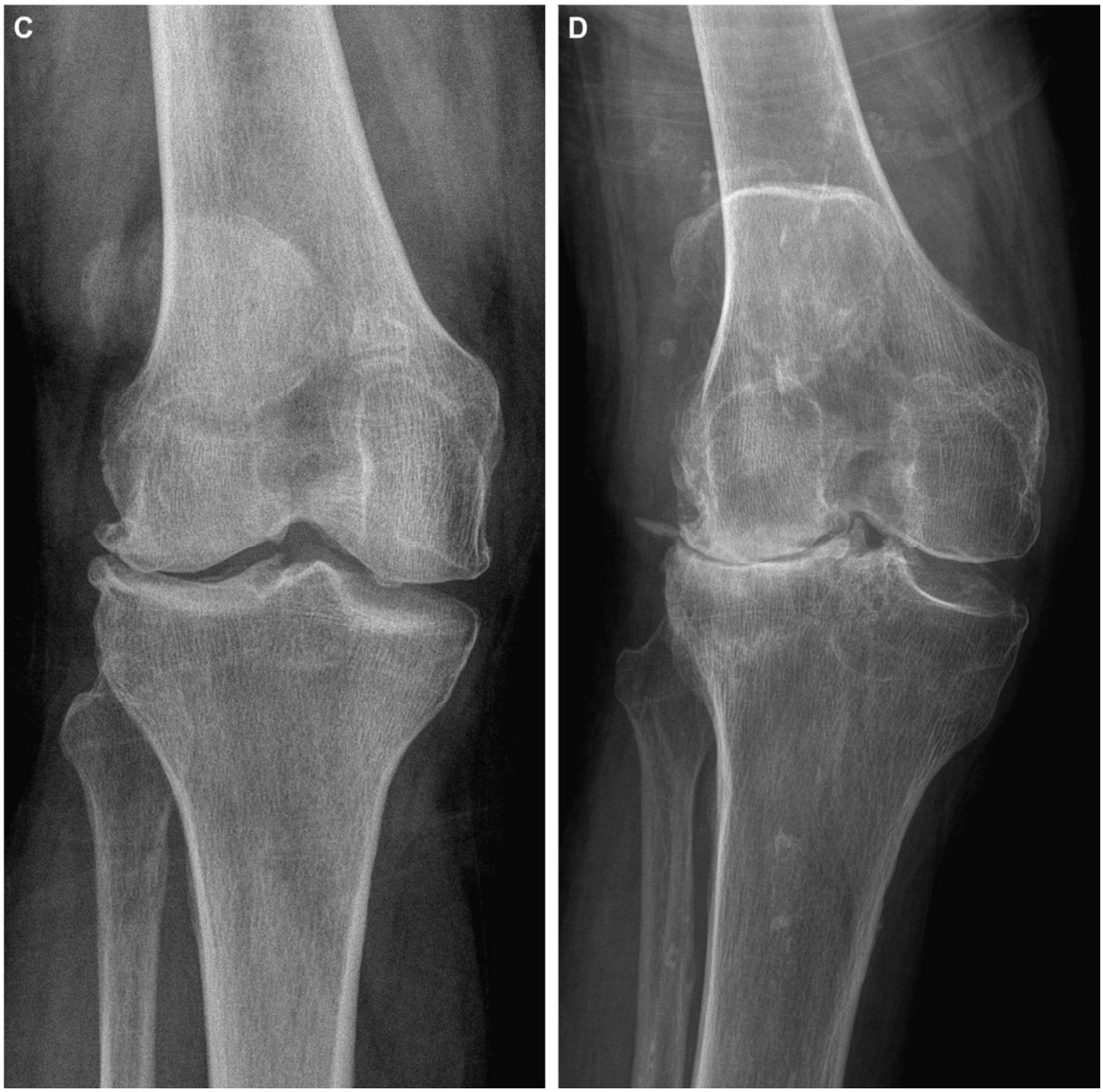


C) Grade 2 Medial tibiofemoral narrowing
D) Grade 3 Medial tibiofemoral narrowing

Lateral Tibiofemoral Joint Space Narrowing

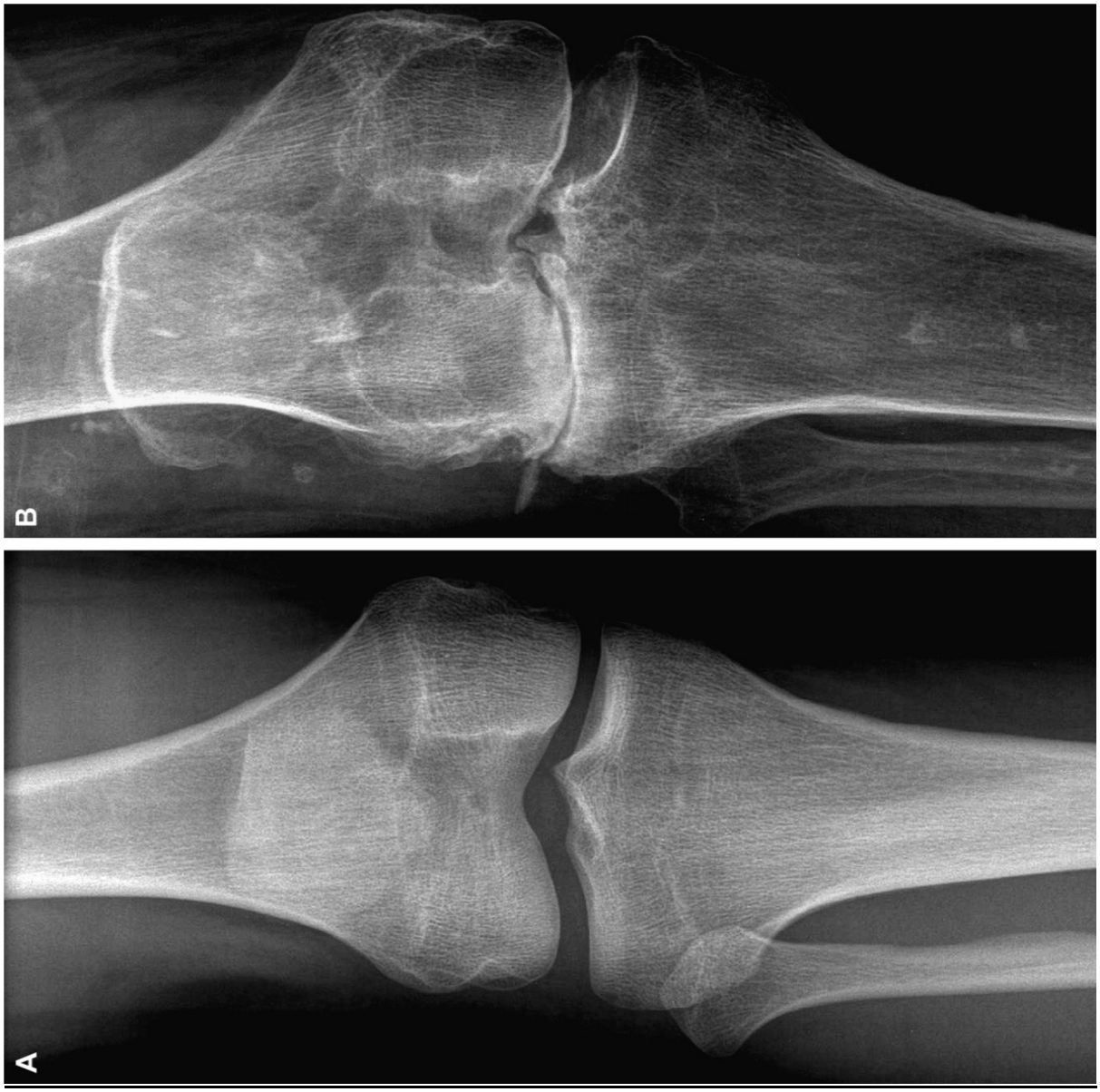


A) Grade 0 Normal
B) Grade 1 Lateral tibiofemoral narrowing

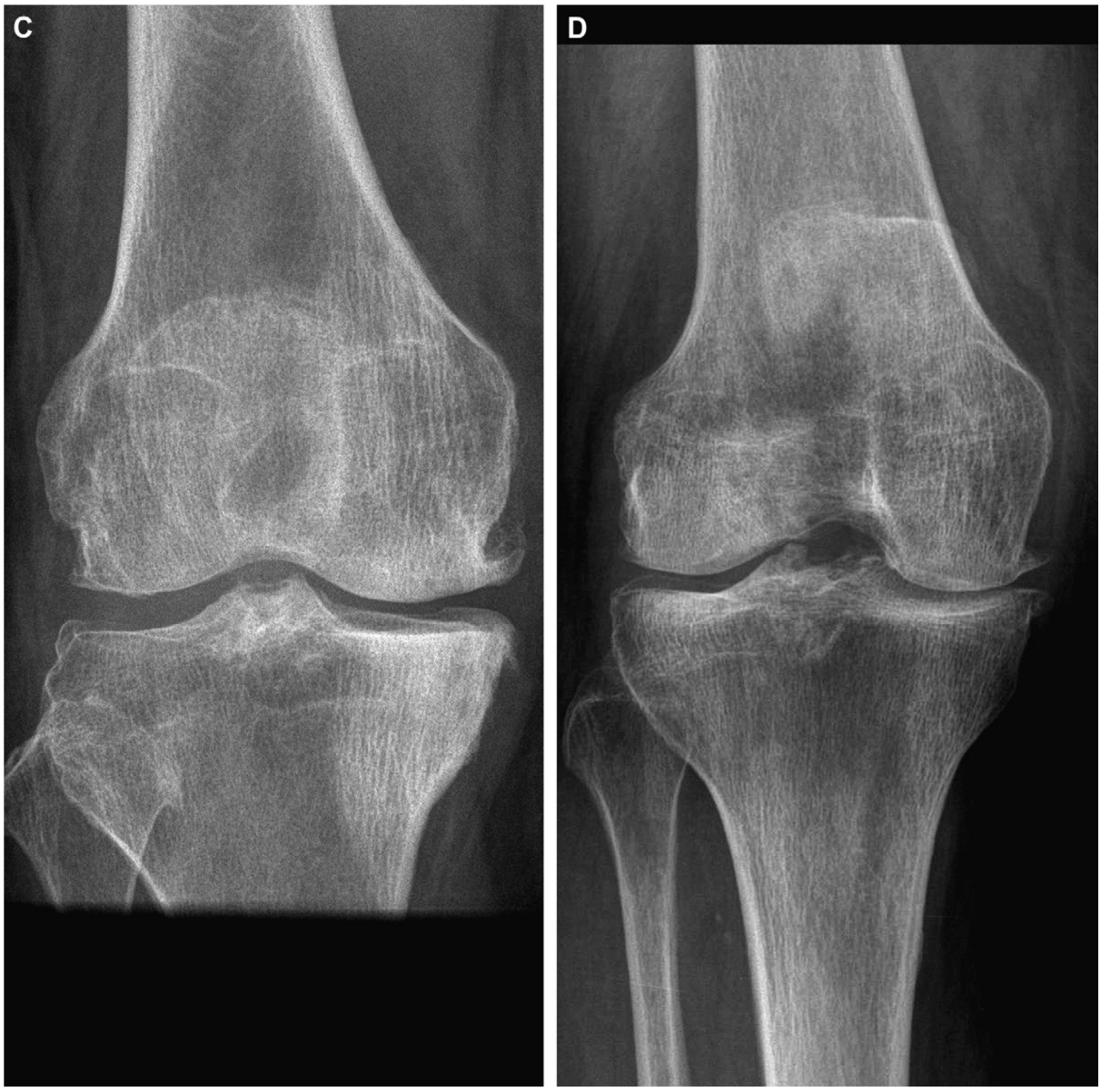


C) Grade 3 Lateral tibiofemoral narrowing
D) Grade 4 Lateral tibiofemoral narrowing

Attrition and Sclerosis



A) Normal
B) Medial tibial attrition



C) Medial tibial sclerosis
D) Lateral femoral sclerosis

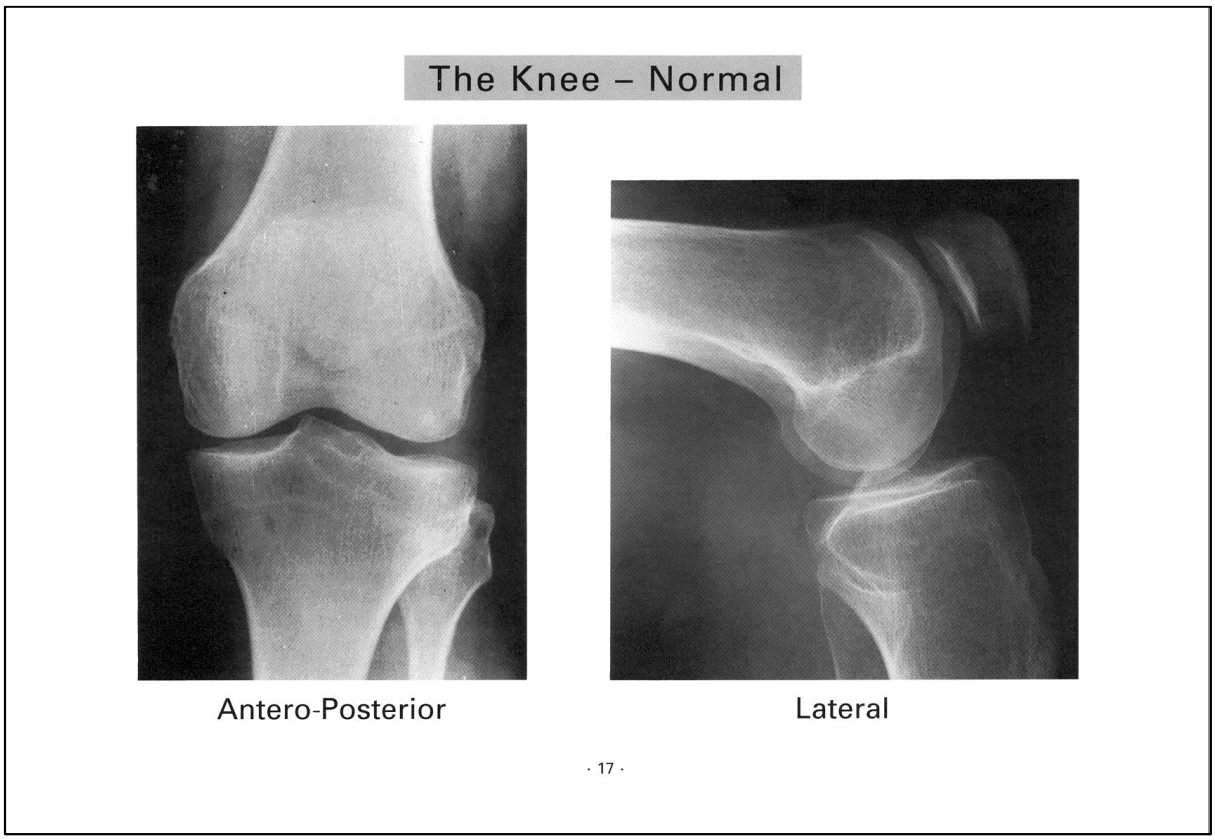
SPECTOR ATLAS

Reference: Images from Spector 1992

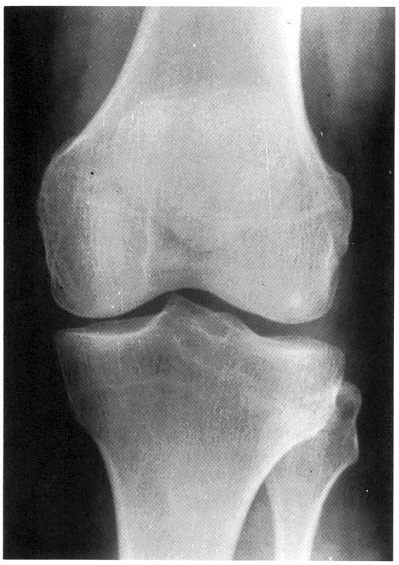
Additional Information:

1. JSN Grade 1 – possible/minimal narrowing
2. JSN Grade 2 – definite narrowing
3. JSN Grade 3 – bone on bone (anterior rim of tibial plateau)
4. Osteophyte Grade 1 – small definite osteophyte (not possible osteophyte as used in K&L grade 1)
5. Tibial spiking – should be scored as present (1) if at least one tibial spine appears abnormal

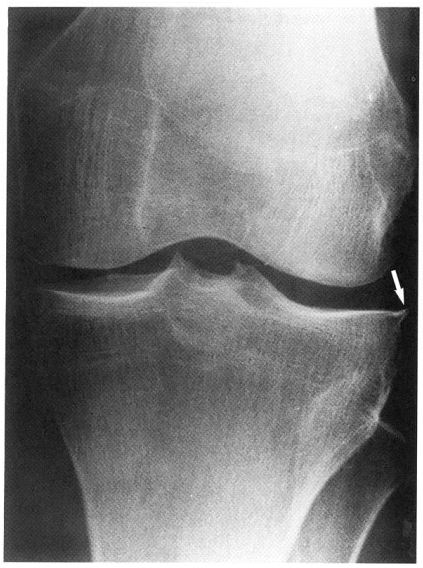
(Spector 1992)



The Knee – Tibiofemoral Osteophyte



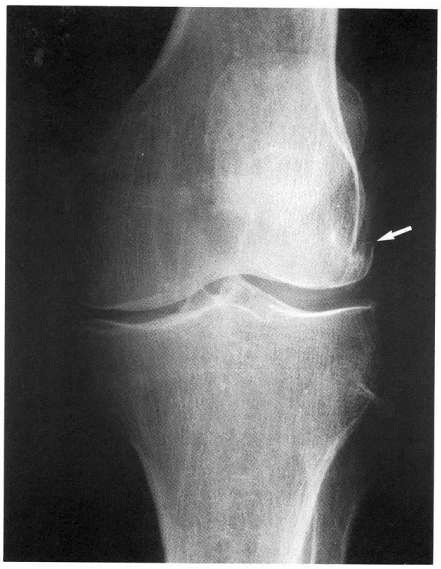
Grade 0



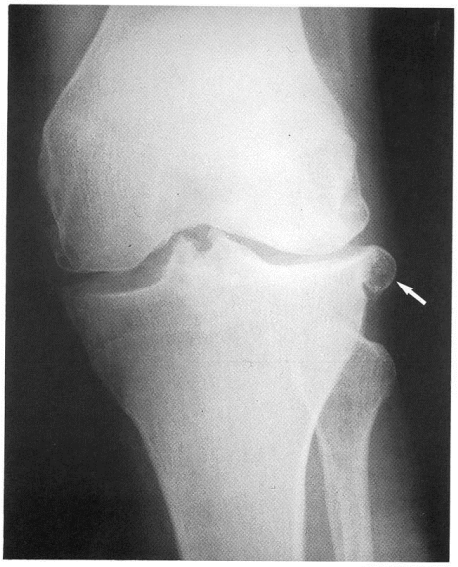
Grade 1

· 18 ·

The Knee – Tibiofemoral Osteophyte



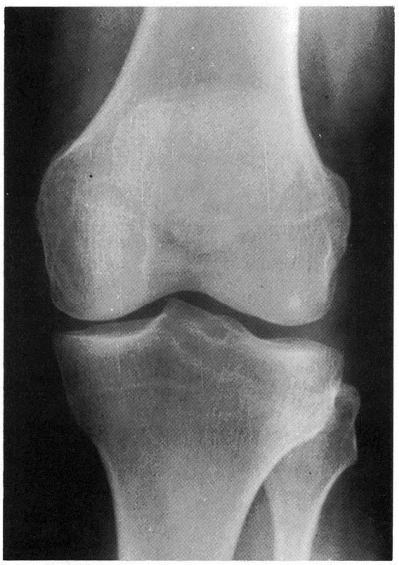
Grade 2



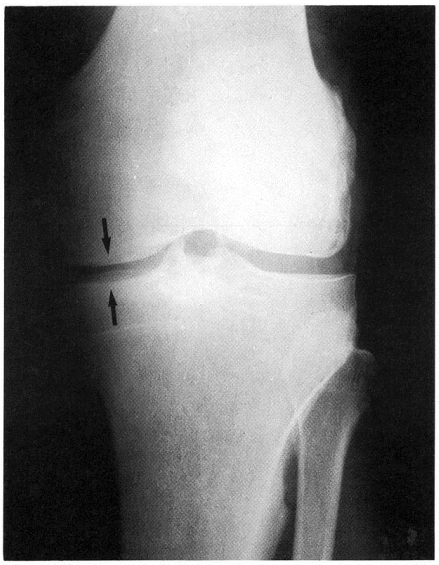
Grade 3

· 19 ·

The Knee – Tibiofemoral Narrowing



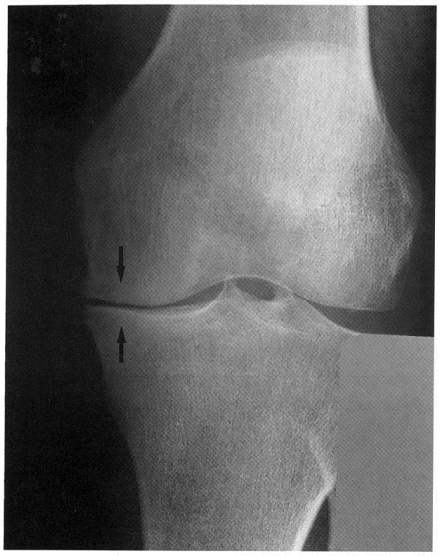
Grade 0



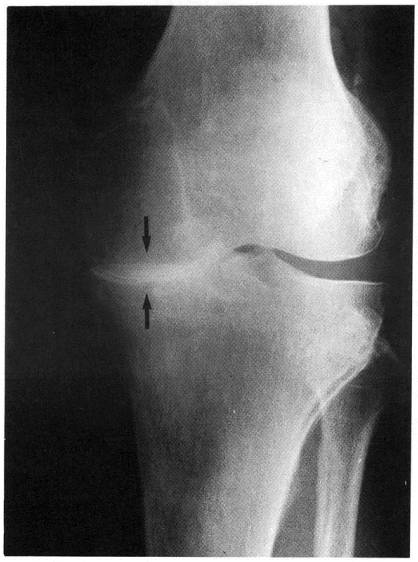
Grade 1

· 20 ·

The Knee – Tibiofemoral Narrowing



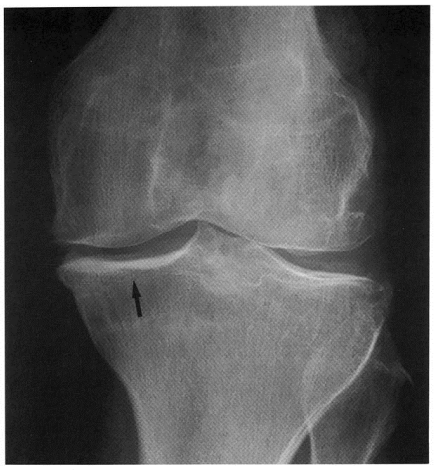
Grade 2



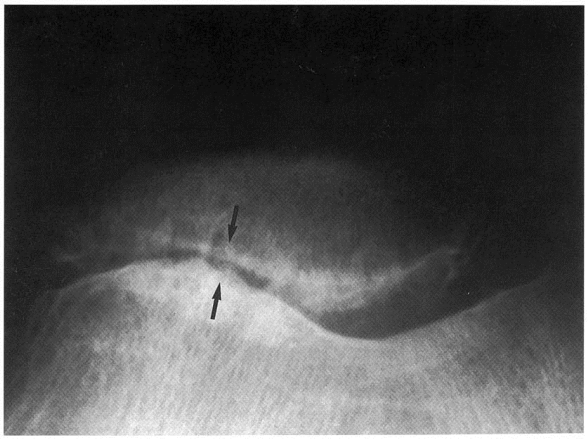
Grade 3

· 21 ·

The Knee



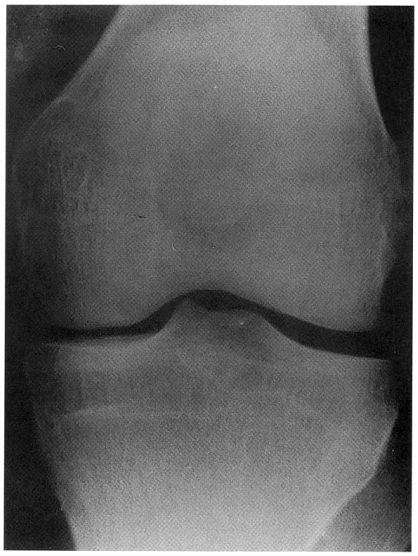
Tibiofemoral Sclerosis
Grade 1



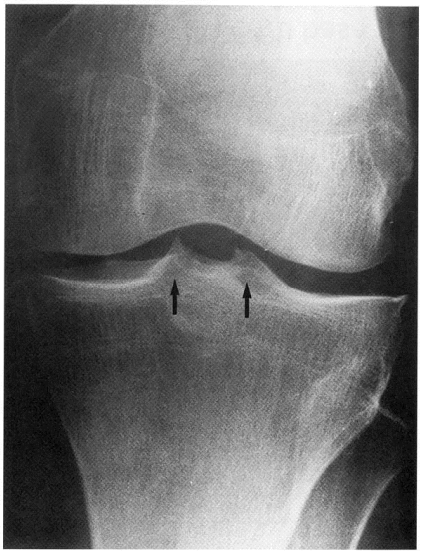
Patellofemoral Sclerosis
Grade 1

· 30 ·

The Knee – Tibial Spiking



Grade 0



Grade 1

· 31 ·

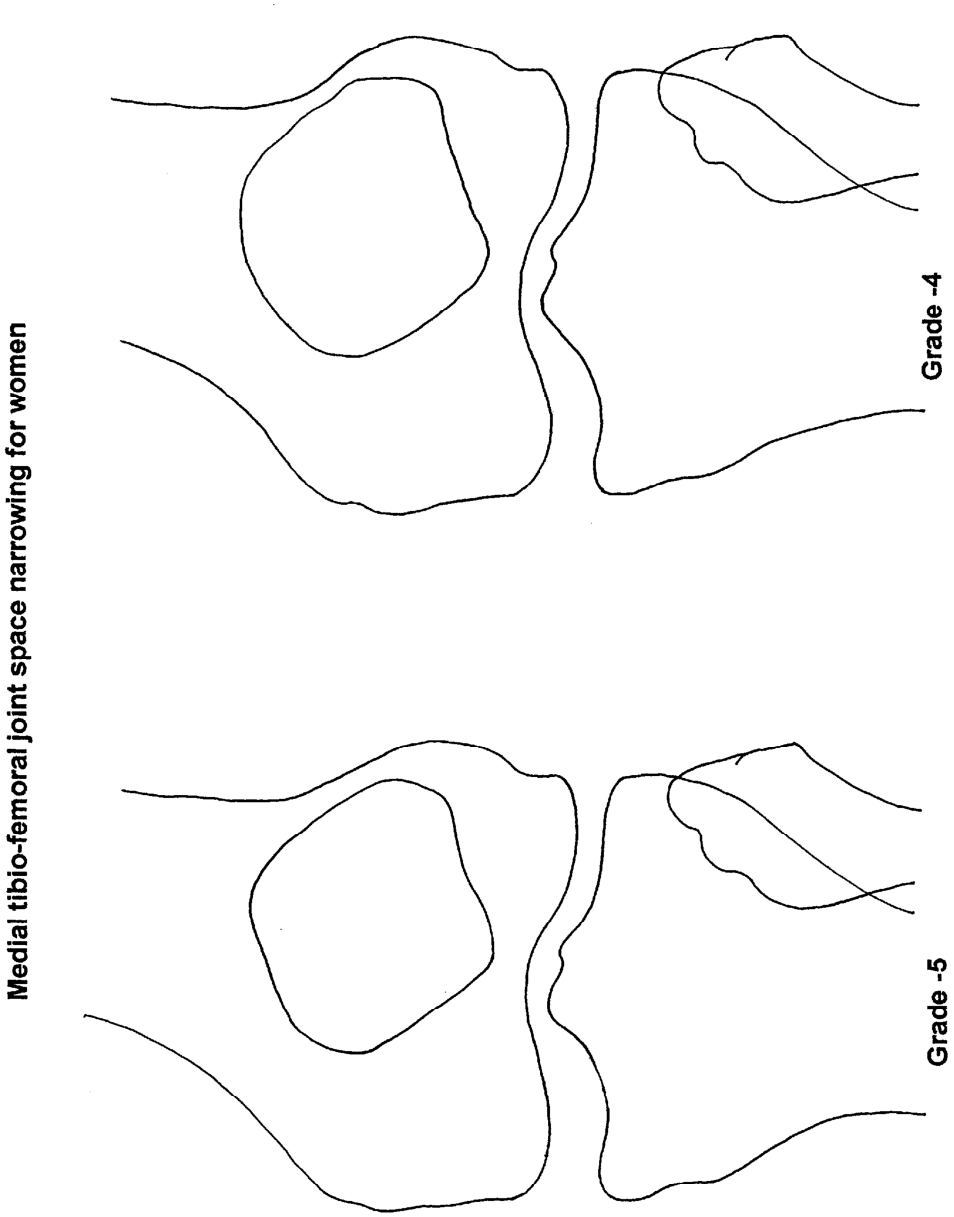
LINE DRAWING ATLAS

Reference: Images from Wilkinson 2005

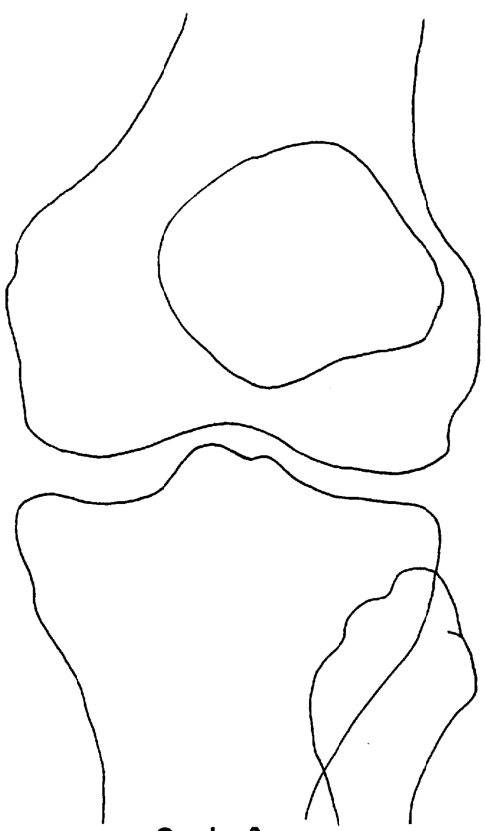
Additional Information:

1. JSN Grade 0 = Does ***not*** necessarily equal normal joint space as in other scoring methods
2. Osteophyte Grade 1 = Small definite osteophyte (not a possible osteophyte as used in K&L)
3. JSN = Anterior rims of the tibial plateau should be used to compare with line drawings

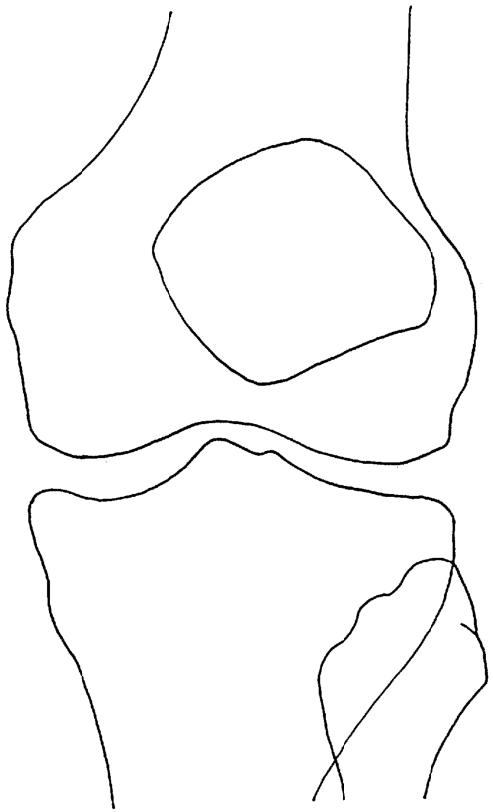
(Wilkinson 2005)



Medial tibio-femoral joint space narrowing for women

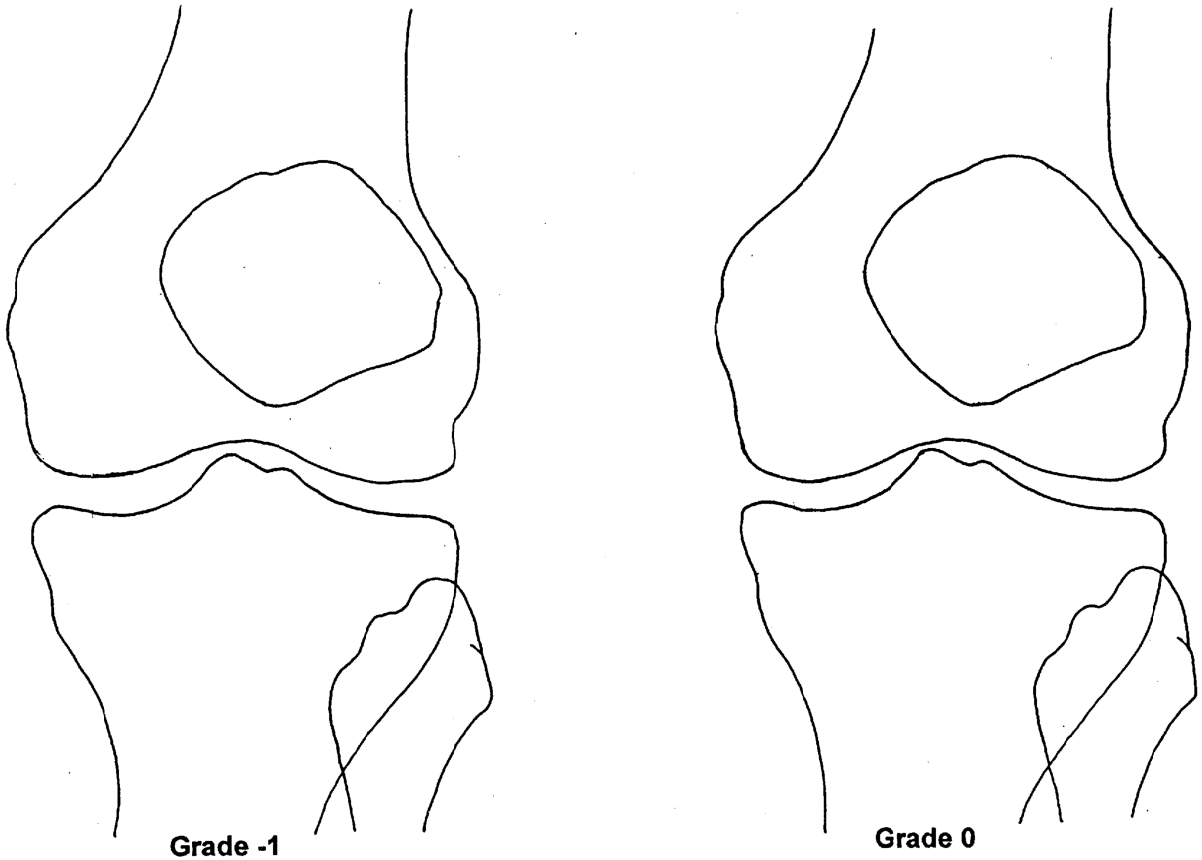


Grade -3



Grade -2

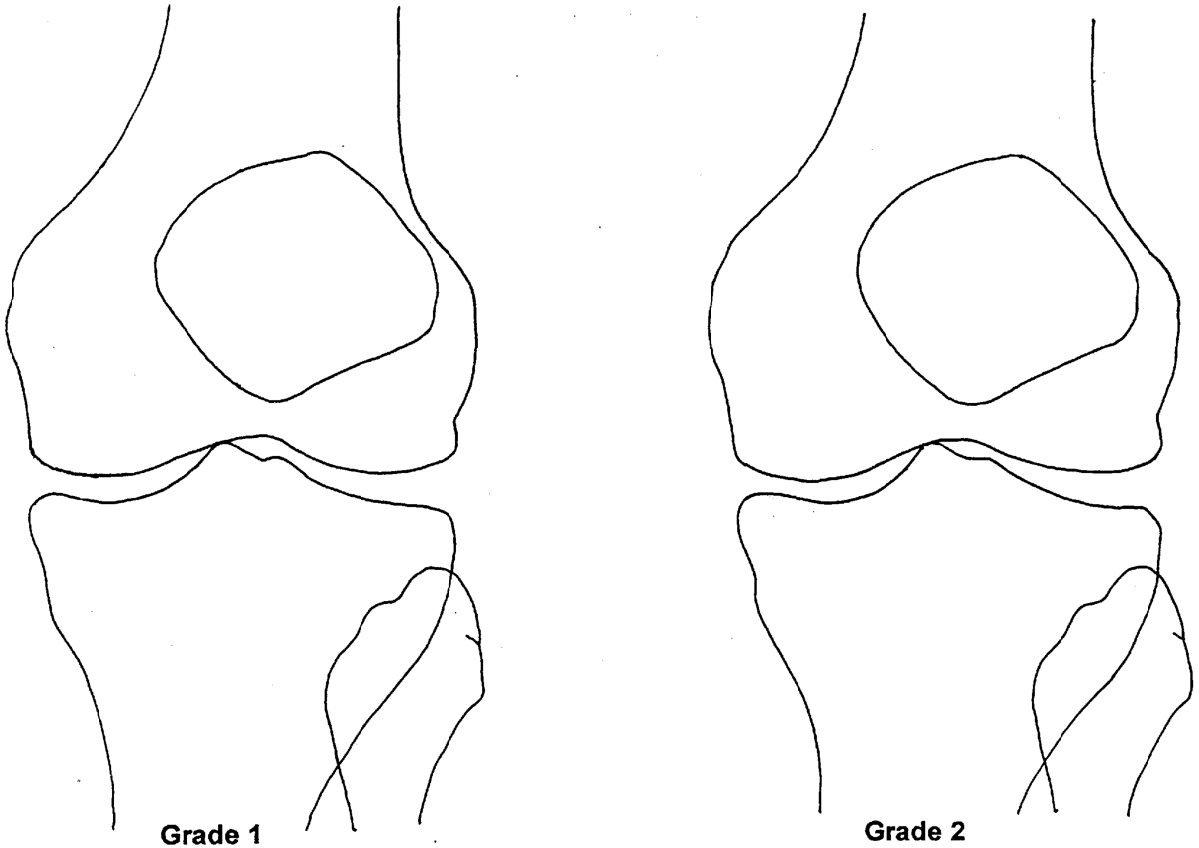
Medial tibio-femoral joint space narrowing for women



Grade -1

Grade 0

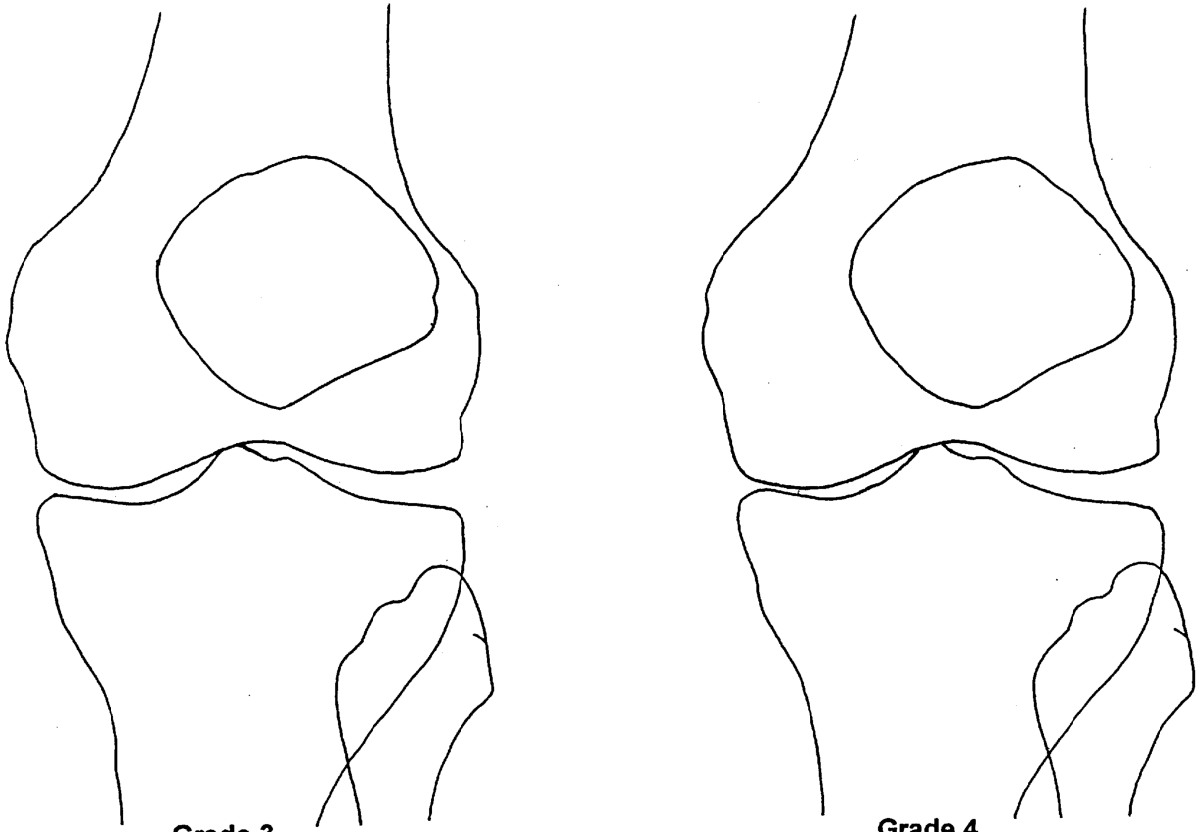
Medial tibio-femoral joint space narrowing for women



Grade 1

Grade 2

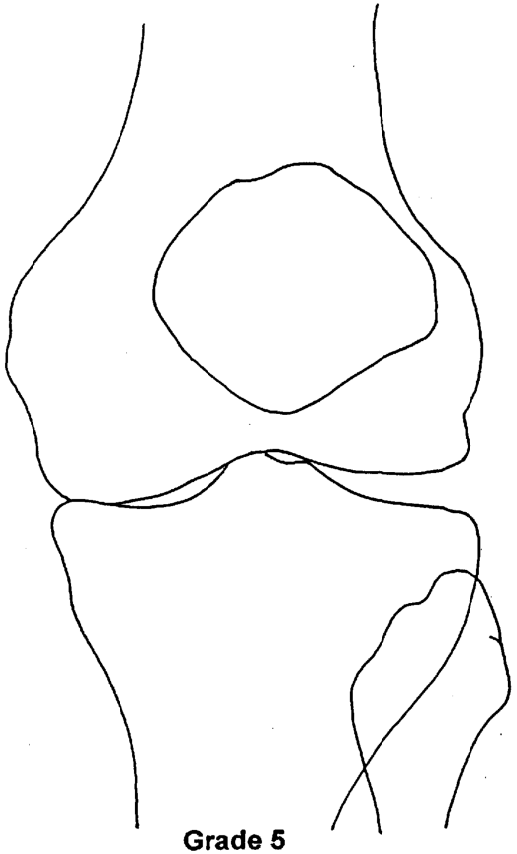
Medial tibio-femoral joint space narrowing for women



Grade 3

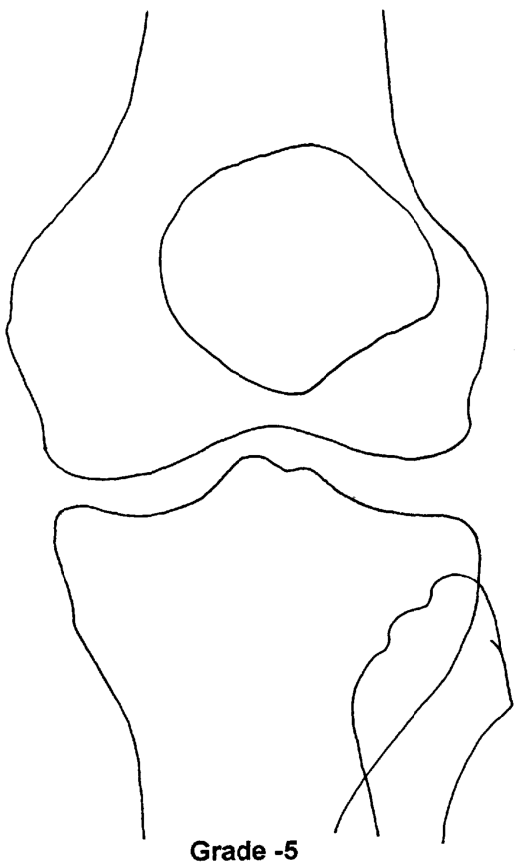
Grade 4

Medial tibio-femoral joint space narrowing for women

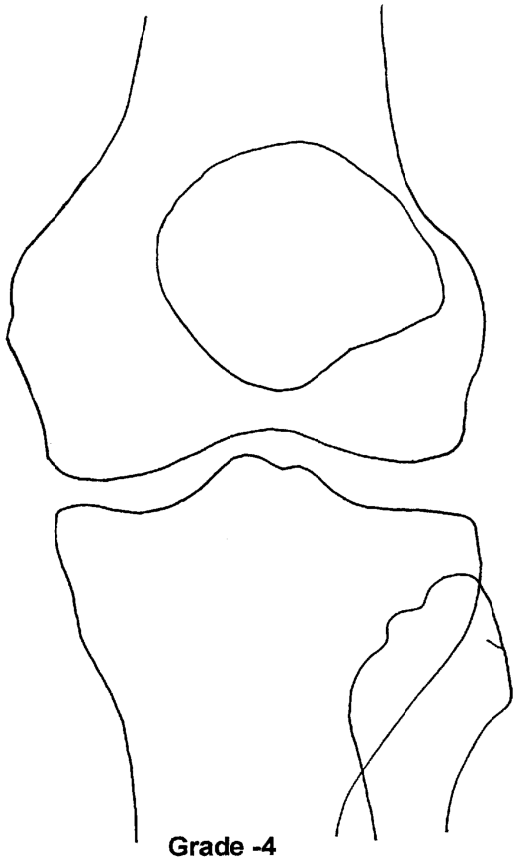


Grade 5

Lateral tibio-femoral joint space narrowing for women

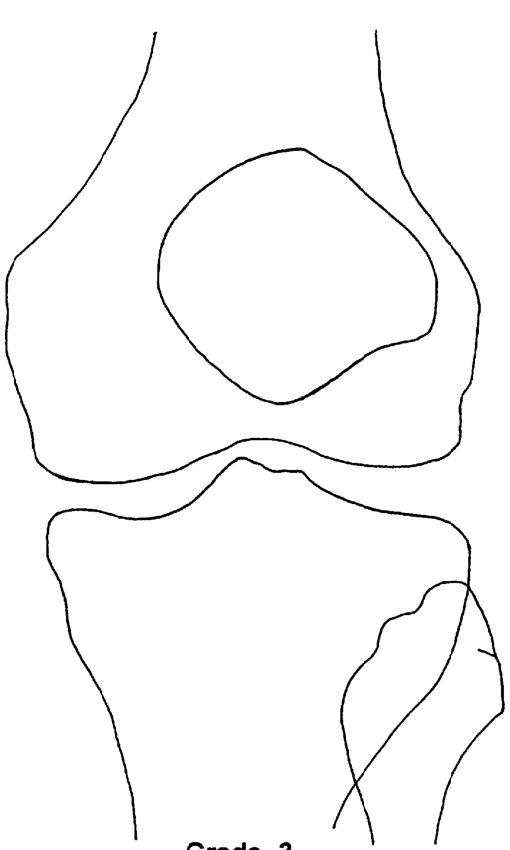


Grade -5

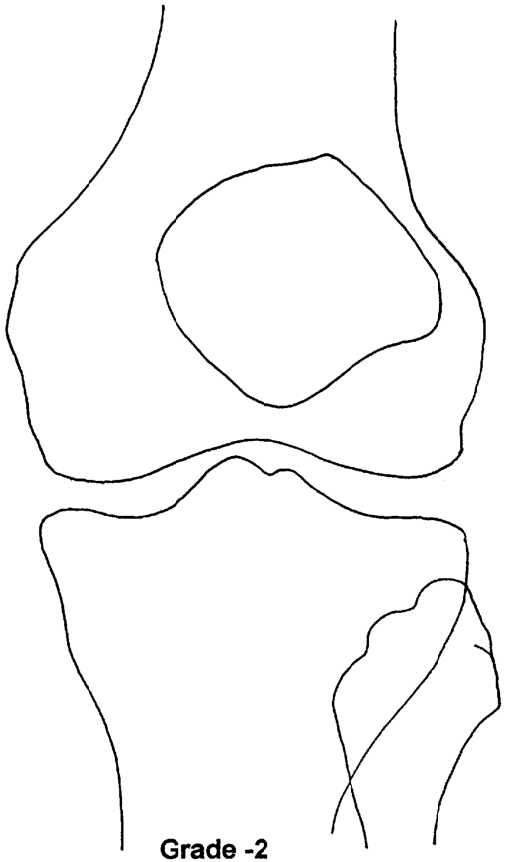


Grade -4

Lateral tibio-femoral joint space narrowing for women

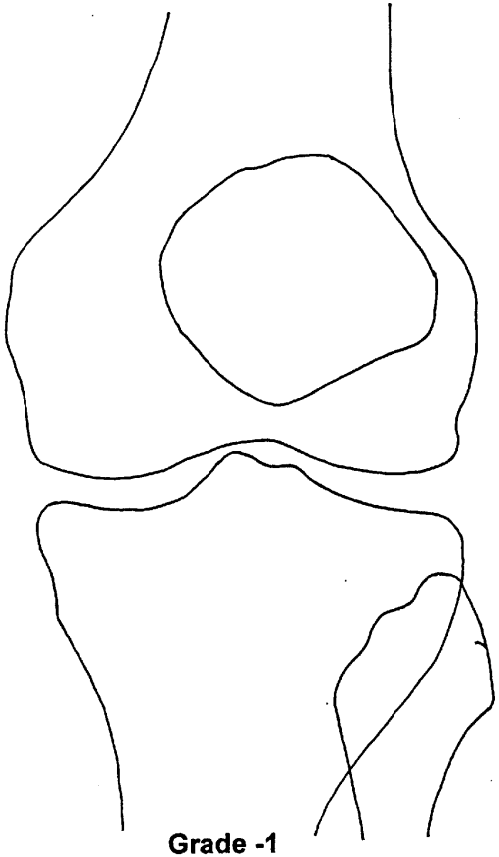


Grade -3

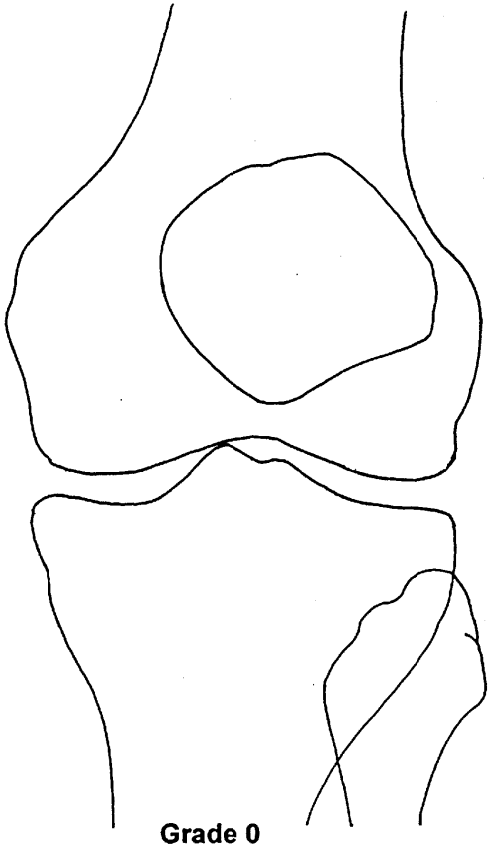


Grade -2

Lateral tibio-femoral joint space narrowing for women

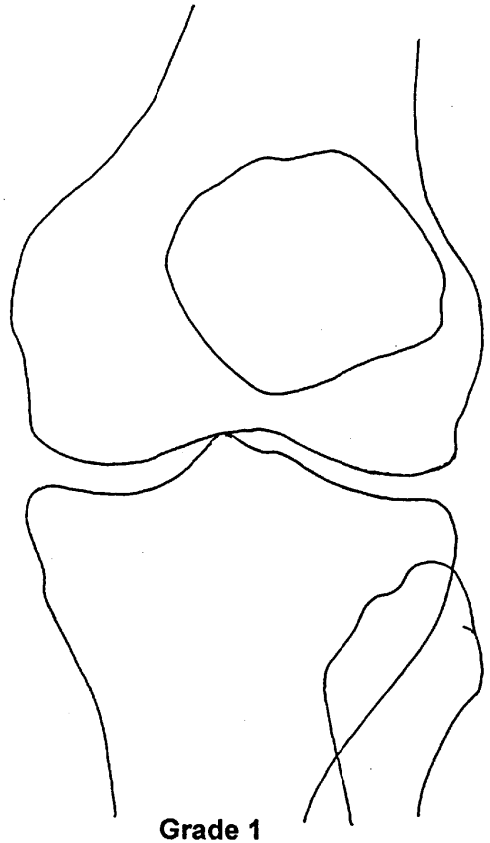


Grade -1

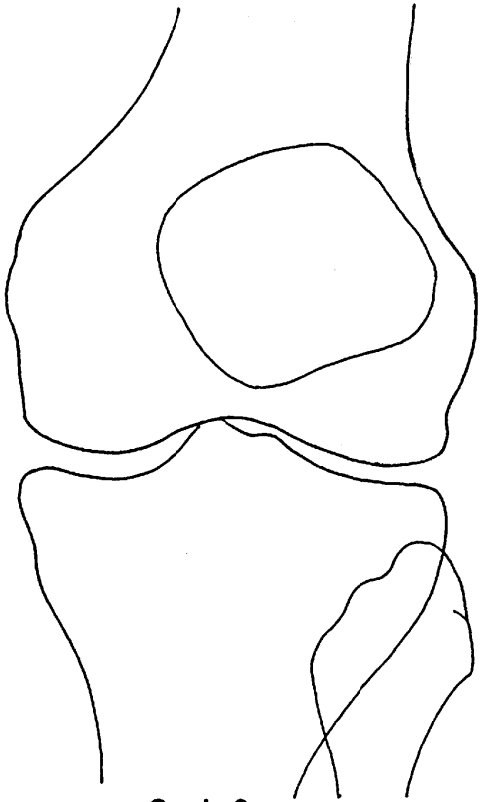


Grade 0

Lateral tibio-femoral joint space narrowing for women

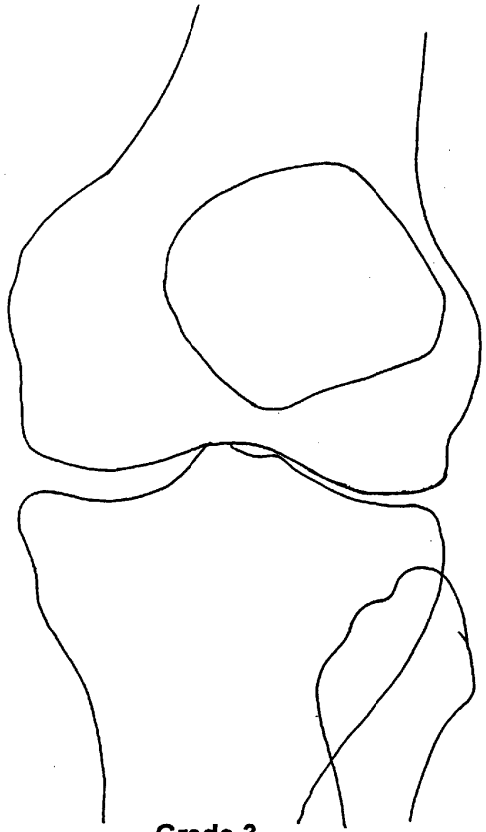


Grade 1

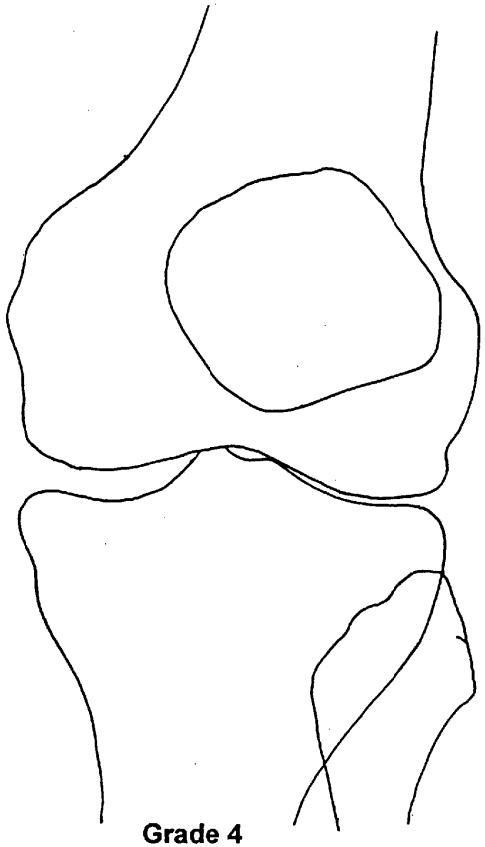


Grade 2

Lateral tibio-femoral joint space narrowing for women

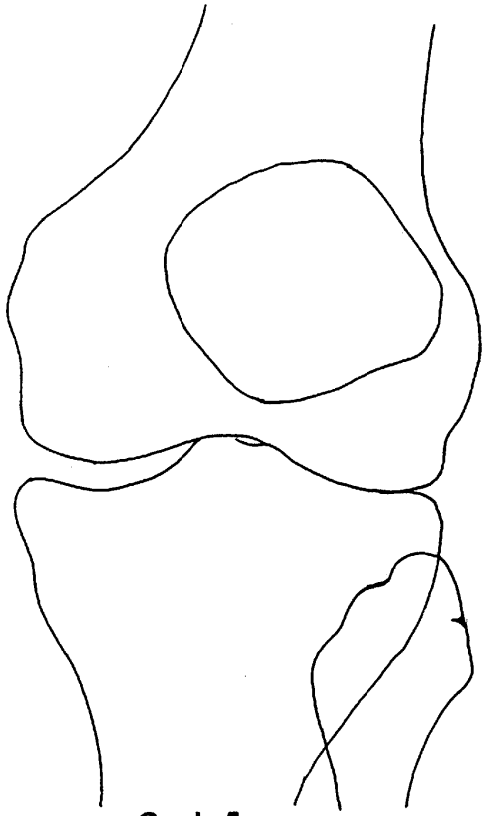


Grade 3



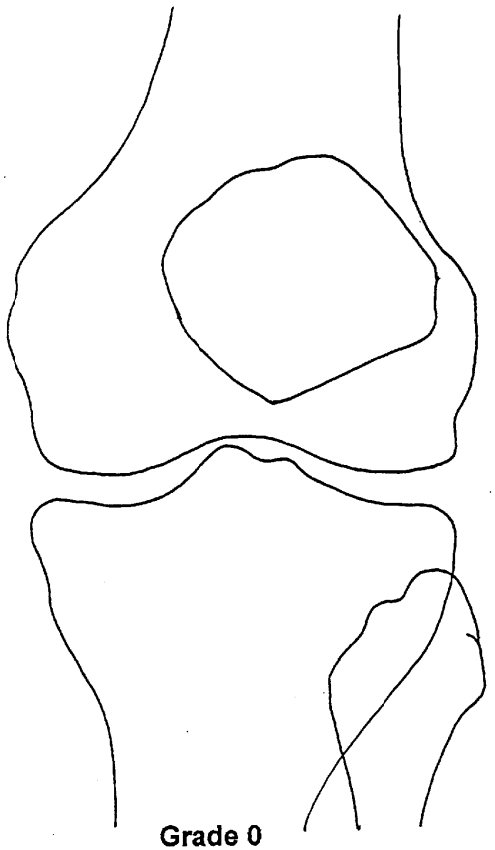
Grade 4

Lateral tibio-femoral joint space narrowing for women

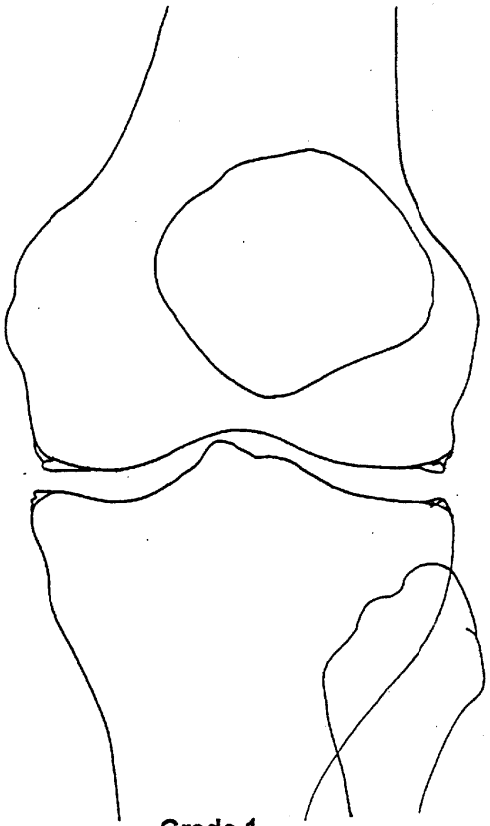


Grade 5

Osteophytes in all tibio-femoral sites

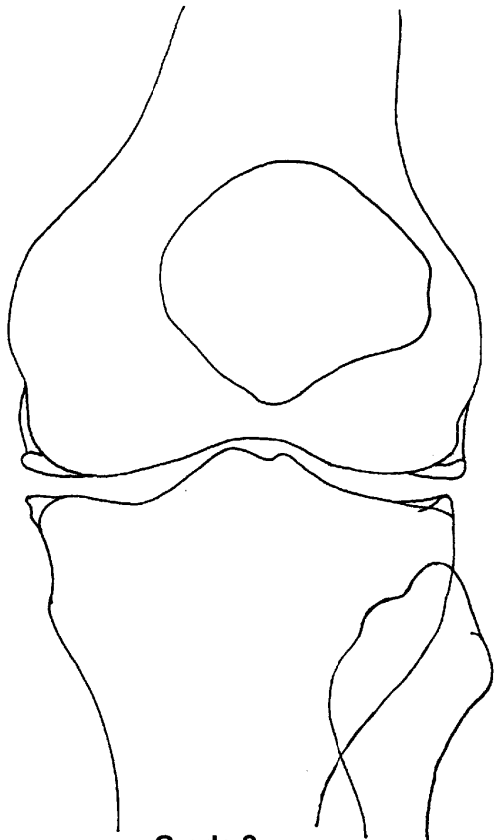


Grade 0

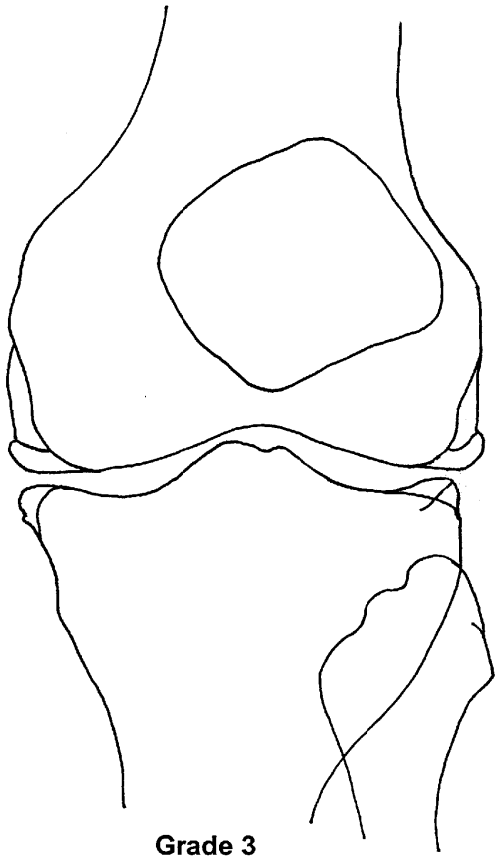


Grade 1

Osteophytes in tibio-femoral sites

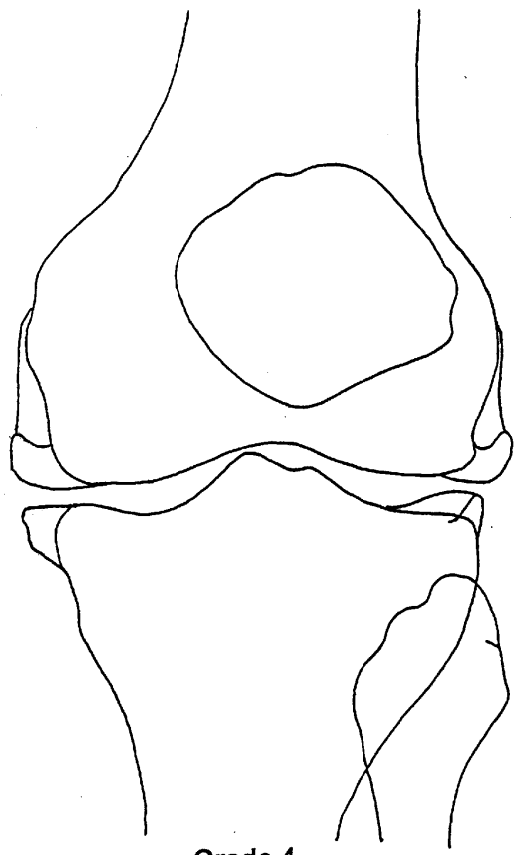


Grade 2

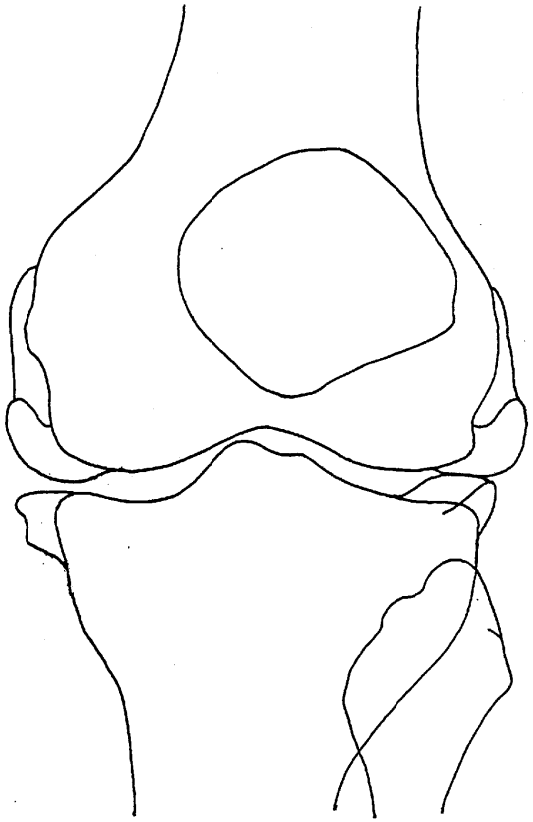


Grade 3

Osteophytes in all tibio-femoral sites

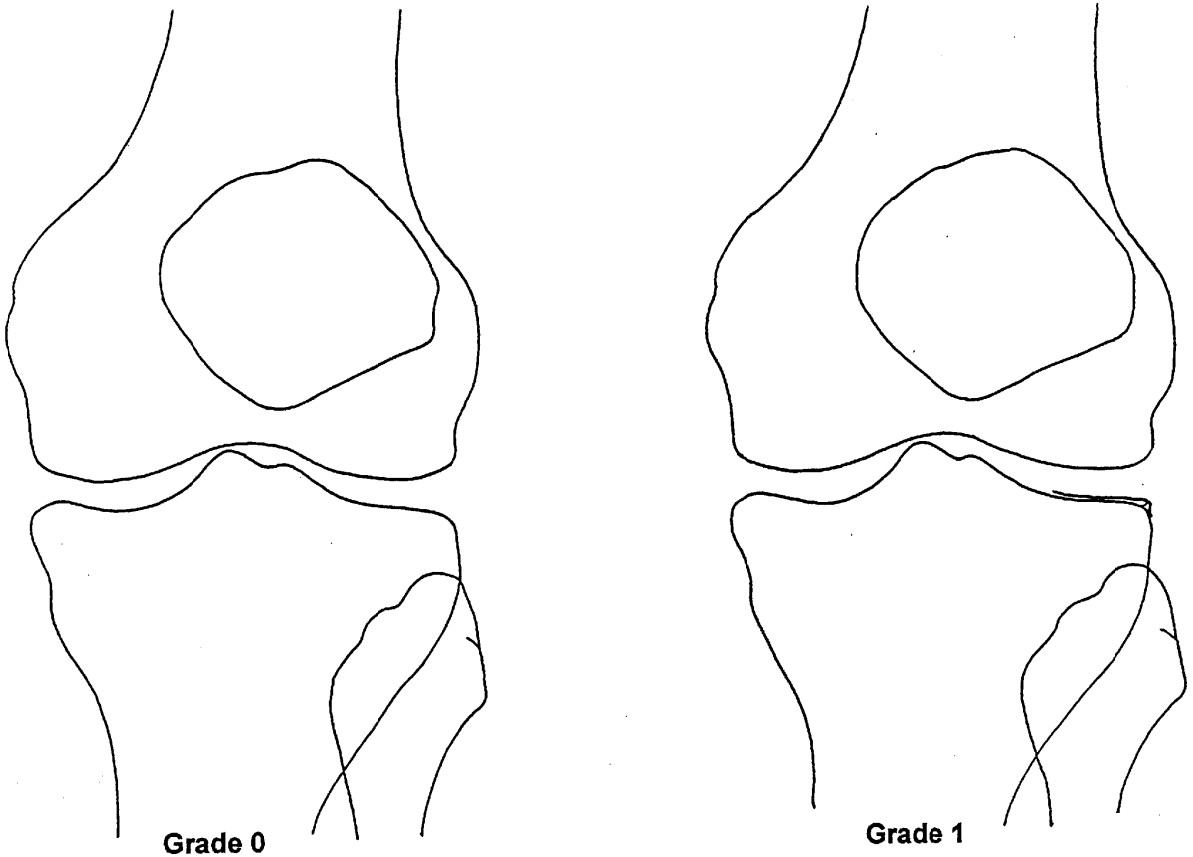


Grade 4



Grade 5

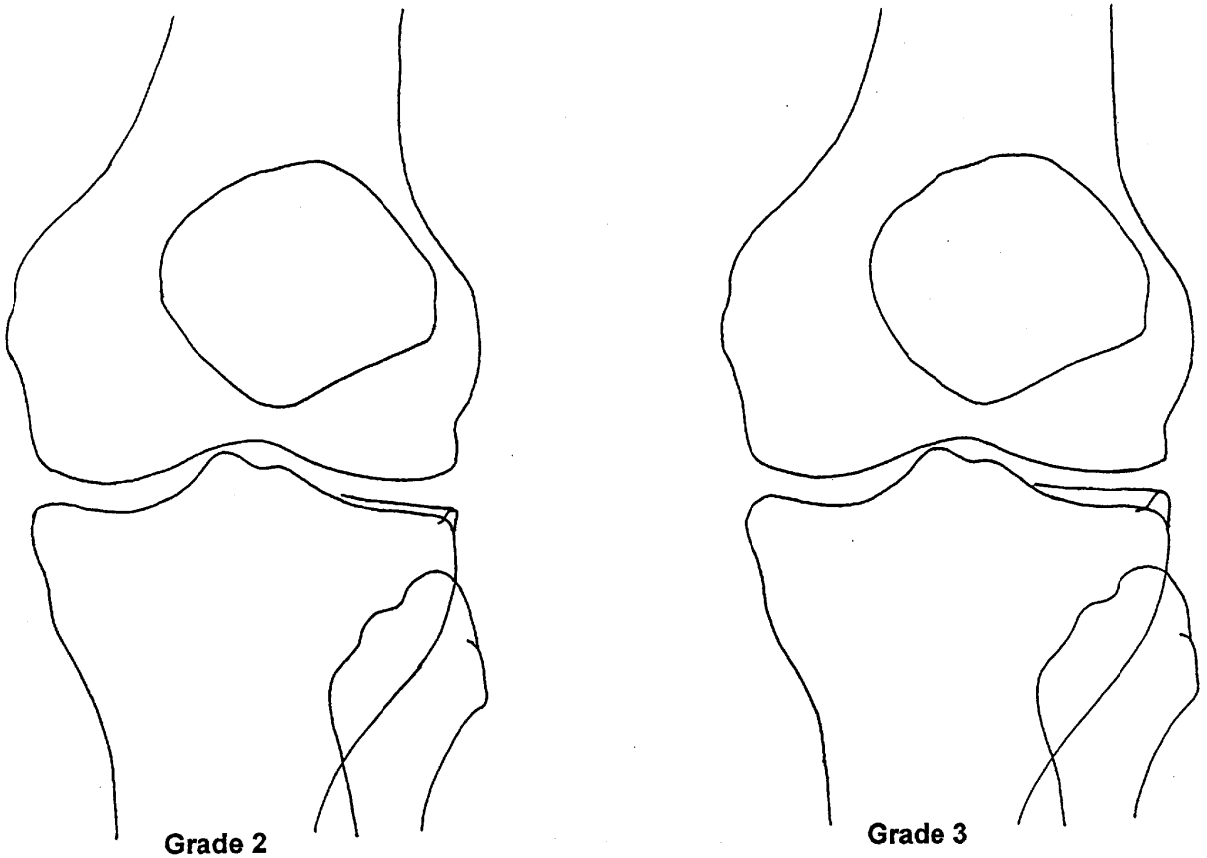
Lateral tibial osteophyte (optional shape)



Grade 0

Grade 1

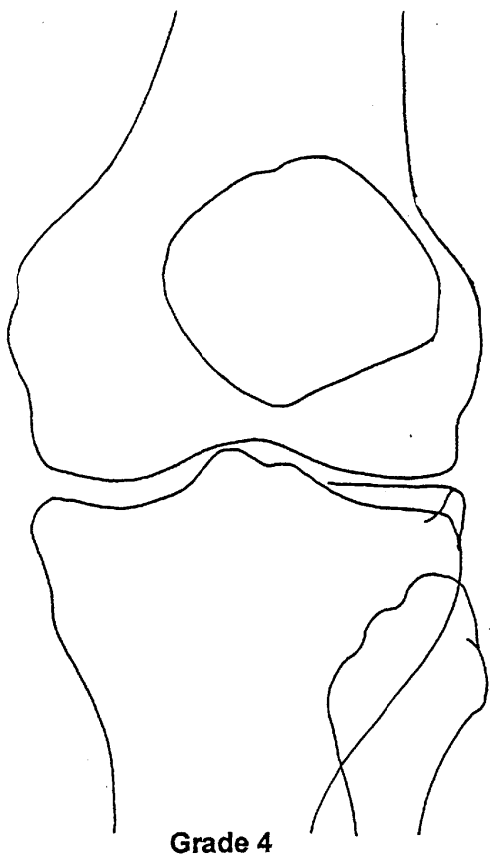
Lateral tibial osteophyte (optional shape)



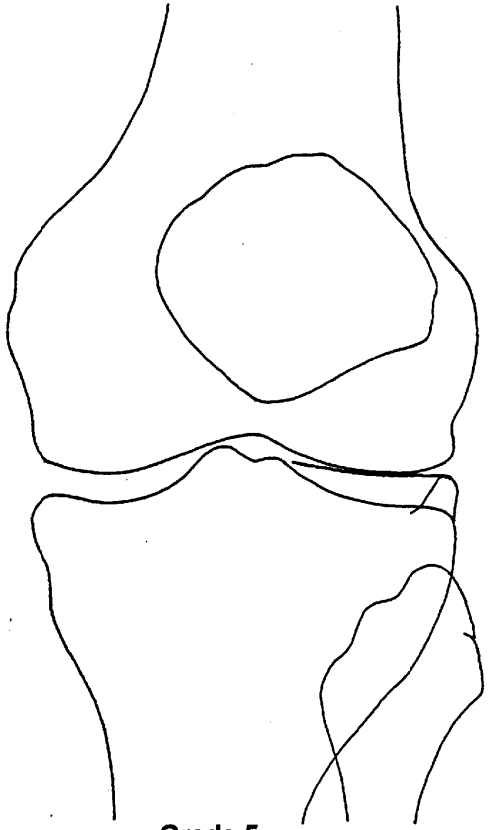
Grade 2

Grade 3

Lateral tibial osteophyte (optional shape)



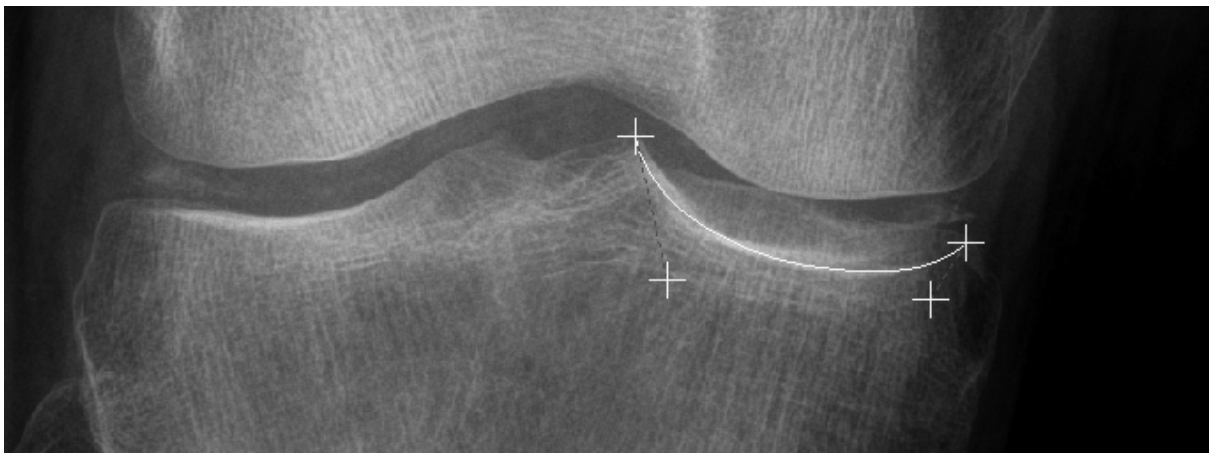
Grade 4

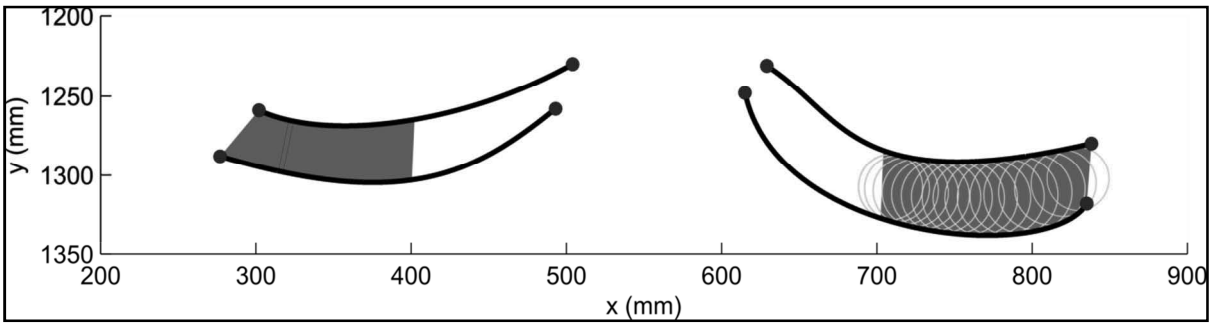


Grade 5

KneeMorf Manual

(Version 1.1)





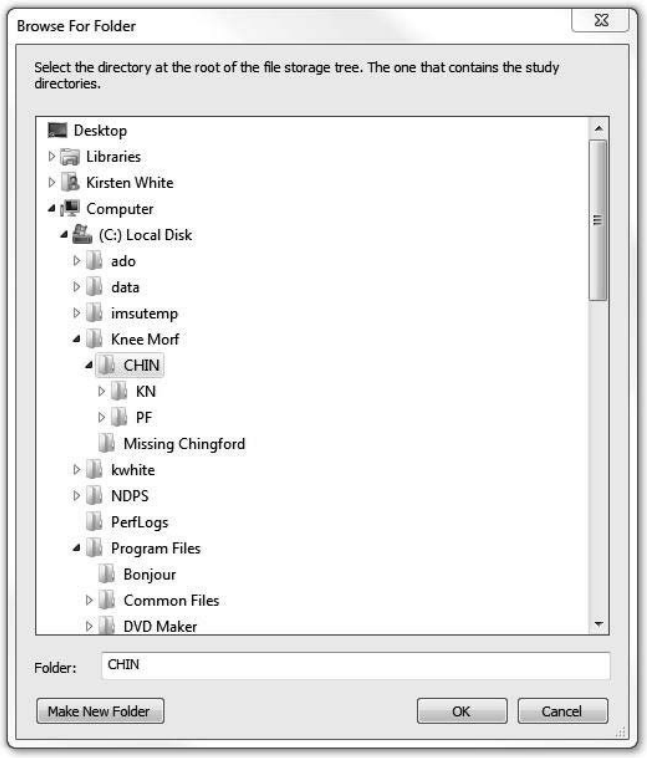
Quick Start Guide and Point Placement

Kirsten Leyland and Dr. David Hunter

University of Oxford

Quick Start Guide

Browsing for the Image Folder



Browse to the folder which is the ***first*** of the folder tree containing x-ray images

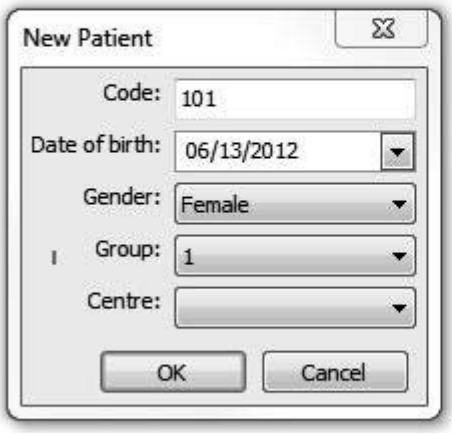
Select the folder labelled Knee Morf and click OK

i.e. Knee Morf - CHIN – KN – V020 – (xrays)

Individual x-rays will be labelled similarly:

CHINKN000020V020B.DICOM

Selecting a New Patient



Code – Patient ID

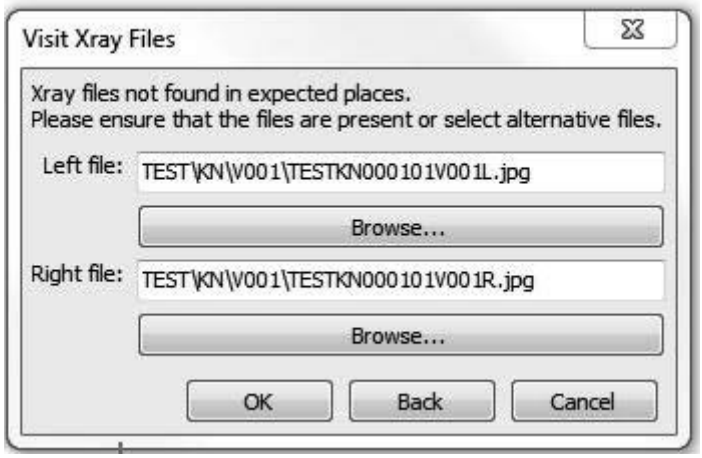
DOB – Date in American Format (MM/DD/YEAR)

Gender – Female/Male

Group – Selected when setting up the study

Centre – Not a required field

Manually Selecting X-rays

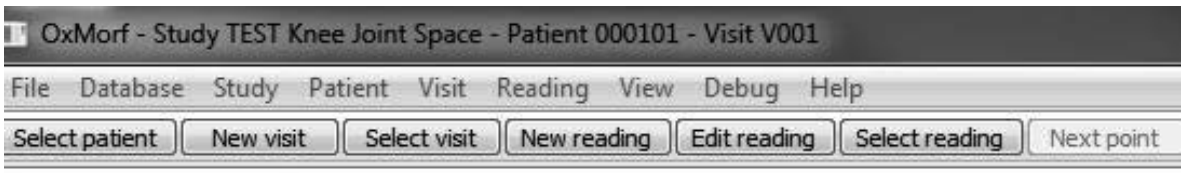


This box will appear when KneeMorf cannot find an x-ray that matches the *study name, joint, visit date and patient number*

It will prompt you to manually browse for the file

N.B. This will also appear when there is more than one file for the subject (normally labelled 1 of 2, etc.)

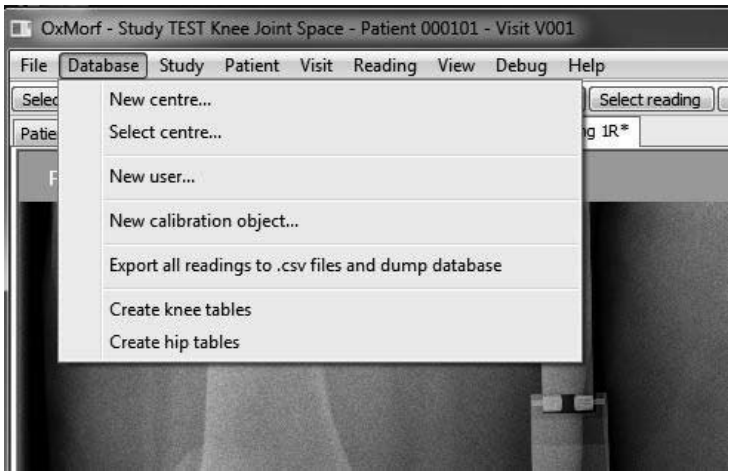
Selecting a New X-ray Reading

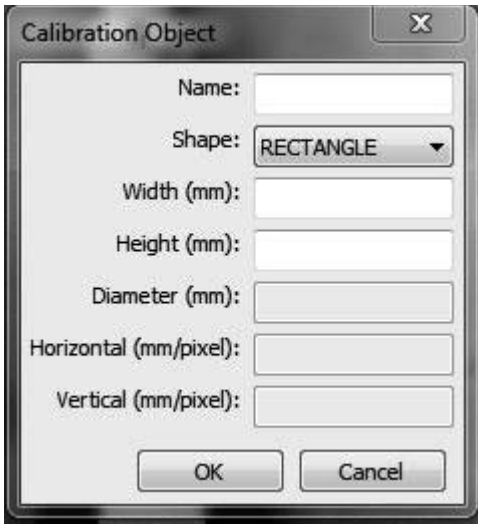


Number/Side: Options will appear for unread knees. For a new x-ray, choices will be 1Both, 1Right, 1Left. If the only the right has been completed by a reader, choices will be 2Right, 1Left.

Date, **User** and **Centre**: should be set and do not have to be adjusted

Creating Calibration Object for Digitised Plain Film X-rays





For Digitised Plain Film X-rays (i.e. any Chingford x-rays other than year 20), a calibration object must be created

Go to **Database**, **New Calibration Object**

Enter a **Name** specific to the Year/type of x-ray (e.g. Chingford Year 1)

Select a **Shape: Fixed** for entering pixel size

Enter **Horizontal** and **Vertical** pixel size in mm as originally given with the x-rays (0.04233333 for digitised Chingford)

Once created, the new calibration object will appear in the dropdown menu of the 'Select calibration

Selecting Calibration Object for DICOM (Digital) x-rays

For Digital x-rays (Chingford year 20), use the option "calibrate from DICOM tag: "Imager Pixel Spacing", which will automatically be present in the calibration object popup box

Keyboard Shortcuts

Display Options

Option	***Command***
Zoom in	Ctrl + left drag
Zoom out	Ctrl +A
Adjust contrast	C +left drag (horizontal)
Adjust brightness	C + left drag (vertical)
Scroll around image	Right drag
Toggle point visibility	P
Toggle point label visibility	L
Toggle graphic line visibility	G

General Point Placement

Option	***Command***
Create current point (if doesn't already exist)	Left click
Move current point	Left drag
Set point and move on to next one	1. Press <enter> OR 2. Press 'Next point' on button bar
Return to previous point	1. Ctrl + Z OR 2. Press 'Previous point' on button bar

Caliper Placement

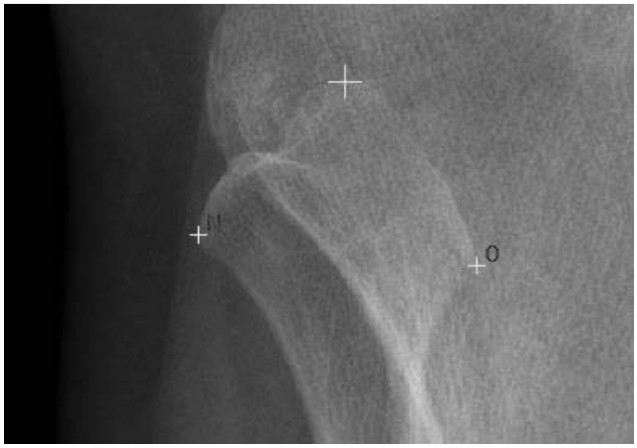
Option	***Command***
Adjust caliper width	Alt + left drag
Rotate caliper (only works if rotation is enabled for the current point, otherwise caliper will turn red)	Alt + right drag

Bezier Curve Placement

Option	***Command***
Move curve control point	Left drag (moves closest point)
Move entire curve	Alt + right drag
Add additional point	+/= key

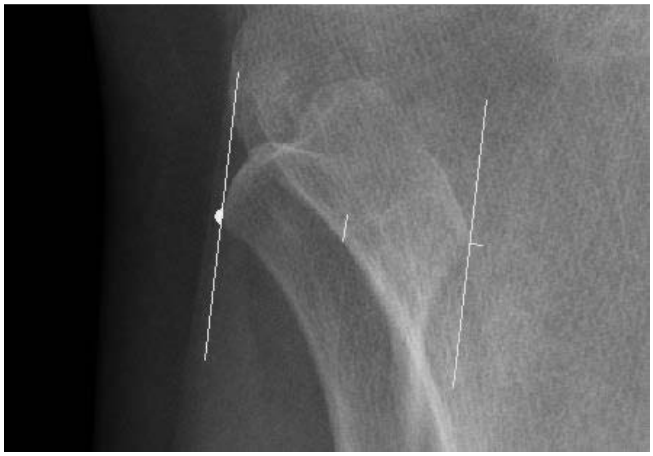
Joint Space Point Placement

Point 1: Top of fibular head (point)



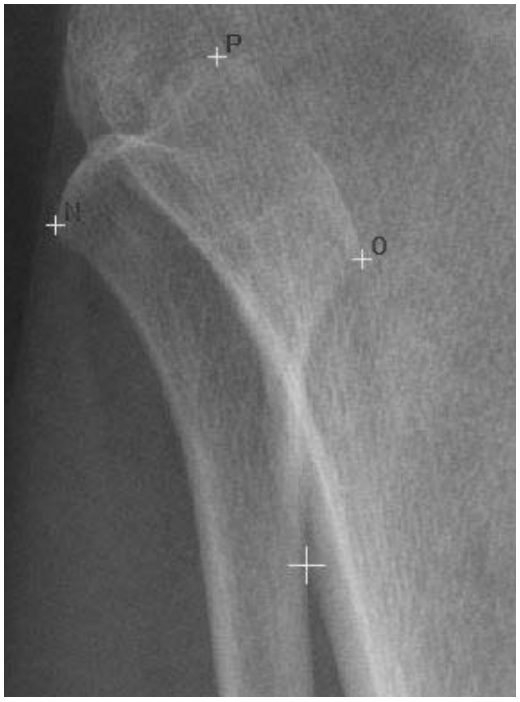
- Click to place on top most part of fibular head
- This point may be hidden behind the tibia (but still visible)

Point 2 and 3: Lateral and medial edge of fibular head (caliper)



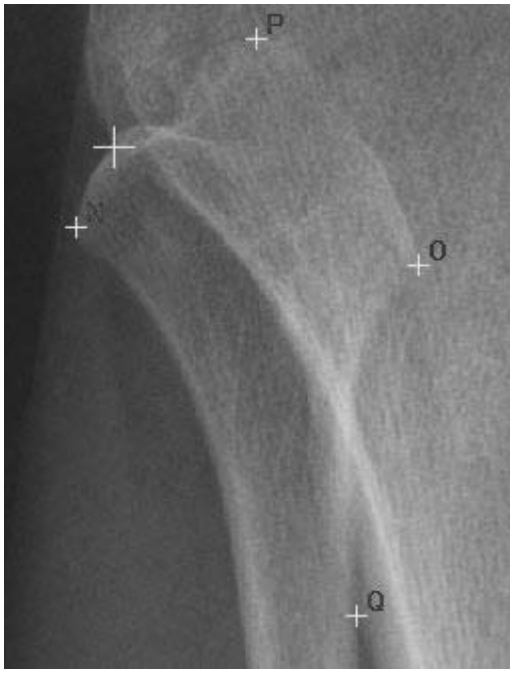
- Doesn't matter which side is medial or lateral
- Orient the calipers so that the centre of each line is on the widest part of the fibular head (***Alt +Right Mouse***)
- Press ***return*** to set both points

Point 4: Distal tibia-fibular overlap (point)



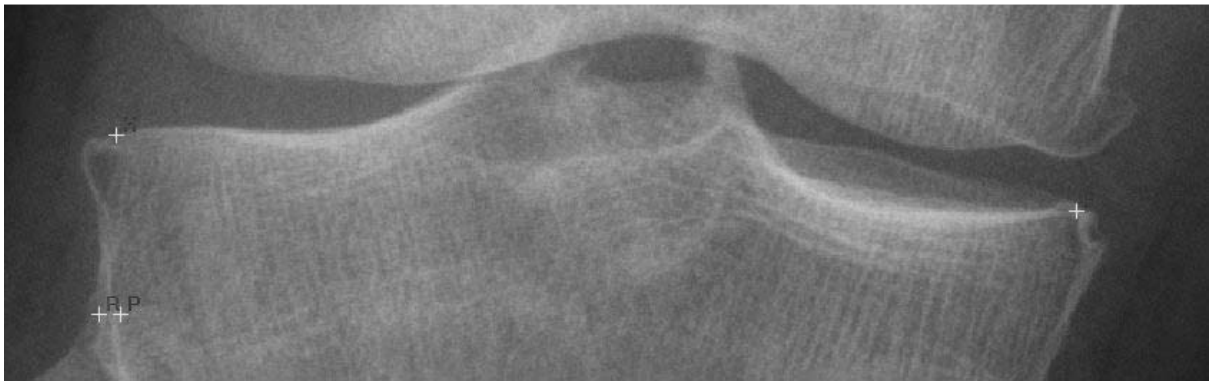
- Place point where the fibula and tibia meet, furthest from the top of the fibula (distally)
- This point may vary from being along the shafts or up near the middle of the fibular head

Point 5: Proximal tibia-fibular overlap (point)



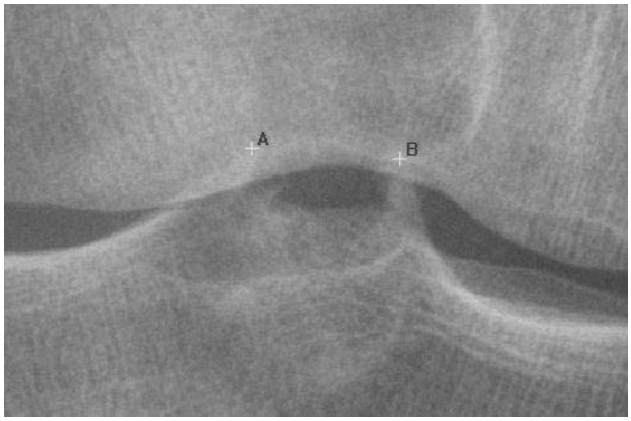
- Place the point at the uppermost point where the fibula meets the tibia
- This point is normally toward the top of the fibular head

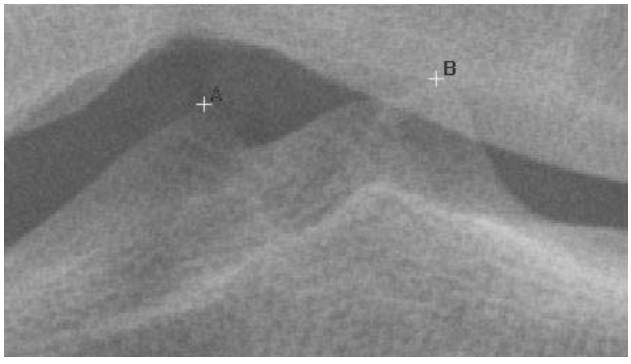
Point 6 and 7: Medial and lateral tibial plateau edge (point)



- Place point at the edge of the tibial plateau, in the middle of the sclerotic line (indicating the plateau floor, if present), NOT including marginal osteophytes
- Keep in mind that these points are used later to define the tibial bezier curve and the top of the tibial metaphysis

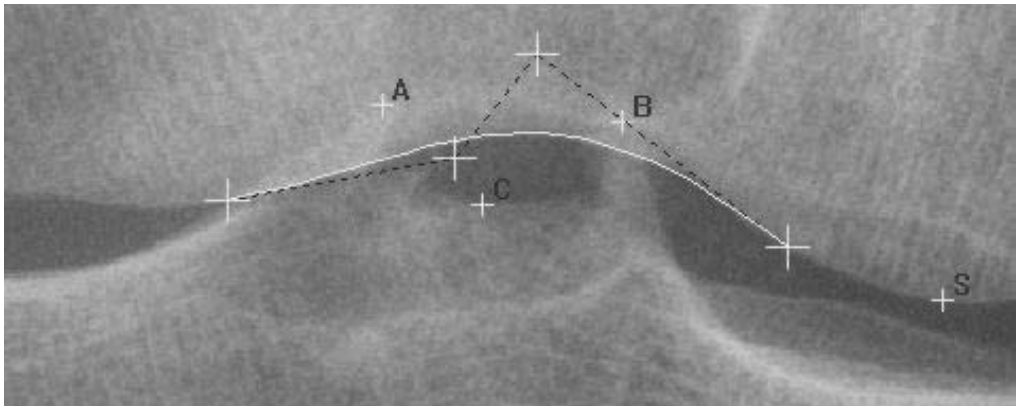
Point 8 and 9: Medial and lateral tibial spine (point)





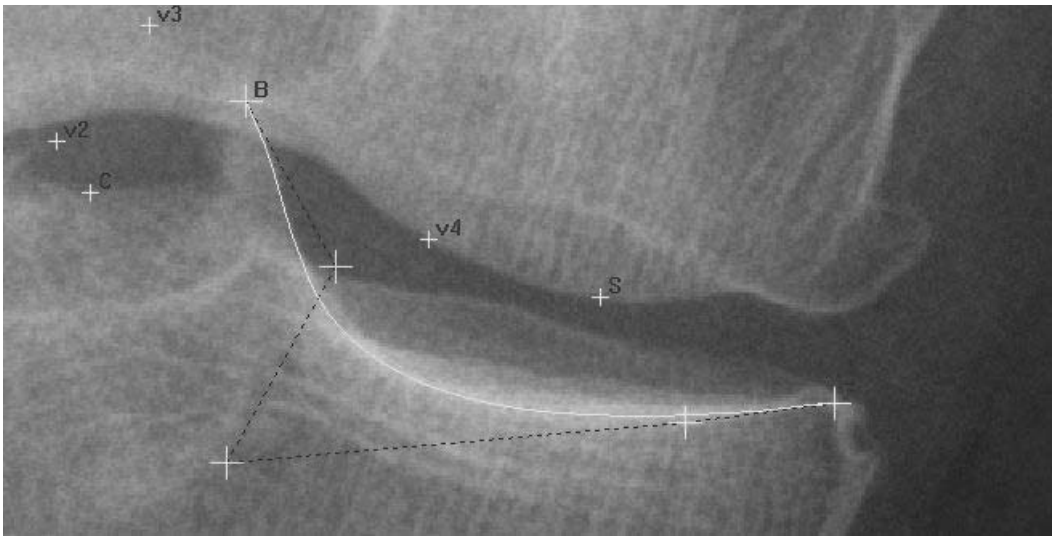
- Click the tip of the medial and lateral spines
- Some spines may appear taller or split into more than one tip – click on the tallest peak of this type of spine

Points 10-15: Curve of the femoral intercondylar notch (bezier curve)



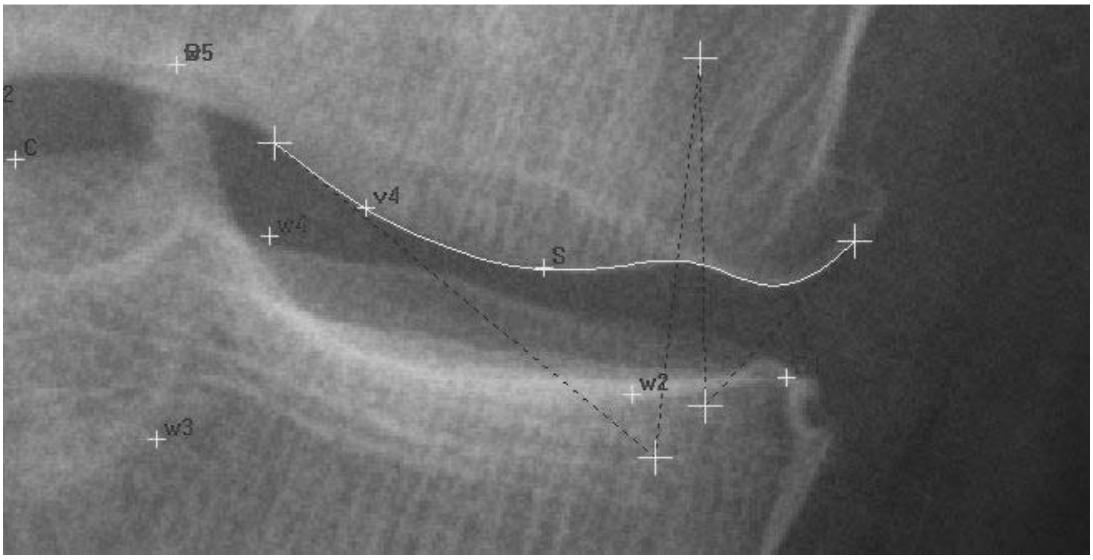
- Click on edge of lateral condyle where it starts to flatten out to place the first point
- Drag the point on the opposite side of the curve to the same spot on the medial condyle
- Manipulate the two intermediate points to reshape the curve using the ***Left Mouse*** button placed near the point you want to move
- Click the ***plus sign key (+)*** to add extra points if the curve needs additional inflexion points

Points 16-21: Curve of medial tibial plateau (bezier curve)



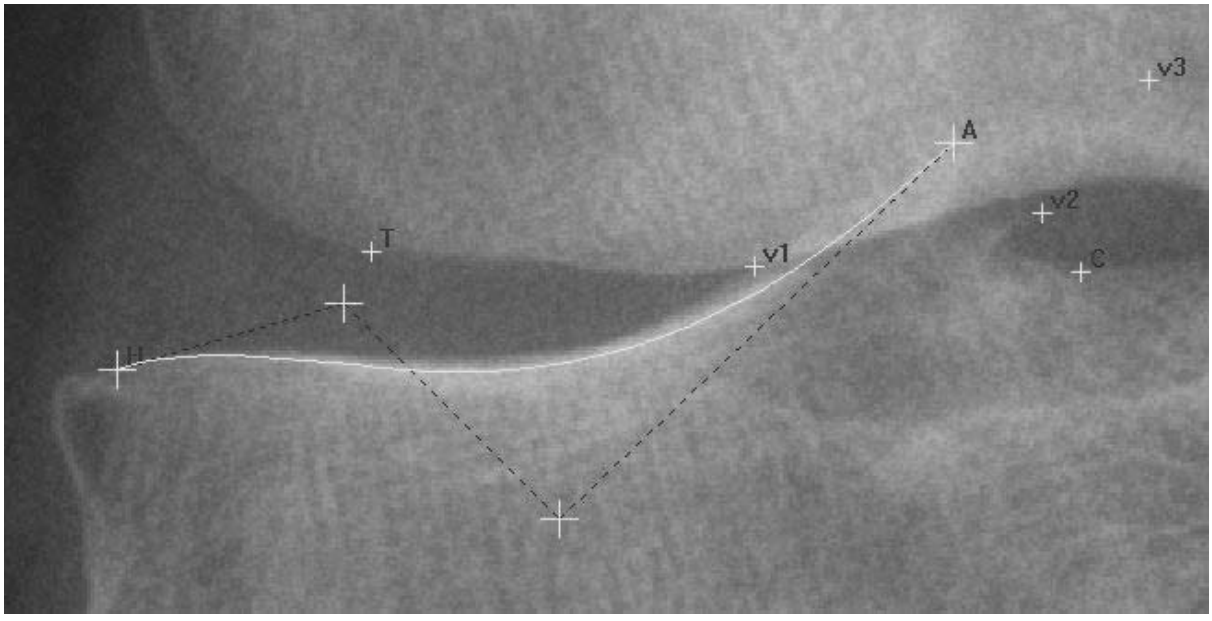
- End points of the curve will automatically be set where the tibial spine point and the tibial plateau edge points were placed
- Manipulate the two intermediate points to reshape the curve using the ***Left Mouse*** button placed near the point you want to move
- Click the ***plus sign key (+)*** to add extra points if the curve needs additional inflexion points
- ***Follow the curve along the middle of the lowest brightest (sclerotic) line***

Points 22-27: Curve of medial femoral condyle (bezier curve)



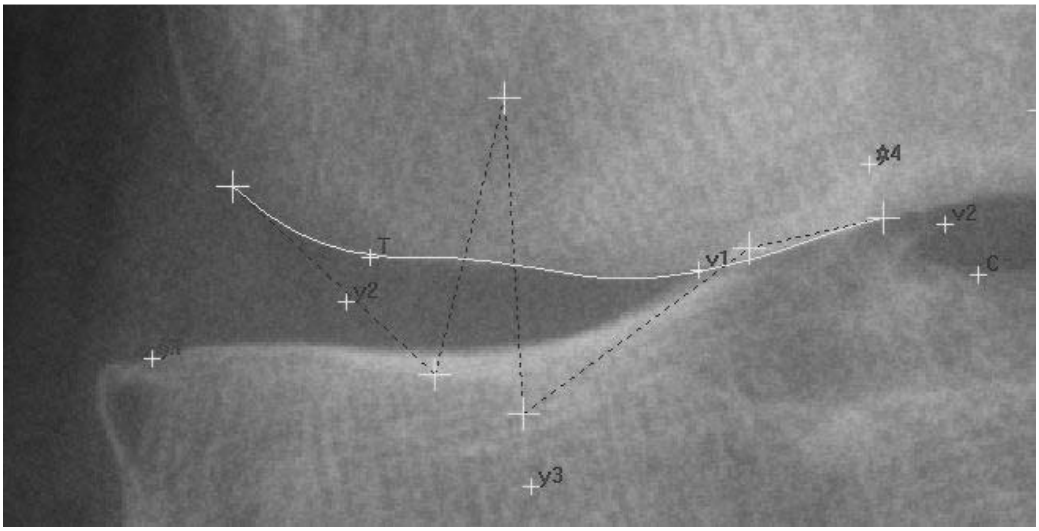
- Place the curve to capture a bit of the outer curve of the joint as well as some of the femoral notch
- DO include osteophytes if present
- Include 'dips' or small curves if well-defined

Points 28-33: Curve of lateral tibial plateau (bezier curve)



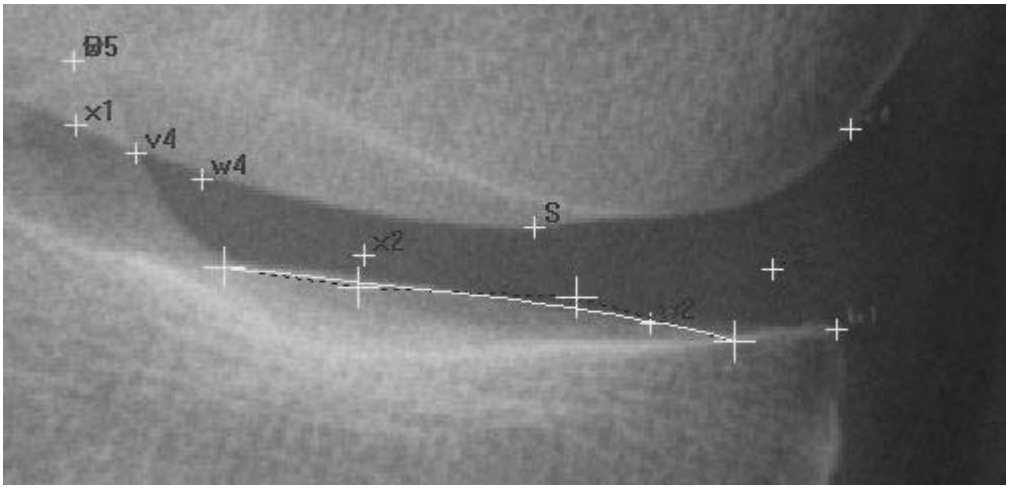
- Outermost points will automatically be placed based upon previous points
- Curve should follow the middle of the brightest (sclerotic) line or edge of the uppermost part of the plateau (whichever is present)

Points 34-39: Curve of lateral femoral condyle (bezier curve)



- Place the curve to capture a bit of the outer curve of the joint and as much of the inner part of the femoral notch as possible
- DO include osteophytes
- Place curve on the cortical edge of the condyle

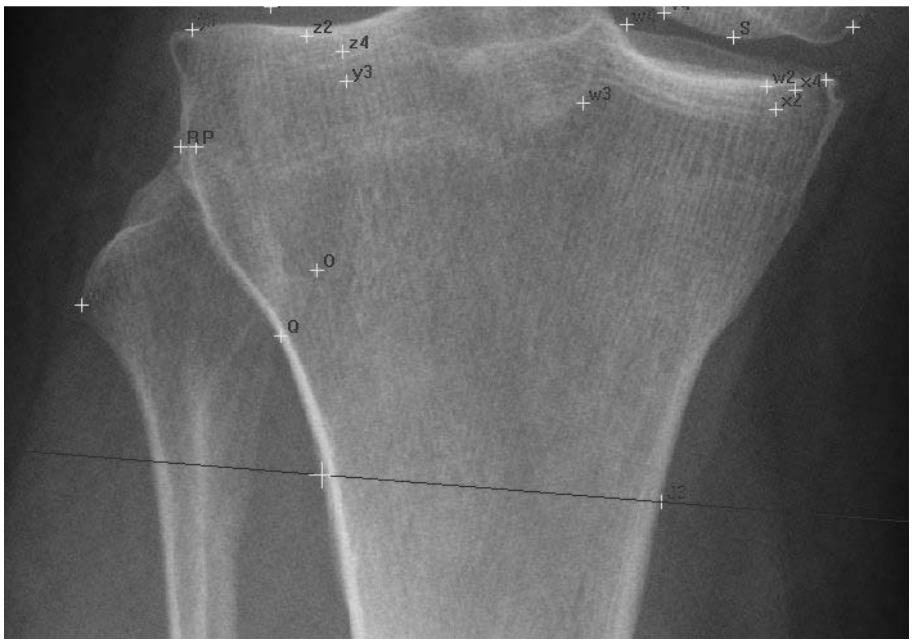
Point 40-45 medial tibial rim (and/or point 46-51 lateral rim)



- Place end points where the rim meets the tibial plateau curve
- Use standard keyboard shortcuts (see X) to set curve

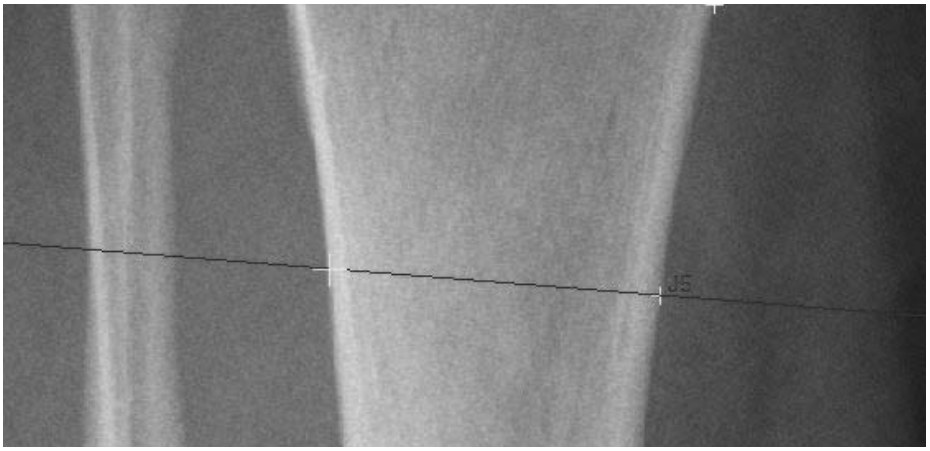
Alignment Point Placement

Points 52 and 53: Medial and distal edge of tibial shaft – 70mm (constrained points)



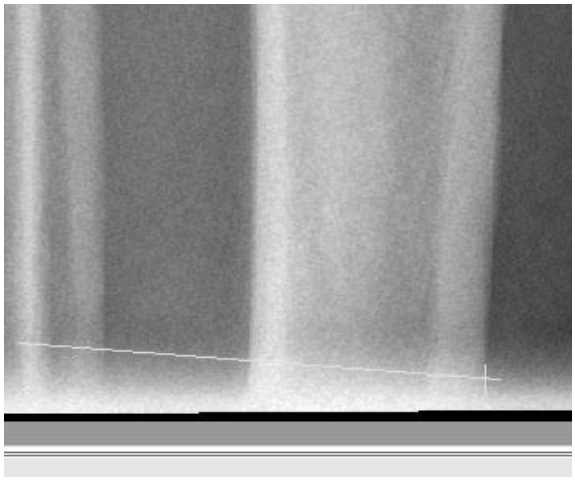
- Place on the outermost edge of the bone NOT on the middle of the brightest line
- Check to make sure the medial and lateral points are on the correct sides

Points 54 and 55: Medial and lateral edge of tibial shaft – 100mm (constrained points)



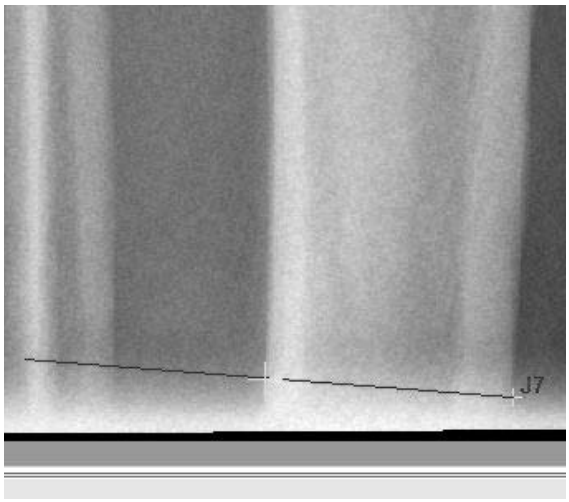
- Place on the outermost part of the bone NOT on the middle of the brightest line
- Check to make sure the medial and lateral points are on the correct sides as per the point description
- Click Skip Point button on top of page to not place points if the constrained line is off the edge of the x-ray

Point 56: Most distal edge of medial tibial shaft (point with levelling guide)



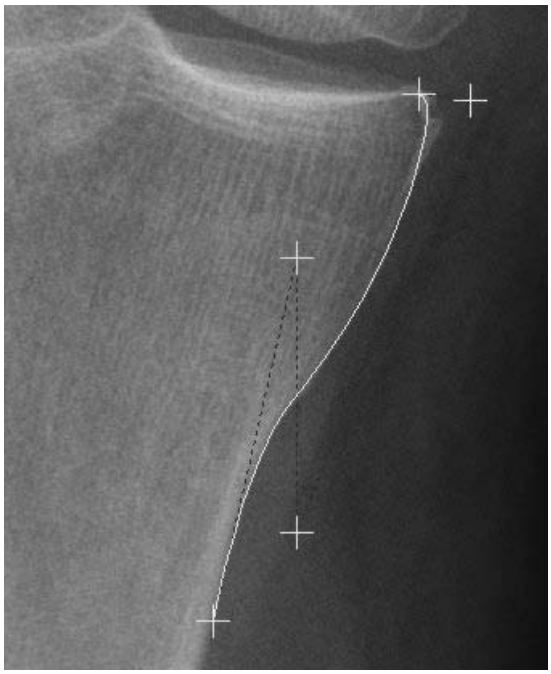
- Use Check box in right hand panel to select Tib Shaft Override =1 to set this point instead of the 100mm points
- Place point on the edge of the medial shaft as far down as possible (where bone edges are clearly visible)
- Make sure the guideline can be clearly seen on both sides of the shaft and is not obscured

Point 57: Most distal edge of lateral tibial shaft (constrained point)



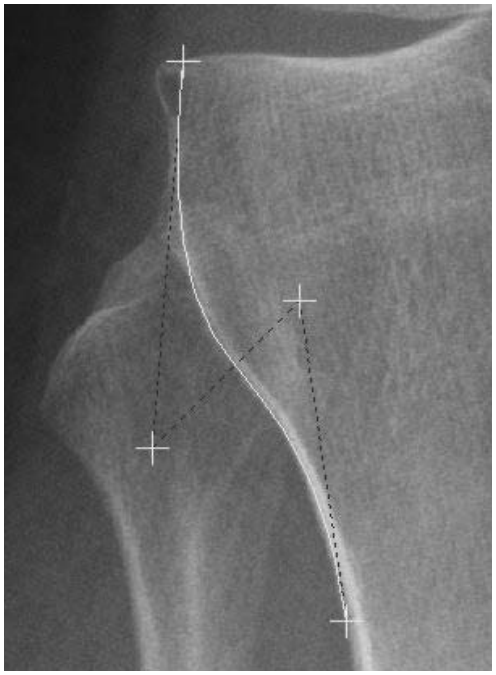
- Place point on the lateral edge of the tibial shaft along the line
- Do not place on the brightest line, but at the very edge of the bone

Points 58-63: Curve of medial metaphysis edge (bezier curve)



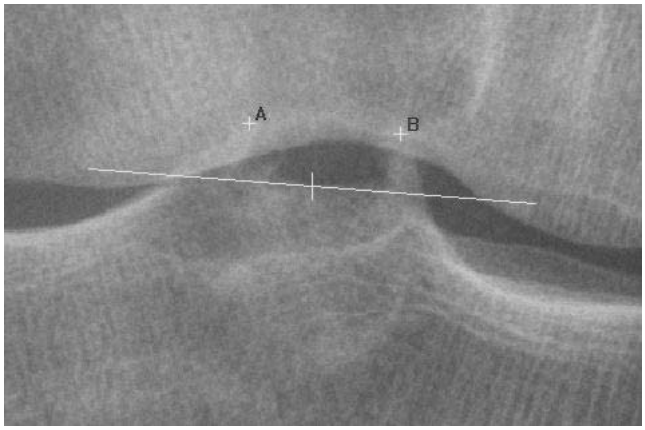
- The two end points will automatically set
- Follow the outline of the edge of the bone NOT including osteophytes
- Don’t worry about very small indents
- Stay on the outer edge of the bone (not in the middle of bright lines)

Points 64-69: Curve of lateral metaphysis edge (bezier curve)



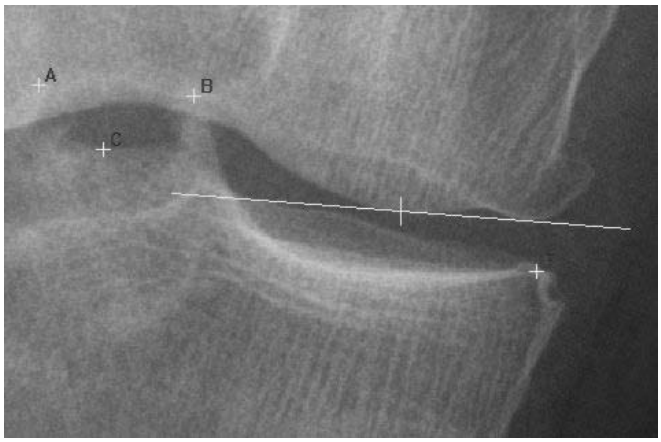
- The two end points will automatically set
- Follow the outline of the edge of the bone NOT including osteophytes
- Don’t worry about very small indents
- Stay on the edge of the bone (not in the middle of bright lines)

Point 70: Tibial spine bony surface (floor between spines) (point with levelling guide)



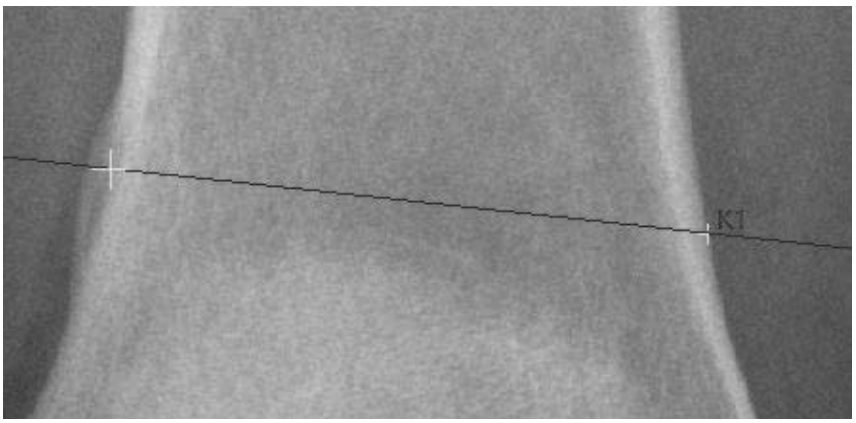
- Place on the lowest surface of the bone between tibial spines
- This point may not be perfectly centred between the spines

Point 71 and 72: Most inferior points of medial and lateral condyle (point with levelling guide)



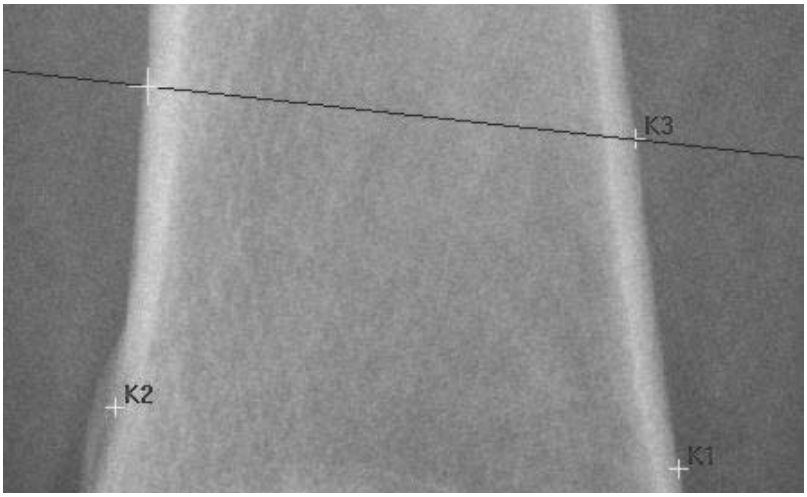
- Use the line as a guide to show the lowest (most distal) edge of the femoral condyle
- If the guide line overlaps bone on either side of the centre point, then the point may not be in the optimal location

Point 73 and 74: Medial and lateral edge of femoral shaft – 70mm (constrained point)



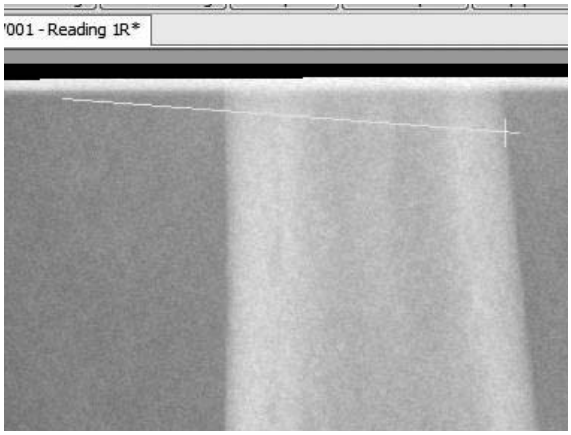
- Place on the outermost part of the bone NOT on the brightest line
- Check to make sure the medial and lateral points are on the correct sides
- Do not include bulges from the back of the femur

Point 75 and 76: Medial and lateral edge of femoral shaft – 100mm (constrained point)



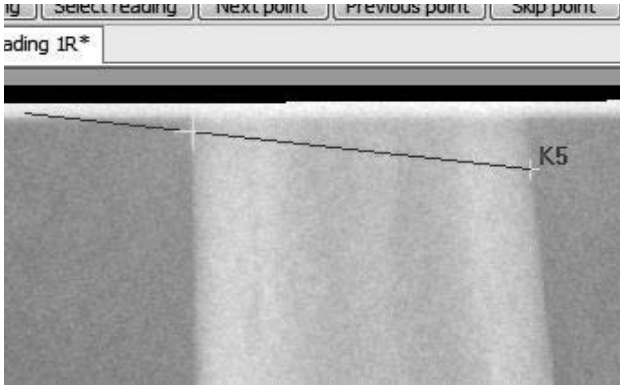
- Place on the outermost part of the bone NOT on the middle of the brightest line
- Check to make sure the medial and lateral points are on the correct sides

Point 77: Most distal medial edge of femoral shaft (point with levelling guide)



- Use Check box in right hand panel to select Fem Shaft Override =1 to set this point instead of the 100mm points
- Place point on the edge of the medial shaft as far up as possible (where bone edges are clearly visible)
- Make sure the guideline can be clearly seen on both sides of the shaft and is not obscured

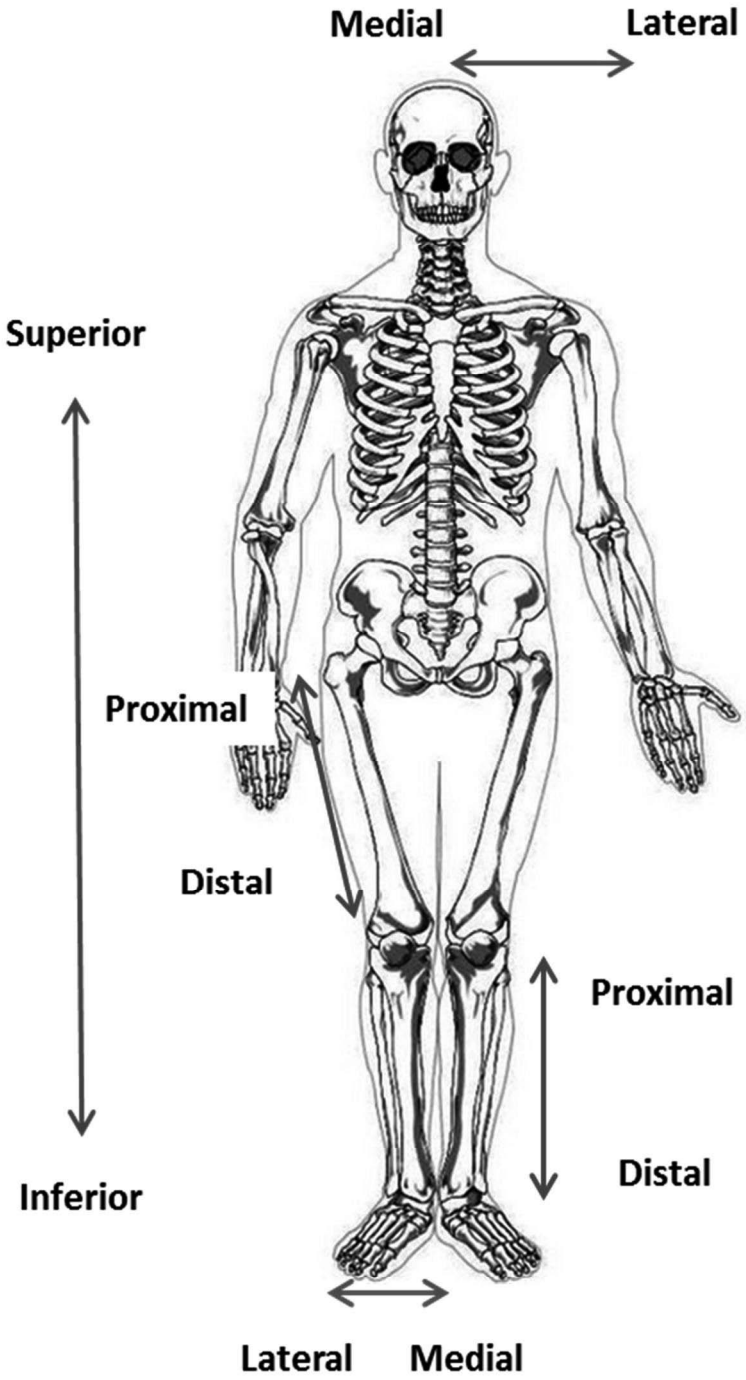
Point 78: Most distal lateral edge of femoral shaft (constrained point)



- Place point on the lateral edge of the femoral shaft along the line
- Do not place on the brightest line, but at the very edge

Reference

Anatomical Orientation



Bones/Anatomical Features