UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

MRC Lifecourse Epidemiology Unit

Human Development and Health

Volume 1 of 1

Definition of trajectories of joint space narrowing in knee osteoarthritis in the presence of measurement error

by

Camille Michelle Parsons

Thesis for the degree of Doctor of Philosophy

June 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

Epidemiology

Thesis for the degree of Doctor of Philosophy

Definition of trajectories of joint space narrowing in knee osteoarthritis in the presence of measurement error

Camille M Parsons

Osteoarthritis (OA) is the most prevalent joint disease in older adults, with the knee being the most commonly affected joint. The disease progression is slow, often over decades. During progression the breakdown of cartilage occurs within the affected joint, causing the joint space to narrow such that ultimately the joint will fail. This process leads to great pain and disability on an individual level, and an ever increasing economic burden to health services. The gold standard measurement currently used within epidemiological studies to monitor the progression of the disease is joint space narrowing. However, changes in joint space measurements over time are small and sensitive to measurement error. This makes identification of real deterioration difficult, and maybe masking the identification of factors associated with progression. Therefore the research aims of this thesis are to develop statistical methods that identify change in continuous knee joint space measurements over time within individuals diagnosed with knee osteoarthritis.

Two datasets were used in this thesis; the Strontium ranelate Efficacy in Knee Osteoarthritis trial (SEKOIA), a 3-year international, multicentre, double-blind, placebo-controlled phase 3 trial, and the Osteoarthritis Initiative (OAI), an open-access longitudinal dataset following study participants in the United States of America. Three different statistical methods were applied to both datasets; the reliable change (RC) index, frequentist linear mixed effect (LME) modelling and Bayesian hierarchical modelling, to determine how each method accounted for measurement error within the knee joint space width measurements.

When compared with crude differences in knee joint space width, implementation of the novel approach of the RC index dramatically reduced the proportion of study participants in both datasets that were identified as having statistically reliable change after accounting for measurement error. The RC index allows for calculation of a magnitude of change threshold that gives a figure above which it can be said reliable change had occurred; the threshold was found to be approximately 1mm in a year in both SEKOIA and the OAI. Using the frequentist techniques of LME modelling, on average, joint space width decreased by 0.14mm per year in SEKOIA and

0.08 in the OAI, and results from Bayesian hierarchical modelling indicated a similar pattern. Each statistical method used within this project handles measurement error in different ways. The RC index only considers two time points at a time, but accounts for measurement error by using the variability of the measurement within the calculation. Both the frequentist linear mixed effect and Bayesian modelling account for measurement error in a similar way, by using all available time points from the total population within the modelling process to 'smooth' individual trajectories of change.

These methods could prove valuable in monitoring of disease progression and in identification of risk factors for the progression of knee OA which have not yet been discovered. The RC index threshold of change could be directly applied within a patient population to aid clinicians in monitoring progression of knee OA. In a research environment, the simpler LME modelling technique could be implemented to 'smooth' data for measurement error and obtain individual estimates of change. The more complex Bayesian methodology can be used to investigate phenotypes for disease progression. Additionally, application of these statistical methods in other clinical areas may aid in monitoring of other conditions.

Table of Contents

List of tables

List of figures

List of Equations

DECLARATION OF AUTHORSHIP

I, Camille Michelle Parsons, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Definition of trajectories of joint space narrowing in knee Osteoarthritis in the presence of measurement error

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed: .....

Date:.....

Project Outputs

Awards

- I was awarded a University of Southampton, 3 year PhD scholarship of £21,000 in 2012

- I was awarded a Worldwide Universities Network Research Mobility Programme award for £3,000 in 2015

- I was awarded a bursary from AGE UK to attend the Longitudinal Studies: Maximising their Value for Ageing Research meeting in 2015

Findings presented in this work

- Chapter 5: **Parsons C**, Judge A, Leyland K, Bruyère O, Petit Dop F, Chapurlat R, Reginster J-Y, Edwards MH, Dennison EM, Cooper C, Inskip H, And the SEKOIA Study Group. Novel approach to estimate Osteoarthritis progression – use of the reliable change index in the evaluation of joint space loss. *Under peer review*

Related publications to which the candidate has contributed

- **Parsons C**, Fuggle NR, Edwards MH, Goulston L, Litwic AE, Jagannath D, van der Pas S, Cooper C, Dennison EM and the EPOSA research group. Concordance between clinical and radiographic evaluation of knee osteoarthritis. *Aging clinical and experimental research* (2017): 1-9.

- **Parsons C**, Eymard F, Edwards MH, Petit-Dop F, Reginster J-Y, Bruyère O, Chevalier X, Cooper C, Richette P. Statin use and knee osteoarthritis progression: Results from a post-hoc analysis of the strontium ranelate efficacy in knee osteoarthritis trial. *Joint Bone Spine* (2017)

- Edwards MH, **Parsons C**[1], Bruyère O, Dop FP, Chapurlat R, Roemer FW, Guermazi A, Zaim S, Genant H, Reginster JY, Dennison EM. High Kellgren-Lawrence Grade and Bone Marrow Lesions Predict Worsening Rates of Radiographic Joint Space Narrowing; The SEKOIA Study. *The Journal of Rheumatology*. 2016 Jan 15:jrheum-150053.

- **Parsons C**, Clynes M, Syddall H, Jagannath D, Litwic A, van der Pas S, Cooper C, Dennison EM, Edwards MH. How well do radiographic, clinical and self-reported

[1] Joint first authorship

diagnoses of knee osteoarthritis agree? Findings from the Hertfordshire cohort study. *SpringerPlus*. 2015 Apr 15;4(1):177

- Edwards MH, van der Pas S, Denkinger MD, **Parsons C**, Jameson KA, Schaap L, Zambon S, Castell MV, Herbolsheimer F, Nasell H, Sanchez-Martinez M. Relationships between physical performance and knee and hip osteoarthritis: findings from the European Project on Osteoarthritis (EPOSA). *Age and ageing.* 2014 Jun 10;43(6):806-13.

Acknowledgements

I am very grateful to my supervisors, Professor Cyrus Cooper, Professor Hazel Inskip and Associate Professor Andy Judge for their support and guidance with this doctoral thesis. I am also very grateful to the Worldwide University Network who provided me with funding to complete a research visit to the University of Auckland, New Zealand. Without this funding and expert guidance from Professor Thomas Lumley and Associate Professor Renate Meyer I would not have been able to develop my Bayesian statistical skills. I would also like to thank my colleagues at the MRC Lifecourse Epidemiology Unit.

I am grateful to the study participants, and to the study teams of SEKOIA and the OAI. If I had not had access to such rich longitudinal data I would not have been able to learn from and complete this doctoral project.

A massive thank you to my friends, you have provided me encouragement throughout this doctorate journey, and whether it was tea, biscuits, advice, a listening eat, hugs or wine you were always there when required. Thank you to you all. A special mention must go to Suzie, your support and guidance through this journey have been invaluable. A huge thank you to my parents for their eternal support and encouragement, you taught me that I can achieve anything.

Finally I would like to thank Tim, not only have you been my own personal professional printing and binding service but you have provided me with endless support over the years. You are my rock, and without you I would not have the belief I could achieve such things.

Definitions and Abbreviations

ACR	American College Of Rheumatology
BGR	Brooks-Gelman & Rubin
BLUPs	Best Linear Unbiased Predictors
BMI	Body mass index
BUGS	Bayesian inference Using Gibbs Sampling
DMOADs	Disease-modifying osteoarthritis drugs
DP	Dirichlet process
FDA	Food and Drug Administration
GEEs	Generalized estimating equations
GNP	Gross National Product
GP	General Practitioner
JSN	Joint space narrowing
K&L	Kellgren and Lawrence
KOOS	Knee Outcomes in Osteoarthritis Survey
LME	Linear mixed effect
MCMC	Markov chain Monte Carlo
ML	Maximum likelihood
MRI	Magn`etic Resonance Imaging
OA	Osteoarthritis
OAI	Osteoarthritis Initiative

OARSI	Osteoarthritis Research Society International atlas criteria
OCT	Optical Coherence Tomographic
OR	Odds ratio
PSRF	Potential scale reduction factor
RC index	Reliable change index
SD	Standard deviation
SEKOIA	Stontium ranelate Efficacy in Knee Osteoarthritis trial
SEM	Standard error of the measurement
SF12	Medical Outcomes Study Short Form 12
SF-36	Short Form-36
SRM	Standardised response mean
US	Ultrasound
VAS	Visual Analogue Scale
VPC	Variance Portioning Coefficient
WHO	World Health Organisation
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index

Chapter 1: Introduction

1.1 Overall aim

To develop statistical methods to identify change in continuous knee joint space measurements over time within individuals diagnosed with or at risk of developing knee osteoarthritis (OA).

1.2 Overview of chapter

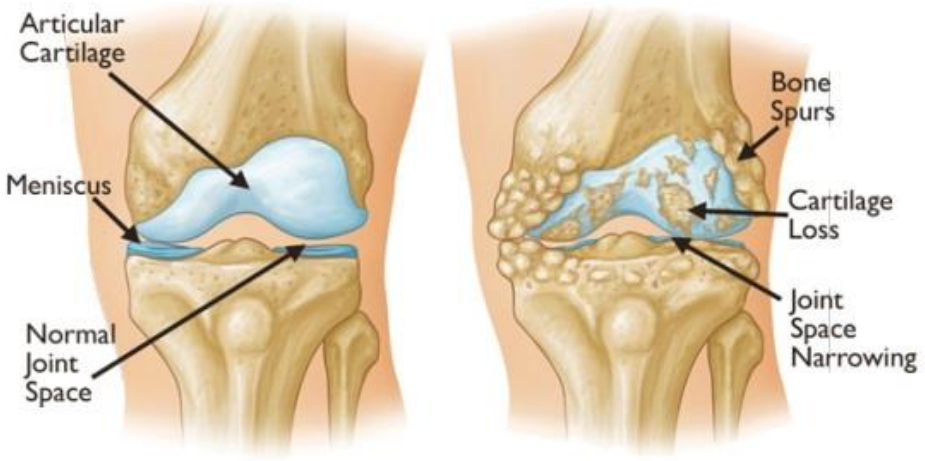
A musculoskeletal condition is defined as being one in which the nerves, tendons, muscles and supporting structure of the body are affected. OA is the most common musculoskeletal condition in older adults and can affect any joint in the body. The aim of this first introductory chapter is to describe the disease OA, and briefly outline the current and potential future burden of the disease both economically and at an individual level. The chapter will also provide an explanation of the different definitions used to diagnose the disease, and explain the gold standard method which is currently used within epidemiological studies to monitor the progression of the disease. This chapter will also provide a discussion of the statistical challenges that arise from monitoring continuous measurements of joint space, leading to the aims and objectives of this doctoral thesis being outlined at the end of the chapter.

1.3 Background of osteoarthritis

Joints in the body, where two ends of bones meet, are coated with a tough connective tissue called cartilage. Cartilage is a firm tissue; however it is softer and more flexible than bone and so the end of each bone being covered in cartilage aids in the joint being able to freely move and rotate. In addition to aiding with movement, cartilage also acts as a shock absorber and spreads the load being exerted on the joint evenly across the entire joint. Unlike other tissue types within the body, cartilage does not have its own blood supply. Within the body it is blood cells that help repair damaged tissue and therefore damage that occurs to cartilage either heals very slowly or does not repair at all. If the cartilage within a joint becomes damaged it may mean

that the joint can no longer move freely due to increased friction. When this occurs the joint disease which results from the body being unable to fully repair cartilage damage within the affected joint is called OA. OA is a degenerative disease that not only involves the cartilage but also many of the surrounding tissues within the affected joint, leading to dramatic structural changes in the joint. This may be because all types of tissues within the affected joint become more active than normal and an imbalance between the breakdown and repair of these tissues can occur. Therefore, in addition to cartilage damage, a joint with OA may also develop other structural changes within the joint such as osteophytes, subchondral bone changes, meniscal alterations and joint swelling. Osteophytes, also known as bone spurs, are where new bone forms around the edge of the affected joint and grows outwards. These bone spurs can be seen in Figure 1-1, which is a diagram of a healthy knee compared with a knee with established OA. Although structural changes do not necessarily affect all tissues in the affected joint, OA is considered a whole joint disease. Ultimately, as the disease progresses the joint affected by OA will fail, meaning that the structural changes within the joint have been so dramatic the joint is no longer able to function. The process of joint failure leads to debilitating effects on the sufferer.

Figure 1-1 Illustration of the structure of a healthy knee joint and a joint with OA



1.4 Definitions of knee osteoarthritis

Any joint in the body can be affected by OA, but the most commonly affected joint is the knee and so OA at the knee site will be the focus of this PhD thesis. OA in the knee can restrict mobility; limit walking, climbing stairs, bathing, personal care and driving a car. In severe cases, OA is an extremely painful condition and a substantial barrier to people's mobility and independence, and significantly compromises their wellbeing and quality of life. Disability due to OA rose by 16% in the UK between 1990 and 2010, with this trend expected to continue (1). The estimates of the prevalence of knee OA vary, depending on the definition being used to diagnose those suffering with the disease, as there are various different ways in which those with knee OA can be identified. Currently OA can be diagnosed clinically, through assessment of structural changes or subjectively through a patient self-report.

Various diagnostic criteria have been developed for knee OA (2). The primary aim of all diagnostic criteria is to differentiate knee OA from different forms of arthritis, such as rheumatoid arthritis. Knee OA diagnosed using the clinical definition occurs by using the medical history of the patient alongside a physical examination of the affected joint. A subjective, self-reported, definition of OA is driven by the patient. While a diagnosis made focusing on the structural changes tends to use radiographic images to visualise structural changes within the joint. Therefore both clinical and subjective OA definitions have symptomatic components while radiographically-defined OA is only concerned with visible structural changes within the knee.

Knee symptoms are common manifestations of knee OA (3) and are often the point at which a patient with knee OA will seek clinical help. Pain tends to be the most widely reported symptom of OA (4) and tends to be made worse by moving the affected joint. However, in addition to pain, the other primary symptoms of OA include stiffness in the joint and reduced range of movement. Those suffering from OA will also tend to report a grating or grinding sensation in the joint, known as crepitus, and swelling of the joint. Symptoms of OA will vary from person to person, and can even vary in severity over time in a person diagnosed with OA. Although, there is no clear reason as to why severity of symptoms vary, people diagnosed with OA often have better and worse periods of disease (5).

One of the most recognised criteria for the diagnosis of clinical knee OA was developed in the early 1990s by the American College of Rheumatology (ACR) (2). The ACR definition of knee OA is based on clinical characteristics of the disease although the definition is primarily driven by the presence of pain; under the ACR OA definition a person must have experienced pain on at least half of the days of the previous month. In addition to pain, an individual must also satisfy 3 of the following 6 criteria:-

- Aged 50 years or over
- Morning stiffness that lasts less than 30 minutes
- Crepitus on active motion (sensation of crunching or creaking which is commonly felt on movement of the knee joint)
- Bony tenderness (pain felt on touching of the knee joint)
- Bony enlargement (hard tissue felt when examining the knee joint)
- No palpable warmth (no warmth is felt when touching the knee)

The most commonly recognised process for identifying structural changes in the knee is using radiographic images, though some studies have begun assessing the use of Magnetic Resonance Imaging (MRI), ultrasound (US) and optical coherence tomographic (OCT) as a methods to assess joint morphology and diagnose knee OA (6). However, there are currently no well-established diagnostic criteria for use within these new imaging methods and radiography remains the most accessible and utilised tool. There are many different radiographic classification systems for diagnosing knee OA, and currently no single gold standard is used in either epidemiological studies or within clinical settings across the world. The two most widely used radiographic classification systems are the Kellgren and Lawrence (K&L) grading system (7) and the Osteoarthritis Research Society International (OARSI) atlas criteria (8, 9).

The K&L grading system was originally described in 1957 and provides an overall score that determines the severity of OA. These scores range from 0 to 4 with a score of 0 representing a normal knee without any structural evidence of knee OA and higher scores representing worsening OA features. Classification occurs by examination of structural changes on radiographic images, as can be seen in Figure 1-2 and described in Table 1-1, and relates to the presence of certain structural changes of interest, namely osteophytes, joint space narrowing, sclerosis and cysts.

Figure 1-2 Radiographic images from patients with no knee OA and K&L grades 0 - 4

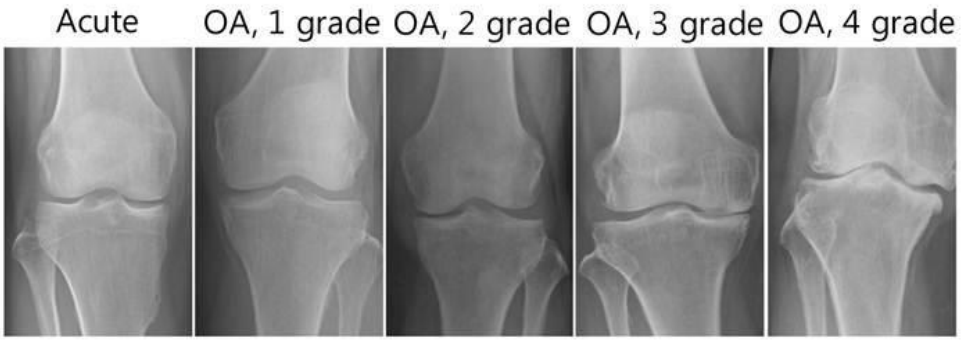


Table 1-1 Knee OA grading using K&L grading

Grading	K&L description
0	No changes
1	Doubtful JSN and possible osteophytic lipping
2	Definite osteophyte and possible JSN
3	Moderate multiple osteophytes, definite JSN, some sclerosis and possible deformity of bone ends
4	Large osteophytes, marked JSN, severe sclerosis, and deformity of bone ends

JSN joint space narrowing

Unlike the K&L grading system that uses a combination of structure changes within each score, the OARSI atlas uses a semi-quantitative separate scoring system to isolate features and compartments of knee OA separately. The OARSI atlas was developed after the K&L grading system, in 1995, and includes scores for four areas for osteophytes and two compartments for joint space narrowing. Scores in each area or compartment are scored from 0 to 3, with higher grades representing more severe features. To be diagnosed with knee OA under the OARSI atlas criteria an individual's knee must have either OARSI atlas grade score for joint space narrowing of 2 or above in either in medial or the lateral compartment; an ORASI atlas grade score for osteophytes of 2 or above in either the medial femoral condyle, medial tibial plateau, lateral femoral condyle or the lateral tibial plateau; or a joint space narrowing score of 1 and a osteophyte score of 1, as described in Table 1-2.

Table 1-2 Knee OA grading using OARSI atlas criteria

OARSI atlas knee OA criteria
JSN of grade 2 or higher
Sum of osteophytes grades greater than or equal to 2
Grade 1 JSN in combination with grade 1 osteophytes

Epidemiological studies of knee OA have often been based on a radiographic definition of OA (10) and the use of K&L grades has been adopted by the World Health Organisation (WHO) as the standard method for defining OA. Conventionally in such studies, an individual is classified as suffering from knee OA if their knee radiographic image has a K&L score of grade 2 or above (11).

More recently, some epidemiological studies have implemented a self-reported, subjective, definition of knee OA (12, 13). In instances where knee OA is defined using a subjective definition, study participants have been asked to self-report whether they believe they have OA in the joint of interest, by asking such questions as 'Do you have knee OA?'.

Each method used to define and identify those with knee OA has its strengths and weaknesses, and it is always important to consider the population and context in which the method of defining knee OA is being used. For example, a radiographic definition of knee OA, regardless of whether using the K&L or OARSI atlas definition, requires a knee radiograph to be performed and the image to be examined and classified by a trained individual. This may be a costly and timely method, whereas a self-reported definition of knee OA is simply a single question asked of the study participant. Previous studies have found modest associations between a radiographic diagnostic approach to knee OA, using the K&L definition, and both clinical and self-reported diagnoses of knee OA (14, 15). However it has been shown that the use of the OARSI atlas to define radiographic knee OA is much more likely to result in a diagnosis of knee OA than using the K&L grading system (16).

Physical symptoms were found to correlate poorly with radiographic knee OA (17). This is not surprising as radiographic OA is purely dependent on structural changes within the knee joint and is not based upon assessment of any physical OA symptoms. Therefore using a radiographic definition of OA may highlight structural changes within the knee joint that occur at an early stage of disease before symptoms have been experienced by the patient. Hence it is common practice, when using K&L grades to classify a knee joint as OA positive when grade 2 or above is observed or when using the OARSI atlas to classify the presence of knee OA when one of the three separate criteria is met. It is likely that, at this point, in addition to structural changes, the patient has begun to experience symptoms.

1.5 Burden of disease

As the different definitions capture slightly different aspects of knee OA disease, it is unsurprising that prevalence of knee OA varies greatly depending on the definition used. Due to the sensitivity of radiographically-defined OA to detect structural changes within the knee joint, highest prevalence levels occur when OA cases were defined radiographically. When comparing different definitions of radiographic knee OA one study has found that rates of knee OA were almost twice as high when using the OARSI atlas criteria than with the traditional K&L grade 2 plus knee OA cut-off (16).

In addition to the definition used, prevalence also varies greatly by gender, age and geographical area being considered. A previous study by Yu at al highlighted than within England 9 new cases of OA are diagnosed for every 1000 at-risk adults (18), and current estimates within European populations report that severe radiographic changes affect nearly 50% of those aged 75 years and above (19). In addition, the WHO report 'Global burden of Osteoarthritis in the year 2000' states that approximately 10% of the world's population who are 60 years or older experience symptomatic problems that are attributable to OA (20). In 2013, it was estimated that 4.11 million people in England had knee OA and this number is estimated to rise to 6.5 million by 2020 (Arthritis Research UK -2013: Osteoarthritis in General Practice). When considering those aged less than 45 years, men tend to have higher prevalence of knee OA than women. However, after the age of 55 years, women tend to be more frequently affected than men (21). Globally, knee OA is ranked as the 11th highest contributor to disability (22). Regardless of the definition

of knee OA being used to describe prevalence, the burden of knee OA is large. The progression of knee OA disease leads to a great deal of pain and disability which, in turn, leads to a large economic burden of disease. The economic burden of disease is categorised into different types of costs, which are described as direct, indirect or intangible costs. Direct costs come in the form of pharmacological and nonpharmacological treatments, caregiver time, surgery, hospital resources, and costs of side effects from treatments or research. Indirect costs include loss of productivity, absenteeism, premature mortality and disability payments/benefits. Intangible costs are defined as the pain and suffering experienced by the patient as a result of knee OA, decreased quality of life and the potential suffering from depression and/or anxiety.

In 2006, the Royal College of General Practitioners (GP) estimated that, in the UK, over 1 million adults consulted their GP each year with symptoms of OA, with the resulting cost per GP consultation estimated at £36 for a 12-minute consultation (23). In the 8th annual report by the National Joint Registry, released in 2011, it was reported that a total number of 81,979 knee replacements were performed. Although the cost of a knee replacement varies between hospitals in the UK, the national tariff price for a total knee replacement in 2010/2011 was set at £5198. Therefore the estimated cost of knee replacements in 2011 was £426 million (24).

It is particularly difficult to estimate the indirect cost of knee OA. Gross national product (GNP) is a broad measure of a nation's total activity, and is the market value of all that the citizens within a country have produced (goods and services) within a year. Previous studies have shown that OA has a significant negative impact on the UK economy, of around 1% GNP (24). It was estimated by the Department of Work and Pensions that, in 2002, 36 million work days were lost because of OA[2]. Whilst it is almost impossible to put an economic estimate on the intangible costs of patients' pain, suffering and impact on life, it is clear that there is a large economic burden of disease. This is true not only for the cost of joint replacement once the knee joint has failed, but also for the cost of ongoing treatment during the slow progression of the disease. Unfortunately the economic burden of disease will only continue to rise, due to the aging population and increased life expectancy.

[2] UK Department for Work and Pensions, Disability Living Allowance cases in payment caseload by Gender of claimant

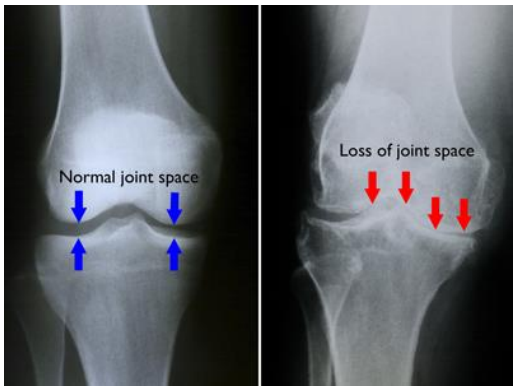
1.6 Disease progression

Current strategies for the care of those diagnosed with knee OA focus on trying to relieve the symptoms and improving the function of the knee joint. This is done through non-pharmacological and pharmacological approaches, and surgical procedures. Non-pharmacological treatments include physical therapy, exercise strategies such as general, pool and strengthening exercises, heat and cold, patient education and in some circumstances weight loss. Surgical approaches include arthroscopic lavage and debridement (smoothing of rough cartilage), osteotomy (shortening of the bone to change the alignment of the joint to allow cartilage wear to occur at a different portion of the joint) and bone marrow stimulation (drilling smalls holes through the joint into bone marrow in the hope that a form of scar cartilage grows over the defect). Ultimately, as the disease progresses, the only option for treatment is total joint replacement. Pharmacological care strategies include analgesic, anti-inflammatory agents, intra-articular corticosteroids, or glucosamine sulphate. However, to date there are no pharmacological treatments which are registered as structure modifying. That is, there is currently no drug treatment available which has consistently been shown to reverse or even slow down the process of structural deterioration within the knee joint.

To be able to offer the best and most appropriate care to a patient with knee OA it is important that assessment of any functional impact of the OA and severity of the disease are monitored. Different methods are used to monitor disease progression within different settings. In a clinical setting, as the disease progresses, patient symptoms will tend to worsen. Therefore the clinician will tend to focus on managing a patient's worsening symptoms, and will monitor worsening of the clinical components of knee OA. In epidemiological studies of knee OA, monitoring of disease progression using a radiographic definition of OA (10), by focusing on structural changes rather than symptoms, is most commonly used. To date many epidemiological studies have monitored radiographic progression by assessing whether a patient's K&L categorisation has changed over time. Regardless of the radiographic knee OA definition being used within a study, a limitation of both classifications outlined above is their reliance on a subjective assessments of the knee OA features present on the radiographic image (16). A more objective approach to monitoring the natural progression of knee OA is to track the joint space width. In clinical trials assessing the potential of disease-modifying osteoarthritis drugs (DMOADs), joint space width is currently the Food and Drug Administration's (FDA) only approved endpoint (25). Knee joint space width

is a measurement used as an outcome marker to determine cartilage loss. It is a continuous measure obtained from the radiographic image, by measuring between the ends of each bone within the knee joint, also known as inter-bone distance. As OA within the knee progresses, the joint space width decreases, see Figure 1-3, as cartilage damage occurs. In the past, this measurement was obtained by hand using specially designed rulers overlaid on the knee radiographic image (26), but more recently the measurement process has become semi-automated for use on digitised radiographic images (27).

Figure 1-3 Illustration of a radiographic image showing a normal and a reduced joint space within the knee



Use of knee joint space measurements, as a continuous measure, is useful for monitoring the natural disease progression (28), both in a research and clinical setting, because the continuous measurement of joint space width is more sensitive to identifying change than the categories used within the K&L categorisations (19) or within OARSI atlas grading. This is mainly due to the nature of the disease; OA progresses slowly, and changes in knee joint space may vary from 0.1 to 0.6mm per year (29-31). It has been demonstrated that in knees from healthy individuals maximal joint space measurements are around 8mm (32), but it has been estimated that joint space measurements could be in error by up to 1mm (33). Therefore, when disease progression is such a slow process, the presence of measurement error may make it difficult, over short periods of time, to distinguish between those individuals who have experienced real deterioration, which is a worsening of their knee OA, and an apparent change that is simply due to error within the measurements.

Traditionally, in epidemiological studies that focus on disease progression through monitoring of continuous knee joint space measurements, OA disease progression has been compared between groups by calculating the mean difference between measurements, and then testing whether group differences are significant using such statistical techniques as paired t-tests (34). However, such statistical techniques will only reveal differences in means between groups or indicate whether a population mean joint space has changed over time; such methods give no information about changes within individuals. An individual's change is the observed difference between two knee joint space measurements taken at different times, and this may be dominated by measurement error in either or both measurements. In addition to obscuring true knee OA disease deterioration, measurement error may indicate an apparent increase in observed joint space over time. Increases in knee joint space are only biologically possible in extremely rare clinical circumstances or if there has been an intervention, such as joint replacement or joint distraction, and so it is highly likely that any observed increases in differences are attributable to measurement error. Therefore it is important, in both research and clinical settings, to be able to remove the effect of measurement error so as to identify differences that are more likely to be due to real change in disease. In a clinical setting, it is important to ensure that patients are on correct treatment pathways and are not incorrectly identified as having had an improvement or deterioration in structural components of knee OA. In a research setting it is important to ensure that any interventions or risk behaviours that are being related to disease progression are correctly identified. Although several risk factors for knee OA have been firmly identified as being associated with increased prevalence of the disease, such as obesity (35), increasing age (36, 37) and gender (38), there is still debate in other areas as to whether certain modifiable risks are linked to worsening of knee OA. This may in part be due to the real change over time being obscured by measurement error and therefore weakening the possibility of finding associations.

1.7 Aims and Outline

It is clear that there is an ever-increasingly large economic burden of knee OA disease, as well as a debilitating disease effect at an individual level. The disease progression is slow, often over decades, but ultimately the progression of the disease will lead to replacement of the knee joint. Currently, care strategies for those diagnosed with knee OA focus on relieving symptoms, until the joint fails and total joint replacement occurs. Not only is there an economic burden of joint

replacement but the operation can be traumatic for the patient. There is an added complication of joint replacements often having a restricted lifespan, therefore the younger the patient receiving the total knee replacement the higher the likelihood of the replacement knee joint failing during their lifetime, and therefore requiring more surgery and a possible further replacement.

It is therefore preferable to be able to accurately identify factors that may slow the progression of knee OA, or even stop the progression. To do this it is necessary to be able to identify reliable changes in knee OA disease progression, rather than changes which have only occurred through measurement error. Therefore the major research foci of this doctoral thesis are:-

1) Identification of reliable changes in continuous knee joint space width measurement, using such methods as the reliable change index to quantify how much change in knee joint space width is required to exceed measurement error.

2) Identification of novel applications of established statistical modelling techniques, such as linear mixed effect modelling and Bayesian modelling, to model change in knee joint space width as an outcome after accounting for measurement error.

3) A comparison of the different statistical methods identified for determining change, in order to be able to identify the single most appropriate statistical methods to use for monitoring change.

The three statistical methods that are explored within this doctoral project are the reliable change index, linear mixed effect modelling and Bayesian modelling. The reliable change index is a method of identifying whether an observed change between two time points within an individual is meaningful in the presence of measurement error. However both the frequentist linear mixed effect and Bayesian modelling are methods that are able to utilise all longitudinal measurements. The application of frequentist linear mixed effect modelling is a well-established longitudinal modelling technique, and in the context of this thesis allows for the modelling of change in knee joint space width. Whereas, although becoming more widely applied, Bayesian

modelling is a statistical method that is not currently widely used within musculoskeletal research.

To aid in the understanding of how well the identified statistical methods are able to detect reliable changes in knee OA disease progression, this thesis applied the methods in two datasets, SEKOIA and the OAI study. The SEKOIA study was a 3-year international trial and this doctoral project uses the placebo arm data. The OAI study is an open-access longitudinal dataset following study participants from America. The study participants entering the SEKOIA trial already had established knee OA and the knee joint space width measurements were obtained under strict study protocol. The OAI study participants already had established knee OA or were identified as at risk of developing knee OA on entry into the study. Therefore the OAI study data contains a wider spectrum of knee joint space width measurements than SEKOIA.

1.8 Summary

This chapter has provided a brief explanation of knee OA, and its ever increasing economic burden and effects on individuals. As explained there are a number of different definitions that can be used to diagnose knee OA, but the gold standard for monitoring disease progression within knee OA patients is the continuous joint space measurement. Nevertheless, using such a measurement provides methodological challenges, and so the overarching aim of this doctoral study is to develop statistical techniques that account for such challenges. Chapter 2 of this thesis will present a review of the statistical methods which have been used in previous studies that have monitored the progression of knee OA using continuous joint space measurements. Chapter 3 provides an outline of the study design of the SEKOIA study and Chapter 4 outlines the OAI study. In both Chapters 3 and 4, a description of the information collected during both studies is provided and descriptive statistics of the study participants are presented to give a broad description of the study population's that are the focus of this doctoral project.

Chapter 5 provides a description of the reliable change index and presents results from the novel application of the reliable change index in estimating OA progression within SEKOIA and the OAI study.

Chapter 6 provides a brief description and results of the frequentist linear mixed effect modelling technique. Chapter 7 provides a brief introduction to Bayesian methodology with a particular focus on Bayesian Hierarchical modelling, and presents results from Bayesian analysis. Chapter 8 gives a comparison of the different statistical methods used to monitor disease progression while considering measurement error. In chapter 9 the epidemiological application of the single proposed statistical method for monitoring progression is outlined. Chapter 10 reflects on the work that has been completed in this doctoral thesis, and, in addition to a summary of the findings, also provides a discussion on how the statistical methods outlined during the work could be applied within a clinical and epidemiological setting.

Chapter 2: Literature review

2.1 Overview of chapter

This chapter provides a review of statistical methods that have been used in previous studies to monitor progression of knee OA. All studies reviewed used continuous knee joint space measurements to identify change in joint space width over time. The chapter begins with a description of the methods used to complete the literature review, the results of the literature search follow which summarise the statistical methods used in the literature to monitor knee joint space loss, paying special attention as to whether any methods account for the issue of measurement error.

2.2 Literature review methods

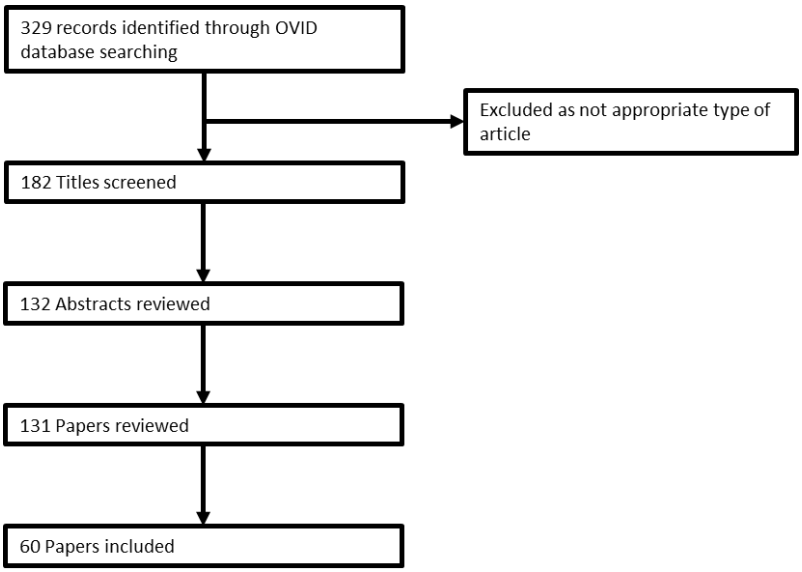
A literature search was conducted for articles published up to April 2017, using the search engine OVID which uses the databases Embase and Medline. Eligible studies included those with repeated continuous measurements of knee joint space width measured using radiographic images, and with statistical analysis assessing longitudinal change between the measurements. Studies were excluded if they focused on animals or were not in the English language. Conference abstracts were also excluded, as often it was not possible to clearly determine the statistical methods used to analyse change in knee joint space measurements over time. Both free text and MeSH terms were applied and the search terms used were 'knee osteoarthritis', 'joint space' and 'change' (Table 2-1). Screened studies were excluded if they did not report change in knee joint space in millimetres, and instead used joint space narrowing classified using a categorical or grading scheme. The focus of the literature search was to collate information about the statistical methods which are currently used within musculoskeletal research to assess change, rather than focusing on studies that assess the reliability of the radiographic protocol.

Table 2-1 Search strategy and MeSH terms used for the literature search

	Search strategy and MeSH terms
1	(*Osteoarthritis, Knee/ or (osteoarthritis adj2 knee).ti,ab)
2	"joint space".ti,ab
3	"change".ti,ab
4	1 and 2 and 3
5	Remove duplicates from 4
6	(5 and humans/) or (5 not(humans/ or animals/))
7	Limit 6 to English language

The literature search identified 329 articles, Figure 2-1 , and after excluding those articles which did not meet the eligibility criteria, 182 titles were screened. Of those, 132 abstracts were extracted and screened for relevance, and a total of 131 studies were read for inclusion. Of those studies reviewed, a total of 60 articles were found to assess change in knee joint space over time using a continuous radiographic measurement.

Figure 2-1 Flow chart of the screening process for articles included in the literature review



2.3 Literature review results

As previously discussed, the continuous measurement for knee joint space obtained from radiographs is commonly used to assess progression of knee OA, and is regarded as the gold standard measurement for assessing anatomical changes within the knee joint in clinical trials (25). The large proportion of the excluded manuscripts assessed change in joint space narrowing through grading systems. For example, Uitterlinden et al (39) used joint space narrowing grades from the K&L grading system and assessed whether individuals had changed score during the study period. The use of grading systems will be less sensitive to narrowing of joint space widths over time, and therefore measurement error will not be as prominent a problem when using such schemes. The remaining manuscripts reviewed and excluded used images obtained from MRI, such as Wirth et al (40). The use of MRI to visualise joint structures is a reasonably recent development and will continue to undergo evaluation, but this evaluation is beyond the scope of the work in this thesis.

The majority of the 60 articles included within the literature review for this thesis focused on relating change in knee joint space width to potential risk factors for progression or assessing the impact of an intervention. In the vast majority of studies, change was simply defined as difference between the time points under review. For example, Vilim et al calculated change in

knee joint space width over the three year duration of their study by subtracting the baseline knee joint space width from the knee joint space width measured at year 3 (41). Within Vilim's study, change in joint space width over the 3-year duration was correlated with serum levels of cartilage oligomeric matrix protein at baseline to determine if any association existed between levels of this protein and change in knee joint space width. During their study, Vilim et al considered change between 2 time points, baseline and 3 years later. Of all the 60 papers reviewed the vast majority of studies, 78%, monitored knee joint space width over only two time points, Table 2-2.

Table 2-2 Frequencies of study visits in papers contained in literature review

Number of time points	Number of studies
2	46
3	5
4	3
5	2
6	1
7	1

In the 46 studies with two repeated measures of knee joint space width, all but 3 studies used either a crude difference of change, calculated by subtracting one time point from another, or a rate per year in mm, calculated by dividing the difference by the number of years of follow-up. Regardless of whether mm per year or crude change was reported as the measure of change, 21 studies reviewed used a t-test to investigate the magnitude of change between different groupings of interest. For example, Hunter el at (42) followed study participants over an 18 month period to determine if a combination of dieting, and/or exercising had a positive effect in obese adults with symptomatic knee OA. To determine if any of the interventions were successful, change in joint space width between baseline and 18 month follow-up was calculated. The mean change in joint space width was then compared across the different intervention groupings. The problem with using such simple methods to determine change is

that they will only reveal differences at the population level and give no information at the individual level.

In those papers reviewed that contained two repeated measures and did not use crude difference or mm per year, the measures of change that were used were standardised response mean (SRM) and percentage change. For example, in a study assessing novel methods for measuring bone area by Bowes et al the change in minimum joint space width over a 2-year period was reported as percentage change, i.e. a reduction of 4.3% in joint space width among the 352 study participants was reported (43). The advantage of summarising change in this way is that the percentage can give an easily interpretable quantifiable figure for the amount of change. However this simple method is still reducing a large amount of information into one figure, thus losing detail. In addition to reporting change as a percentage, Bowes et al also reported the SRM. The SRM is calculated as the mean change divided by the standard deviation (SD) of change. Bowes et al were not the only study reviewed that reported change using the SRM technique. Conrozier et al calculated the SRM over a 1-year period, which allowed for direct comparison of the change in joint space width that occurred between the two different study groups of interest (44). Once calculated the SRM gives an effect size index, which provides an indication of the responsiveness of scales to clinical change, or, in the contexts of the literature reviewed for this thesis, the SRMs reported give an indication of the responsiveness of change in knee joint space width. The advantage of a SRM is that direct comparisons can be made across groups to determine which SRM is larger, and easily determine whether greater levels of change have occurred. However the SRM is still only revealing differences at the population level.

In the 12 studies with more than two repeated measures, the total number of study visits varied from 3 to 7. In a study by Bruyère in 2013 repeated radiographic images were obtained at baseline and then annually over a 3 year period (45), resulting in 4 measures per study participants within the study design. The focus of the study was to compare the change in joint space between study participants with knee OA and between study participants who had a knee joint replacement. Despite the methods of the paper highlighting that multiple radiographic images were obtained, the only associations that were investigated were the change in joint space width over the 3 year period since baseline. Thus the repeated joint space width measurements were not fully utilised within the remit of the study presented.

Of all the studies reviewed, Fukui et al (46) contained the greatest number of repeated measurements. The methods of this study stated that radiographic images were obtained at baseline and every 6 months during the 3 year follow-up period, therefore giving the possibility of 7 knee joint space measurements. However, within this study, change was reported using joint space width narrowing rate (mm/year), which appears to only have been calculated using the baseline and 3 year joint space measurements and therefore not making full use of the repeated measures. A similar approach was used by Pavelka et al in 2000, as radiographic images were obtained at baseline and then annually over a 5 year period (47). In their paper, the mean joint space width was presented for each study year, and the change in joint space across each year and annual joint space narrowing were plotted graphically. However no statistical analysis was carried out that utilised all available repeated measurements, just an in-depth description of joint space width measurements at each study visit.

Regardless of how change was defined in the 60 articles contained within the literature review for this thesis, in all but a few studies basic statistical methods are used to assess change. The most commonly used statistical method within the articles reviewed was a t-test, for example Eckstein et al used a paired t-test to compare minimum joint space width change over 2 years in those who went on to receive a knee replacement and in matched controls (48). Using a paired t-test the authors of this study found a statistically significantly greater difference in change in joint space width between the two groups. Other simple statistical techniques that were used in the articles reviewed were the nonparametric Wilcoxon test, ANOVA and ANCOVA. However a handful of those studies reviewed did use more advanced statistical analysis. In a study undertaken by Hellio le Graverand, the authors used a mixed model containing a random intercept and a random slope to compare joint space measurements obtained from radiographic images obtained at three time points between a placebo and treatment group (49). Eaton et al also used a linear mixed models to explore the association of skin fluorescence with progression of knee OA (50), as did Tourville et al to characterise joint space width changes occurring in the knee 4 years after a knee injury occurred (51). Generalised estimating equations models (GEEs) were used by Lapane et al to estimate whether long term nonsteroidal anti-inflammatory usage relieves symptoms and delays disease progression (52). These studies were the only studies reviewed that fully utilised repeated measurements within one statistical analysis technique.

Not only did the statistical techniques used to analysis the continuous joint space measures vary across the 60 studies but so did the study duration. As can be seen from Table 2-3, the shortest study duration of all studies was 2 weeks and the longest follow-up period was 8 years. The most frequent study durations were 1 year, 2 years and 3 years with 12, 16 and 13 studies respectively, following study participants up over these time frames. The time frame is an important element of study design to consider when monitoring knee OA disease progression, as research by Reichmann et al demonstrated (53). Reichmann and team performed a systematic review to gain a more accurate understanding of joint space width measurements of responsiveness to change. Their research concluded that studies with a follow-up period of 2 years or more showed greater levels of responsiveness to change. Although measurement error is not discussed as a consideration within Reichmann's review, it is likely that change becomes more difficult to identify within studies with short follow-up periods. This is because of the small magnitude of change that would be expected over shorter periods, which may be as little as 0.6mm per year (31), as compared with the magnitude of measurement error, which could be as much as 1mm (33).

Table 2-3 Frequencies of study durations in papers contained in literature review

Study duration	**Number of studies**
2 weeks	1
6 weeks	1
12 weeks	1
1 year	12
18 months	1
2 years	16
30 months	2
3 years	13
4 years	6
5 years	3
8 years	2

Error within knee joint measurements can arise due to a variety of reasons including: error at the measurement process level, limitation of the measurement instruments, and because of fluctuations in measurements that may occur over time (54-56). Error at the measurement process level occurs while obtaining the measurement, such as changes in the positioning of the knee during radiographic imaging. An example of measurement error due to limitation of the measurement instrument would be the technique used to measure the radiographic image. Previously joint space width measurements were obtained by hand using a specially designed ruler overlaid on knee radiographic images (26), more recently the measurement process has become semi-automated for use on digitised radiographic images (27). Fluctuations in joint space widths can also occur due to biological variation of the measurements, for example previous studies have demonstrated that cartilage thickness measurements varied depending on the time of day they were obtained (57, 58).

Despite there being multiple reasons that could lead to measurement error within knee joint space measurement, of all the articles reviewed only one study discussed measurement error and the impact this may have on assessment of change in joint space width (59). Ravaud and colleagues in 1999 considered the effect that measurement error may have on assessing the smallest detectable change over time in continuous structural outcome measures. The study developed statistical methods for determining cut-offs above which the smallest detectable change could be identified. To do this they proposed a method that involved estimating the variability of the measurement error from observations obtained over a period short enough that no real change would have occurred, and used this to determine cut-offs. Although the method of cut-point calculation within this study still led to crude differences being the focus of analysis, it does provide a way of identifying individuals whose change is not considered real and who can be removed from analysis. But the study did not propose methods for adjusting for measurement error in circumstances where repeated measurements are used, and, as highlighted by Gustafson, it would be misleading to simply ignore the concept of measurement error, as measurement error causes bias and so associations can be attenuated (60, 61).

2.4 Summary

The majority of analyses used to assess change in joint space measurements over time in the sixty studies reviewed used simple statistical techniques. Often the studies summarised the repeated measurements into one change variable, and then used techniques such as t-tests to determine the association between magnitude of change and the factor of interest. Previous studies have shown that a study design with a follow-up period of 2 years or more has greater ability to detect change (53). It may be possible to detect change with shorter follow-up periods, although the greater the follow-up period the greater the chance that any change in knee joint space width identified is real change and not measurement error. The assumption made by the majority of investigators in the studies reviewed was that change in knee joint space is constant, therefore rates of change were assumed to be additive. However, it may be possible that different rates of change in joint space occur at different points of disease progression. Research by Felson et al found that structural progression of OA had a pattern of inertia, with knees having periods of 'stability' and 'deterioration' (62), and Eckstein and colleague assessed trajectories of cartilage loss within four years of a knee replacement and found that there was an accelerated loss of cartilage in the two years prior to surgery (63).

Of all studies reviewed, only one mentioned the issue of measurement error (59) and identified that without accounting for measurement error the differences observed may not be 'true organic change'. Despite raising the issue of measurement error, the method developed within the paper was a technique for determining cut-points above which change could be considered real. Therefore, to date no statistical technique has been developed that would identify change in continuous knee joint space width measurements, after accounting for measurement error. Neither has methodology been developed that would identify change in knee joint space width using repeated continuous knee joint space measurements across more than two time points while taking account of the issue of measurement error.

Chapter 3: The SEKOIA Study

3.1 Overview of chapter

This chapter describes one of the two datasets used within this doctoral thesis, the placebo arm of the SEKOIA study. The SEKOIA study was a 3-year international trial during which a large battery of information about study participants was collected. The chapter outlines the study design, and gives a description of the information collected and descriptive statistics about the study participants.

3.2 SEKOIA population and subject selection

The Strontium ranelate Efficacy in Knee Osteoarthritis triAl (SEKOIA) was a 3-year international, multicentre, double-blind, randomised placebo-controlled phase 3 trial that was established to assess the structure-modifying effect of a drug treatment, strontium ranelate, on radiological and clinical progression of OA in the knee joint (64). Study participants were recruited into the trial from 98 study centres across 18 different countries and upon recruitment were randomised to either a drug regime of strontium ranelate 1 gram per day, strontium ranelate 2 grams per day, or a placebo treatment. For the purpose of this doctoral project, access has been granted to data on participants in the placebo arm of the trial who received no active treatment.

Study participants were recruited from secondary care establishments where they were already receiving outpatient care for knee OA. To be eligible for entry into the SEKOIA trial, study participants had to be Caucasian males or females aged over 50 years. They had to be ambulatory, capable of walking, and capable of reading and writing to enable full understanding of the trial documentation. Study participants had to have a primary diagnosis of knee OA, and had to meet a number of criteria in regards to the diagnosis of knee OA, which was based upon ACR clinical criteria, described earlier (section 1.4). However, in brief, for entry into SEKOIA a study participant had to have had pain on at least half of the days of the previous month. For the purpose of the study this was defined as an intensity of 40mm or above on a 100-mm visual

analogue scale (VAS). The VAS global knee pain scale (65) is a measure of pain intensity in the knee, and was assessed by the patient marking the level of pain on a 100mm scale where 0mm represented no pain and 100mm represented extreme pain. In addition to pain, the patient must have also had any 3 out of the 6 following criteria:-

- Age 50 years or over
- Crepitus (sensation of crunching or creaking which is commonly felt on movement of the knee joint)
- Morning stiffness lasting no longer than 30 minutes
- Bony tenderness (pain felt on touching of the knee joint)
- Bony enlargement (hard tissue felt when examining the knee joint)
- No palpable warmth (no warmth is felt when touching the knee)

In addition to having clinical knee OA, participants also had to have a K&L grade of 2 or 3. Thus included patients had definite osteophytes and possible narrowing of joint space (grade 2 K&L) or moderate multiple osteophytes, definite narrowing of the joint space and some sclerosis (hardening of tissue within the joint), and possible deformity of bone ends (grade 3 K&L). Participants also had to have a knee joint space width measured between 2.5mm and 5mm. On entry into SEKOIA one target knee was selected to be studied, and if both knees fulfilled the entry criteria the target knee was selected as the knee which was the most painful.

As the aim of SEKOIA was to assess the efficacy and safety of strontium ranelate in the treatment of knee OA, exclusions included those who had already had a knee replacement, recent injections into the knee joint and recent treatment acting on cartilage metabolism (66). The study was approved by the ethics committee of every site included within the trial and conformed to the principles of the Declaration of Helsinki. Written informed consent was provided by all study participants before randomisation. The first study participant visit took place on 28th April 2006, and recruitment ended on 10th March 2008. The last study visit took place on 17th February 2011.

3.3 Data collected

A large battery of information was collected about each study participant throughout the duration of the 3 year trial. This included demographic, lifestyle, anthropometry and clinical

measurements. A brief description of the data collected during the 3 year trial is described in sections 3.3.1 and 3.3.2.

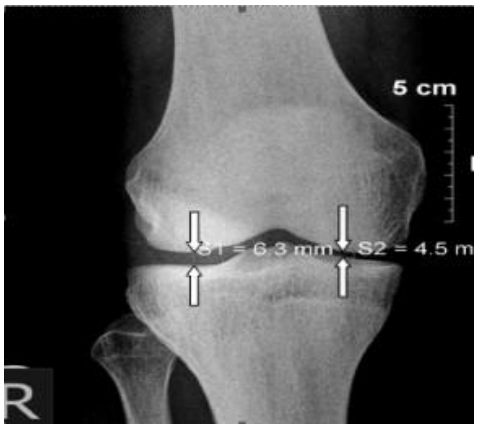
3.3.1 Individual characteristics

At the SEKOIA baseline study visit, study participants answered questions about their individual demographic characteristics including age, sex, smoking habits, alcohol consumption and level of physical activity. A detailed medical history was also obtained and was used to determine knee OA disease duration. Weight in kg and height in cm was recorded at entry and then annually at each study visit during the trial and used to determine each study participant's body mass index (BMI). A knee physical examination was performed at baseline and at each study visit, to determine physical signs of knee OA (such as swelling, and warmth).

3.3.2 Clinical outcomes

As the primary outcome of the SEKOIA study was to determine the efficacy of strontium ranelate as a structure-modifying treatment for knee OA, knee radiographs were performed at baseline, and then annually on the target knee using a standardised technique across all centres (66). All of the radiographic images were measured centrally by a single reader blinded to the study participant and treatment allocation. Each blinded post-baseline radiographic image was measured in comparison with the study participant's baseline image to optimise reproducibility and sensitivity. Although this means that radiographic images were not completely blinded this method was implemented in order to reduce variability. In contrast, although the readers had sight of the baseline radiographic image, they were blinded to treatment assignment and study participant identification. The knee joint space width was measured using a standardised computer-assisted method. To do this the digital image of the knee was magnified and the mid-portion of the medial and lateral compartments of the knee identified. Thus, the mid-point of the inner and outer portions of the knee were identified, similar to the illustration in Figure 3-1. The distance between the tibial and femoral bone was measured, and the joint space width was defined as the smallest distance measured between the two bones.

Figure 3-1 Illustration of a digital knee radiographic image used to measure knee joint space width



Although OA can affect both the medial and lateral aspects of the knee joint, the most commonly affected compartment of the knee is the medial. Therefore the joint space width measurements in the medial compartment of the knee were the focus of the recorded knee joint space assessments in the SEKOIA study. A summary and comparison to the OAI of radiographic image acquisitions elements is contained in the appendices.

Other clinical findings which were recorded during the 3-year trial included detailed reports of pain using VAS to ascertain global knee pain, and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), to understand knee function. The version of WOMAC used in SEKOIA was a self-administered visual analogue, on the same 100mm scale as VAS, questionnaire designed to assess health status and health outcomes in OA of the knee (67). The questionnaire contains 24 questions targeting areas of pain (5 questions), stiffness (2 questions) and physical function (17 questions). The lower the WOMAC score the better status, i.e. lower scores represent study participants who have less pain, stiffness and restricted physical function. VAS was obtained at inclusion into SEKOIA and then at 6 monthly intervals.

3.4 SEKOIA descriptive statistics

3.4.1 Study visits

The primary outcome of the SEKOIA study, knee joint space width, was assessed at baseline and then every 12 months until completion of the 3-year study. Therefore the target was 365 days between knee radiographic images. Table 3-1 shows the mean, standard deviation (SD) and minimum and maximum days between study participants' radiographic images. The mean number of days between baseline images and year 1 radiographic images was 389 days. The shortest time between two study visit radiographic images was 70 days between a year 1 visit and a year 2 visits, and the maximum duration between visits was 615 days.

Table 3-1 Days between study visits for knee joint space width measurements in SEKOIA

	Mean	SD	Minimum	Maximum
Baseline – year 1	388.9	42.7	127	615
Year 1 – year 2	360.2	55.3	70	579
Year 2 – year 3	359.0	48.4	115	586

3.4.2 Individual characteristics

A total of 559 patients were randomised to the placebo arm of the SEKOIA study and demographic characteristics of these study participants are presented in Table 3-2. Descriptive statistics were used to summarise individual characteristics. Just under 30% of study participants were men and the mean (SD) age was 62.8 (7.5) years. On average, the men were significantly older than the women. The women were shorter than the men, mean (SD) 160.3 (6.7) cm and 174.7 (7.7) cm respectively, and the men were significantly heavier than the women. The men and women had similar mean BMI, mean (SD) 29.8 (4.1) kg/m^2 and 29.8 (5.5) kg/m^2 respectively. More than 60% of the total participants reported having never smoked, but significantly fewer men than women reported never smoking. More than 70% of men reported drinking some level

of alcohol, which was considerably higher than in women, and the majority of study participants reported doing medium or intense levels of physical activity.

Table 3-2 Demographic characteristics of SEKOIA study participants

	All (n=559)		Men (n = 167)		Women (n = 392)		P-value
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	
Age (years)	62.8	7.5	63.8	7.8	62.3	7.3	0.03[a]
Height (cm)	164.6	9.6	174.7	7.7	160.3	6.7	<0.01[a]
Weight (kg)	80.8	16.1	90.8	13.9	76.5	15	<0.01[a]
BMI (kg/m^2)	29.8	5.1	29.8	4.1	29.8	5.5	0.98[a]
	n	**%**	**n**	**%**	**n**	**%**	
Smoking							
No	359	62.2	79	47.3	280	71.4	
Ex	142	25.4	67	40.1	75	19.1	
Yes	58	10.4	21	12.6	37	9.4	<0.01[b]
Alcohol							
No	284	50.8	39	23.4	245	62.5	
Has stopped	8	1.4	4	2.4	4	1.0	
Yes	267	47.8	124	74.3	143	36.5	<0.01[c]
Physical activity							
Low (occasional)	64	15.4	22	16.8	42	14.7	
Medium (regular)	292	70.0	96	73.3	196	68.5	
Intensive (competitive)	61	14.6	13	9.9	48	16.8	0.18[b]

All continuous variables were normally distributed; [a]p-value for t-test; [b]p-value for Chi-square ;[c]p-value for Fisher's exact

3.4.3 Clinical outcomes

Mean joint space width and 95% confidence intervals, obtained from radiographic images at yearly intervals, and are presented in Figure 3-2. At baseline men had a significantly larger knee joint space width than women, mean 3.65mm (95% CI 3.52mm to 3.78mm) and 3.45mm (95% CI 3.36mm to 3.53mm) respectively. During the study, men continued to have larger knee joint width space measurements than women. However, after baseline, the mean knee joint space width for men and women was found not to be significantly different across the remaining study years. By the end of the study, men had a mean knee joint space width of 3.21mm (95% CI 2.99mm to 3.42mm) while women had an average knee joint space width of 3.11mm (95% CI 2.98mm to 3.24mm).

Figure 3-2 Mean joint space width (mm) and 95% confidence interval for each study visit in SEKOIA by gender

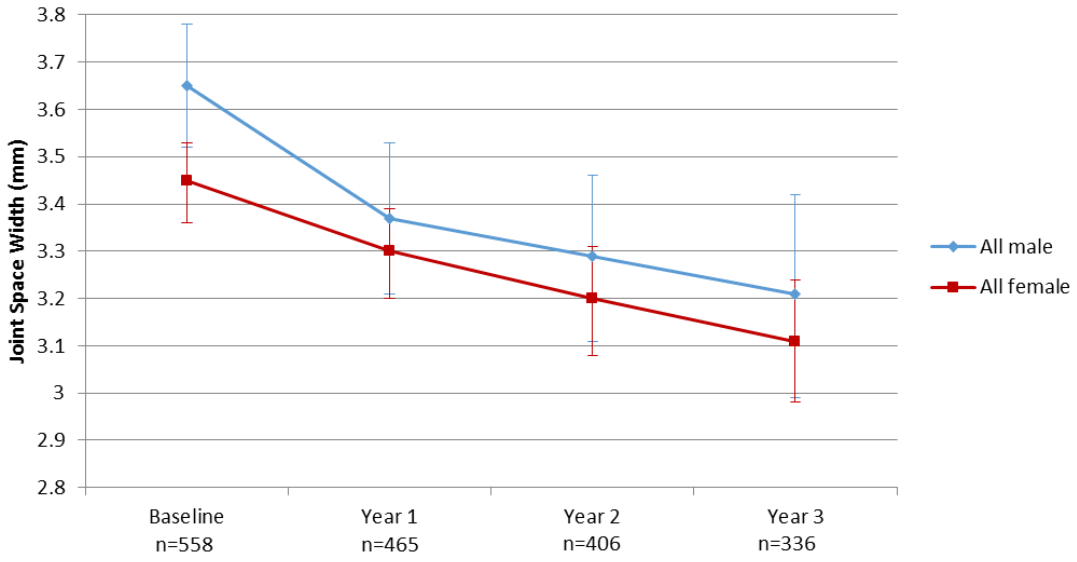


Table 3-3 contains descriptive statistics of VAS global knee pain, WOMAC component subscores, WOMAC global score and K&L grades for the total population and by gender at baseline. Higher scores of VAS knee pain indicate higher level of knee pain. Lower WOMAC scores indicate better status.

Table 3-3 Clinical characteristics of SEKOIA study participants at baseline

	All (n=559)		**Males (n=167)**		**Females (n=392)**		**p-value**
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	
VAS knee pain	53.7	22.4	48.4	22.5	56.1	22	<0.01[a]
WOMAC Component Score							
Pain	211.2	108.0	194.9	105.0	218.1	108.6	0.02[a]
Stiffness	91.1	50.4	83.2	48.0	94.4	51.1	0.02[a]
Physical function	694.7	380.3	618.4	384.0	727.4	374.5	<0.01[a]
WOMAC Global score	998.5	509.3	899.2	513.6	1040.4	502.3	<0.01[a]
	n	**%**	**n**	**%**	**n**	**%**	
Kellgren and Lawrence Score							
2	350	62.6	103	61.7	247	63.0	
3	209	37.4	64	38.3	145	37.0	0.77[b]

All continuous variables were normally distributed; [a]p-value for t-test; [b]p-value for Chi-square

At baseline, on average, women experienced significantly more knee pain than men. A similar pattern was seen in the component WOMAC scores, as higher scores indicate a worsening of the component score. Therefore women reported having significantly higher levels of pain, greater levels of stiffness and greater levels of physical function limitation when compared with men. As reported earlier within the chapter, study participants had to have a K&L grade of 2 or 3 to enter SEKOIA, and thus not all possible K&L grades are represented within SEKOIA. Around 60% of study participants had a K&L grade of 2 at baseline and the proportions were similar in men and women.

3.4.4 Change in joint space width

One method that can be used to monitor the progression of knee OA is to calculate change in knee joint space width; this is the crude difference between continuous knee joint space width measurements. Within SEKOIA this can be calculated across the yearly intervals. The average change in knee joint space width for all study participants between baseline and year 1 was -0.20mm, Table 3-4, which means that on average study participants knee joint space width reduced by 0.2mm between baseline and year 1. Over the course of the first SEKOIA study year, on average, men had a greater reduction in knee joint space width than women and then a similar pattern of loss for men and women was observed across all remaining study years. However, the differences in the magnitude of change in knee joint space width between baseline and year 1 never reached statistical significance when comparing men and women. On average, study participants had a greater change in knee joint space width between baseline and year 1 than across remaining study years.

Table 3-4 Change in knee joint space width between SEKOIA study years by gender

	All			Men			Women			p-value
	n	Mean	SD	n	Mean	SD	n	Mean	SD	
Change in joint space width between:-										
Baseline and year 1 (mm)	465	-0.20	0.46	145	-0.25	0.51	320	-0.18	0.44	0.12[a]
Years 1 and year 2 (mm)	400	-0.11	0.43	119	-0.11	0.50	281	-0.10	0.40	0.96[a]
Year 2 and year 3 (mm)	329	-0.11	0.45	100	-0.11	0.52	229	-0.10	0.42	0.88[a]

All continuous variables were normally distributed; [a]p-value for t-test

3.5 Summary

In this chapter the study design of the 3 year international SEKOIA study was outlined and the large battery of information about each study participant was described. Descriptive statistics of the number of days between study visits, main demographic, lifestyle, anthropometric and clinical measurements were presented. During the 3-year duration of the study, on average, men lost more knee joint space between study years than women. A reduction in knee joint space width was identified across all study years, both in the total population and within men and women.

Chapter 4: The Osteoarthritis Initiative

4.1 Overview of chapter

Chapter 4 describes the open-access dataset OAI, which is the second of the two datasets used within this doctoral project. The OAI dataset is a longitudinal study that followed study participants from America that had been identified as at risk of developing knee OA or already had established knee OA. This chapter outlines the OAI study design, gives a description of the broad spectrum of information collected and presents descriptive statistics about the study participants.

4.2 The OAI study population and subject selection

The OsteoArthritis Initiative (OAI) is a multi-centre, longitudinal, prospective observational study focusing primarily on knee OA. The overarching aim of the OAI study was to improve public health through the prevention or alleviation of pain and disability from OA. It was believed that this would be achieved by developing a research resource available for the scientific evaluation of biomarkers for OA, and investigation of the natural history of, and risk factors for, knee OA onset and progression (68). Participants were recruited and enrolled at four study centres across America; Ohio State University, University of Maryland School of Medicine, University of Pittsburgh School of Medicine and Brown University School of Medicine. On recruitment, the study participants were divided into three different study populations: the subcohort of progression, whereby the study participants already had symptomatic knee OA, the subcohort of incidence, if they had been identified as being at elevated risk of developing knee OA; or a small group of 'reference' study participants. The 'reference' subcohort contains a small number of study participants who at baseline did not have established knee OA or an elevated risk of developing knee OA. These participants were followed up using the same protocol as those within the progression or incidence subcohorts but provide normal reference data. For the purpose of this PhD project, all available participants from the OAI are used regardless of subcohort allocation.

To be eligible for entry to the OAI, study participants had to be aged between 45 and 79 years on entry, and recruitment goals were set to ensure that the numbers of men and women within each decade of age and subcohort were approximately equal. All ethnic groups were eligible for the study, with a recruitment goal that 23% of the cohort should be made up of study participants from ethnic minority groups. For study participants to be eligible for the 'progression' subcohort, they had to have frequent knee symptoms in the past 12 months before recruitment, which were defined as "pain, aching or stiffness in or around the knee on most days" for at least one month during the past 12 months. In addition to knee symptoms, study participants also had to be diagnosed with radiographic knee OA, which was defined using the OARSI atlas. Therefore, on radiographic images of the knee an ORASI atlas grade for osteophytes of between 1 and 3 was required. For study participants to be eligible for inclusion within the 'incidence' subcohort, study participants could not have symptomatic knee OA as defined within the progression subcohort in either knee at baseline, but they had to have characteristics that placed them at increased risk of developing of developing knee OA during the OAI study. These characteristics were defined as:-

- Knee symptoms in the past 12 months, defined as either
    - pain, aching or stiffness in or around the knee on most days for at least a month
    - frequent use of medication to treat knee symptoms on most days for at least a month
    - pain, aching or stiffness in or around the knee at any time in the past 12 months but not on most days for at least one month
- Overweight, which was defined using gender and age-specific cut-points for weight
- A history of knee injury causing difficulty walking for at least a week
- Any previous knee surgery
- Family history of knee OA
- Heberden's nodes (bony swellings that form on the hands as a result of OA)
- Repetitive knee bending in current daily activities at work or during leisure activities

The number of characteristics a study participant had to satisfy was different depending on the age of the participant. The age-specific eligibility criteria for inclusion within the 'incidence' subcohort were:-

- For those aged 45 to 49 years, frequent knee symptoms, or frequent use of medications for treatment of knee symptoms, or infrequent knee symptoms; AND one or more of the other identified risk factors
- For those aged 50 to 69 years, frequent knee symptoms, or frequent use of medications for treatment of knee symptoms, or be overweight, or have two or more identified risk factors

- For those aged 70 to 79 years, frequent knee symptoms, or frequent use of medications for treatment of knee symptoms, or one or more of the identified risk factors

Those within the 'reference' subcohort had to report no pain, aching or stiffness in either knee in the past year, have no radiological findings of knee OA in either knee and have no risk factor characteristics.

As the aim of the OAI study was to investigate the natural history of knee OA, those with diagnoses of rheumatoid arthritis or inflammatory arthritis were excluded from the study. Individuals were also excluded from entry if they had already had a total knee replacement, severe joint space narrowing (OARSI joint space narrowing grade 3 or bone-on-bone), unable to provide blood sample or a current participant in a double-blind randomised controlled trial. Study participants were recruited between February 2004 and May 2006. All study participants signed informed consent, and the OAI study was approved by all the institutional review bodies involved in the study.

The OAI study data is an open access source, and all OAI data used in the thesis were downloaded from a central repository between October and December 2016. Data were stored separately for individual study visits within the central repository, and so once downloaded were then merged using STATA 14 (69) and cleaned to remove any implausible values.

4.3 Data collected

The OAI study collects a wide spectrum of data via questionnaires, interviews, clinical examinations, biological sampling and imaging. Collected data include demographics, health behaviours, knee symptoms, blood and urine samples, MRI scans and radiographs. A brief description of the information collected is described in the following, section 4.3.1 and 4.3.2.

4.3.1 Individual characteristics

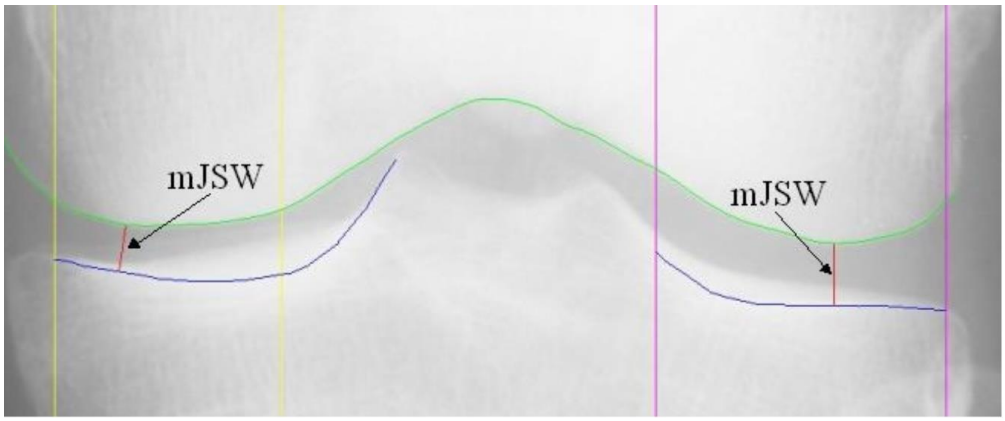
On enrolment in the OAI study, participants were interviewed to obtain demographic characteristics including age, sex, marital status, household occupancy, education, smoking history and current alcohol consumption. A medical history was also obtained, detailing such things as fracture history and co-morbidity index. Weight in kg and height in cm were recorded at each study visit during the OAI study and used to calculate each study participant's BMI. The physical activity scale for the elderly (PASE) scale was used to assess levels of activity by combining information on leisure, household and occupational activity (70), these scores range from 0 to 793 and higher PASE scores indicating greater physical activity.

4.3.2 Clinical outcomes

As the aim of the OAI study was to gain an understanding of the natural history of knee OA, a wide spectrum of knee symptoms and clinical examinations was recorded. Knee radiographs were performed at baseline, and at 12, 24, 36, 48, 72 and 96 months using the 'fixed flexion' knee radiography protocol (71) in all OAI study centres. All radiographic images were read centrally, and were examined in pairs of study participants' radiographic images but blinded to chronological order of the image. The joint space width in the knee was measured using a customized software tool. The software identifies the end of the femur bone above the knee joint, known as the femoral condyle, and the surface of the shin bone (tibial plateau). This enables the software to draw margins between the edges of the bones and measure the distance to give the joint space width, as can be seen in Figure 4-1. The minimum distance within the medial compartment (inside section of the knee) of the knee joint was recorded, and these measurements are the joint space widths that will be used from the OAI in this thesis.

If both knees met the inclusion criteria for the OAI study, both knees were entered into the observational study. However for the purpose of this thesis and in line with the SEKOIA data only one knee per study participant was included in analysis. To select the knee to be included an inclusion criteria similar to SEKOIA were used, namely the knee with the highest K&L grade and the lowest joint space width for each study participant. A summary and comparison to SEKOIA of radiographic image acquisitions elements is contained in the appendices.

Figure 4-1 Illustration of output from software used in the OAI on a digital knee radiograph



mJSW minimal joint space width

In addition to radiographic images, other clinical findings were captured through questionnaires and examinations. These outcomes included details of knee pain, aching and stiffness, knee swelling, tenderness, limitation of motion, use of medication for joint pain and arthritis, arthritis in other joints and physical disability due to knee pain and arthritis. Knee pain and function questionnaires included WOMAC (as described within chapter 3), the Knee Outcomes in Osteoarthritis Survey (KOOS) (72) and the Medical Outcomes Study Short Form 12 (SF12) (73).

4.4 The OAI descriptive statistics

4.4.1 Study visits

A total of 3469 study participants were enrolled into the OAI study. Knee joint space width measurements were recorded at baseline, and at 12, 24, 36, 48, 72 and 96 months in the OAI study and the target days between study visits varied between 365 or 730 days, as outlined in Table 4-1. The mean number of days between the initial enrolment visit and the 12-month visit was 418 days, and the maximum number of days between a study participants baseline and 12-month visits was 651 days. The mean numbers of days between the 12- and 24-month, the 24- and 36-month and the 36- and 48-month study visits were very close to the target of 365 (362, 364 and 361 respectively). Similarly, the mean numbers of days between the 48- and 72-months and the 72- and 96-month study visits were very close to the target of 730, but the variability was quite wide.

Table 4-1 Days between study visits for knee joint space width measurements in the OAI study

	Mean	SD	Minimum	Maximum	Target
Baseline - 12 months	417.4	41.0	302	651	365
12 - 24 months	362.0	42.9	151	549	365
24 - 36 months	363.9	39.4	162	571	365
36 - 48 months	360.9	35.9	182	539	365
48 - 72 months	725.3	48.7	387	1085	730
72 - 96 months	720.4	46.5	339	966	730

4.4.2 Individual characteristics

Descriptive statistics for the demographic characteristics of 3469 study participants enrolled in the OAI study are presented in Table 4-2. Forty-one percent of study participants were men and the mean (SD) age was 61.6 (9.1) years; no significant difference was found in the average age of men and women. The men were on average taller and heavier than the women, however the men and women had similar mean (SD) BMI, 29.1 (4.1) kg/m^2 and 29.0 (5.3) kg/m^2 respectively. The men were on average significantly more physically active than the women, PASE score 175.6 and 150.2 respectively. Just over 40% of the study population reported being current smokers, with a higher percentage of the men in the OAI study smoking on enrolment into the study than the women, 42.8% and 38.8% respectively. The majority of the OAI study population reported drinking some level of alcohol on entry into the study, 81.1%, and the men were more likely to drink some level of alcohol than the women.

Table 4-2 Demographic characteristics of the OAI study participants

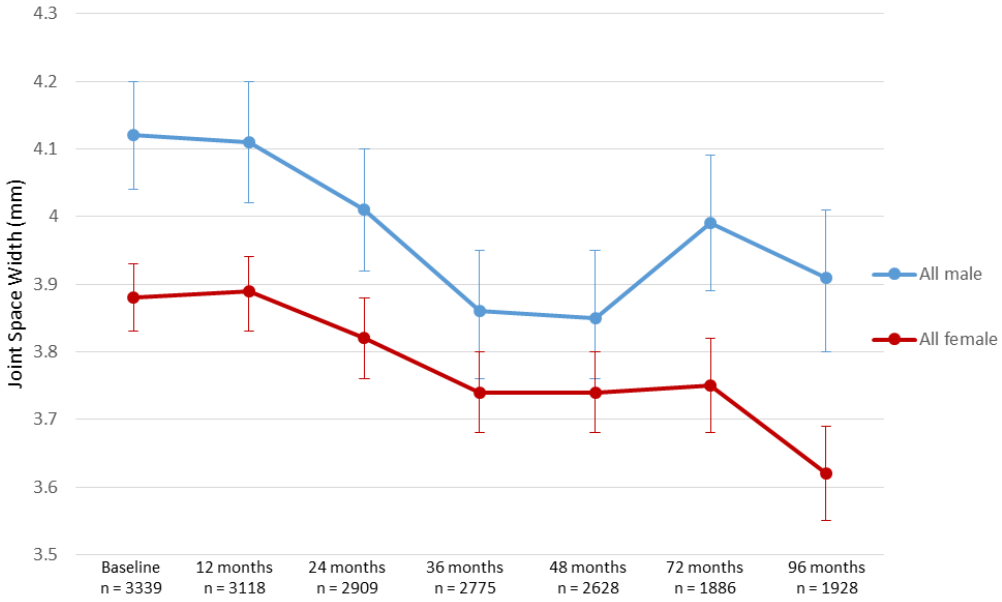
	All (n=3469)		Men (n = 1427)		Women (n = 2042)		p-value
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	
Age (years)	61.6	9.1	61.5	9.4	61.7	9.0	0.56[a]
Height (cm)	168.2	9.2	176.3	6.4	162.6	6.1	<0.01[a]
Weight (kg)	82.5	16.1	91.0	14.3	76.6	14.7	<0.01[a]
BMI (kg/m^2)	29.1	4.8	29.1	4.1	29.0	5.3	0.67[a]
PASE score	160.6	81.5	175.6	86.9	150.2	75.7	<0.01[a]
	n	**%**	**n**	**%**	**n**	**%**	
Smoking							
No	1829	53.4	707	50.1	1122	55.6	
Ex	211	6.2	99	7.0	112	5.6	
Yes	1387	40.5	604	42.8	783	38.8	<0.01[b]
Alcohol							
No	652	18.9	217	15.3	435	21.5	
Yes	2790	81.1	1202	84.7	1588	78.5	<0.01[b]

PASE – Physical activity scale for the elderly. All continuous variables were normally distributed; [a]p-value for t-test; [b]p-value for Chi-square

4.4.3 Clinical outcomes

As previously described, knee joint space width measurements were obtained from radiographic images at baseline, and at 12, 24, 36, 48, 72 and 96 months during the OAI study; the mean joint space width and their 95% confidence intervals are presented in Figure 4-2. Throughout the study, women had a consistently narrower joint space width than men. On entry into the OAI study, men had an average knee joint space width of 4.12mm (95% CI 4.04mm to 4.20mm) and women had a mean joint space width of 3.88mm (95% CI 3.83mm to 3.93mm). Average joint space width decreased through the first 36 months of the study. But between 48 and 72 months an increase in the average joint space width occurred. As an increase in joint space is only biologically plausible in very rare circumstances, it is likely the increase observed is attributable to measurement error and will be investigated in further analysis reported later in this doctoral project.

Figure 4-2 Mean joint space width (mm) and 95% confidence interval for each study visit in the OAI study by gender



The clinical characteristics at baseline of study participants in the OAI study are reported in Table 4-3. The median KOOS pain score was 86.1, and the median WOMAC pain component score was 2.0. No difference was found in the pain level reported by men and women when using either pain score. However, significant differences were found between men and women for baseline levels of stiffness and disability, reported using WOMAC, with women having higher median scores in both components. At baseline just under 16% of study participants in the OAI study were found to have no evidence of radiographic knee OA with a K&L grade score of 0. When considering a radiographic definition of knee OA, among those with a K&L grade score of 2 or more, 73% of study participants on entry into the OAI study had knee OA.

Table 4-3 Clinical characteristics of the OAI study participants at baseline

	All		Males		Females		
	Median	**Interquartile range**	**Median**	**Interquartile range**	**Median**	**Interquartile range**	**p-value**
KOOS pain score	86.1	69.4-97.2	86.1	71.9-97.2	86.1	69.4-97.2	0.66[a]
WOMAC Component score							
Pain	2.0	0.0-5.0	2.0	0.0-4.0	2.0	0.0-5.0	0.17[a]
Stiffness	2.0	0.0-3.0	1.0	0.0-3.0	2.0	0.0-3.0	<0.01[a]
Disability	5.0	0.0-16.0	4.5	0.0-14.0	5.3	0.0-17.0	<0.01[a]
WOMAC global score	8.3	1.0-22.9	8.0	1.0-20.0	9.0	1.1-24.3	<0.01[a]
	n	**%**	**n**	**%**	**n**	**%**	
Kellgren and Lawrence Score							
0	545	15.9	220	15.5	325	16.1	
1	374	10.9	136	9.6	238	11.8	
2	1343	39.1	496	34.9	847	42.0	
3	883	25.7	395	27.8	488	24.2	
4	292	8.5	173	12.2	119	5.9	<0.01[b]

All continuous variables were skewed; [a]p-value for Wilcoxon-ranksumt; [b]p-value for Chi-square

4.4.4 Change in joint space width

Crude differences between continuous joint space measurements were calculated between study visits in the OAI study. The average change in knee joint space width between baseline and 12 months was -0.03mm, which means on average study participants' knee joint space width reduced by 0.03mm between baseline and 12 month visit. The average change between 48 to 72 months, and 72 and 96 months looks larger than across previous study visits, Table 4-4, below but these mean changes in joint space width are across a 24 month time frame rather than the 12 month time frame of other pairs of study visits. Therefore if considering these changes in knee joint space width over a 12 month period the change in joint space widths would be comparable to the other mean changes over the same duration.

Over each time point of the OAI study, on average, men had a greater reduction in knee joint space width than women. However, the differences in the magnitude of the change between men and women were only large enough to reach statistical significance between the 12 and 24 months study visits, and the 24 and 36 month study visits.

Table 4-4 Change in knee joint space width between the OAI study years by gender

	All			Male			Female			p-value
	n	Mean	SD	N	Mean	SD	n	Mean	SD	
Change in joint space width between:-										
Baseline and 12 months (mm)	3091	-0.03	0.54	1272	-0.04	0.60	1819	-0.02	0.49	0.38[a]
12 and 24 months (mm)	2765	-0.12	0.51	1160	-0.15	0.56	1605	-0.10	0.48	0.01[a]
24 and 36 months (mm)	2594	-0.13	0.49	1078	-0.16	0.51	1516	-0.10	0.47	<0.01[a]
36 and 48 months (mm)	2458	-0.06	0.52	1029	-0.06	0.48	1429	-0.06	0.54	0.80[a]
48 to 72 months (mm)	1781	-0.11	0.55	727	-0.12	0.54	1054	-0.11	0.56	0.64[a]
72 to 96 months (mm)	1774	-0.14	0.56	729	-0.13	0.54	1045	-0.14	0.57	0.74[a]

All continuous variables were normally distributed; [a]p-value for t-test

4.5 Summary

In this chapter the study design of the longitudinal observational OAI study was outlined, and the large battery of information collected over the 96 months of the study were described. A statistical description of the number of days between study visits, demographic and lifestyle characteristics and clinical measurements were presented. During the 96-month duration of the OAI study, an increase in the average knee joint space was observed between 48 and 72 months and increases in knee joint space width are typically biological implausible. Statistical methods to explore and understand these measurement errors in knee joint space width will be examined in the chapters to follow.

Chapter 5: Reliable change index

5.1 Overview of chapter

Longitudinal data are increasingly being collected within both clinical and research environments. Within a clinical environment this allows the tracking of chronic diseases. In a research setting this allows for the study of what characteristics or risk factors may be associated with disease deterioration. This chapter will explain the reliable change index and present results from the novel application of the reliable change index in estimating OA progression within SEKOIA and the OAI study.

5.2 Methods

When assessing change the easiest method is to calculate the difference between two time points,

$$d_i = y_{2i} - y_{1i},$$

where i is a patient/study participant, y_{2i} is the measurement of interest at time point 2 and y_{1i} is the measurement of interest at time point 1. Traditionally, if these change measurements have been shown to be normally distributed, statistical techniques such as paired t-tests are used to test for group differences (17). If the change measurements are not normally distributed, then non-parametric rank comparisons are used to compare change (74). However, such analyses will only reveal differences in means between groups, so for example will only indicate whether a population mean joint space has changed over time; giving no information on changes within individuals. In many studies, the researchers aim to relate change in joint space over time to modifiable risk factors, or to gain evidence that a treatment is effective. In such studies it would be appropriate to investigate such associations in relation to individual change. An individual's change is the observed difference between two measurements taken at different times, and this may be dominated by measurement error in either or both measurements. The more variable a measurement, the greater the chance of it appearing that change has occurred, even though little of the observable change may be attributable to true

change. For example, a study by Buckler et al has shown that height can fluctuate by as much as 2.8cm during the day, with people being taller in the morning than the evening due to the compression of joints that happens in the spine during the day (75). Therefore it could appear that a study participant had a 2.8cm change in height when in fact the change could be attributable simply to variation within the measurement if the observations had been obtained at different times of the day.

The reliable change (RC) index is a statistical method for identifying whether an observed change within an individual is meaningful in the presence of measurement error (76). The RC index provides a method of determining whether an individual's change is within the bounds of measurement error or whether the change exceeds the bound. The greater the error in the measurement under investigation, the lower the likelihood that an observed change is attributable to a true effect.

The RC index was developed in 1991 by Jacobson and Truax (76), and is a popular method in the field of clinical psychology (77). The principle behind the index is to determine whether the magnitude of change observed in a study participant provides evidence of change, i.e. the change observed is more than could be explained by the unreliability of the measure. Several variations of the RC index have been proposed (78), but all variations identify the extent to which study participants' current measurements differ from their previous measurements. All the variations of the RC index follow the same fundamental expression:

$$RC\ index = \frac{Y - Y'}{SE}$$

where Y is the study participant's measurement at the later time point, Y' represents the predicted measurement for the study participant at the later time point of interest and SE is the standard error of the measurement. The different approaches to the RC index vary in how they determine the different elements of the RC index principle, and the assumptions that are required to be met for the results to be valid. The version of the RC index that will be explored

within this thesis is the version developed by Christensen and Mendoza (79), and the RC index formula that produces a standardised score for each study participant is shown in Equation 1 below.

Equation 1 Calculation of reliable change index scores

$$RC\ index = \frac{Y_{2i} - Y_{1i}}{\sqrt{S_1^2 + S_2^2 - 2S_1S_2r_{x_1x_2}}}$$

The predicted score is represented by the study participants' measurement at time point 1, Y_{1i}, and the same study participants actual measurement at time point 2 is Y_{2i}. The standard error of the difference between the observations is derived using S_1^2 and S_2^2 which are the variance of the measurements at time points 1 and 2 respectively, and S_1 and S_2 are the standard deviations of the measurements at time points 1 and 2 respectively. $r_{x_1x_2}$ is the Pearson's correlation coefficient between the measurements at the two time points.

As highlighted by Maassen the assumption of the traditional RC index developed by Jacobson and Truax, Equation 2, is that the variances of the measurements at the differing time points is equal (80). However, even if on visual inspection the variances of the measurements across time points appear to be equal, formal statistical testing may often indicate departure from equality. However the adaptation of the RC index developed by Christensen and Mendoza does not require equal variance in measurement time points this allows for departure from this assumption.

Equation 2 Jacobson and Truax reliable change formula

$$RC\ index = \frac{X_2 - X_1}{\sqrt{2(S_1\sqrt{1 - r})^2}}$$

The RC index calculation yields a standardised z-score (i.e. the scores have a mean of 0 and a SD of 1). Therefore, following the convention of using 5% level of statistical significance an RC index score of in excess of 1.96 denotes in magnitude a statistically significant reliable change, indicating that the change observed reflects more than the fluctuation in the measurement procedure. Each study participant's RC index score can be categorised into one of three categories: an increase (RC index >1.96), a decrease (RC index<-1.96) or stable (RC index between -1.96 and 1.96).

All RC index analyses were undertaken using STATA 14 (69).

5.3 RC index results in SEKOIA

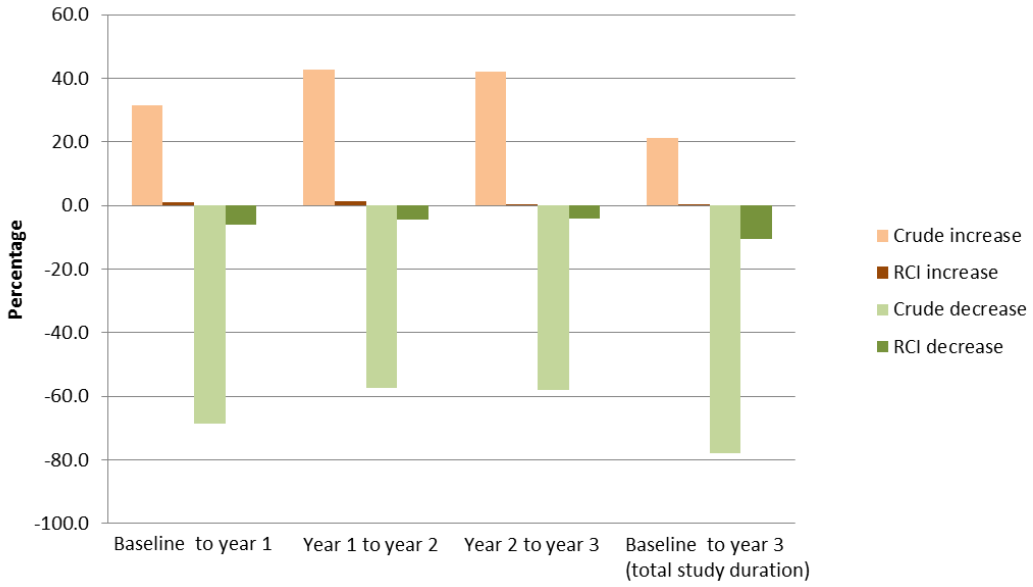
Of the 465 study participants who had knee joint space width measurement at baseline and year 1, 319 (68.6%) had either no change or an apparent decrease in joint space width over the year when assessing using crude change, and this figure was nearly 60% between the remaining study years, Figure 5-1 and Table 5-1. An RC index value was calculated for the differences in measurements between each SEKOIA study visit for each study participant. These indicated that 28 (6.0%) study participants had an RC index less than -1.96 when assessing the observed difference between baseline and year 1, Figure 5-1 and Table 5-1. Thus it is only in these 28 study participants that a statistically reliable decrease in joint space width was observed that was larger than would be expected through fluctuation in the joint space measurement or measurement error. A similar pattern was observed between year 1 and year 2, and between year 2 and year 3, with 18 (4.5%) and 13 (4.0%) respectively having a statistically significant reliable decrease in knee joint space measurements between these years.

Table 5-1 Changes in joint space widths and RC index results in SEKOIA

	Baseline to year 1		**Year 1 to year 2**		**Year 2 to year 3**		**Baseline to year 3 (total study duration)**	
	N		**N**		**N**		**N**	
Total in study	465		400		329		336	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Crude increase (>0mm)	146	31.4	171	42.8	138	41.9	74	22.0
Crude decrease (<0mm)	319	68.6	229	57.3	191	58.1	262	78.0
RCI increase	5	1.1	5	1.3	3	0.9	1	0.3
RCI decrease	28	6.0	18	4.5	13	4.0	36	10.7

Conversely, 146 (31.4%) study participants were identified as having an absolute increase in knee joint space width measurements between baseline and year 1, and 171 (42.8%) study participants were identified as having an increase between year 1 and year 2, and 138 (41.9%) between year 2 and year 3. Using the RC index calculation, 5 study participants (1.1%) had an RC index score greater than 1.96 when the observed differences between baseline and year 1 were assessed. These five study participants are of note as they appear to have had an increase in joint space width greater than can be explained by the fluctuations of an imprecise measurement procedure. Use of the RC index for measurements between year 1 and year 2, and between year 2 and year 3 indicated that only 5 (1.3 %) and 3 (0.9%) study participants respectively had an increase in joint space width.

Figure 5-1 Percentages of crude differences and RC index results in the SEKOIA study



No study participants were found to have a statistically significant reliable increase or decrease across all the following time periods: between baseline and year 1, between years 1 and year 2, and years 2 and year 3.

Of the 336 study participants with measurements at baseline and year 3, 262 (78%) had a crude decrease in joint space width measurements over the 3-year duration, whereas only 36 (10.7%) were identified as having a reliable decrease when using the RC index. When considering those study participants who were identified as having an increase in joint space width measurements between baseline and year 3 using a crude difference 74 (22%) study participants were identified, whereas only 1 (0.3%) was identified as having an increase when using the RC index score.

Notably the percentage of study participants identified as having a reliable increase using the RC index between baseline and year 3 was less than the percentage across all other combination of study time periods. But the percentage of study participants identified as having a statistically significant reliable decrease in joint space width was much greater between baseline and year 3 than when considering change across individual study years.

A magnitude of change in millimetres (change threshold) was calculated by transforming the RC index results to give a change in joint space width above which it can be said that statistically reliable change occurred. Between baseline and year 1, change greater than 0.91mm was considered statistically reliable using the RC index to calculate a magnitude of change, while between years 1 and 2 and years 2 and 3 the magnitude of change threshold was calculated as 0.85mm and 0.88mm respectively. The largest threshold for statistically reliable change was seen between baseline and year 3 where a change in joint space width had to reach 1.23mm.

5.4 RC index results in the OAI

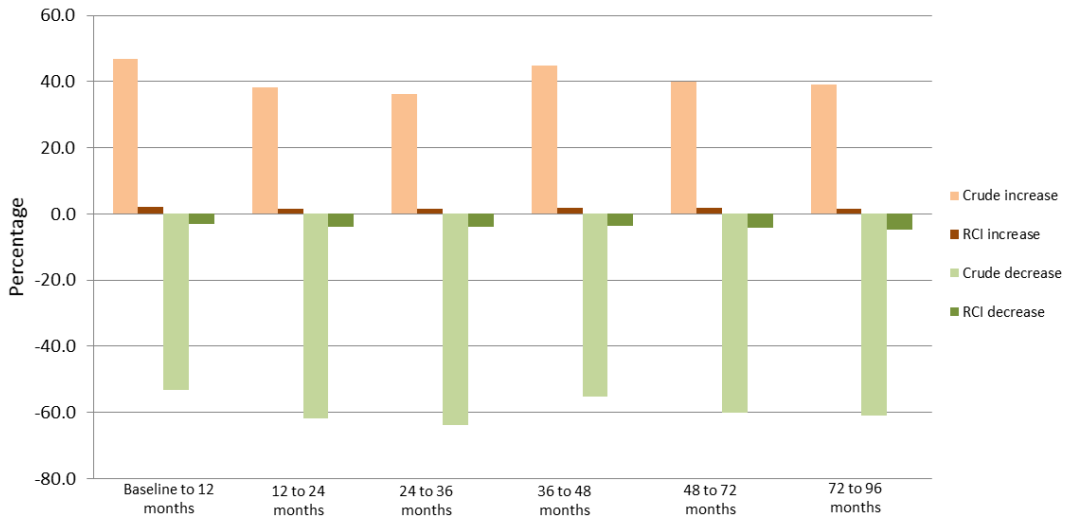
Table 5-2 contains the RC index and crude change results across all pairings of study visits in the OAI. There were 3091 study participants who had knee joint space width measurements at both the baseline and 12 month study visit, and of those 1646 (53.3%) had either no change or an apparent decrease in joint space width when assessing change using crude differences. The RC index indicated that only 96 (3.1%) of study participants has a statistically reliable decrease in knee joint space width between baseline and 12 month, Figure 5-2. A similar pattern was observed between all the OAI study visits; between 12 and 24 months, 24 and 36 months, and 36 and 48 months; 111 (4.0%), 101 (3.9%) and 86 (3.5%) of study participants had a statistically reliable decrease in knee joint space width respectively. The greatest percentage of study participants who were identified as having a reliable decrease in knee joint space width occurred for the change between the 72- and 96-month study visits, at 4.9% (n=87).

Table 5-2 Changes in joint space widths and RC index results in the OAI

	Baseline to 12 month		**12 to 24 month**		**24 to 36 month**		**36 to 48 month**		**48 to 72 month**		**72 to 96 month**		**Baseline to 96 month (total study duration)**	
Total in study	**N** 3091		**N** 2765		**N** 2594		**N** 2458		**N** 1781		**N** 1774		**N** 1918	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Crude increase (>0mm)	1445	46.7	1054	38.1	942	36.3	1100	44.8	711	39.9	691	39.0	476	24.8
Crude decrease (<0mm)	1646	53.3	1711	61.9	1652	63.7	1358	55.2	1070	60.1	1083	61.0	1442	75.2
RCI increase	63	2.0	42	1.5	40	1.5	47	1.9	32	1.8	25	1.4	11	0.6
RCI decrease	96	3.1	111	4.0	101	3.9	86	3.5	76	4.3	87	4.9	178	9.3

When considering increases in knee joint space width, 1445 (46.7%) study participants were identified as having a crude increase between baseline and the 12-month study visit. Similar patterns were observed between all other study visits, but never less than 36.3% of study participants in the OAI had a crude increase in joint space width when assessed using crude differences. Using the RC index, the percentage of study participants identified as having an increase was dramatically reduced. Only 63 (2.0%) study participants were found to have a statistically reliable increase in knee joint space width between baseline and the 12-month visits, indicating an increase in knee joint space width greater than can be explained by the variabilities in the measurements between both study visits. Similar levels of study participants were found to have an increase in joint space width across all the remaining study visits, as can be seen in Figure 5-2.

Figure 5-2 Percentages of crude differences and RC index results in the OAI



Of the 1918 study participants with measurement at baseline and the 96th month study visit, 1442 (75.2%) had a crude decrease in knee joint space width measurements over the 8 year study duration. Whereas only 178 (9.3%) were identified as having a statistically reliable decrease when using the RC index. Of the 476 study participant who were identified as having a crude increase in joint space width over the 8 year study duration, only 11 (0.6%) were identified as having a reliable increase in joint space width using the RC index.

No study participants in the OAI were consistently identified as having a statistically reliable increase or decrease across all the OAI study visit intervals.

Table 5-3 Magnitude of change in the OAI study

	Magnitude of change (mm)
Baseline to 12 month	1.06
12 to 24 month	1.01
24 to 36 month	0.95
36 to 48 month	1.01
48 to 72 month	1.08
72 to 96 month	1.09
Baseline to 96 month	1.73

The RC index results were used to calculate a magnitude of change in millimetres, a change in joint space width threshold above which it can be said that statistically reliable change occurred, Table 5-3. Between baseline and 12-month study visit the threshold of change was calculated to be 1.06mm, therefore if a study participants had a change in joint space width greater than 1.06mm a statistically reliable change had occurred. The largest threshold for statistically reliable change was seen over the total 8-year duration of the OAI study, where a change of 1.73mm was calculated to be the threshold for reliable change. Using the RC index to calculate a magnitude of change threshold in millimetres across the remaining different combinations of individual study visits under consideration the magnitude varied between 0.95mm and 1.09mm.

5.5 Summary

In SEKOIA, joint space widths were monitored yearly and in the OAI joint space widths were monitored every 12 months for the first 48 months and then every 24 months until the study participants had been followed-up for 96 months.

If the measurements of crude differences in SEKOIA were taken in isolation it would lead to the conclusion that, between baseline and year 1, 70% of those study participants under observation had a worsening of their knee OA. However use of the RC index between baseline and year 1 in SEKOIA indicates that only 6.0% (28) of study participants had a statistically reliable decrease in observed joint space width that was larger than would be expected through measurement error between joint space measurements. Therefore considerably fewer study participants than initially highlighted through simple differences can reliably be considered to have had a decrease in joint space. Similar patterns were observed between year 1 and year 2, and between year 2 and year 3 in the SEKOIA study.

Conversely, around 31% of study participants between baseline and year 1 in SEKOIA, and approximately 42% of study participants between year 1 and year 2, or between year 2 and year 3 in SEKOIA were identified as having an absolute increase in joint space width. As real increases are extremely rare, this shows the impact of measurement error if crude differences are assessed without taking any account of measurement error. Use of the RC index identified a markedly lower number of study participants having an increase in joint space width in SEKOIA, only 5 study participants (1.1%) between baseline and year 1, 5 (1.3%) between years 1 and 2, and only 3 (0.9%) between years 2 and 3.

In the OAI, around 60% of study participants had a crude decrease in knee joint space width between study visits. The highest percentage of study participants found to decrease between study visits was 63.7% between the 24- and 36- month study visits. However implementation of the RC index indicated that considerably fewer study participants had a statistically reliable decrease in knee joint space width, for example only 3.1% (96) study participants were identified between baseline and 12 month OAI study visit as having a statistically reliable decrease. Similar proportions were observed between all other OAI study visits.

When considering increases in joint space widths within the OAI, around 40% of study participants had a crude increase in joint space width between the pairs of study visits under consideration. Use of the RC index dramatically reduced the number of study participants that were identified as having an increase, for example between 72 and 96 month study visits only

25 (1.4%) were identified as having a statistically reliable increase in joint space width after accounting for measurement error compared with the 691 (39.0%) that were identified using crude differences.

Using the RC index to account for measurement error, it was possible to calculate a magnitude of change threshold that gives a change figure above which it can be said that statistically reliable change had occurred. Within the SEKOIA study the magnitude of change threshold varied between 0.85mm and 0.91mm for different combinations of individual study years under consideration. While a change of 1.23mm was considered statistically reliable when assessing change over the 3 year study duration of SEKOIA. Interestingly, between consecutive study visits in the OAI the change threshold was found to be reasonably consistent, ranging between 0.95mm and 1.09mm, and similar to the thresholds calculated for the SEKOIA study. The magnitude of change threshold across the 96-month study period of the OAI was 1.73mm.

5.6 Discussion

The aim of this chapter was to assess the effectiveness of the RC index as a novel approach to estimating OA progression, which takes account for the presence of measurement error through longitudinal assessment of continuous knee joint space widths. When compared with simple differences, implementation of the RC index dramatically reduced the proportions of study participants that were identified as having statistically reliable change after accounting for measurement error. The more variability in the measurement under consideration, the greater the chance of it appearing that change has occurred. To date, no studies have been identified that apply the RC index methodology within musculoskeletal research not only to monitor joint space measurements but also assess disease deterioration. The RC index has, however been successfully applied within psychological and neurological research. For example Ferguson et al used the RC index to determine clinically significant change between pre- and post-intervention Short Form-36 (SF-36) scores. These scores are continuous and provide a measure of patient health (77). In the study by Ferguson and earlier studies by Kendall it was highlighted that the RC index is an important technique to determine change, as assessing crude differences alone does not provide reliable information about whether an intervention had a clinically meaningful effect (77, 81).

There are other statistical techniques and metrics currently used within musculoskeletal research to identify whether change has been significant such as the standard error of the measurement (SEM) or SRM. However neither of these techniques are appropriate for assessment at the individual level, rather they give information about change at the population level. Therefore an advantage of using the RC index rather than calculating SEM or the SRM is that reliability of an individual study participant's change can be determined, and additionally the estimate of the standard error used within the RC index calculation can be used to quantify the joint space width change above which change could be considered statistically reliable, within the population under consideration.

Although the RC index has its merits there has also been much debate and criticism of the technique (78, 82). One of the major criticisms is that, although all variations of the calculation can be simplified to the same fundamental expression, each approach differs slightly in how the elements of the RC index are calculated. For example, the RC index developed by Jacobson and Truax which is often referred to as the 'original' RC index (76) requires an externally-derived test-retest reliability coefficient to be able to calculate the standard error, and this version of the RC index also assumes equal variance in the measurements at both time points. Hinton-Bayre has made a comparison of the different RC variations but there is currently no consensus as to which RC index is preferred (78). However, as explained within the methods section of this chapter, the variety of the RC index used within this thesis is the formula proposed by Christensen and Mendoza. The advantage of using this RC method is that equal variance in measurements between time points does not have to be assumed, and all elements of the formula are calculated internally.

The use of the RC index within this thesis aims to identify true progression of joint space narrowing in those with knee OA from measurement error. As previously mentioned the more variable a measurement the more likely it is that change is due to measurement error rather than true change. Therefore the RC index will identify fewer people with statistically reliable change when the measurement is more variable, as demonstrated by application of the RC index within the SEKOIA and the OAI studies. The SEKOIA data had less variability in knee joint space widths than the OAI as the SD's for knee joint space width measurements in SEKOIA ranged from 0.83 to 1.04 while those in the OAI ranged from 1.28 to 1.40. The RC index performed as

expected, as, on average, the proportion of study participants identified as having a statistically reliable decrease in joint space width was lower in the OAI than in SEKOIA.

No SD is too large to enable the application of the RC index within a dataset, and the same holds true for change. No change in measurement is too large to enable a RC index score to be calculated. However a disadvantage of the RC index methodology is that although it identifies individual statistically reliable change regardless of the magnitude of the SD or change, the resulting RC index score does not provide any information about the reason for change. Therefore the RC index is unable to distinguish actual disease progression from large changes in joint space widths due to errors in the process of obtaining the measurement, such as variability in the radiographic positioning of the knee. Although the use of joint space width longitudinally is the current gold standard for monitoring disease progression, previous studies have shown that inconsistent knee positioning during radiographs can cause a systematic shift in joint space width (83) and so any change in joint space width may be attributable to change in positioning of the knee during radiograph rather than disease progression. However previous studies have shown that the use of inter-margin distance is optimal for assessing joint space width and reducing variation in joint space width due to knee positioning (84). The minimal joint space width at the medial tibiofemoral compartment, the inter-margin distance was measured in SEKOIA annually from radiographs obtained under a strict study protocol (66, 85). Therefore the SEKOIA data were collected with all the associated safeguards around methodology and training, and all radiographs were assessed by one reader thus reducing measurement error. The OAI radiographs were also obtained using the same fixed flexion frame as used within the SEKOIA study, and the minimum joint space width was measured in the medial tibiofemoral compartment. Thus it is likely that the joint space measurements collected during SEKOIA and the OAI contain less measurement error than routine clinical measurements. As there are different radiographic techniques that can be used to obtain knee radiographs it would be important to assess the use of the RC index in data where other methods have been used, though this is not within the remit of this thesis.

A further assumption of the RC index is stability of measurements between time points and, to date, RC index scores have not be used to assess deterioration within a measurement. The natural disease progression of OA is a slow process, often taking many years. Therefore the

assumption would be that on an annual basis little or no change in joint space width in a study participant would have occurred. This can be seen within the RC index results in this thesis, as the greater the duration between joint space width measurements the greater the proportion of study participants who were identified as having statistically reliable change. This in part may be explained by the greater time for disease progression to have occurred, allowing potentially greater deterioration, which can be more easily distinguished from the measurement error that is still present.

Although there are many different methods to assess change, either at the population or individual level, there is no 'gold standard' method that has been identified to assess statistical significant change over time, especially when assessing OA progression. Therefore there is no gold standard against which to compare the RC index, but the use of this novel approach does allow for identification of statistically reliable individual change after accounting for measurement error, unlike the calculation of simple crude differences. The formula is also simple enough that summary statistics from the study population are sufficient to allow assessment of individual study participant reliable change. In chapters 6 and 7, the established yet more complex statistical modelling techniques of linear mixed effect modelling and Bayesian modelling will be used to model change in knee joint space width as an outcome while accounting for measurement error. In chapter 8, all these statistical methods used for determining change will be compared to identify the strengths and weakness of each technique for monitoring change and identifying measurement error.

Chapter 6: Linear mixed effect modelling

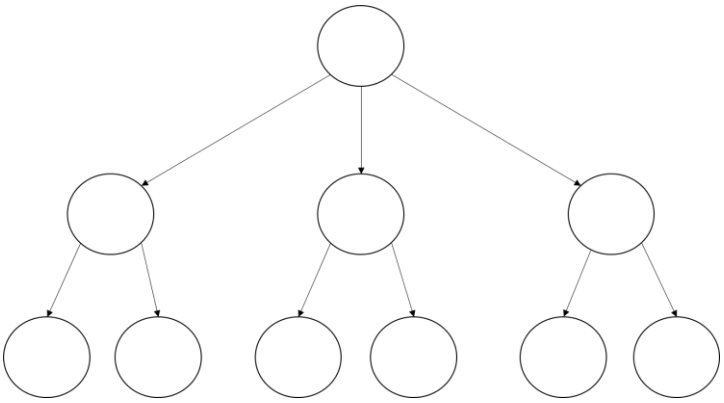
6.1 Overview of chapter

In both a research environment and a clinical setting where knee joint space width has been monitored over multiple time points, full use is rarely made of all repeated measures. Often repeated measures are summarised into one change variable, as discussed in chapter 2. In the previous chapter, chapter 5, the simple statistical method of the RC index was assessed to determine its application in monitoring OA progression, always focusing only on differences between two points. In this chapter the more complex but established frequentist technique of linear mixed effect modelling is explored. This method can account for the clustered nature of the repeated measures data. The results of analysis completed using this technique in both SEKOIA and the OAI are presented, followed by a discussion of the results and the statistical technique.

6.2 Methods

The focus of this doctoral thesis is monitoring of knee joint space width measurements of individuals over time, and the technique described within this chapter will be implemented within the SEKOIA and the OAI datasets. Both datasets of interest within this thesis have a data structure with multiple levels, as there is the clustering of repeated knee joint space width measurements over time within study participant. Although the focus of the thesis is clustering arising from repeated measures, it should be noted that there are other ways in which nested data structures can occur such as multiple observations within a geographical area such as hospitals. These types of nested data structures are often said to have a hierarchical or multilevel structure, and this is illustrated graphically in Figure 6-1. Within this illustration each node represents a data point of interest, for example the lower nodes could be individual children's heights that are clustered within schools, the middle nodes, within an education authority, the top level node.

Figure 6-1 Graphical representation of hierarchical data structure



It is important to recognise the clustered nature of the data as it is expected that two observations belonging to the same group may be more alike than two observations drawn from different groups. This is especially so in the context of this PhD thesis as the repeated measurements of interest contain a time element, thus there is an ordering to the repeated measurements for each study participant. This means that not only are measurements more likely to be similar within study participants but there will be an ordering of the repeated measurements. Ignoring such correlations between observations can lead to incorrect inferences, as observations are no longer independent. For example, it could lead to the inference that a characteristic of interest has a 'real' association with the outcome when in fact the effect could be ascribed to chance, i.e. there is a high risk of Type 1 error if correlation between observations is ignored. Therefore dependencies between observations need to be accounted for during analysis to avoid incorrect statistical inferences. In a frequentist framework, multilevel models can be used to take account of the lack of independence between observations. Multilevel models may take many different names and these include hierarchical linear models, linear mixed effect models, random effects models and variance components models (86). However regardless of terminology used, if the clustered nature of the data is ignored and multilevel models are not used to analyse such data, the standard errors of model estimates will be underestimated and consequently confidence intervals will be too narrow and p-values too small. Multilevel modelling techniques allow for the clustering of observations by incorporating the nature of between-group variability and the effects of group-level characteristics on individual outcomes. The frequentist statistical modelling technique that will be applied within this PhD thesis is linear mixed effect (LME) modelling. The basic principle behind this type of multilevel model in longitudinal analysis is to allow regression coefficients to

differ between participants, thus allowing participant-level effects to be determined in addition to exploring sample level effects. The simplest form of LME modelling is when only the intercept is allowed to vary, as shown in Equation 3.

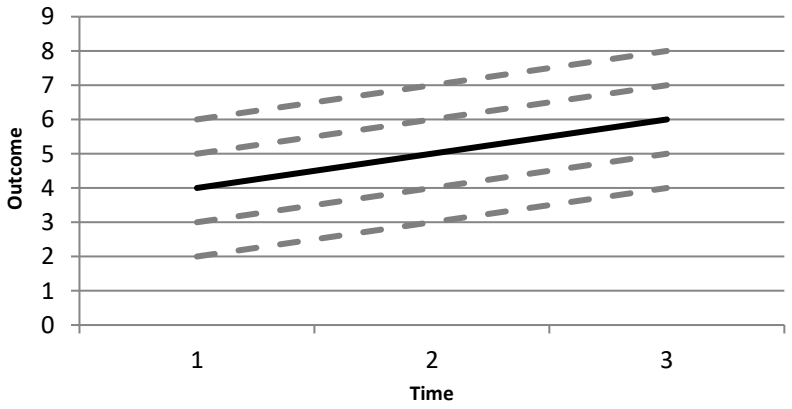
Equation 3 Linear mixed effect random intercept model

$$Y_{ij} = \beta_{0i} + \beta_1 X_{ij} + \varepsilon_{ij}$$

$$\beta_{0i} = \beta_0 + U_{0j}$$

In this equation, Y_{ij} is the measurement for observation i in the jth study participant, β_0 is the intercept and can also be described as the overall mean of Y across all study participants. β_1 is the slope, X_{ij} an indicator variable for the occasion of measurement (time variable) and U_{0j} is the group-level residuals, which can also be referred to as group random effect. ε_{ij} is the error for participant i in group j. This model can be represented graphically as seen in Figure 6-2, where the solid black line is the overall population average and the grey dashed lines are for individual participants. The mean of Y for group j, the intercept, is $\beta_o + U_{oj}$ which gives a different intercept value for each group, i.e. a random intercept.

Figure 6-2 Graphical representation of LME random intercept model



In two level hierarchal LME models the residual values are split into two components corresponding to the two levels within the data structure, the group-level residuals, U_j, and the

individual level residuals, ε_{ij}. As with single-level regression models, the residuals are assumed to follow normal distributions with a mean of zero, $U_j \sim N(0, \sigma_u^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_e^2)$. This then allows the overall variance to be separated into the two components, the between-group variance σ_u^2 and the within-group σ_e^2, where group in the context of this thesis refers to study participants. The between-group variance summarises the departures of group means from the overall population mean, and the within-group variance gives an indication of the variability of observations within a group. For example, if the within-group variance were zero this would indicate that all observations were the same within the same group. Whereas if the between-group variance were zero this would indicate that no differences existed between groups, and all the groupings had the same mean as the overall population mean. Both of these scenarios are highly unlikely and therefore to measure the proportion of the total variance that is due to differences between groups the variance partition coefficient (VPC) can be calculated, Equation 4.

Equation 4 Variance Partition Coefficient

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

The VPC takes values between 0 and 1, where 0 represents no group differences and 1 represents no within-group differences.

The LME random intercept model assumes that the relationship between each group and the outcome is the same, i.e. the slope is fixed across groups. However this can also be allowed to vary between groups leading to a model that allows for random intercept and random slopes (87), as seen in Equation 5 often described as LME random intercept and random slope models.

Equation 5 Linear mixed effect random intercept and slope model

$$Y_{ij} = \beta_{0i} + \beta_{ij} X_{ij} + \varepsilon_{ij}$$

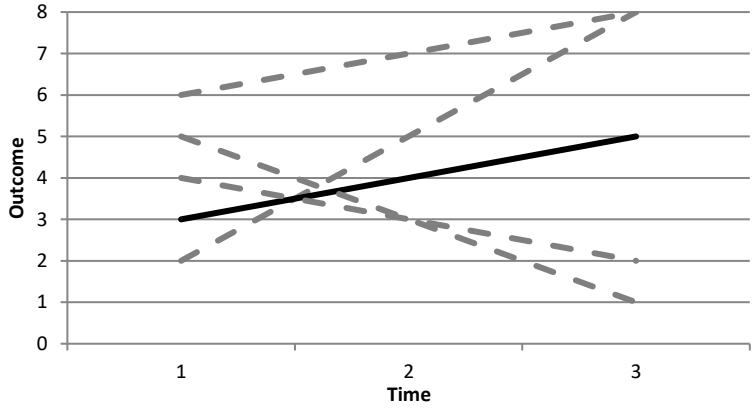
$$\beta_{0i} = \beta_0 + U_{0j}$$

$$\beta_{ij} = \beta_1 + U_{1j}$$

In this equation Y_{ij} are measurements for observation i in the jth group, β_{0i} is the random intercept and is obtained by adding the overall intercept term to the group level intercept. β_{ij} is the random slope term, which is obtained by adding the β_1 to the individual level slope residuals, to give a different slope value for each group. This model is represented graphically in Figure 6-3 where each group's variable has a different rate and different baseline value. The grey dashed lines in this figure represent individuals, while the solid black line represents the overall population average.

This doctoral thesis is concerned with trajectories of knee joint space narrowing and so, within the application of LME random intercepts and random slope models. In this thesis, y_{ij} represents the joint space measurement at time *i* for individual study participant *j*, β_0 is the average joint space width for all study participants, β_1 is the average change in joint space width for one unit of time for all study participants.

Figure 6-3 Graphical representation of LME random slope model



The random effect residuals, u_{0j} and u_{1j}, within the LME random intercept and random slope model are assumed to follow a normal distribution with a mean of zero and variances of σ_{u0}^2 and σ_{u1}^2 respectively, as shown in Equation 6.

Equation 6 Random effect terms for linear mixed effect model

$$U_{0j} \sim Normal(0, \sigma_{u0}^2)$$

$$U_{1j} \sim Normal(0, \sigma_{u1}^2)$$

The covariance σ_{u01} is the covariance between the group intercepts and the slope. If the estimates obtained for β_0 and β_1 are positive then a covariance value that is positive implies that groups with a high intercept residual tend to have high slope residuals, i.e. a group with a high intercept value, greater than the population average, tends to have a slope that is steeper than average. Conversely, if the same positive estimates of β_0 and β_1 were obtained but the covariance value was negative, it would imply that those groups with high intercept residuals tend to have small slope residuals. Thus groups with a larger than average intercept tend to have shallower slopes than average.

It is important once an LME model has been fitted to carry out model diagnostics to check the distributional assumptions of the residuals and check that there are not any unusual observations that affect the model fit. The techniques for model diagnostics are well established for standard linear regression models (88) but, due to LME modelling being more complex than standard single level models, model diagnostics are more difficult to perform and interpret. The modelling assumptions of LME models are similar to most statistical regression modelling assumptions, which are that residuals should have a constant variance and are normally distributed. As LME models are multilevel, the residuals' assumptions must hold true for both levels, so residuals at level 1 and level 2 must be assessed to ensure the assumptions of normality and constant variance hold true. Additionally when using LME models there is the assumption of linearity, namely that the relationship between the dependent and independent variables must be a linear. During the LME modelling process, fitted values, conditional residuals, standardized conditional residuals and best linear unbiased predictors (BLUPs) can all be calculated to enable both formal and informal checking of the model fit and residual assumptions, and for any outliers.

The fitted values are the estimated responses from the model, the conditional residuals are the difference between the observed value and the conditional predicted value. Conditional residuals are at level 1 of the LME model and so there is a conditional residual calculated for every data point within the dataset. Standardized conditional residuals are the conditional residuals rescaled using the residual standard deviation. BLUPs reflect the difference between the predicted responses for the level 2 units and the population average (89). As the focus of this thesis is to model change in knee joint space width measurements over time, the level 1 residuals of interest will be at the knee joint space width measurement level, ε_{ij}, and the level 2 residuals concern study participants, u_{0j} and u_{1j}. The assumption of constant variance for the level 1 residuals can be determined by plotting the standardised conditional residuals against the fitted values, and, if the assumption of constant variance held, a completely random pattern in the scatter graph would be observed. To assess whether the residuals are normally distributed Normal Q-Q plots can be used at both levels using the conditional residuals and the BLUPs. Finally to assess model fit, the fitted values can be plotted against the observed values. If the two values should be similar, indicating that the model is accurately predicting new values, thus the model fits the observed data.

By fitting a LME model to both the SEKOIA and the OAI data the assumption of linearity has been made, i.e. it has been assumed that the relationship between time and joint space width is linear. Thus it is assumed that time has an additive effect, and joint space width changes by the same amount for every additional year.

All LME modelling and residual diagnostics for this thesis were undertaken using STATA 14 (69).

6.3 LME Results in SEKOIA

Only 324 study participants of the 559 from the placebo arm of SEKOIA had joint space width measurements at all of the four study visits. However this type of longitudinal analysis allows study participants who don't provide an outcome at the final time point but are observed at an earlier point to contribute to the analysis. Therefore to ensure no selection bias occurred, data from all 559 study participants within the placebo arm were used in the LME modelling. As

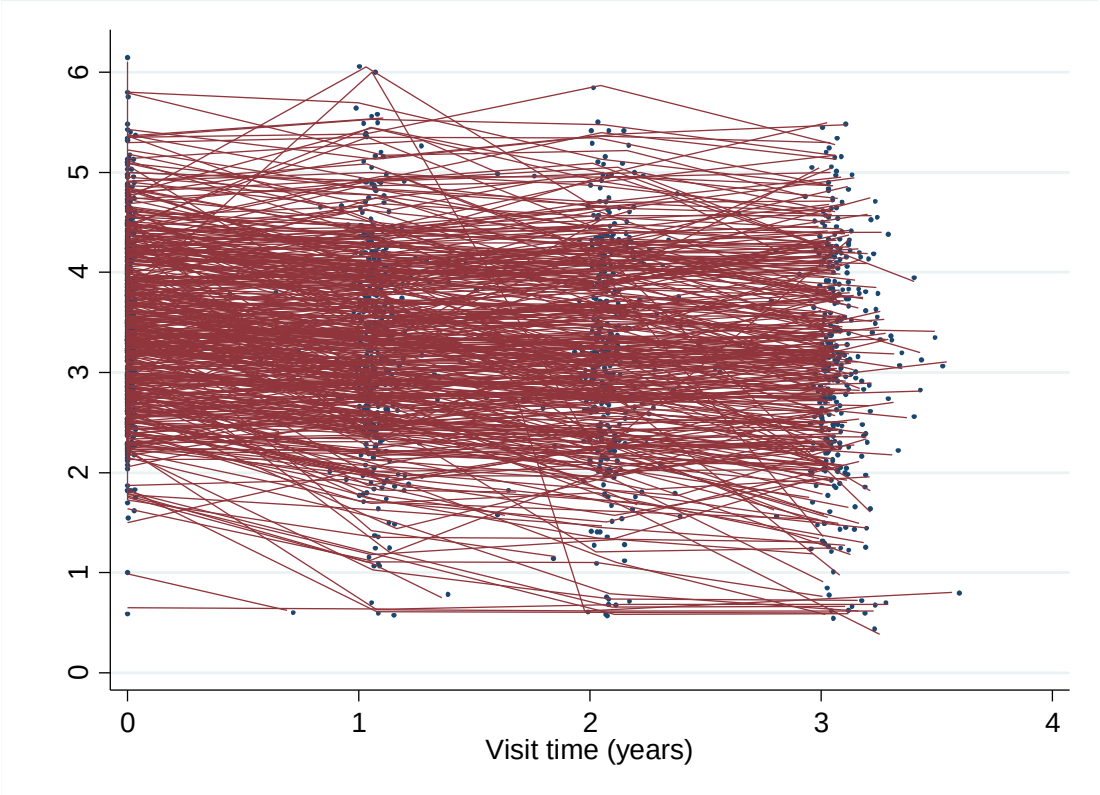
described within section 3.4.1, although the study schedule for SEKOIA was for radiographic images to be obtained at 12 monthly intervals, the exact time points varied. Therefore to ensure the model was as accurate as possible, time was modelled in the LME analysis using the exact years since the study participant's baseline visit radiographic image was obtained, treating the baseline visit as time 0.

One of the major assumptions of LME modelling is the linearity of the relationship between the dependent and independent variable, or, in the case of this thesis, between the individual knee joint space width measurements and time. Assessing this assumption informally can be done graphically or numerically. Graphically each study participant's knee joint space width measurements are plotted against the study visit time, Figure 6-4., and there appears to be no funnelling over time of knee joint space width measurements. Numerically the percentage change in joint space width measurements are calculated, Table 6-1, and it appears that similar levels of joint space width are lost across all combinations of study visits in SEKOIA, but it is very difficult to confirm the assumption of linearity.

Table 6-1 Percentage change in joint space width between study years in SEKOIA

Change in joint space width between:-	All
Baseline and year 1 (%)	-5.63
Years 1 and year 2 (%)	-3.30
Year 2 and year 3 (%)	-3.39

Figure 6-4 Individual SEKOIA study participants knee joint space width plotted over time



To further assess whether fitting a LME model is appropriate the VPC can be calculated. The VPC indicates that 87% (VPC: 0.65/0.75 = 0.87) of the variance in joint space measurements can be attributed to differences between study participants. Therefore indicating that a large proportion of the variance in knee joint space width measurements is attributable to differences between individuals and so a multilevel modelling approach that accounts for such clustering of data is an appropriate way to analysis the SEKOIA data.

Table 6-2 contains the average overall estimates for joint space and joint space change over time from the LME model analysis. The overall mean joint space was estimated to be 3.49mm and the average change in joint space, regression line, for all study participants was -0.14mm, therefore on average joint space decreased by 0.14mm for each successive SEKOIA study year.

Table 6-2 Estimates of fixed effects from LME modelling in SEKOIA

Parameter	Coefficient	95% Confidence Interval		p-value
β_0	3.49	3.42	3.56	<0.001
β_1	-0.14	-0.16	-0.12	<0.001

β_0 is in mm, β_1 is mm per year

Estimates for the random effects of the LME model are presented in Table 6-3. As seen in Table 6-3, the covariance between the random effect of the slope and intercept, σ_{u01}, is 0.027. As the covariance is close to 0 it indicates that there is minimal relationship between a study participant's average joint space width and the rate at which their joint space changes. The estimate for within-group variance, σ_e^2, is reasonably small, 0.094, which is further evidence that fitting an LME to the data is an appropriate analysis technique as small within-group variances suggest that the majority of the variation occurs between groups, or individual study participants in this example. If this estimate were large, it would indicate that observations within individuals were considerably different from each other and therefore lead to the conclusion that clustering is not an important factor of the data structure. The estimate of the variance among individual study participants' joint space width was, σ_{u0}^2, 0.632. When assessing this in comparison to the variance of individual regression slope, σ_{u1}^2, and observation variance, σ_e^2, it becomes clear that the majority of variance within the data occurs in the random effects for the intercept.

Table 6-3 Estimates of random effects from LME modelling in SEKOIA

Parameter	Estimate	95% Confidence Interval	
σ_{u0}^2	0.632	0.554	0.721
σ_{u1}^2	0.022	0.016	0.029
σ_{u01}	0.027	0.010	0.044
σ_e^2	0.094	0.085	0.104

The estimates provided in Table 6-3 are the average random effects, as linear mixed modelling technique focuses on calculating population level effects. Therefore these values are the average effects for all study participants within the LME. During the modelling process individual random effects are not directly estimated, but it is possible to 'predict' individual random effects, which are the BLUPs that were previously discussed within the methods section of this chapter. The BLUPs represent the individual random effects U_{0i} and U_{1i}, and have been plotted as histograms in Figure 6-5 and Figure 6-6 respectively.

Figure 6-5 Histogram of predictions for individual study participants slope (mm per year) from LME modelling in SEKOIA

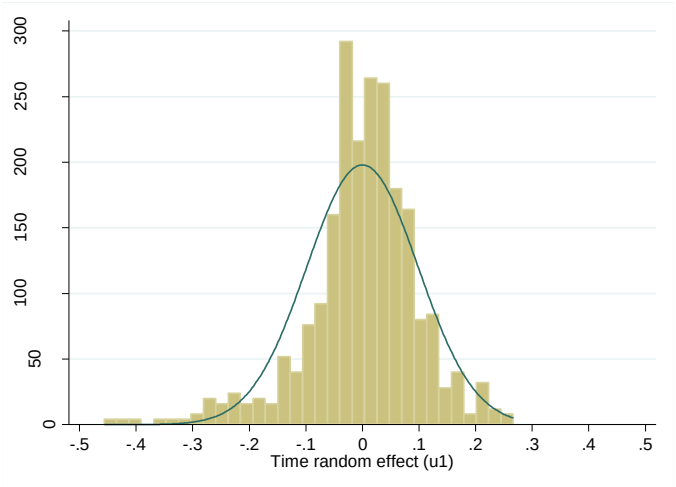
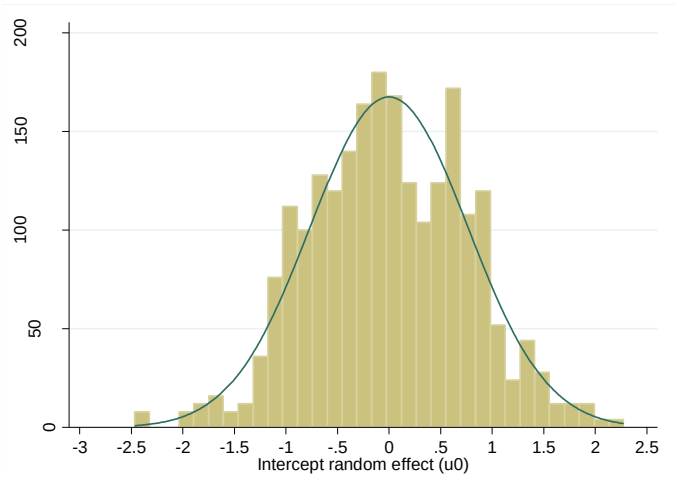


Figure 6-6 Histogram of predictions for individual study participants intercept (mm) from LME modelling in SEKOIA



If the individual BLUP estimate is greater than 0, it means that the estimate for the individual is greater than the model average, and, vice versa, if the estimate is less than 0, the individual's random effect is less than the model average. Therefore in Figure 6-5, all those individuals with a BLUP greater than 0 lose more joint space width on average over the duration of the study than the average for all study participants. A similar process applies when considering the random effect of the intercept; those individuals in Figure 6-6 whose BLUP estimate is less than 0 have an average knee joint space width that is less than the overall population average. However, as the BLUPs represent U_{0i} and U_{1i}, it is possible, using Equation 5, to calculate individual specific beta's for each study participant by adding the model average obtained during the LME process to each individual BLUP.

Figure 6-7 Histogram of individual slope effects (mm per year) from LME modelling in SEKOIA

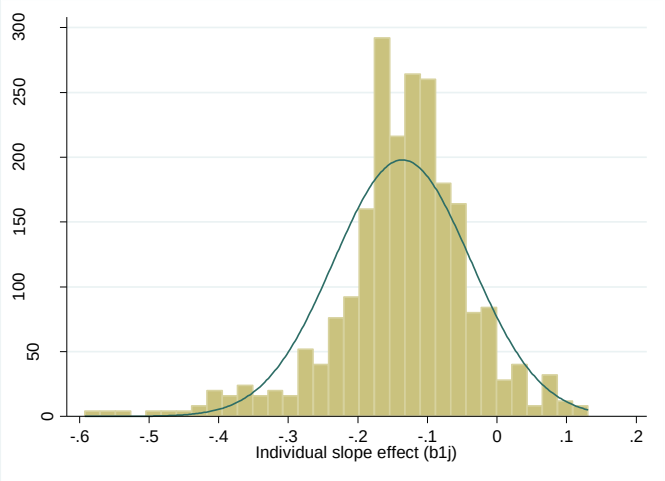


Figure 6-8 Histogram of individual intercept effects (mm) from LME modelling in SEKOIA

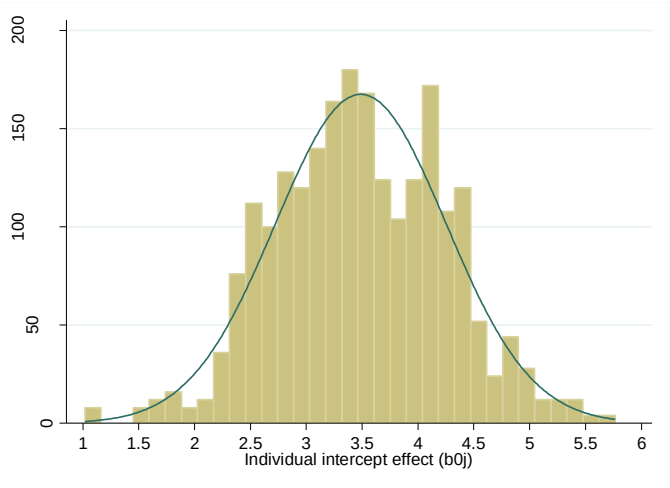
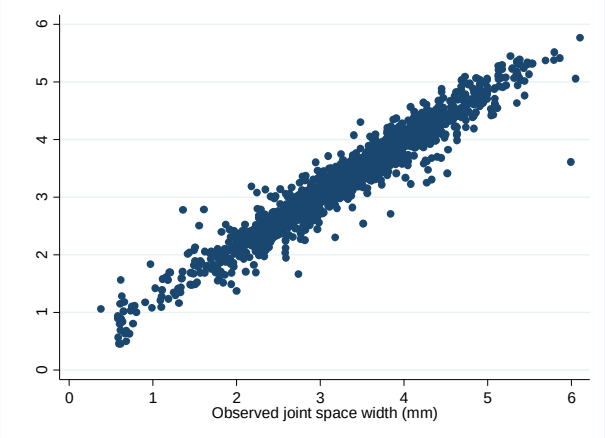


Figure 6-7 and Figure 6-8 are histograms of the individual slope and intercept effects from the LME modelling. As one would expect, given the results from LME modelling, the individual intercept effects are centred around 3.49mm, with a minimum value of 1.02mm and a maximum value of 5.77mm. The individual slope effects from the LME are centred around -0.14mm, with the maximum individual slope effect being 0.13mm and the minimum individual slope effect being -0.59mm. Of the 559 study participants included within the LME modelling analysis, 32 had an individual slope effect that was greater than 0. Thus, using the LME modelling techniques 5.7% of study participants were identified as having an apparent increase in joint space across the study duration.

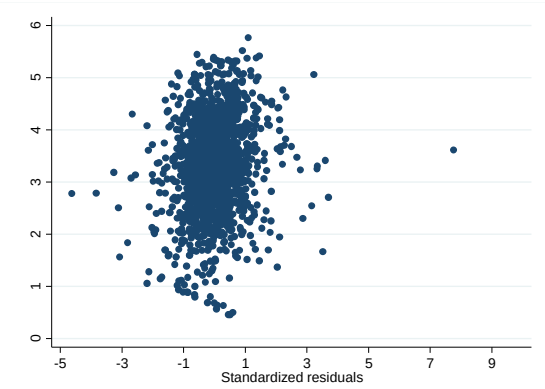
Once the LME model has been fitted, it is important to assess the appropriateness of the model fitted to the data and ensure the assumptions of the residuals are not violated. The assumption of linearity has already been assessed using scatter graphs to visual inspect the data, but, to further assess goodness of fit of the model the observed knee joint space width measurements were plotted on a scatter graph against their corresponding fitted values from the LME model, Figure 6-9. There is relatively good agreement between the observed and predicted values, with the possible exception of an outlier when the observed joint space was around 6mm but the predicted joint space width was just under 4mm.

Figure 6-9 Assessment of goodness of fit for LME model in SEKOIA



To assess the constant variance assumptions of the residuals the standardised conditional variables were plotted against the fitted values, Figure 6-10. There appears to be no pattern to the scatter graph, and so the assumption of constant variance holds for the LME model fitted to the SEKOIA placebo dataset.

Figure 6-10 Homoscedasticity of residuals from SEKOIA LME model



In addition to constant variance, the residuals must have a normal distribution. To ensure this assumption holds the conditional residuals and the BLUPs were plotted on Q-Q plots, Figure 6-11 and Figure 6-12. At the observation level, Figure 6-11, there appears to be a small level of deviation at the extremes of the distribution and in particular one residual that appears to

deviate dramatically from the line. However the majority of the residuals appear to follow the line of the Q-Q plot and so at level 1 the residual assumption of normality holds.

Figure 6-11 Normal Q-Q plot of the conditional residuals in SEKOIA

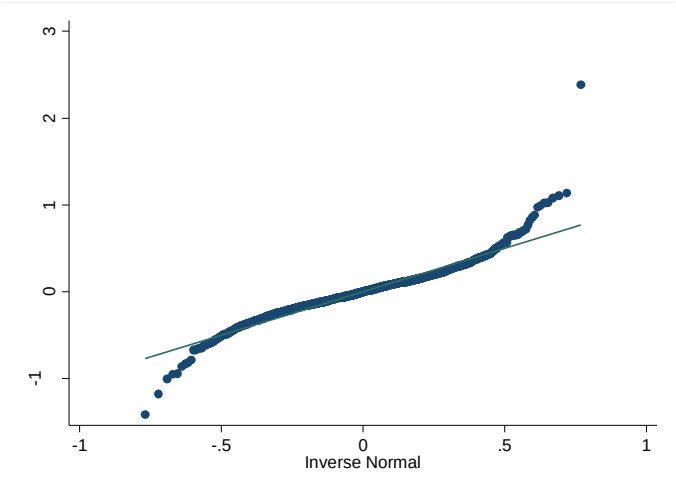


Figure 6-12 contains the Q-Q plots for the random effects, level 2 residuals. The plot on the left is the residuals of the intercept random effect and the plot on the right is a Q-Q plot for the slope random effects. The random effects for the intercept follow the Q-Q plot line very well, however there is a small amount of deviation at the lower end of the distribution in the random slope effects. However the majority of the observations lie on the Q-Q plot line and so the assumption of normal residuals at level 2 hold.

Figure 6-12 Normal Q-Q plots for the BLUPs of the random effects in SEKOIA

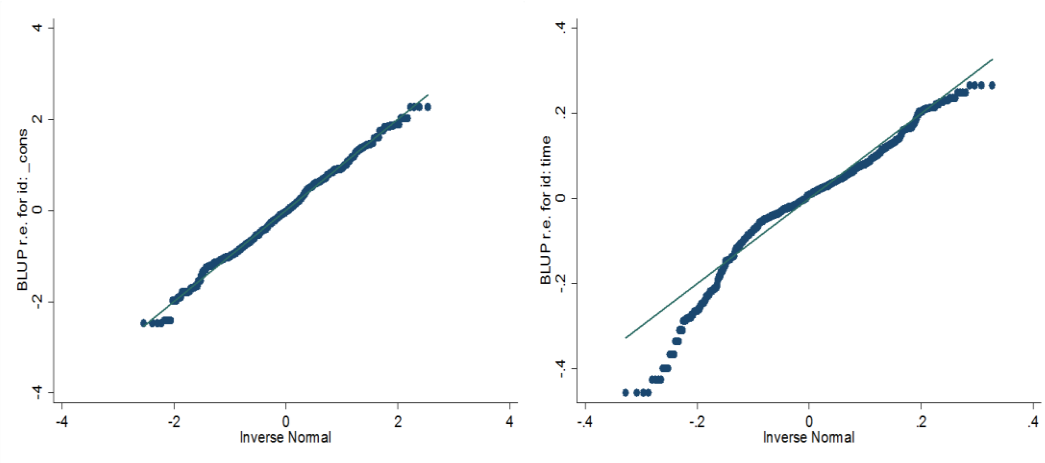
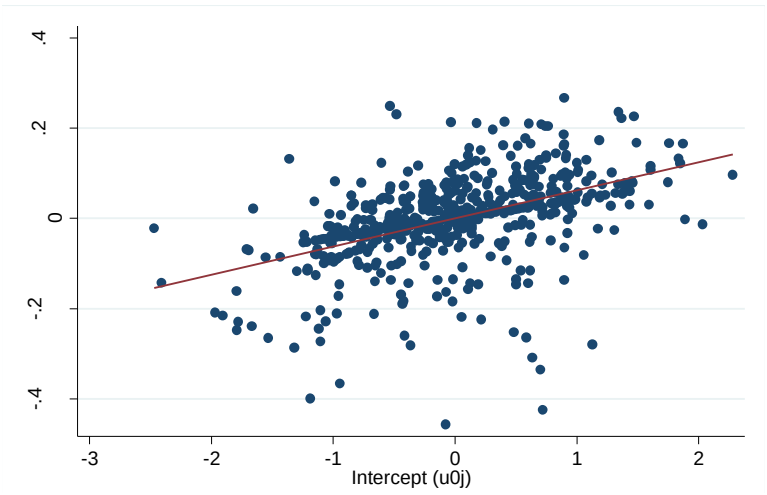


Figure 6-13 plots the BLUPs of the intercept and the slope random effects against each other. There appears to be an upward trend indicating that those study participants who start with smaller joint space widths on entry into SEKOIA, smaller intercept values, i.e. have greater rates of annual joint space reduction.

Figure 6-13 Scatter plot of individual study participants intercept (mm) and slope estimates (mm per year) from LME modelling in SEKOIA



It appears from a number of the graphics used to visually inspect the goodness of fit and residual assumptions of the LME model fitted to the SEKOIA data, Figure 6-9, Figure 6-10, Figure 6-11

and Figure 6-12, that a potential outlier in the data may exist. To determine how much influence this outlier has on the model, the study participant was identified and the LME analysis was re-run excluding the study participant to assess how the results may change. Removing the study participant had little to no effect on the estimates produced during the modelling process or on the fit of the model. Therefore this study participant's data did not unduly affecting the model and so they were left within the analysis.

6.4 LME results in the OAI

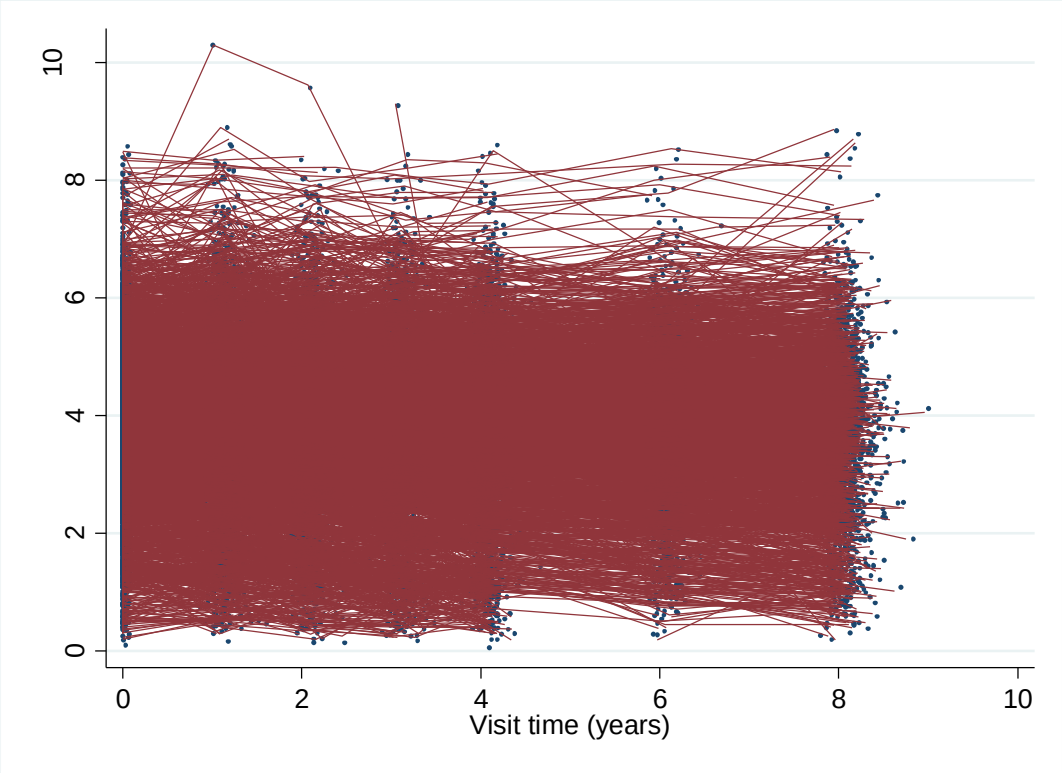
Of the 3469 study participants that entered the OAI less than half had a knee joint space width measurement at all 7 study visits, 1505 study participants. However since this type of longitudinal analysis allows participants who provide at least one outcome to contribute to the analysis all study participants were used in the LME modelling, regardless of whether they had an incomplete series of knee joint space width measurements. Using all available knee joint space width measurements reduces any selection bias that may occur when only using those study participants with complete data. The time variable that was used in the LME modelling was the number of years since baseline that a study visit occurred, with the baseline visit of each individual OAI study participant being time 0.

To ensure the estimates obtained during the LME modelling analysis in the OAI were valid the assumption of linearity must be checked. To assess this assumption informally, the percentage change in joint space width measurements between OAI study visits was calculated, Table 6-4, and each study participant's knee joint space width measurement was plotted against the study visit time, Figure 6-14. Joint space width between baseline and 12 months reduced by 0.75%, which appears slightly lower than across study visits. When assessing the change in joint space width patterns graphically, Figure 6-14, it appears that no obvious funnelling of knee joint space width measurements occurs over time, but with so many study participants within the analyses a distinct pattern is difficult to determine.

Table 6-4 Percentage change in joint space width between study visits in the OAI

Change in joint space width between:-	All
Baseline and 12 months (%)	-0.75
12 and 24 months (%)	-3.00
24 and 36 months (%)	-3.32
36 and 48 months (%)	-1.58
48 to 72 months (%)	-2.89
72 to 96 months (%)	-3.62

Figure 6-14 Individual study participants knee joint space width in the OAI plotted over time



The VPC can be used to further assess whether a LME model is an appropriate method to use to analysis the OAI, by quantifying the percentage of variation observed at the study participant level. In the OAI the VPC indicates that 92.5% (VPC: 1.77/1.92=0.925) of the variance in knee joint space width measurements can be attributed to differences between study participants. So, differences between individuals should be accounted for, and a multilevel modelling approach is suitable to analyse the OAI longitudinal joint space width measurements.

The estimates for the population level average joint space and joint space change over time from LME modelling are presented in Table 6-5. The mean joint space width at baseline, the intercept parameter β_0, was estimated to be 3.97mm. The average change in knee joint space width per year, the slope parameter β_1, was estimated to be -0.08mm. Therefore, on average, knee joint space width decreased by 0.08mm per year in the OAI data.

Table 6-5 Estimates of fixed effects from LME modelling in the OAI

Parameter	Coefficient	95% Confidence Interval		p-value
β_0	3.972	3.926	4.018	<0.001
β_1	-0.080	-0.085	-0.076	<0.001

β_0 is in mm, β_1 is mm per year

Table 6-6 contains the estimates for the random effects of the LME model. The estimate of the covariance between the random effect of the slope and intercept, σ_{u01}, is very small, 0.027. This indicates a minimal relationship between the average knee joint space width measurement of a study participant and the rate at which their joint space changes over time. The within group variance, σ_e^2 , was estimated to be 0.144, suggesting that the majority of the variation observed between knee joint space widths in the OAI data is between study participants. The parameter σ_{u0}^2 gives an estimate of the variability among average individual study participants joint space width. When comparing this parameter of 1.765 to the estimate of variation in change over time in knee joint space width, σ_{u1}^2, of 0.011 it is clear that the majority of the variation observed in the data occurs in the random effects for the intercept.

Table 6-6 Estimates of random effects of LME modelling in the OAI

Parameter	Estimate	95% Confidence Interval	
σ_{u0}^2	1.765	1.679	1.856
σ_{u1}^2	0.011	0.010	0.012
σ_{u01}	0.027	0.019	0.034
σ_e^2	0.144	0.140	0.147

The estimates in Table 6-6 are the average random effects and therefore represent the population, study participant, level effects. Figure 6-15 and Figure 6-16 contain the individual random effect, U_{0i} and U_{1i}, which are predictions from the modelling process, known as BLUPs. If the BLUP estimate is greater than 0 then the random effect for the individual is greater than the model average, and, vice versa, if the estimate for the BLUP is less than 0, the individual's random effect is less than the model average.

Figure 6-15 Histogram of predictions for individual study participants slope (mm per year) from LME modelling in the OAI

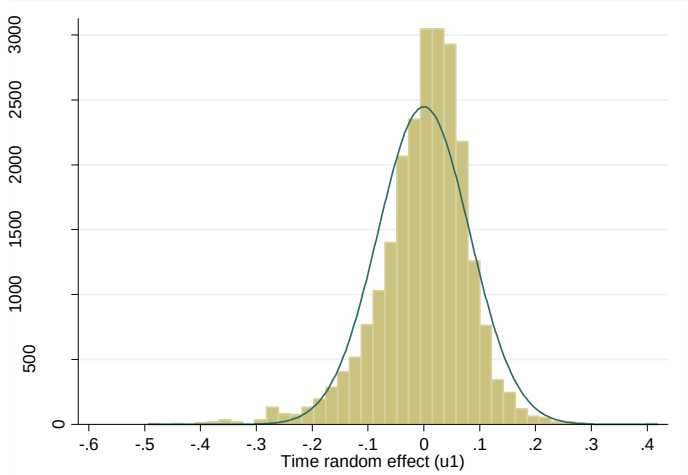
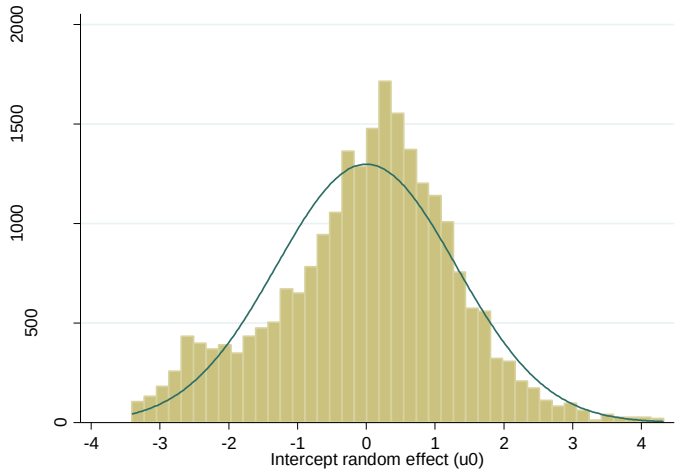


Figure 6-16 Histogram of predictions for individual study participants intercept (mm) from LME modelling in the OAI



The individual random effect BLUPs for time random effect are presented in Figure 6-15. All individuals with a BLUP random effect greater than 0 lose more knee joint space width per every 12 months in the study than the average for all study participants in the OAI. A similar interpretation can be made in relation to the individual random effect estimates, BLUPs, for the intercept. Those study participants in the OAI with an intercept BLUP greater than 0 had an average knee joint space width that was greater than the overall average. These BLUPs give estimates of individual random effects for each person, though it is possible to calculate beta's for each individual using the formula from Equation 5.

Figure 6-17 Histogram of individual slope effects (mm per year) from LME modelling in the OAI

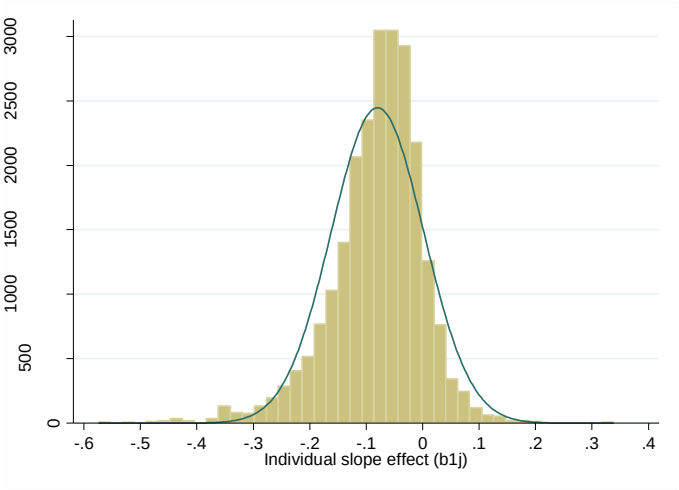


Figure 6-18 Histogram of individual intercept effects (mm) from LME modelling in the OAI

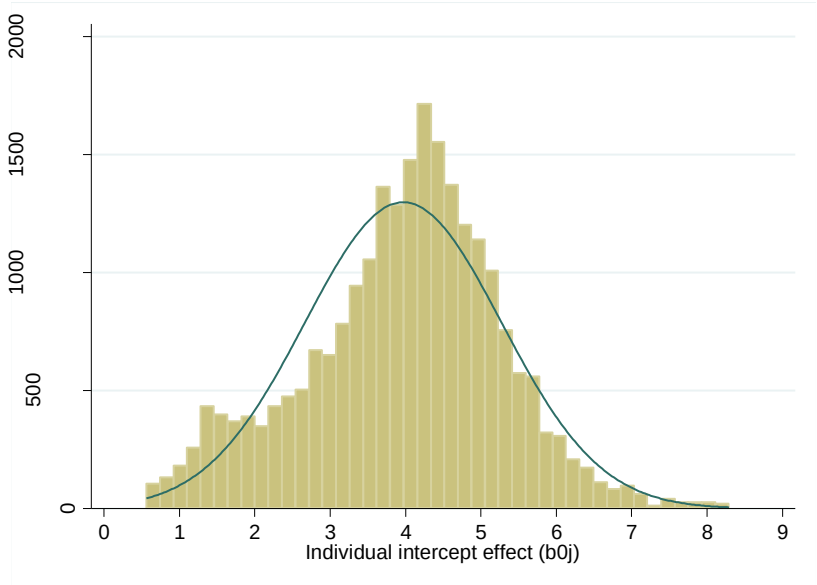
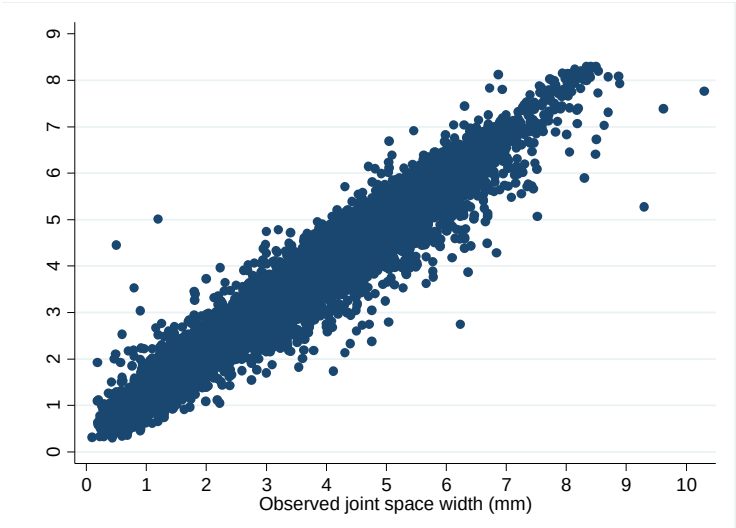


Figure 6-17 and Figure 6-18 contain the individual beta estimates for slope and intercept effects from the LME modelling in the OAI data. The individual slope effects, Figure 6-17, have a mean of -0.08mm, with the maximum individual slope effect being 0.34mm and the minimum slope effect being -0.57mm. Of the 3469 OAI study participants included within the LME modelling analysis, 400 had an individual slope effect that was greater than 0. Thus, 11.5% of study participants were identified as having an apparent increase in knee joint space for every 12 months within the OAI study. The mean value for individual intercept effects is 3.97mm and the minimum intercept effect is 0.56mm and the maximum is 8.29mm.

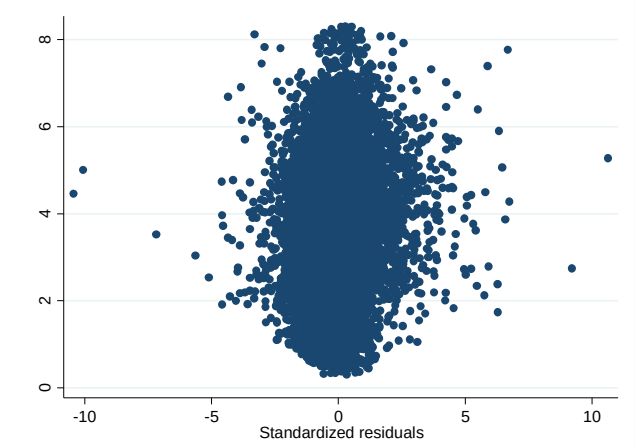
As previously highlighted, it is important, once the LME model has been fitted, to assess the appropriateness of the model fit and ensure that the assumptions of the residuals hold. The assumption of linearity was assessed using scatter graphs to visually inspect the relationship. However the goodness of fit of the LME model can be further assessed by comparing the observed knee joint space width values to the fitted knee joint space width values, Figure 6-19. There is relatively good agreement between the observed and fitted values, and therefore an LME model was an appropriate model to use to analysis the OAI data.

Figure 6-19 Assessment of goodness of fit for LME model in the OAI



The constant variance of the residual assumption can be assessed using Figure 6-20, in which the standardised conditional residuals are plotted against the fitted values. There does not appear to be any particular pattern in the scatter graph, i.e. residuals do not seem to systematically increase in size as the fitted values do. However, there are a number of large standardised conditional residuals, values greater than ±5, so these needed to be checked to ensure that they were not unduly influencing the model estimates.

Figure 6-20 Homoscedasticity of residuals from LME modelling in the OAI



In addition to the constant variance and linearity assumptions, the residuals must also be assessed to ensure that they have a normal distribution. The assumption of normally distributed residuals can be visually assessed using Q-Q plots, Figure 6-21 and Figure 6-22.

Figure 6-21 Normal Q-Q plot of the conditional residuals from LME modelling in the OAI

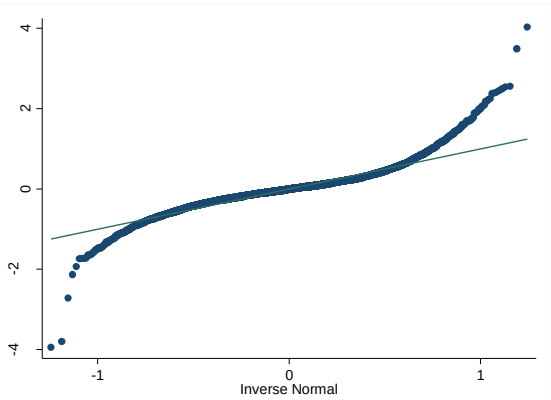
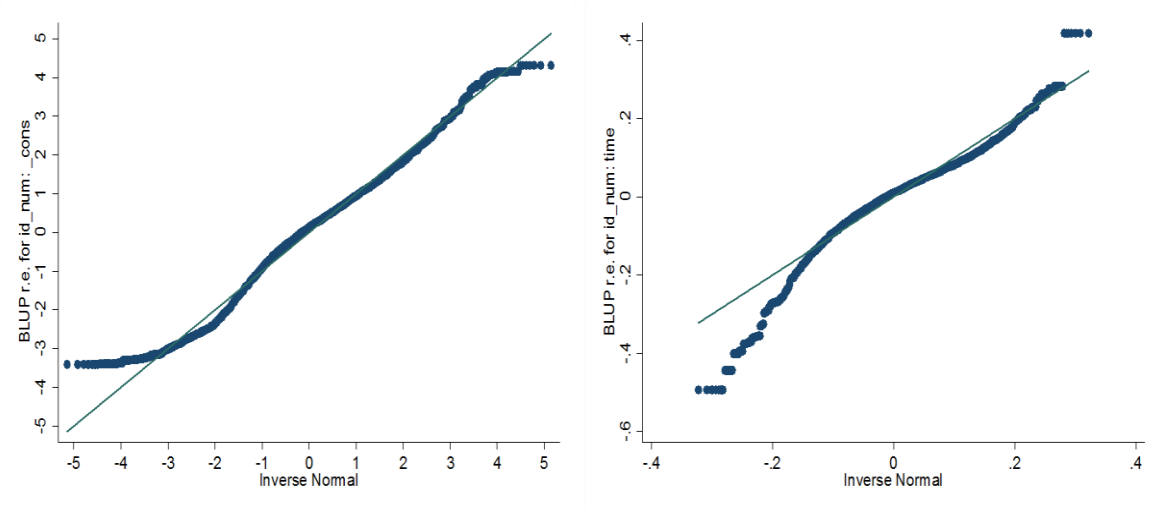


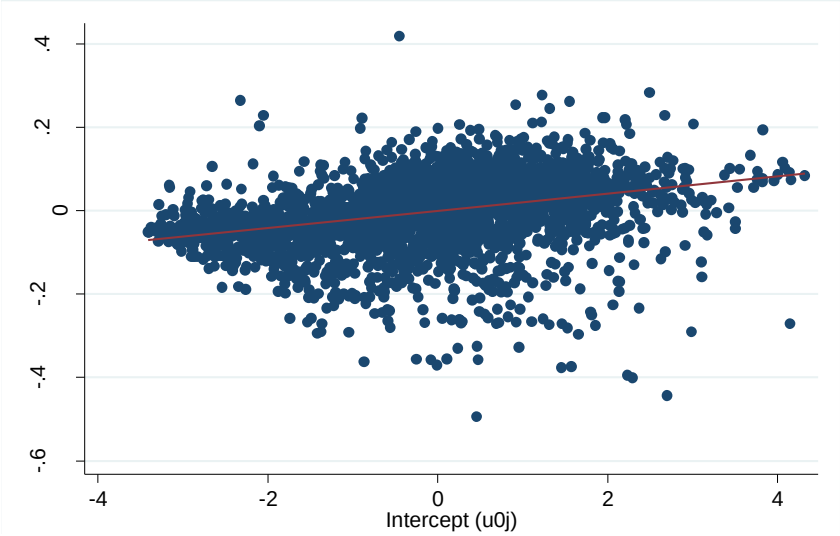
Figure 6-22 Normal Q-Q plots for the BLUPs of the random effects in the OAI



The conditional residuals, at the individual observation level, seem to deviate slightly at the extremes of the distribution, but the majority of the residuals fall along the line, Figure 6-21. A similar pattern can be seen when assessing the BLUPs, and the extremes of the distribution seem to curve away from the normal distribution though the majority of the observations fall along the line. Thus it appears that the assumption that the residuals follow a normal distribution holds.

As reported earlier in this results section, the correlation between the random effect of the slopes and random effect of the intercept is very small. This relationship can be further examined by plotting the BLUP estimates for the intercept and slope against each other, Figure 6-23. The line of best fit has a small incline, and therefore there is a very weak relationship between the BLUP for the slope and the BLUP for the intercept.

Figure 6-23 Scatter plot of individual study participants intercept (mm) and slope estimates (mm per year) from LME modelling in the OAI



As highlighted earlier, there were a number of large standardised conditional residuals and these could indicate observations which are unduly influencing the estimates obtained during the LME modelling. To determine how much influence these outliers had, the 4 observations that had standardised residual values greater than 7 were removed and the LME analysis re-run. The results of the LME model remained unchanged when removing these observations and so these observations did not unduly influence the model and so the outliers were retained within the analysis.

6.5 Summary

In SEKOIA joint space width measurements were obtained at baseline and then yearly during the 3-year duration of the clinical trial. In the OAI, joint space widths were measured every 12 months for the first 48 months and then every 24 months until the study participants had been followed up for a total period of 96 months.

It is important to ensure that a multilevel modelling approach is a suitable way to approach the longitudinal analysis of both datasets within a frequentist framework. Thus it is important to ensure that the majority of the variation is observed at the group level, indicating that it is important for clusters to be accounted for and thus a multilevel modelling approach should be conducted. In SEKOIA the VPC indicated that 87% of the variance observed in the knee joint space width measurements was attributed to differences between individual study participants, and 92.5% of the variation in knee joint space width measurements in the OAI was attributable to differences between study participants. Therefore, in both SEKOIA and the OAI it was appropriate to account for the clustered nature of the data by using a multilevel model. In addition to assessing the percentage of variation attributable to differences between clusters, to further assess the goodness of fit of a LME model the assumption of linearity must hold. This was assessed informally, by plotting each study participant's knee joint space width measurement against the corresponding study visit time. Due to the number of study participants and the relatively narrow range of joint space width measurements, it was difficult to confirm a linear relationship informally in either dataset.

The overall model estimate for the average change in knee joint space width per year in the SEKOIA study was -0.14mm, therefore on average joint space decreased by 0.14mm for each successive year in the SEKOIA study. The covariance between the random effect of the slope and intercept was almost 0, 0.027, and so there is little to no relationship between a study participant's average knee joint space width in SEKOIA and the change in joint space width. The estimates for individual random effects, calculated by using the BLUPs and estimates of overall population effects, indicated that, as expected, the individual random effects for the intercept were centred around 3.5mm, with a minimum value of 1.02mm and a maximum value of 5.77mm. When looking at the estimates for individual random slope effect the values were

centred around -0.1mm and had a maximum value of 0.13mm and a minimum value of -0.59mm. All assumptions for the residuals of the LME model held in the SEKOIA LME analysis, and no individual study participant's data was found to unduly influence and affect the model fit.

Results from the LME modelling estimated that the average change in knee joint space width per 12 months in the OAI study was -0.08mm. So, on average, knee joint space width decreased by 0.08mm per 12 months in the OAI study. The estimate of the covariance between the random effect of the slope and intercept was very small, and interestingly similar to the covariance in the SEKOIA data 0.027, indicating that there was a negligible relationship between OAI study participant's baseline knee joint space width and the rate of change. However, the estimate of the variance for the random effect of intercept was much larger than other estimates of variation within the LME model, signifying that the majority of the variation was observed in the random effect of the intercept. The maximum individual random slope effect was 0.34mm per 12 months and the minimum individual random slope effect was -0.577mm per 12 months. The mean value for the individual random intercept effect was 3.97mm, the minimum intercept effect was 0.56mm and the maximum random intercept effect was 8.29mm. When assessing the residuals all the assumptions held, and so the estimates from the LME modelling process in the OAI dataset are reliable, and no observations were found to be unduly affecting the estimates obtained during the modelling.

6.6 Discussion

There are many different statistical techniques that can be used to assess change over time. The aim of this chapter was to outline and assess one of these methods, frequentist LME modelling, as an approach to monitoring knee OA progression. One of the major strengths of using sophisticated statistical techniques, such as LME modelling, is that all measurements obtained can be used within the same analysis. However, when repeated measurements are obtained for the same study participants, the measurements are not independent, and so statistical techniques such as linear regression cannot be used for the analysis s as one of the main assumptions of these analysis techniques is independence between observations (90). Ignoring the clustered nature of the data and using standard regression techniques to analyse the data will lead to incorrect inferences as there is a high risk of Type 1 errors. Ignoring the correlated

nature of the data during the analysis process can lead to underestimation of the standard error of any time-independent predictors within the model and an over estimation of any time-dependent predictors. Using LME models allows for the correlated nature of longitudinal data to be accounted for by allowing regression coefficients to vary across clusters, or, as in the context of this thesis study participants. In other words, LME models allow for heterogeneity across study participants by allowing random effects to be estimated during the modelling process (91) .

One of the main assumptions of using LME modelling is that the relationship between the independent and dependent variable is linear. This assumption was assessed informally in both datasets by plotting individual study participants' knee joint space width by study time. Due to the large number of observations and narrow range of knee joint space width measurements within each dataset this assumption is difficult to assess informally. However, making the assumption that the relationship between time and joint space width is linear means that change over time is modelled using a monotonically decreasing pattern and ensuring that change can only decrease over time. Whereas if time were treated as a categorical variable, where time was grouped together by scheduled study visit time, variations in the timing of visits could lead to the change in knee joint space width between each study visit being over- or under-estimated.

Being able to model random effects for both the intercept and regression coefficient is an advantage of using LME models, especially within the context of modelling OA disease progression. Including a random effect for time within the model used in this thesis allows for different trajectories of change for each study participant, this is particularly important as study participants were at different stages of knee OA severity. Within SEKOIA study participants already had established knee OA, and those in the OAI had established disease or were at risk of developing knee OA. As OA is a degenerative disease that progresses over many years, rates of progression of the disease vary across the life course and can depend on the presence or absence of associated factors (92). Therefore creating a LME model allowing the study visit time to be modelled using a random effect allows for different rates of progression that may occur between different study participants.

Using LME models also has the advantage of being able to model a random effect for the intercept, in other words LME models that contain a random intercept allow all clusters being modelled to have a different average value. Therefore, by having a random intercept in the models when analysing the OAI and SEKOIA study populations, it is possible to allow all study participants to have a differing average knee joint space width. Biologically this would be the advantageous way to model the data, as it would be extremely unlikely that all study participants would have the same average knee joint space width due to the study participants in the OAI or SEKOIA study having different severities of disease.

As discussed, the major aim of this chapter was to assess the effectiveness of the frequentist approach of LME models in estimating OA disease progression accounting for the presence of measurement error. Individual study participant change estimates were calculated post-hoc, as, during the modelling process, all measurements were used to calculate average population level change, before BLUPs were estimated to give subject specific regression coefficients. Therefore during this process extremely large and small values would be 'smoothed' towards the population average, thus 'smoothing' individual change trajectories for measurement error. By adjusting each individual study participant's trajectory of change for measurement error in this way it is then possible to use these individual trajectories, BLUPs, in further research to gain a better understanding of factors that may be associated with rates of knee OA disease progression.

Unlike other forms of longitudinal analysis, such as repeated measures ANOVA, LME models offer a very flexible framework in terms of the structure of the data. For example the design of a study does not have to be balanced to use LME models, therefore there is no need to have the same number of observations per study participant. The repeated measure ANOVA, is one of the least complex statistical methods for analysing longitudinal data but it is one of the more restrictive methods, owing to the data needing to have a balanced design with no missing data and the measurements have to be obtained at fixed time points (93, 94). LME models can account for data that have been obtained at study visits that may have occurred at different times or have different time periods between them (95). The flexible framework offered by LME models arises because of the way the parameters contained in the model are estimated. During the analysis process, parameters are estimated using iterative maximum likelihood (ML)

estimation, which is an optimization process that estimates the value of the parameter which gives the observed data the maximum probability. For example no noticeable difference in level of computational time was required to run LME models on the SEKOIA data with a maximum of 4 time points for 559 study participants, compared with a maximum of 7 time points and 3469 study participants within the OAI.

Although LME models have their advantages, there are other sophisticated statistical models within the frequentist statistical framework which also allow for within-subject correlation to be accounted for. Other than LME models, the other technique frequently used within epidemiological research is GEE (96). GEE accounts for correlation within clustered observations by using a 'working' correlation structure for the clustered observations. Different statistical software packages allow for specification of different structures, with the most commonly available structures being exchangeable and unstructured. An exchangeable structure is where it is assumed that the correlation between subsequent measurements are all the same, whereas unstructured correlation is where no particular structure is assumed. The choice of correlation matrix can change the magnitude of the results obtained, and therefore one of the advantages of LME over GEE is that it has been argued that it is easier to choose which coefficients within the analysis should be modelled using a random effect rather than making the correct choice of the correlation structure (90). LME models containing differing random effects can be compared using the log-likelihood test, however changing the correlation structure within GEE analysis changes the number of parameters which are estimated during the modelling process. The more complex the correlation structure in GEE, such as using unstructured correlation, the greater the number of parameters and this in turn can reduce the statistical power of the analysis.

Another marked difference between the two different longitudinal modelling techniques is the way in which the regression coefficients are calculated. GEE models are sometimes referred to as 'population average' models and this is because the regression coefficient obtained during the analysis is the average of all the individual regression coefficients (97). Whereas the regression coefficients calculated during the LME modelling process can be assessed at the subject level giving a subject specific regression coefficient.

Despite the differences between the two methods, both GEE and LME modelling are well established statistical techniques within the frequentist framework, and many statistical software packages have inbuilt procedures, such as the 'mixed' command in Stata (98). Therefore the decision as to which longitudinal technique to use may not always be a simple one. However a previous study by Twisk demonstrated that when considering continuous outcomes, as in the context of this thesis analysis, GEE and LME models are extremely comparable (90). As the question of interest within the context of this thesis was assessment of individual change in knee joint space over time, the statistical longitudinal technique of LME modelling was chosen over GEE, as it enabled greater focus on individual estimates of change rather than population average change.

Use of LME models identified that on average in the SEKOIA study participants knee joint space width reduced by 0.14mm per year, and by 0.08mm per year in the OAI study. This indicates that on average over the duration of SEKOIA, study participants' knee joint space width reduced by nearly half a millimetre, 0.42mm, and over the duration of the OAI study participants on average lost a total of 0.64mm. These results were obtained without consideration of any potential confounders and so with further work these estimates of change per year could be further refined. Although LME models have many advantages and offer a very flexible framework to model change over time, especially when assessing OA progression, in the presence of measurement error, unfortunately LME models do not allow for direct quantification of the magnitude of measurement error. However, using LME models allows for 'smoothing' of the estimates of individual study participants' change trajectories. LME models, although more complex statistical models, are an established methodology for handling longitudinal data and thus statistical packages have built-in commands to handle such analyses. In a clinical setting it would not be possible to make use of LME models to aid in the management of patient diagnosed with knee OA. But, in a research setting the use of LME models could enable risk factors to be identified which have not been previously discovered due to change estimates being masked by measurement error. For example, the individual annual change estimate for each study participant could be used as an outcome measure in further analysis and due to the LME modelling process these estimates have been 'smoothed' to take account of some, at least, of the measurement error. This novel extension to the LME modelling is discussed in greater detail in chapter 9.

Chapter 7: Bayesian Hierarchical Modelling

7.1 Overview of chapter

As discussed in the previous chapters, longitudinal data are being ever increasingly collected within both a clinical setting and a research environment. Methods such as the RC index are simple to apply but do not make full use of the richness of the data, as such methods only use two time points rather than all available repeated measurements. The hierarchical repeated nature of data can be accounted for using frequentist techniques, as in chapter 6, or by Bayesian methodology. This chapter will outline Bayesian hierarchical modelling methodology, then the results of Bayesian analysis for the SEKOIA and the OAI studies are presented. Finally the chapter will conclude with critical evaluation of the Bayesian methodology implemented.

7.2 Methods

7.2.1 Background

Bayesian modelling is a technique based upon using prior information obtained from previous observations. The overriding principle of Bayesian analysis is that inference is made about what is not known given previous knowledge and observations, thus it is a statistical procedure which endeavours to estimate parameters of an underlying distribution based on observed distributions. Bayesian analysis offers a very flexible framework that is different from frequentist techniques. This is because the rationale behind Bayesian analysis is that the only sensible measure of uncertainty is probability. Frequentist techniques treat any parameter being estimated, e.g. θ, within the analysis as fixed with the confidence interval around the estimated parameters being random and giving an indication of the percentage of such confidence intervals that include the fixed estimate, θ.

All Bayesian inference and modelling is based on Bayes' theorem, which was initially presented by Reverend Thomas Bayes in the mid-seventeen hundreds in "An essay towards solving a

problem in the doctrine of chances" (99). Bayes' theorem, sometimes referred to as Bayes' rule, is a way of calculating conditional probabilities, i.e. probabilities whose values depend on the values of other probabilities. The practical application of conditional probability occurs considerably more often than generally thought. For example, in a clinical setting, a doctor often uses conditional probability when weighing up the probability that a patient has a disease given that they have presented with a set of symptoms. This is conditional probability, as the probability of having a disease is dependent on the probability of having a set of symptoms. Using conditional probability for the discrete case of two events A and B, Bayes' theorem can be expressed as in Equation 7.

Equation 7 Bayes' Theorem

$$P[A|B] = \frac{P[B|A]P[A]}{P[B]}$$

In this formula P[A] is the probability of A prior to having information about event B and is called the prior probability of A. P[B] is the probability of B and is sometimes referred to as a normalizing constant. P[B|A] is the probability of B assuming that event A as occurred, and this is called the likelihood function for B given A. P[A|B] is the probability of A given that event B has occurred and is known as the posterior probability.

Bayes' theorem can be extended from the discrete case and is generalised as shown in Equation 8.

Equation 8 Generalised Bayes' Theorem

$$posterior\ probability = \frac{prior\ \times likelihood}{normalizing\ constant}$$

Therefore, the basic idea of Bayesian analyses is to estimate the posterior probability or distribution from the combination of a prior distribution and a likelihood function. In other

words the posterior probability represents the modified belief of the parameter of interest after observations have occurred. The options for calculating the posterior distribution include:-

- Using algebraic calculations to give an exact numeric result, known as exact analytics
- Using algebraic calculations to calculate the posterior distribution numerically to arbitrary precision, known as exact numeric
- Using approximations of distributions to obtain results, known as approximate analytic
- Using appropriate functions of random numbers, generate a large random sample from which estimates of the property of interest can be calculated, known as computer simulation. One technique that can be used for such simulations is known as Markov Chain Monte Carlo (100)

Under the Bayesian framework, any unknown parameter is treated as a random variable, unlike frequentist methods where the unknown parameter is treated as fixed. Exact analytical, exact numeric and approximate techniques are only possible in discrete or very simple Bayesian analysis. However as analysis becomes more complex in high dimensions, such as hierarchical modelling, it becomes very difficult, if not impossible, to use these methods to approximate the posterior distribution. Therefore computer simulation techniques are used to estimate the posterior distribution, especially when conducting Bayesian hierarchical analysis. This thesis will focus on the Markov Chain Monte Carlo (MCMC) simulations. MCMC is a method that can be used to draw samples from a target posterior density, ζ (θ |x), whereby each sample is drawn sequentially but only depends on the previously drawn sample. These sample draws, also known as iterations, start with an approximate target density and with each step of the sequential procedure the approximations are improved. The main concept of Markov Chain simulations is to devise a Markov process such that the iterations converge to the target distribution. The number of iterations completed during the simulation process must be large enough to ensure that all the samples are sufficiently close to the target density; once this occurs the process is then said to have converged. The process for a Markov chain involves creating a sequence θ_t, by beginning at θ_0 and at the tth stage selecting a value from a transition function $Q_t(\theta_t|\theta_{t-1})$, such that the value of θ_t only depends on the previous value via the transition function. The starting value θ_0 is usually based on a good approximation to the target density, but in order for the starting value to converge to the target density, the transition function must be selected with care. Metropolis-Hastings is the general name given to the methods of choosing appropriate transition functions (101), and one special case of Metropolois-Hastings used in this thesis is Gibbs Sampling (102). Gibbs sampling is a MCMC algorithm, which is a conditional

sampling method which generates samples from joint posterior distributions by sampling in a cycle from each of the full conditional posterior distributions. The software program 'Bayesian inference Using Gibbs Sampling' (BUGS) (103) was a project that began in the 1989 and has been specially designed to contain language for specifying complex Bayesian models using the Gibbs sampling methodology. All Bayesian analyses for this thesis were undertaken using WinBUGS 14 (103) implemented through the statistical package RStudio (104).

7.2.2 Bayesian Hierarchical Modelling

The reasoning behind producing a hierarchical model is the same whether completing the analysis within a Bayesian or frequentist framework; the intention is to attempt to model and understand data with a complex data structure. Within a Bayesian framework this is done by making inferences on parameters $\theta_1, \dots, \theta_n$ measured on N units which are related or connected. Before preparing a Bayesian model, assumptions about the relationships between parameters need to be made. Three approaches to this are: -

- Identical parameters
- Independent parameters
- Exchangeable parameters

The identical parameters' assumption assumes that all θ's are identical, where θ represents group level effects. This assumption is very rarely appropriate in data with a hierarchical structure as this assumptions means that all group level effects being estimated are identical regardless of the observations clustered within the group. Specifically, the rationale for producing a hierarchical model is because it is believed that group differences exist, and the identical parameter assumption assumes equivalence across groups.

The independent parameters' assumption is the complete opposite to the identical parameters', as it assumes that all θ's are independent. Therefore under this assumption all θ's are completely unrelated, and can result in estimates that are highly variable, as estimates are made only using the units contained within each grouping, regardless as to whether there are very small numbers of observations in any cluster. Therefore under the independent parameter assumption it is presumed that the groups have no similarities, and in this circumstance the

analysis of the individual clusters could be completed separately as there is nothing to be learned from analysing all the clusters together.

The third assumption, exchangeable parameters, assumes all θ's are similar to each other and therefore the label given to the grouping conveys no information. In broad terms, the exchangeable parameters are mathematically equivalent to assuming all θ's are drawn from a common prior distribution with unknown parameters. For example, consider house prices in a geographical area containing a number of different neighbourhoods. Under the identical parameter assumption it would be assumed that the mean house price is the same in all neighbourhoods within the geographical area being studied, which is highly unlikely due to prices often varying depending on the features of the house or neighbourhood. Under the independent parameters assumption it would be assumed that the mean house price in one neighbourhood was completely independent of that in another, and therefore estimates of mean house price in a particular neighbourhood would be based only on houses contained within that neighbourhood. Under the exchangeable assumption it is assumed that the neighbourhoods are similar and therefore it is possible to borrow strength from the likelihood contributions for all neighbourhoods within the geographical area, which can lead to global smoothing of means and improved precision of estimated means. The hierarchical models analysed within this thesis will use the exchangeable parameters assumption as it seems to be the most appropriate for the data being considered.

7.2.2.1 Priors

Using Bayesian analysis, observable outcomes from hierarchical data are modelled conditional on certain parameters. A Bayesian hierarchical model is specified in Equation 9.

Equation 9 Bayesian hierarchical model

$$y_{ij} \sim Normal(\mu, \tau^2)$$

In this equation y_{ij} is the measurement *i* of group *j*, μ and τ^2 are unknown parameters to be estimated and are referred to as hyperparameters. μ refers to the mean of the normal distribution for observations y_{ij} and τ^2 is the precision; the variance of the distribution for y_{ij} is calculated as $\sigma^2 = {}^{1}/_{\tau^2}$. All unknown hyperparameters, μ and τ^2, need to be assigned a prior distribution, for example

$$\mu \sim Normal(0, 100)$$

$$\tau^2 \sim Gamma(0.001, 0.001)$$

The assignment of prior information about distributions of all unknown parameters is a crucial component of Bayesian analysis. Prior information can be based on background literature, historical data or can be a subjective guess. There are two types of priors, 'non-informative' and 'informative', with different types of priors being used within different types of Bayesian analysis. Informative priors are used in Bayesian analysis where explicit and definite information is known, and therefore the analysis is based upon more than the current data. Non-informative priors, which are also referred to as 'reference' or 'objective' (105) priors, are used in analyses where scientific objectivity is at a premium. That is non-informative priors are used where the intention is to ensure that no pre-conceived knowledge or bias is introduced into the analysis through specification of the prior distributions. However the term non-informative prior is misleading as all priors contain some level of information and it would be better to consider them as 'vague' priors. Even when using non-informative priors it is still important to ensure that estimates calculated are not unduly influenced by the choice of priors, and so it is important that the prior distribution specified has a plausible range of values. Therefore even for non-informative priors there may never be an agreed 'objective' prior as different statisticians will apply different priors. In some circumstances, different priors can lead to different conclusions. To try to overcome such problems a combined approach to the choice of priors is reasonable. Therefore it becomes important to distinguish between primary parameters of interest and secondary structure parameters that have less influence on the results.

When the primary parameters of interest in the Bayesian analysis are location parameters, which are quantities such as means and regression coefficients, a vague non-informative prior should be used. For example a uniform prior distribution, $\theta \sim Uniform(-100, 100)$, with a wide

range or a normal prior distribution, $\theta \sim Normal(0,100000)$, with a large variance could be used (106). As long as the scale of the prior distribution is related to the data being modelled, using non-informative priors on primary parameters will ensure that parameter estimates are not unduly biased by choice of priors. However, vague priors for variances, whether at individual observations or group level, are harder to specify to ensure that any inferences are not dependent on how the model was initially parameterized. One choice to overcome such a problem in analysis using binomial and normal populations was proposed by Sir Harold Jeffreys, who developed special cases of non-informative priors for multiparameter models. Jeffreys' prior (107) for the mean of normally distributed data is the flat prior, $\mu = 1$, and for scale parameters within the model it can be shown that the sample variance is estimated by the 'inverse' prior $\frac{1}{\sigma^2}$.

Regardless of whether a non-informative or informative prior is being used within Bayesian analysis, there is no such thing as the one 'true' prior. Therefore it is vital to gain an understanding of how much influence the choice of prior distribution has on the results through sensitivity analyses in which alternative priors are specified. Performing such procedures will ensure that results obtained from Bayesian analyses are robust to the selection of priors and are not an extreme finding due to improper specification of the prior distribution.

7.2.2.2 Hierarchical model specification

In a similar manner to frequentist regression techniques, Bayesian hierarchical models can have different intercepts and slopes for each cluster contained within the data. As an example, groups might be hospitals and individual observations may be patients. In this situation the parameters at the hospital level would not vary independently between patients and can therefore be constrained to have the same distribution. In the context of this thesis the observations are knee joint space width measurements taken over time and these are clustered within individual study participants. A longitudinal hierarchical Bayesian regression model with random slopes and intercepts for repeated measurements within a Bayesian framework would be specified as in Equation 10.

Equation 10 Bayesian random slope and random intercept model

$$y_{it} \sim Normal(\mu_{it}, \tau) \qquad i = 1, \ldots, n; t = 1, .., p$$

$$\mu_{it} = \theta + \alpha_i + \beta_i x_t$$

$$\theta \sim Normal(0, 0.001)$$

$$\alpha_i \sim Normal(0, \tau_\alpha)$$

$$\beta_i \sim Normal(0, \tau_\beta)$$

In this equation *n* is the number of study participants and *p* the number of time points for each study participant. It is assumed that θ is constant and can be interpreted as the overall average. Therefore y_{it} is the measurement of study participant *i* at time *t,* α_i is the random effect for the intercept, and $\sigma_\alpha^2 = {}^1/_{\tau_\alpha}$is the variance component that measures the variability of the observations between the various participants. β_i is the random effect for the slope, and $\sigma_\beta^2 = {}^1/_{\tau_\beta}$is the variance component that measures the variability between time points. $\sigma^2 = {}^1/_{\tau}$ measures the overall variability of all observations. Although modelling the data in this way allows for each participant to have an individual slope component, the assumption is that the slopes are all similar enough to be from the same distribution with a common mean slope. Therefore they are all sampled from the same distribution, under the exchangeable parameter assumption discussed earlier within this chapter. So in the context of this thesis analysis y_{it} is the joint space measurement for individual i in study year t, θ is the overall intercept, and is interpreted as the average joint space width for the total study population. α_i is the random intercept parameter allowing for a different knee joint space width for each study participant at the beginning of the study, β_i is the random effect allowing a different slope, interpreted as change in knee joint space width, for each study participant and x_t is the time of each study visit in years since the baseline study visit.

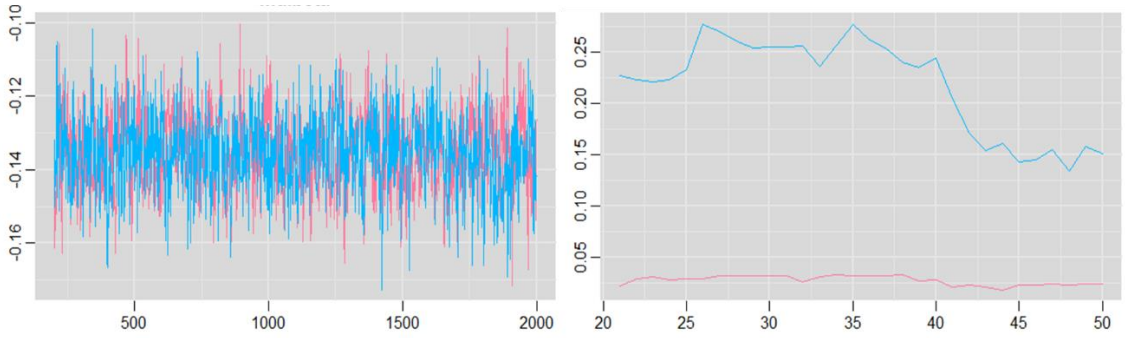
7.2.2.3 Model computation and evaluation

To compute estimates for the model parameters, all simulated draws are collated and used to summarise the posterior density. In other words, estimates from all MCMC iterations are used to compute model estimates. Therefore in order to ensure inferences can be made from the

Bayesian modelling, the MCMC process must contain enough iterative simulations to represent the target distribution. Before the MCMC process can begin, the Markov Chain must be initialised by providing initial values for each unknown parameter in the model. To ensure that initial values do not have any undue influence on any summaries calculated, an initial burn-in period is performed. These early iterations in the MCMC process are discarded and are not used when summarising model parameters. The number of iterations discarded will vary in different analyses, as the iterations discarded should be all the initial non-stationary portion of the Markov chain. Hence all initial iterations that occur while the simulations are jumping around the stationary distribution will not be included. Once iterations have become stable around the desired posterior distribution the analysis is said to have converged. Convergence can be determined by eye or using formal detection methods. Whether using formal methods or assessing convergence by eye it is recommended that two Markov Chains are simulated at the same time during the modelling process. Running multiple MCMC chains, rather than one long run MCMC chain, allows for very different, over dispersed, initial values to be selected. Thus by running multiple chains in parallel a quicker and more simpler assessment can be made of how much the posterior estimates are being affected by the selection of the initial values.

One of the simplest methods of determining convergence is by examining trace plots. A trace plot is a continuous line joining successive values of a parameter plotted against iteration number. A model is said to have converged once the continuous line plotted on the trace plot looks like random scatter around a stable mean value and the appearance of this is often referred to as a “fat hairy caterpillar”. Because hierarchical models may contain a large number of parameters, it would not always be realistic to inspect all individual parameters for convergence visually. In such circumstances the primary parameters of interest should be monitored to determine when convergence occurs. An example of a trace plot representing a model that is yet to reach convergence alongside a model which has converged can be seen in Figure 7-1. Each line in the trace plots represents a MCMC chain, in the plot on the right the two lines can be clearly seen and the estimates do not overlap across any of the iterations, indicating that running 50 iterations are not enough for convergence of the model to have occurred. In the trace plot on the left 2000 iterations have been run, and the two chains in the trace plot are mixed together. This mixing of both MCMC chains indicates that similar estimates are being obtained for the monitored parameter and thus convergence of the model has occurred.

Figure 7-1 Bayesian trace plot examples



In addition to assessing convergence by eye, a number of techniques have been implemented in Bayesian statistics software that allows convergence to be assessed formally. Different methods examine different features of the Markov Chains, but all techniques that have been developed to assess convergence are based on the null hypothesis of convergence, i.e. they have been developed to assess non-convergence in the model. One such technique is the diagnostics developed by Brooks-Gelman and Rubin (BGR) (108). The BGR diagnostic is a method that relies on two parallel chains being run so that formal assessment can be made of whether the chains converge to the same posterior distribution. Once the burn-in iterations of the analysis have been discarded the between-chain and within-chain variance is calculated. Then posterior marginal variance is calculated, which is a weighted average of the between-chain variance and the within-chain variance. If the chains have reached the target distribution then the posterior marginal variance should be very close, thus the ratio of the posterior marginal variance and the within-chain variance should be very close to 1. The square root of this ratio is known as the potential scale reduction factor (PSRF). Large values of the PSRF indicate that the between-chain variance is substantially greater than the within-chain variance and thus further iterations should be run. Whereas if the PSRF is close to 1 it can be concluded that each chain has stabilised and it is likely each chain has reached the target distribution, thus the model can be said to have converged.

No single method of assessing model convergence can provide completely conclusive evidence that convergence has occurred, and so a combination of both formal and visual methods to determine convergence are used to assess model convergence in this thesis.

7.3 Bayesian results in SEKOIA

All 559 study participants within the SEKOIA placebo arm were included in the analysis, to avoid introduction of any bias into the posterior estimates obtained during the Bayesian modelling. For clarity the prior assumptions assigned to the hyperparameters of the Bayesian model are set out in the equations contained within Equation 11. The numerical values chosen within these hyperparameters have been chosen in the hope of exerting as little influence over the models posterior estimates as possible.

Equation 11 Hyperparameters for Bayesian SEKOIA analysis

$$\theta \sim Normal(0, 0.001)$$

$$\tau \sim Gamma(0.001, 0.001)$$

$$\alpha_i \sim Normal(0, \tau_\alpha)$$

$$\beta_i \sim Normal(\mu_\beta, \tau_\beta)$$

$$\mu_\beta \sim Normal(0, 0.001)$$

$$\sigma_\alpha \sim uniform(0,1)$$

$$\sigma_\beta \sim uniform(0,1)$$

As previously described τ is the precision parameter at the observation level, precision parameters for all joint space measurements, and the residual error variance is $1/\tau$. τ_α is the precision at the study participant level and is calculated as $(1/\sigma_\alpha)^2$, where σ_α is the standard deviation of the random intercepts, τ_β is the precision of random slopes and is calculated as $(1/\sigma_\beta)^2$, where σ_β is the standard deviation of the random slopes.

The two different sets of initial values for all unknown parameters in the model are provided in Table 7-1. Both sets of initial values are different to ensure dispersion across the priors being used in the model, and therefore to ensure that the results are as reliable as possible, by

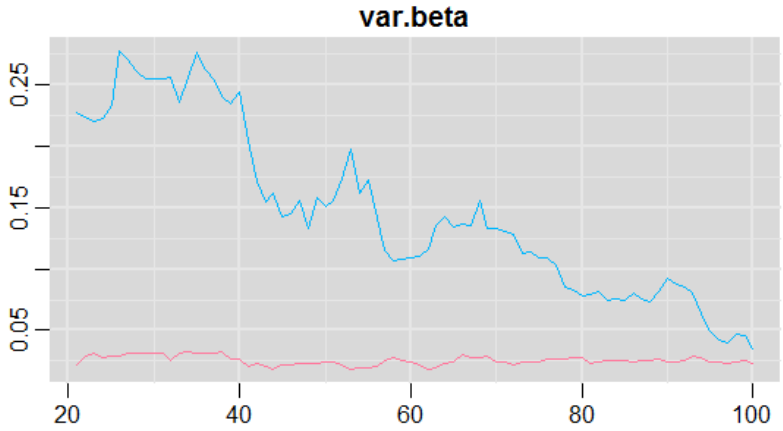
ensuring that posterior estimates do not differ dramatically depending on the initial values used within the MCMC chain. The different chains should eventually converge to similar values, but if not, then the initial values are having undue influence on the results. The initial values for chain 1 used within this analysis were selected randomly, within the boundaries of the distributions of the hyperparemeters. Once the initial values for chain 1 had been selected, the initial values for chain 2 were selected to be different from those within chain 1.

Table 7-1 Initial values for both MCMCs for SEKOIA analysis

Parameter	Initial value for chain 1	Initial value for chain 2
α_i	Value of 0 for every study participant	Value of 0.5 for every study participant
σ_α	1	0.5
σ_β	1	0.5
θ	0	5
τ	1	10
μ_β	1	10

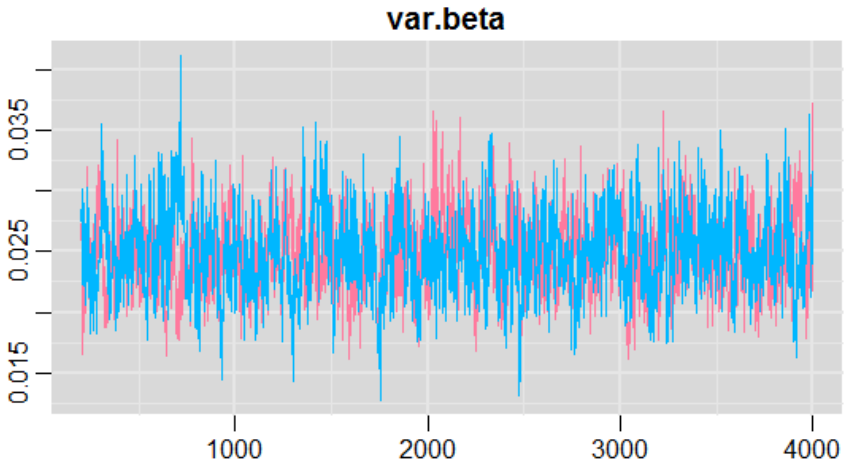
As discussed in the methods section, when conducting Bayesian analysis it is essential that convergence of MCMC chains is assessed. Initially 100 iterations were run with a burn-in period of 20 iterations, on visual inspection of the trace plots it was clear that convergence had not occurred and a greater number of burn-in iterations was required, see Figure 7-2 for one of the many parameters, the variance of beta $1/\tau_\alpha$, in the hierarchical model for knee joint space measurements in SEKOIA.

Figure 7-2 Trace plot demonstrating lack of convergence while modelling SEKOIA data



A total of 4000 iterations for each MCMC was completed with a burn-in period of 200, and, on visual inspection of the trace plot, convergence appeared to have occurred. Trace plots for all parameters were inspected and all appeared to be mixing well, thus representing "fat hairy caterpillars". As an example the trace plot for the variance of the beta, $1/\tau_\alpha$, parameter can be seen in Figure 7-3. In this figure the two chains, represented by blue and pink lines, no longer appear distinct and at this point both MCMCs vary around a similar distribution, therefore the Bayesian hierarchical model has converged.

Figure 7-3 Trace plot for visual inspection of convergence in SEKOIA



Once convergence had been assessed using the visual inspection method of trace plots, the formal method of BGR diagnostics was assessed. As described within the methods section, to be able to formally determine convergence the PSRF must be close to value of 1. All PSRFs across all parameters contained within the SEKOIA hierarchical Bayesian model were calculated to be within a very narrow range close to 1, namely 1.00 – 1.01. Therefore both visual and formal methods for assessment of convergence confirmed that using 4000 iterations within a burn-in period of 200 enabled the model to converge.

The posterior estimates for the Bayesian analysis are contained within Table 7-2. The mean variance across all knee joint space measurements was 0.092 with a credible interval of 0.084 to 0.102. The variance for the random intercept was 0.668 and the variance for the random slope was 0.025. Therefore, the portioning of the variance within this model indicates that joint space width varies more between different individual study participants than between joint space measurements within the same person or across all joint space width measurements. Thus fitting a hierarchical model allowing for clustering at the individual level and random effects for the slope and intercept is an appropriate model to fit. The posterior estimate for the overall intercept value is 3.48mm.

Table 7-2 Posterior analysis of joint space width measurements in SEKOIA

Parameter	Posterior Mean	95% Credible Interval
θ	3.484	3.409, 3.556
α_i	0.003	-1.447, 1.496
β_i	-0.135	-0.393, 0.073
$1/\tau$	0.092	0.084, 0.102
$1/\tau_\alpha$	0.668	0.587, 0.758
$1/\tau_\beta$	0.025	0.019, 0.031

Figure 7-4 and Figure 7-5 are histograms of posterior estimates for individuals' random slope effects and random intercept effects obtained from the SEKOIA Bayesian analysis. The mean estimate for the random effect of the intercept is 0.003, and, as expected this term is close to 0.

The minimum value for individual posterior estimate for the random intercept is -2.56 and the maximum value is 2.30. The mean estimate for the random slope effect is -0.135, so, on average, study participants lose 0.135mm of joint space per year. The minimum individual posterior estimate is -0.64 and the maximum individual posterior estimate is 0.15.

Figure 7-4 Histogram of individual posterior estimates for random intercept (mm) from Bayesian modelling in SEKOIA

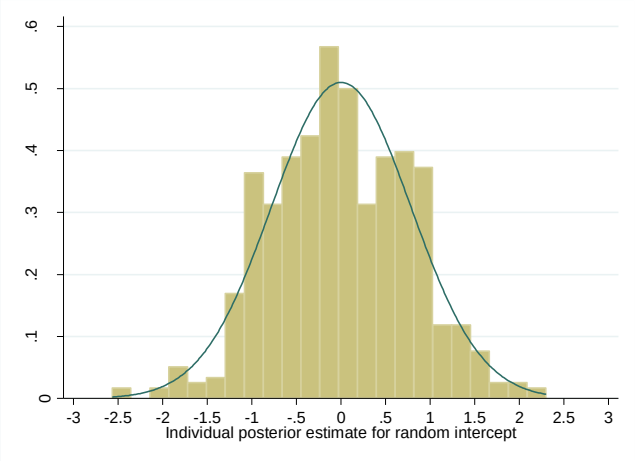
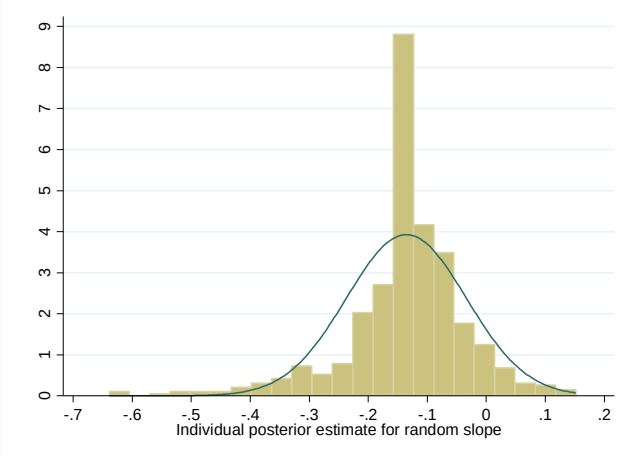


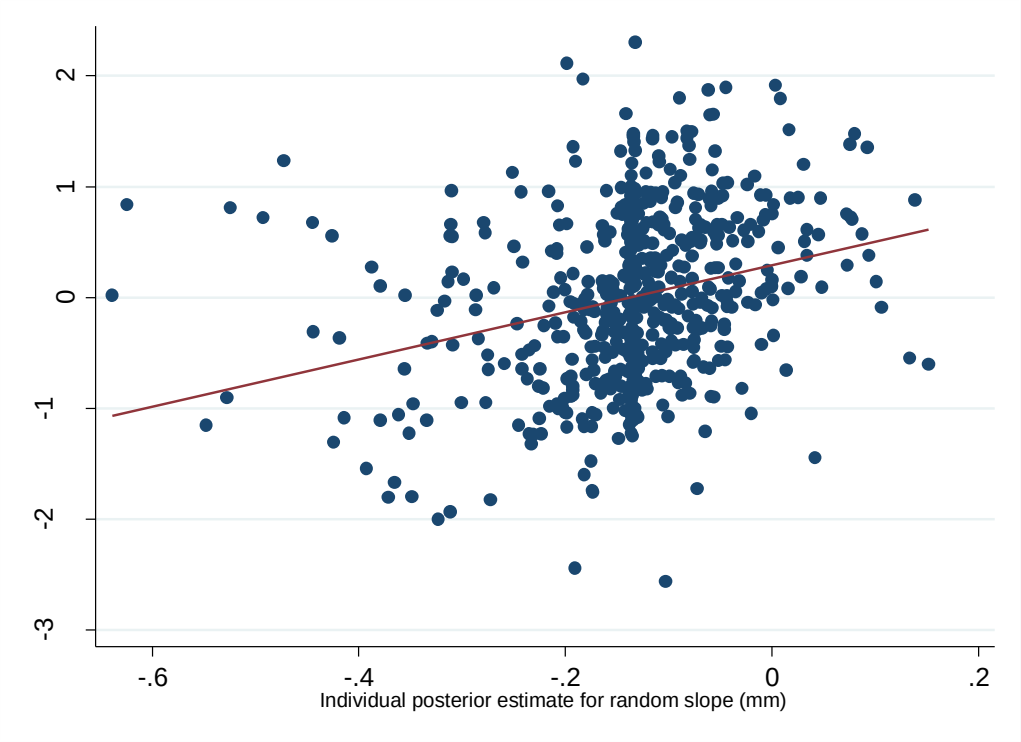
Figure 7-5 Histogram of individual posterior estimates for random slope (mm per year) from Bayesian modelling in SEKOIA



The scatter graph below, Figure 7-6, plots each study participant's random slope effect against their random intercept effect, with a line of best fit through the scatter plot. The line of best fit has a positive gradient; therefore as posterior estimates for an individual's random slope

increase so does an individual's posterior estimate for their random intercept effect. In other words the larger an individual study participant's posterior estimate of average knee joint space is the larger their posterior estimate of change in knee joint space is.

Figure 7-6 Scatter plot of individual posterior estimates for random slope and random intercept from Bayesian modelling in SEKOIA



7.4 Bayesian result in the OAI

A total of 19491 knee joint space width measurements across the 3469 study participants were included within the Bayesian model. The Bayesian hierarchical model with a random slope and random intercept fitted is presented in Equation 10, and the prior assumptions assigned to the hyperparameters within the hierarchical model are laid out below, Equation 12.

Equation 12 Hyperparameters for Bayesian OAI analysis

$$\theta \sim Normal(0, 0.001)$$

$$\tau \sim Gamma(0.001, 0.001)$$

$$\alpha_i \sim Normal(0, \tau_\alpha)$$

$$\beta_i \sim Normal(\mu_\beta, \tau_\beta)$$

$$\mu_\beta \sim Normal(0, 0.001)$$

$$\sigma_\alpha \sim Uniform(0,1)$$

$$\sigma_\beta \sim Uniform(0,1)$$

The τ parameter is a measurements of the level precision at the individual observation level, or in the context of the OAI data a measure of precision for knee joint space width measurements, and this precision parameter is used to calculate residual error variance, $1/\tau$. The distribution of the random intercept, α_i, is assigned the prior assumption of normality with a mean of 0 and a precision of τ_α, which is calculated as $(1/\sigma_\alpha)^2$. The prior assumption for the random slope effect within the hierarchical model, β_i, follows a normal distribution with a mean of μ_β and a precision of τ_β.

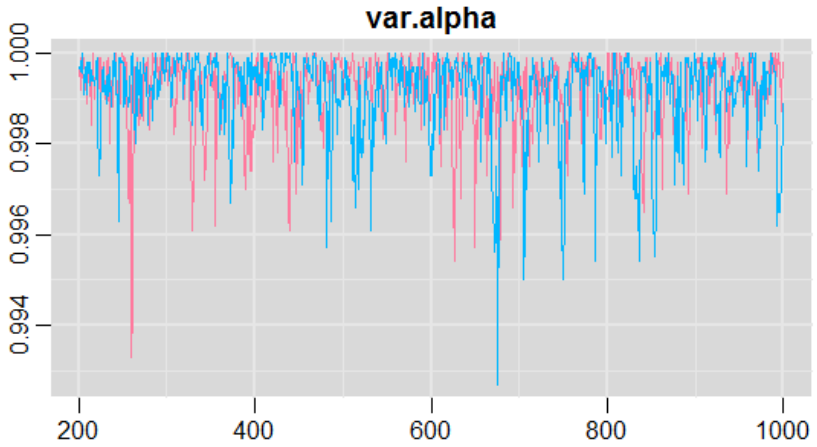
All the unknown parameters contained within the hierarchical model were given initial values, and to ensure best assessment of convergence two MCMCs were used. To ensure dispersion across the model prior's two different sets of initial values were used, and these are presented in Table 7-3. The same initial values selected for the SEKOIA analysis were used within the OAI Bayesian analysis. So the initial values in the first MCMC were selected randomly, within the boundaries of the prior distributions of the hyperparameters. Following selection of the initial values for chain 1, the initial values for the second MCMC were selected to ensure the values were markedly different from chain 1 while still being within the boundaries of the prior distributional assumptions.

Table 7-3 Initial values for the two MCMCs used in the OAI analysis

Parameter	Initial value for chain 1	Initial value for chain 2
α_i	Value of 0 for every study participant	Value of 0.5 for every study participant
σ_α	1	0.5
σ_β	1	0.5
θ	0	5
τ	1	10
μ_β	1	10

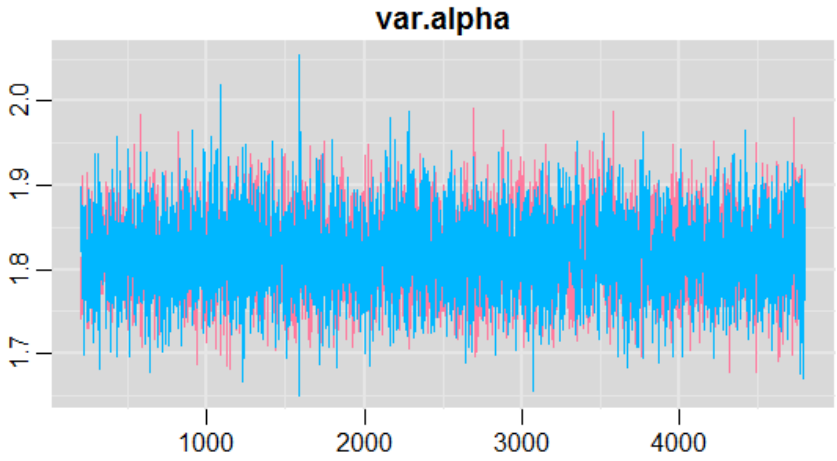
Trace plots can be used to visually inspect both the appropriateness of the prior assumptions for each hyperparameter and the convergence of the model. To visually inspect the appropriateness of the hyperparameters, laid out in Equation 12, and to visually inspect convergence of the Bayesian hierarchical model a total of 1000 iterations were run with a burn-in period of 200 iterations. Figure 7-7 is a trace plot for the variance of the random intercept, and each MCMC is represented by a different colour in the figure. It can be seen from the trace plot that although the two MCMCs appear to be mixing reasonable well there appears to be a ceiling effect at 1. The variance of α_i is calculated from the SD of the random intercept, which was given a prior assumption of the distribution $Uniform(0,1)$. Therefore it is clear from the trace plot that the magnitude of the prior assumptions for the distribution needs to be increased to ensure the model fits the OAI data. Consequently the prior assumption for α_i was changed to $Uniform(0,5)$ and then the model re-run to visually inspect whether this change had improved the model fit.

Figure 7-7 Trace plot demonstrating lack of convergence while modelling the OAI data



To enable visual inspection of convergence of the two MCMCs and to assess whether changing the prior assumption for the standard deviation for the variance of the random intercept was appropriate, a further 3800 iterations for each MCMC was completed. Thus in total each MCMC was run for 4800 iterations with a burn-in period of 200. The trace plot for the variance of α_i can be seen in Figure 7-8. It can be seen that there is no longer a ceiling effect of the estimates at 1, therefore changing the prior has improved the fit of the model. It can also be seen that both chains are mixing well and the trace plot represents a "fat hairy caterpillar" and so the model has reached convergence.

Figure 7-8 Trace plot for visual inspection of convergence in the OAI



The posterior estimates for the Bayesian hierarchical analysis of the OAI data is contained in Table 7-4, and these were calculated using all 9200 iterations that were completed across both chains. The mean variance across all joint space width observations was 0.143 with a credible interval of 0.139 to 0.146. The variance for estimates of the random intercept effect is 1.818 and the variance of the estimates for the random slope effect is 0.011. These posterior estimates indicate that the majority of the variance within this model is on the random intercept effect, thus there is a considerable amount of variability between the average knee joint space width measurements of study participants in the OAI. The portioning of the variance also indicates that fitting a hierarchical model to the OAI data was an appropriate model, because there is a greater level of variance between different individual study participants than between joint space width measurements within the same study individual or across all knee joint space width observations. The posterior estimate for the overall intercept is 3.972mm, with a credible interval of 3.928 to 4.022.

Table 7-4 Posterior estimates of joint space width measurements in the OAI

Parameter	Posterior Mean	95% Credible Interval
θ	3.972	3.928, 4.022
α_i	-0.002	-2.770, 2.446
β_i	-0.078	-0.276, 0.057
$1/\tau$	0.143	0.139, 0.146
$1/\tau_\alpha$	1.818	1.730, 1.912
$1/\tau_\beta$	0.011	0.010, 0.012

Individual posterior estimates for each study participants random intercept and random slope effect for the OAI data can be seen in Figure 7-9 and Figure 7-10. The mean posterior estimate for individual study participants' random intercept is -0.002, and the minimum individual estimate is -3.43 and the maximum individual random intercept estimate is 4.37. The mean estimate for individual posterior random slope effect is -0.078, so on average study participants lose 0.078mm per year of knee joint space width during the duration of the OAI study. The minimum posterior estimate of random slope is -0.60 and the maximum individual estimate is 0.37.

Figure 7-9 Histogram of individual posterior estimates for random intercept (mm) from Bayesian modelling in the OAI

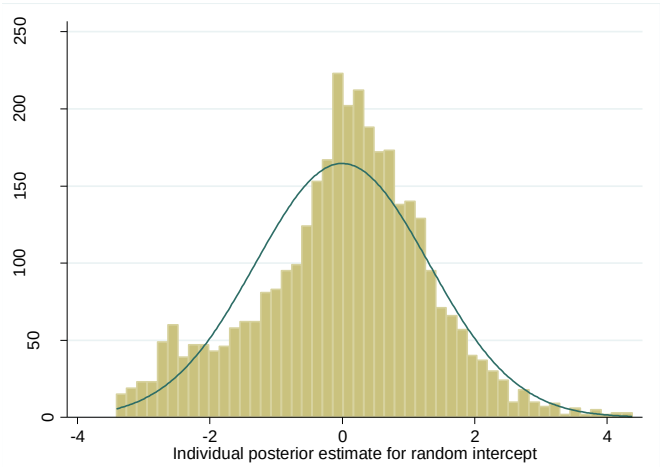


Figure 7-10 Histogram of individual posterior estimates for random slope (mm per year) effect from Bayesian modelling in the OAI

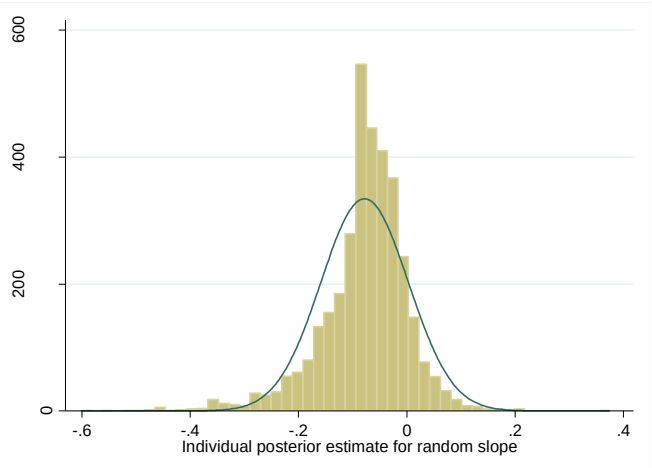
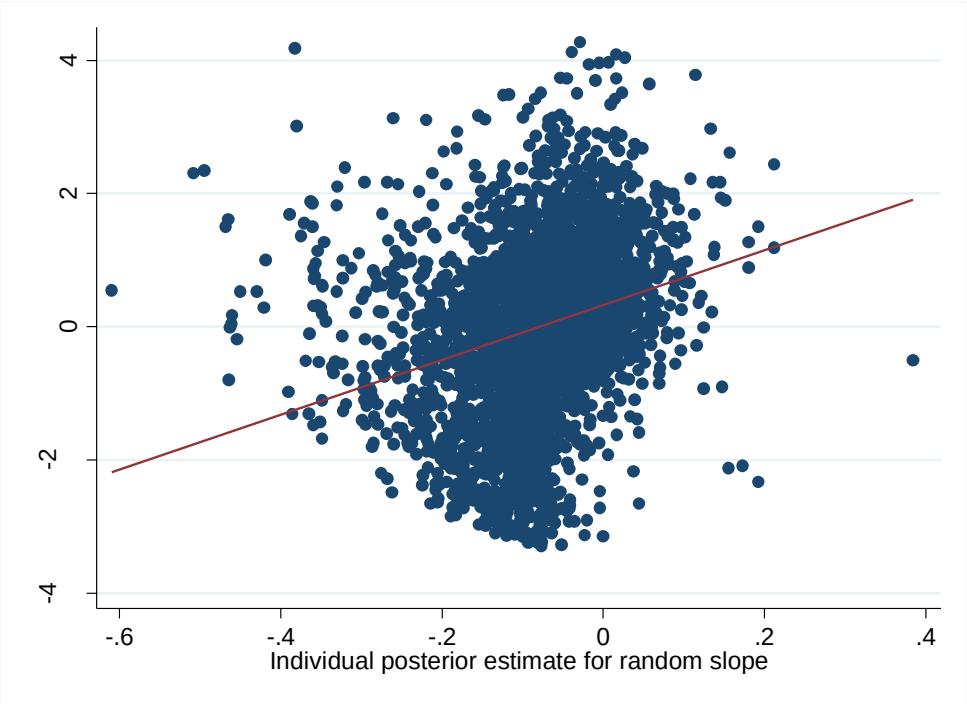


Figure 7-11 is a scatter plot of individual posterior estimates for the random intercept plotted against posterior estimates for the random slope effect, with a line of best fit plotted on the scatter plot. There were 3469 study participants contain in the OAI hierarchical Bayesian model and so it is quite difficult due to the number of individuals being plotted in the scatter plot to determine any particular pattern, however the gradient of the line of best fit is positive. This indicates that those individuals with a greater random intercept estimate tend to have a higher posterior estimate of the random slope effect.

Figure 7-11 Scatter plot of individual posterior estimates for random slope and random intercept effects from Bayesian modelling in the OAI



7.5 Summary

In order to be able to make inferences from the posterior estimates calculated during the Bayesian modelling process, the Bayesian model must have enough burn-in iterations run and the model must have converged. Convergence was assessed visually and formally using the BGR diagnostic. In SEKOIA a total of 1765 knee joint space width observations across 559 individual study participants were included within Bayesian hierarchical modelling. On visual inspection of the trace plots convergence appeared to occur after a total of 4000 iterations with a burn-in period of 200 for each MCMC. Convergence was then confirmed formally using the BGR diagnostic as the PSRF for each parameter had a very narrow range between 1.00 and 1.01.

As previously described, during the Bayesian analysis every parameter contained within the hierarchical model is given a prior distribution. As little is known about change in knee joint

space width measurements over time, the parameters contained within the model were assigned non-informative priors. When convergence was initially assessed in the hierarchical model used to analysis the 19491 radiographic images from the 3469 study participants in the OAI it became clear from the trace plots that the prior assumptions placed upon the parameters used to calculate the variance of the random intercept were too narrow and resulting in an ill-fitting hierarchical model. After changing the prior assumptions and running a total of 4800 iterations with a burn-in period of 200 on visual inspection convergence of the Bayesian hierarchical model had occurred.

The posterior estimate for the overall estimate, θ, in the hierarchical model used to analyse longitudinal change in knee joint space width in the SEKOIA data was 3.48mm. The posterior estimate for the random slope parameter was -0.14mm. Therefore according to the results from the Bayesian analysis, on average, knee joint space width decreased by 0.14mm per year during the SEKOIA study. The minimum individual posterior estimate for the random slope effect was -0.64 and the maximum was 0.15. So one individual study participant within the SEKOIA data had a posterior estimate that indicates on average their knee joint space width reduced by 0.64mm per year, while the maximum posterior estimate for the random slope effect indicates that one study participant on average gained 0.15mm of knee joint space width per year. The mean posterior estimate for the random intercept effect was 0.003, with a maximum posterior estimate for the random intercept of 2.30 and a minimum of -2.56. When plotting the individual study participants random intercept estimates against their random slope estimate the line of best fit has a positive gradient. Thus those with larger random intercept values on average have larger random slope effect. When comparing the variance across all observations with the variance across the random slope and random intercept parameters the magnitude of the variance for the random intercept parameter is the largest.

Results from the Bayesian hierarchical modelling in the OAI estimate that on average knee joint space width reduced by 0.078mm per year, as the mean posterior random effect estimate was -0.078. The minimum individual posterior estimate for the random slope effect within the hierarchical model was -0.60 and the maximum posterior individual estimate was 0.37. The posterior mean estimate for the θ parameter was 3.97. The mean posterior random intercept estimate was -0.002, and the maximum individual estimate for the random intercept effect was

4.32 and the minimum individual estimate was -3.43. When plotting each study participant random intercept posterior estimate against their random slope effect the gradient of the line of best fit was positive. Therefore the higher an individual estimate of the random intercept parameter is the larger the posterior estimate of the individual random effect is. The variance at the observation level was 0.143, the variance of the random intercept parameter was 1.818 and the variance of the random slope parameter was 0.011.

7.6 Discussion

The aim of this chapter was to give a basic description of Bayesian modelling methodology, with a particular focus on hierarchical Bayesian modelling and assess the methodology as an approach to monitoring knee OA progression over time. During the literature review performed for this thesis, chapter 2, no previous studies were identified that applied Bayesian methodology to monitor change in knee joint space width over time. Thus although Bayesian hierarchical modelling was not newly developed for this thesis the application of the pre-existing statistical technique is novel in the assessment of change in knee joint space width over time. One of the many strengths of using the sophisticated statistical method of Bayesian modelling is that all the longitudinal observations can be utilised within the same analysis. The multiple measurements obtained per study participant do not need to be summarised into one summary measure of change. Bayesian hierarchical modelling allows different relationships between clusters to be modelled, using either the identical, independent or exchangeable parameter assumption. In the context of this thesis the exchangeable parameter assumption was used, thus it was assumed that the random slope effect for each study participant was similar to other study participants and therefore the individual random effect posterior estimates are drawn from the same prior distribution. Modelling under the exchangeable parameter assumption therefore allows for strength to be borrowed from other parameters, which leads to smaller credible intervals of estimates because each cluster's posterior borrows strength from other clusters posterior distribution through their joint influence on the population parameter (106). So using the exchangeable assumption to model change in knee joint space width over time in both SEKOIA and the OAI data allows for each study participant to have a unique trajectory, by modelling trajectory parameters as random effects from a common prior distribution. Another strength of using the exchangeable parameter assumption during the Bayesian hierarchical modelling process is that the assumption leads to global smoothing of uncertainty. This in turn

leads to the width of the credible intervals for each random effect being more equal across all clusters than if modelled using the independent structure assumption.

A further advantage of using Bayesian hierarchical analysis with an exchangeable parameter assumption is shrinkage to the mean which occurs in data with an unbalanced design. Shrinkage to the mean is where clusters with larger number of observations contribute more to the overall mean than those with smaller numbers of observations. Thus those clusters with smaller number of observations are effectively being supplemented with information from clusters with large numbers of observation, as under the exchangeable assumption it is assumed the clusters come from the same common prior distribution. In a similar vein, clusters with extreme values/outliers are pulled towards the overall mean. Therefore this shrinkage to the mean is how Bayesian hierarchical models are able to handle measurements error in the outcome variable.

The characteristic feature of Bayesian analysis is the estimation of the posterior distribution, which is done so from a combination of a prior distribution and the likelihood function using a Markov chain simulation process to sample the posterior distribution until convergence is obtained. Therefore one of the most important elements of any Bayesian analysis is to ensure that convergence has occurred in the MCMC simulations. If convergence has not occurred then the inferences made from the Bayesian analysis will be incorrect (101), this is because if convergence of the MCMC simulation has not occurred when inferences are drawn the estimates will be unduly influenced by the initial values provided to start the simulation process. Methods that have been developed to handle the issue of posterior estimates being affected by lack of convergence including simulating multiple MCMCs with initial values that are well spread through the parameter space, monitoring of convergence, and discarding early iterations of the MCMC simulations. Monitoring of convergence can occur visually or formally, however there is no gold standard to confirm whether convergence has occurred. Currently at least thirteen different diagnostic tools have been proposed to assess convergence. Cowles and Carlin reviewed and compared the performance of the different methods to detect convergence, and they concluded that it is not possible to say with complete certainty that a MCMC simulation is representative of an underlying distribution (109). However Cowles and Carlin recommended that using a combination of strategies to assess convergence would allow for reasonable

certainty of convergence. Thus two MCMC simulations with very different initial values, and then both visual assessment, using trace plots, and formal assessment, using the BGR diagnostic, of convergence were used within this thesis.

Missing values often occur in longitudinal data, and can take different patterns. Sometimes the pattern is due to attrition in that study participants are lost to follow up and no further observations are obtained during the study period, or they can be missing by omission, which is where intermittent observations are available for analysis across the study duration. There is a wide breadth of literature on how to handle missing data in longitudinal studies (110-112), regardless of whether in a frequentist or Bayesian framework, however in-depth consideration of how missing data are handled in longitudinal analysis is beyond the scope of this thesis. However, it will be discussed in brief, as handling missing data is a strength of Bayesian analysis. In Bayesian hierarchical models, if outcomes are missing, during the modelling process estimates for the posterior outcome and posterior predictions of missing observations are simultaneously made under the assumption that the missing mechanism is missing at random. This means that no additional data manipulation or adaptation of the Bayesian model code need to be made if there are missing observations within the outcome variable and the missing at random assumption is appropriate. In the context on this thesis, it was assumed that missing knee joint space width observations were missing at random. Thus no adaptations to the model were required. Of note, the data under consideration in this thesis are missing in pairs. Thus if the study visit time was not available then no knee joint space width measurement was available, and vice versa. However in the wider application of Bayesian hierarchical modelling the Bayesian framework is extremely flexible and can accommodate data missing within the response/time variable. Bayesian hierarchical modelling can also accommodate different missing data mechanisms. Treatment of missing data is an important consideration in statistical analysis, especially within the context of longitudinal analysis, but full consideration of Bayesian methodology in the framework of missing data is outside the primary scope for this thesis.

The choice of priors for Bayesian hierarchical models could be seen as both an advantage and a disadvantage of the methodology. One of the major criticisms of Bayesian analysis is the choice of priors, as different values could result in different posterior estimates even if convergence of the MCMC simulations has occurred. There is scant information about trajectories of change in

joint space width measurements over time and so for the purpose of this thesis non-informative, vague, priors were used as the desire was to let the data direct the analysis. Although this choice of vague priors could be criticised, a meta-analysis carried out by Emrani found a mean rate of joint space narrowing of 0.13mm per year across the 27 studies included in the review (113). This mean annual rate of change is comparable to the posterior estimates of annual change obtained for both SEKOIA and the OAI, -0.14mm and -0.08mm respectively.

Although not a problem in the context of this thesis, a problem arises when using vague priors when there are only small amounts of data contained within the analysis. The intention in using vague priors is to influence the posterior estimates as little as possible, when only small amounts of data are available the results can be sensitive to the choice of priors. To gain a greater understanding of the effect that vague priors would have on analysis Lambert et al carried out a simulation study to assess the effect of 13 different prior distributions on the scale parameters for random effects meta-analysis (114). The study found that the choice of prior was particularly important when the number of clusters, level 2 study participants, was limited. With greater numbers of clusters, the estimated posterior effect size was found not to be biased, but the precision of the estimates varied greatly which leads to the credible intervals differing. Therefore the choice of prior distributions in Bayesian hierarchical modelling allows for a very flexible framework, although sensitivity to the choice of prior distributions should always be assessed when the numbers of clusters are small.

Despite increases in computer power over the previous few decades and the introduction of powerful software packages that were designed to specifically handle MCMC simulation methods, such as WinBUGS (103), a weakness of Bayesian analysis is that the MCMC simulation process can be extremely time consuming (115). The SEKOIA data contained 1765 observations from 559 study participants, and the OAI study contained observations from 19491 observations from 3469 study participants. In epidemiological research neither of these datasets are particularly large in size, however using Bayesian hierarchical models to assess change in knee joint space over time in these dataset took a reasonable amount of time to run. The SEKOIA analysis took roughly an hour to run, but the OAI analysis took over a day to obtain posterior estimates. All the Bayesian hierarchical models were fitted in the WinBUGS software through the R interface, R2WinBUGS (116). The linkage of these two software packages offers a powerful

and adaptable method for performing Bayesian analysis, but it should be noted that the coding language used by R is very different to that of the statistical software Stata (69). Thus the statistical package the researcher is familiar with, will affect their ability to implement the Bayesian analysis. Running the Bayesian hierarchical models using the WinBUGS software took 30 seconds for the SEKOIA data while the OAI data analysis took just under 10 minutes to run, however running the Bayesian hierarchical model for the required number of iterations for convergence to occur is only the first step of the analysis. All posterior estimates and graphical representations of estimates require the WinBUGS iterations to be decoded into a format that R is able to handle, and this took upwards of 30 minutes for the OAI iterations to be decoded and took multiple hours for individual posterior estimates of the random effects to be calculated from the multiple MCMC simulations across all iterations. This could be considered a limitation of Bayesian analysis, as changing a very small simple parameter within the hierarchical model can lead to another long process of obtaining individual posterior estimates.

Results from the Bayesian hierarchical modelling indicate that on study participants in SEKOIA had a reduction of 0.14mm per year on average, and a reduction of 0.08mm per year on average in the OAI. If considering change over the full durations of each study, study participants in SEKOIA on average lost 0.42mm of knee joint space width across the 3 year study duration, and study participants in the OAI on average lost 0.64mm across the total study duration. Of note, these annual change estimates obtained during Bayesian modelling are similar in magnitude to the estimates obtained from LME modelling, however a full comparison of the statistical methods is covered in greater detail in chapter 8.

As discussed, Bayesian hierarchical models offer many advantages when analysing longitudinal data and to date no previous studies have been identified that use such techniques to monitor OA progression, especially in the presence of measurement error. These models offer a flexible framework which allows for the combination of priors' knowledge and likelihood to be used in a simulation process to obtain posterior estimates of particular parameters. Different assumptions for the relationship between clusters can be made, and the use of the exchangeable assumption has the advantage of allowing strength to be borrowed from other clusters to improve posterior estimates in clusters with smaller number of observations. The exchangeable assumption also allows for the 'smoothing' of posterior estimates and therefore

accounting for measurement error in the knee joint space measurements. However an important aspect of Bayesian modelling is to ensure convergence has occurred to ensure posterior estimates can be relied upon, and, as highlighted, there is no current gold standard for monitoring convergence though there are many different methods currently available. A further consideration when using Bayesian hierarchical models to analyse longitudinal data is the computer power and time required to obtain posterior estimates of the parameters of interest. Due to these constraints Bayesian modelling does not have a direct application within a clinical setting, however this method of analysing longitudinal data could prove fruitful within a research environment. Using Bayesian hierarchical models to obtain estimates for individual annual change in knee joint space width means that during the modelling process any measurement error that may exist within the longitudinal knee joint space width measurements has been accounted for. Thus the individual study participant annual change could be used in further research projects, as by using these estimates it may become possible to identify risk factors and patterns of disease progression which have not been previously detected. In addition to using the individual estimates as outcome in further research, Bayesian clustering techniques could be used to group study participants together with similar patterns of change. Thus applying such techniques may allow for identification of previous undiscovered phenotypes for progression. Both these extension to Bayesian hierarchical modelling are discussed in chapter 9.

Chapter 8: Comparison of methods

8.1 Overview of chapter

Chapter 2 reviewed the statistical techniques currently used in musculoskeletal research to monitor the progression of knee OA. This review highlighted that none of the methods that have been previously applied have accounted for measurement error within knee joint space width measurements. In chapters 5, 6 and 7 three different statistical methods that account for measurement error were explored. Within this chapter the three different methods are compared in order to explore the strengths and limitations of each methods.

8.2 Comparison of results

8.2.1 Comparison of frequentist and Bayesian modelling

The frequentist LME modelling and Bayesian hierarchical modelling have a fundamental difference in the way in which probability is defined. Within a frequentist framework observed data are considered to be random variables, which means that if further observations were made under the same conditions the observations may differ. So the frequentist framework considers probabilities as frequencies, as the probability is only meaningful in the context of repeated experiments. Whereas under the Bayesian framework probabilities are considered as degrees of belief, that is probability is a way of quantifying certainty about a particular situation. So under the Bayesian framework observations just 'are', and so are not directly considered as random variables. The difference between the two frameworks is important when comparing the results, as the frequentist model focuses on the likelihood while the Bayesian model focuses on probability.

Regardless of the framework under which the analysis is completed, one of the main reasons for implementing both the sophisticated statistical modelling techniques was to enable change in knee joint space per year to be calculated. In a frequentist context a change in knee joint space

width per year for each individual study participant is calculated with estimates that were obtained during the statistical modelling process, in a post-hoc manner. Whereas the change in knee joint space width per year for individual study participants obtained during the Bayesian hierarchical modelling is calculated during the modelling process, as an average of the random effects for change across all MCMC iterations. Regardless of whether the individual change per year in knee joint space was calculated during the modelling process in the Bayesian analysis, or after the statistical modelling, in the frequentist framework, both provide an estimate of change per year in mm. Neither the LME nor the Bayesian hierarchical modelling are considered a 'gold' standard method. Since the two methods provide estimates of the same parameter, annual change in knee joint space, they can be compared using the Bland-Altman method for limits of agreement (117).

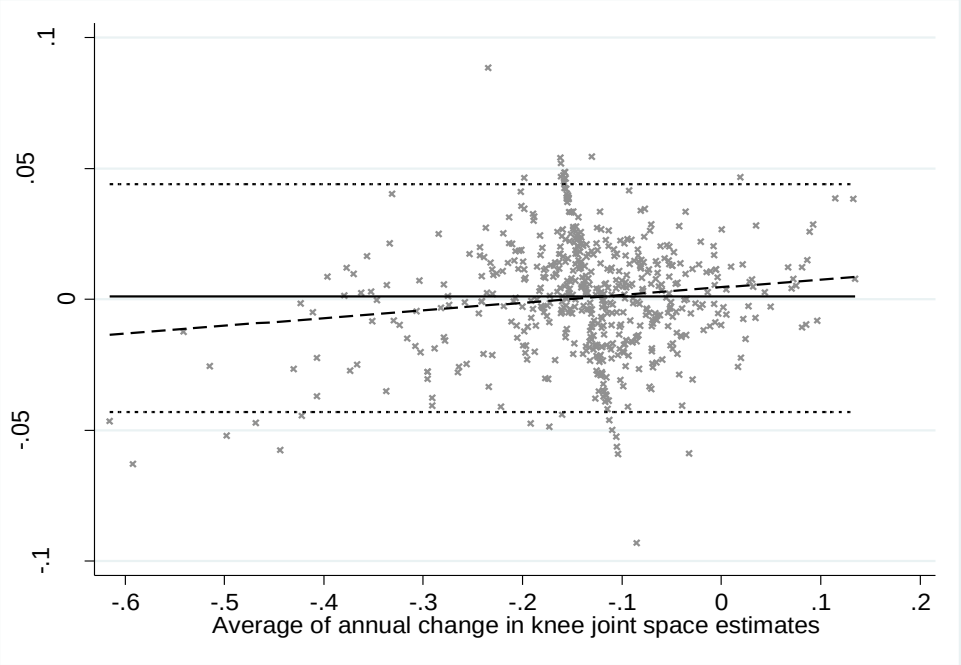
Even though annual change in knee joint space width calculated using LME modelling and Bayesian hierarchical modelling are estimating the same parameter, it is highly unlikely that the different statistical methods would lead to estimates that are identical. When assessing the level of agreement between annual change in knee joint space widths in SEKOIA, Figure 8-1, the estimates seem reasonably similar as there are few extreme outliers and the limits of agreement are narrower. The limits of agreement between the two methods were -0.043mm and 0.044mm, thus the difference in annual change in knee joint space width is less than 0.1mm between the two statistical modelling techniques. The Bland-Altman plot, Figure 8-1, appears to highlight two differences that are noticeably different in magnitude to the majority of estimates, -0.093 and 0.088. When investigating the data for these two study participants further, the study participant with a difference in estimates of -0.093 in SEKOIA had only one knee joint space width measurement at baseline and no follow-up measurement and the study participant with a difference in estimates of 0.088 had only two joint space width measurements, baseline and 12-month measurements.

The mean difference between the two estimates of annual change calculated using the two statistical modelling methods was 0.001mm, and from Figure 8-1 there appears to be a random pattern to the plotted estimates and so no systematic magnitude of difference exists between the two methods when applied to the SEKOIA data. In other words neither the Bayesian nor

frequentist modelling lead to individual estimates of annual change in knee joint space width that are consistently either smaller or larger in either methodology.

On inspection of the Bland-Altman plot for annual change in SKEOIA there appears to be a clear line of estimates. On further investigation into this pattern, it appears that these are the estimates for those study participants who only have a joint space width measurement at baseline.

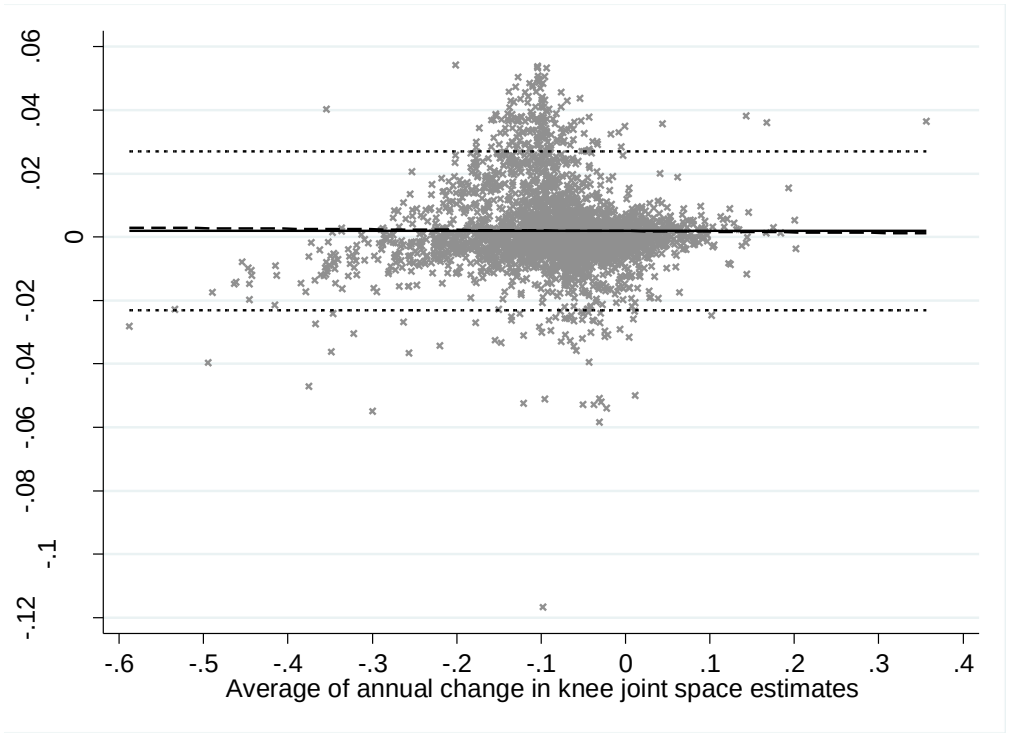
Figure 8-1 Bland-Altman plot for estimates of annual change (mm) in SEKOIA



The average, difference and limits of agreement between the annual change in knee joint space in all 3496 study participants in the OAI data are plotted on a Bland-Altman plot in Figure 8-2. The mean difference between the two estimates is 0.002mm, indicating that on average the estimates of annual change in knee joint space width obtained from the two statistical modelling methodologies are reasonable similar. The limits of agreements, indicated by the thin dashed lines, were -0.023mm and 0.027. So annual change in knee joint space estimates vary by a minimum of -0.023mm and a maximum of 0.027mm per year between the two statistical methodologies. From looking at the Bland-Altman plot there is no systematic trend in the

magnitude of the differences between the two estimates, since there is no apparent pattern to the plotted values, however one study participant appears to have very different annual estimates of change between Bayesian and frequentist modelling as the difference is -0.117mm. During the Bayesian hierarchical modelling the annual change in knee joint space for this study participant was -0.157mm and during the LME modelling the annual change was -0.040mm, thus under the Bayesian analysis the study participant lost 0.12mm of knee joint space per year, compared to an annual loss of 0.04mm calculated using LME modelling. When investigating the data for this study participant in the OAI, they had no baseline joint space width measurement, and although had a measurement for all remaining study visits, an increase of nearly 1mm occurs between the 24- and 36-month visits.

Figure 8-2 Bland-Altman plot for estimates of annual change (mm) in the OAI



The individual level parameters can be used to calculate population level estimates, namely the average knee joint space, the average change in knee joint space width per year across all study participants, the average random intercept, and the average random slope effects. The comparison of these results obtained during frequentist and Bayesian modelling in both SEKOIA and the OAI are contained in Table 8-1 and Table 8-2.

Table 8-1 Comparison of estimates from analysis in the SEKOIA data

Parameter	LME Model		Bayesian hierarchical model	
	Estimate	95% Confidence interval	Estimate	95% Credible interval
Average knee joint space width (mm)	3.487	3.418, 3.557	3.484	3.409, 3.556
Average change in knee joint space width (mm per year)	-0.136	-0.156, -0.116	-0.135	-0.393, 0.073
Variance of the average knee joint space width	0.632	0.554, 0.721	0.668	0.587, 0.758
Variance of change in knee joint space width	0.022	0.016, 0.029	0.025	0.019, 0.031
Model variance	0.094	0.085, 0.104	0.092	0.084, 0.102

The estimates from longitudinal modelling in both the frequentist and Bayesian framework in the SEKOIA data are reasonably comparable, Table 8-1. The average knee joint space width across the duration of the SEKOIA study, random intercept effect, was 3.487mm using the LME models and 3.484mm using Bayesian modelling. The average annual change in knee joint space width across all 559 study participants in SEKOIA was calculated as -0.136 from LME models and -0.135 in Bayesian modelling. So, on average, study participants lose 0.136mm per year of knee joint space width according to the LME modelling results and have a reduction of 0.135mm per year according to posterior estimates from the Bayesian analysis. The portioning of the variance in the longitudinal models was also comparable across the two statistical methodologies. Similar levels of variance in the average knee joint space widths, random intercept, were seen with a variance of 0.632 from LME modelling and 0.668 from Bayesian modelling. The variation around individual estimates of annual change in knee joint space width, random time effect, were almost identical from the two statistical model, with a variance of 0.02 from LME models and 0.03 in Bayesian models, and similarly the estimates for variation at the observations level, the model variance, were also almost identical. Regardless of which statistical framework was used to analyse the longitudinal knee joint space widths, the highest level of variance observed was around the individual estimates for annual change in knee joint space width, indicating that the greatest amount of variation in the statistical model is seen between study participants in the SEKOIA data.

The most marked difference in results between the LME modelling and Bayesian modelling results was seen at the 95% interval around the average annual change in knee joint space width. The 95% confidence interval for the annual change in knee joint space width, random time

effect, from LME models was -0.16 to -0.12, while the 95% credible interval around the population level annual change in knee joint space width was -0.39 and 0.07. Therefore the credible interval obtained during the Bayesian modelling was much wider than the confidence interval obtained during LME regression modelling. The frequentist 95% confidence interval is calculated such that 95% of the time the calculated confidence interval will include the true value, i.e. the parameter is fixed and the confidence interval is random. Whereas the Bayesian 95% credible interval treats the interval as fixed and the parameter as random, in other words 95% of the time the posterior distribution lies within the credible interval. So although both the credible interval and confidence interval are similar, there is a slight difference in the philosophy behind them, and this is reflected in the difference in magnitude of the two different intervals.

Table 8-2 Comparison of estimates from analysis in the OAI data

Parameter	LME Model		Bayesian hierarchical model	
	Estimate	95% Confidence interval	Estimate	95% Credible interval
Average knee joint space width (mm)	3.972	3.926, 4.018	3.972	3.928, 4.022
Average change in knee joint space width (mm per year)	-0.080	-0.085, -0.076	-0.078	-0.276, 0.057
Variance of the average knee joint space width	1.765	1.679, 1.856	1.818	1.730, 1.912
Variance of change in knee joint space width	0.011	0.010, 0.012	0.011	0.010, 0.012
Model variance	0.144	0.140, 0.147	0.143	0.139, 0.146

The results from the OAI longitudinal modelling in both a frequentist and Bayesian framework are comparable, as shown in Table 8-2. The population level random intercept effect, the average knee joint space width, was identical for the two statistical analyses, so both models calculated that study participants have an average knee joint space widths of 3.972mm. The average annual change in knee joint space width, random time effect, was also almost identical, with an estimate of -0.080 from LME modelling and -0.078 from Bayesian hierarchical modelling. So using the results from LME modelling on average knee joint space width reduced by 0.080mm per year, and by 0.078mm per year from Bayesian hierarchical modelling results. In a similar manner to the results from the SEKOIA longitudinal analysis, the majority of the variance within the OAI longitudinal models was at the individual study participant levels. Thus the variance of the random intercept was 1.765 in the LME model and 1.818 in the Bayesian model, compared with much smaller estimates of variance for the annual change, variance of the random time

effect, or at the knee joint space width measurement level, model variance. The most noticeable difference between the estimates obtained from the two modelling methodologies applied to the OAI data was the 95% interval around the average annual change in knee joint space width, namely the random time effect. The 95% confidence interval obtained during LME modelling was -0.085 to -0.076, compared with the much wider 95% credible interval obtained during Bayesian modelling, -0.276 to 0.057.

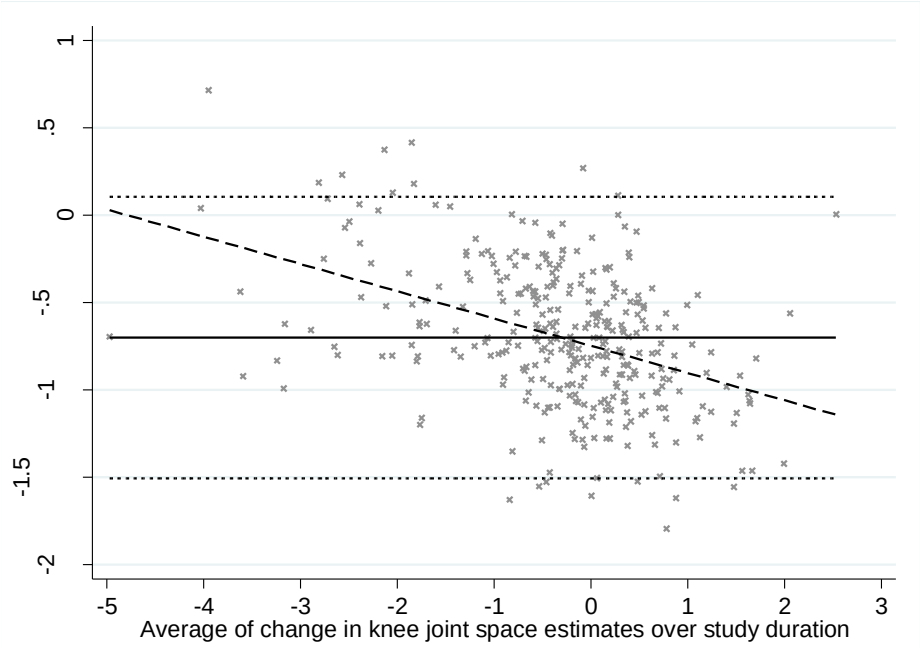
8.2.2 Comparison of RC index and statistical modelling

When assessing longitudinal change in knee joint space width by using LME and Bayesian hierarchical models, direct comparisons can be made in the estimates obtained because both statistical modelling techniques provide estimates of the same parameters. However comparison of the LME and the Bayesian hierarchical modelling results with the RC index results is a little more challenging. As previously stated the RC index is a simple statistical method that allows for the assessment of individual study participant reliable change between two study visits. The RC index gives a standardised z-score that is used to determine whether an observed change is statistically significantly reliable, thus the RC index calculation gives the ability to group individual study participants into three groupings; those who have had a statistically reliable increase in knee joint space width, those who have had a stable knee joint space width between the two study visits and those who have had a statistically significantly reliable decrease in knee joint space width. Therefore direct comparison of the output obtained from the RC index and from the frequentist and Bayesian modelling is not possible. However by standardising the individual BLUP estimates for the random slope effect from the LME models and the posterior estimates for the random slope parameter from the Bayesian analysis these can be compared to the z-scores obtained from the RC index calculations. By doing this it is possible to assess whether those study participants identified as having a statistically significant change in joint space width using the RC index were those identified as having larger yearly change in knee joint space width over the duration of the study using the modelling approaches.

A further consideration when attempting to compare the results from the RC index and the longitudinal statistical modelling is the time frames of data contained within the analysis. The RC index only compares two study visits during the calculation, thus pairs of data are only ever

included in the analysis, while the longitudinal modelling, regardless of frequentist or Bayesian, uses all available data across the study duration. To compare as well as possible, the results from the RC index considering change across the total study duration were compared with the standardised values of the random slope effects; Figure 8-3, Figure 8-4, Figure 8-5 and Figure 8-6. In a similar manor as the comparison of the change estimates obtained during the longitudinal modelling, there is currently no 'gold' standard method. Nevertheless the standardised estimates of change obtained during longitudinal modelling and the z-scores from the RC index are both aiming to provide an estimate of the severity of change over the duration of the study, thus the results are compared using the Bland-Altman method for limits of agreement (117).

Figure 8-3 Bland-Altman plot for estimates of change in SEKOIA from RC index and LME modelling

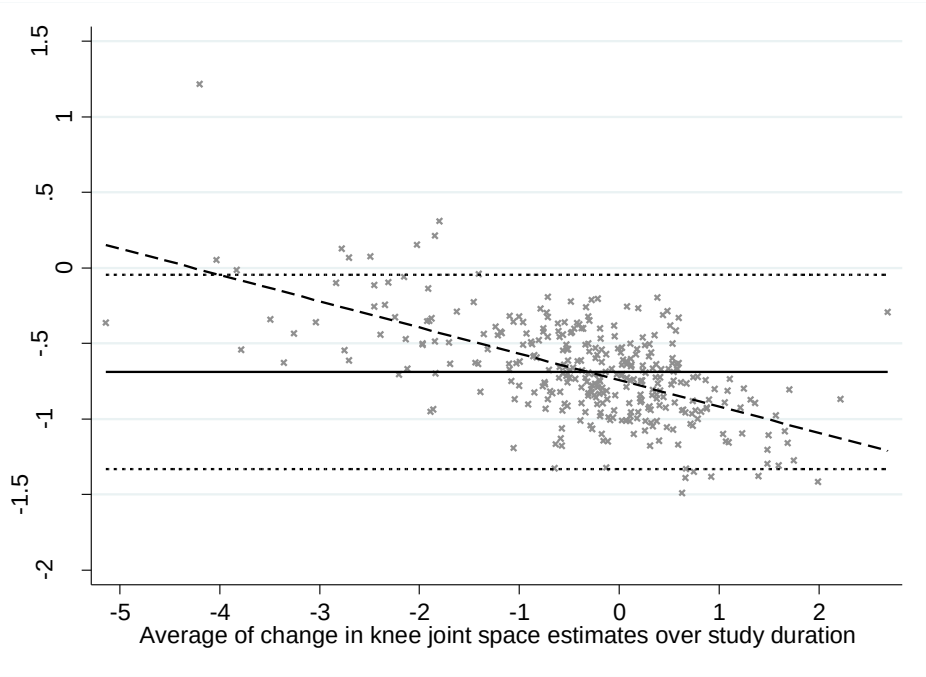


When the level of agreement between the RC index z-score and the standardised estimates of annual change obtained from LME modelling, Figure 8-3, the limits of agreement were -1.505 and 0.106. Thus the mean difference between the two methods was -0.699, meaning that the RC index scores for study participants within the SEKOIA data were consistently larger in magnitude than the standardised estimates of change from the LME modelling. From studying

Figure 8-3 it appears that there is a decreasing pattern to the Bland-Altman plot, indicating that as the differences between the two methods decrease in size the average of the two methods increases.

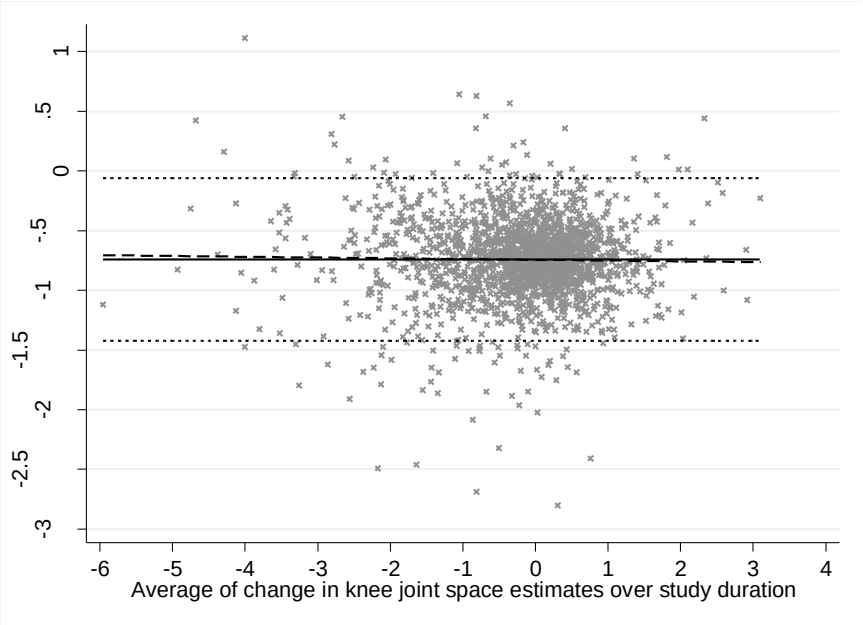
The Bland-Altman plot comparing the RC index z-scores and the standardised estimates of annual change in knee joint space width obtained from the Bayesian modelling in SEKOIA is contained in Figure 8-4. The limits of agreement between the two methods were -1.331 and -0.046, indicating that 95% of the time the difference between the RC index z-score and standardised change from Bayesian modelling was within a range of 1.3. The mean difference between the methods was -0.688, demonstrating that the RC index scores were on average larger in magnitude by 0.688 than the standardised change estimates obtained from the Bayesian modelling in SEKOIA. The plot appears to highlight one difference that is noticeably larger in magnitude than the other comparisons. When investigating the data for the study participant with a difference of 1.22 between the RC index estimate and the standardised Bayesian estimate in SEKOIA, the RC index score for this individual was -3.59 and the standardised Bayesian annual change was -4.81. When looking at the knee joint space width data for this study participants, they have measurements for all study visits over the duration of SEKOIA however an increase of over 2mm occurs in the knee joint space width measurements between the baseline and 12-month study visit. It is highly likely that an increase of such a magnitude is not biologically possible and so is measurement error with the data, however the RC index score used within the Bland-Altman is using joint space widths from baseline and 3-year study visits. Thus the RC index has not had to adjust for the measurement error, as the data considered as being an error was not included within the calculation.

Figure 8-4 Bland-Altman plot for estimates of change in SEKOIA from RC index and Bayesian modelling



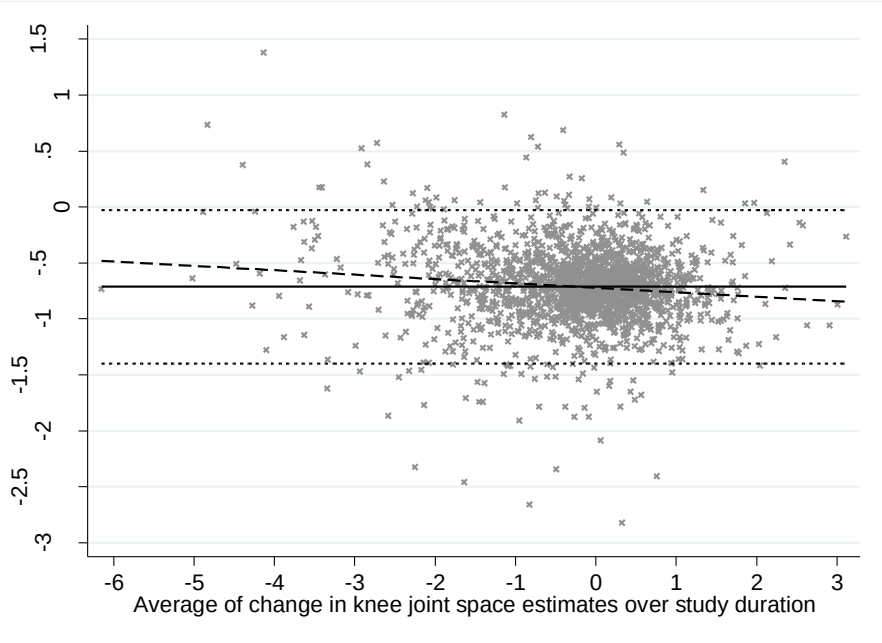
The limits of agreement, the average and the difference between the RC index z-scores and the standardised annual change from LME modelling in the OAI is plotted on a Bland-Altman plot in Figure 8-5. The limits of agreement between the two methods were -1.425 and -0.061, and the mean difference was -0.743. These statistics indicate that 95% of the differences between the two methods lies within a 1.4 range, and that the RC index scores tend to be larger in magnitude than the standardised annual change obtained from LME modelling. From studying the Bland-Altman plot there appears to be no apparent pattern to the plotted values, thus there is no systematic trend between the magnitude of the difference and average of the two methods.

Figure 8-5 Bland-Altman plot for estimates of change in the OAI from RC index and LME modelling



When considering the difference between the RC index scores and the standardised annual change obtained from Bayesian modelling of study participants within the OAI the limits of agreement were -1.400 and -0.027, Figure 8-6. The mean difference between the two methods was -0.713, and so the magnitude of the RC index scores was on average larger in magnitude than the standardised annual change from the Bayesian modelling. No significant pattern appears to exist between the magnitude of the differences between the methods and the magnitude of the average between the two methods, however one study participant appears to have a slightly larger difference between methods than the other OAI study participants as the difference is 1.38. When looking further at this study participants data, although they have a full series of knee joint space measurements across the duration of the study, a number of increases occurs during the study duration. The RC index only using knee joint space width measurements obtained at baseline and the 96th month study visit as a decrease is observed if comparing the baseline and 96th month measurements. So the observations which may contain elements of measurement error are not contained within the calculation.

Figure 8-6 Bland-Altman plot for estimates of change in the OAI from RC index and Bayesian modelling



Across all comparisons the mean difference between the RC index z-scores and the standardised annual change obtained during the longitudinal modelling in both SEKOIA and the OAI is negative. The RC index only assessing whether the absolute change is reliable between two time points, and so no 'smoothing' of measurement error occurs during the analysis. Whereas regardless of whether the longitudinal modelling was completed within a frequentist or Bayesian framework, both methods use all the observations available to obtain change estimates, and so both methods 'smooth' out extreme differences. Given these differences between the RC index scores and the standardised estimates obtained from LME and Bayesian modelling the observed difference between the estimates is no surprise.

8.3 Comparison of strengths and limitations

The analysis of longitudinal data can be considered to have three components,

- The nature of the outcome observations repeated over time
- The structure of the repeated observations
- The dependence or variance between observations over time

These need to be considered when analysing the longitudinal data as each component impacts upon the statistical analysis used, such as unequal intervals between measurement occasions, missing data and non-normality of the outcome observations. To account for the components of longitudinal data three methods to assess change in knee joint space width over time, paying particular attention to the presence of measurement error, have been explored in the previous chapters. Each of these methods has strengths and limitations, especially in the way in which each method accounts for measurement error in the knee joint space width measurements.

All three statistical methods explored within this thesis have the strength that change is assessed on an individual level. The majority of previous epidemiological studies that have focused on assessing change in knee joint space width over time have done so by summarising data at the study population level, whereas assessment of change at the individual level allows a greater understanding of the natural disease progression to be obtained.

The most striking difference, when considering the different methodologies, is the number of knee joint space width measurements that are used to assess change. The RC index uses only two time points, and thus gives information regarding individual change between two study visits. So, in a longitudinal study where multiple measurements have been obtained, use of the RC index only allows comparison of pairs of study visits; thus examination of the full trajectories of change over time using more than two measurements is not possible using this method. Whereas the advantage of using statistical modelling, whether LME models in the frequentist framework or Bayesian hierarchical models, is that all available joint space width measurements can be utilised within the analysis. The ability of the statistical modelling techniques to use all the available data means that a fuller picture of change patterns is assessed, rather than simple the difference between two points. These differences in assessment of change patterns is reflected in the Bland-Altman limits of agreement when comparing the RC index to the standardised annual change obtained from the longitudinal modelling.

The RC index is by far the simplest statistical method of the three explored, as the RC index score is calculated using a simple formula and the version of the RC index used within this thesis is calculated by using summary statistics obtained directly from the study population. But other

versions of the RC index formula, such as the version developed by Jacobson and Truax, require an test-retest reliability statistic, and Cronbach's alpha is often used, which requires multiple observations at the same study visit to allow for an estimate from the study population. Since the RC index is a relatively simple formula, the method has the added advantage that the RC index scores can be calculated without the requirement of any statistical package whereas both the Bayesian and frequentist longitudinal modelling require statistical software packages. Therefore the simplicity of the RC index gives the possibility of the technique being utilised within a clinical setting, as the threshold for reliable change could be applied to future patients presenting from the same hospital population.

LME models are an established methodology for analysing longitudinal repeated measures data and many software packages have built in commands to handle the analysis, such as the 'mixed' command in the statistical software STATA (69). With built-in commands also come manuals and guides, and so there is much literature and various tutorials to aid in the application of LME models to longitudinal data. In comparison to LME modelling and the RC index, Bayesian hierarchical modelling is the most complex and time-consuming analysis of the three methodologies compared, requiring multiple statistical software platforms to be used. However, the introduction of the specialist software WinBUGS in 1989 has enabled MCMC simulations to be implemented more easily, and this has made Bayesian methods slightly more widely used. Despite specialist software being available to implement both statistical modelling techniques, the statistical modelling methods would not be able to be utilised within a clinical setting. Whereas within a research environment, implementation of these methodologies could aid in the identified of risk factors for progression of knee OA.

Another longitudinal data component in which the three different methodologies differ is missing data. The RC index score is calculated across two observations, and both of these measurements are required to enable an individual score to be calculated. So, if a study participant is missing an observation at either time point, an RC index score cannot be calculated. This is thus a limitation of the RC index, as missing data is an inherent problem of longitudinal data. Missing data can lead to biased results, for example if attrition occurs due to worsening of the disease. In contrast, both the Bayesian hierarchical models and LME models are able to handle data with an unbalanced design, so that both statistical modelling

methodologies can handle longitudinal data in which each study participant has a different number of study visit observations. Thus a strength of using either statistical modelling technique over the RC index is that study participants with missing observations do not have to be excluded from the analysis, and use is made of all available data. This is an important consideration as collecting longitudinal data, regardless of whether in a clinical or research environment, is time consuming and expensive. So it is advantageous to use as much of the collected data as possible. However a detailed examination of how to handle missing observations in longitudinal data is not within the remit of this thesis project.

The component of structure of the repeated measurements needs to be considered in any longitudinal data analysis, and this is another area in which the three methodologies differ. Despite the best intentions, each study participants' observations are rarely collected at the exactly the same study intervals. Thus despite the study protocols of both SEKOIA and the OAI stating the period that should exist between study visits, there was variation in the intervals of different study participants, as noted in sections 3.4.1 and 4.4.1. The RC index is unable to account for the different durations between study visits, and the calculation simply focuses on the magnitude of the change between two time points however far apart. Thus the change in knee joint space width measurement of study participants with greatly differing durations between study visits end up being compared within the same analysis. This is a disadvantage of the method as disease deterioration occurs over time, and so, clinically, study participants whose disease had progressed over a longer amount of time end up being included in analyses containing participants with short study durations. This would lead to the data appearing more variable, and therefore the RC index may not identify as many study participants as having a significantly reliable change. In contrast the strength of the modelling techniques compared to the RC index, is that both LME and Bayesian hierarchical models can easily account for different time frames between observations by including a time element. Therefore when modelling longitudinal data to monitor disease progression, being able to incorporate observations that have been obtained at different time points is particularly clinically relevant and advantageous.

The main aim of implementing the three different methodologies in this thesis project was to assess how each method performs when monitoring change in knee joint space width over time in the presence of measurement error. Measurement error is a complex issue, and as indicated

previously can arise due to a number of reasons. Whether error arises during the measurement process, the limitation of the measurement instrument or fluctuations in the measurement over time, the resulting effect is the same, i.e. that change can be masked. The RC index accounts for any measurement error in the observed data by using the SD for the observations at each study visit being compared. The more variable a measurement is, the more likely that any change observed between observations is attributable to measurement error. Thus by using the SD the RC index accounts for the level of variation observed within the study data, and accordingly the more variable a measurement the fewer study participants the RC index would identify as having a statistically reliable change. Both the frequentist LME modelling and Bayesian hierarchical modelling handle measurement error in a slightly different manner to the RC index score. Both modelling methodologies 'smooth' the data, thus both methods smooth extreme values towards the population mean. Therefore both statistical methods use the strength of the study population estimates to smooth individual estimates to account for the presence of measurement error. But unlike the RC index scores, neither modelling technique enables identification of those study participants who had a significant change.

Measurement error due to biological fluctuations in the measurement over time are likely to be relatively small in magnitude when compared to the magnitude of measurement error that may occur during the measurement process. In other words the magnitude of measurement error in the knee joint space width measurement due to fluctuation in the time of the day the observation is made is likely to be considerably smaller than measurement error that is attributable to change in positioning of the knee during the radiograph. This is a consideration when applying the RC index for monitoring change, as the RC index is unable to distinguish large changes that may have occurred due to variability in the measurement process from true change. Thus a large change is likely to have an RC index score which indicates that the change is statistically reliable, whereas extreme change in knee joint space width, especially over a short study duration, is unlikely to be biologically possible.

A strength that the RC index has over the more complex statistical modelling methodologies is that the estimate of the standard error calculated during the analysis process can be used to quantify the joint space width change above which change could be considered statistically reliable. The magnitude of change is not an estimate of the amount of measurement error

observed between two time points, and it is not possible to say that the change in those individuals who did reach this threshold was solely due to measurement error. In a clinical setting the magnitude of the change in knee joint space width could inform if a patient had had a reliable deterioration in their knee OA disease. In contrast, the BLUPs from the LME models and the posterior estimate for the random slope effect give an individual estimate of change for each study participant but do not give information as to whether the individual change is statistically significant. Nor is it possible using either of these methods, to calculate a magnitude of change beyond which it can be said a statistically reliable change has occurred.

8.4 Summary

Both the LME models and Bayesian hierarchical models have proved to have extremely comparable results when simple models were applied to knee joint space width measurements as the outcome, and study visit time as an explanatory variable in both the OAI and SEKOIA data. The Bayesian hierarchical modelling estimates being comparable to the LME model results is as would be expected when providing the Bayesian model with non-informative priors.

Direct comparison of the RC index scores to the estimates obtained during the statistical modelling process was not possible, however standardisation of change estimates obtained during LME and Bayesian hierarchical modelling allowed for comparison of whether the different estimates had similar distributional values. Comparisons between RC index scores and standardised change estimates from both statistical frameworks demonstrated that the RC index scores always provided estimates that were smaller in magnitude.

All three statistical methods explored in this thesis were able to account for the presence of measurement error, either by accounting for the differing level of variation within knee joint space measurements or by smoothing individual estimates of change. But, none of the methods explored within this thesis were able to quantify the magnitude of measurement error. The RC index is able to provide a magnitude of change above which statistically significant change has occurred, but despite being able to calculate a magnitude of change it is not possible to identify those study participants whose measurements contain measurement error. Change in knee joint

space width measurements smaller than the magnitude of change obtained using the RC index could be attributable to measurement error but also attributable to minimal OA disease progression. The frequentist and Bayesian modelling applied in this thesis give an individual estimates of change, but unlike the RC index score it is not possible to determine whether an individual study participant annual change is statistically significant.

Chapter 9: Epidemiological implications and extensions

9.1 Overview of chapter

In the previous chapters the statistical techniques of the RC index, frequentist LME models and Bayesian hierarchical modelling were outlined, and the strengths and limitation of each method discussed. The statistical methodology explored thus far within this thesis have been in the simplest context, and this chapter presents results from extensions to the three statistical methods used in this thesis.

9.2 Epidemiological implications

Knee OA is ranked as the 11th highest contributor to disability (22), and therefore the burden of the disease is vast, whether measured as direct, indirect or intangible costs. At the individual level, knee OA is a debilitating disease which has a slow progression, often occurring over decades. Current strategies for the care of those diagnosed with knee OA focus on relieving symptoms in a hope of improving the function of the knee joint, as there is currently no treatment which has shown to consistently reverse or even slow down the progression of structural deterioration within the knee joint. Ultimately, as the disease progresses, the only available treatment for the disease is a joint replacement. Currently joint replacements have a restricted lifespan, and therefore the younger a patient receiving a joint replacement the higher the likelihood of the replacement knee joint failing during the patient's lifetime. If the replacement joint fails, more surgery will be required to replace the failed artificial joint, and this can only be performed a restricted number of times. So it would be preferable to identify factors that change the progression of knee OA, both protective and risk factors were identified, and to do this it is necessary to be able to identify and monitor reliable change in knee joint space width measurements.

Knee OA is an extremely complex disease, with the cartilage and many of the surrounding tissues of the joint affected and can be diagnosed using various different criteria. Regardless of

definition, the ultimate health care aim (10) is to be able to prevent the disease rather than responding to disease onset. For this aim to be achieved a full and complete understanding on the natural disease progression and risk factors is required.

Several risk factors for knee OA have been identified in previous epidemiological studies, and these risk factors divide into two broad categories: systemic and local. Local risk factors are those which commonly act on the biomechanical nature of the joint (19), and systemic risk factors are genetic or biochemical in nature. So, some of the risk factors that have been previously identified are modifiable and some are beyond control of the patient, such as genetic factors.

A well-established local factor associated with the risk of knee OA is weight, and in particular obesity, with an increase in weight being shown to be linked to increased risk of OA, particularly in the knee. Previous epidemiological studies have shown that obesity, defined as a BMI greater than 30 kg/m^2, can increase the risk of knee OA more than 2-fold, odds ratio (OR) of 2.8 (118). Not only has obesity been found to be significantly associated with knee OA, but generally as weight increases so does the risk of knee OA. Although the positive effect of identifying such a risk factor, is that weight is a modifiable behaviour and weight loss has been shown to improve symptoms of knee OA (119), it is nonetheless hard to achieve significant loss of weight.

Another well-established risk factor of OA, regardless of joint affected, is age, and previous epidemiological studies have shown that the incidence and prevalence of knee OA increase considerable with increasing age (120). The risk factor of age is deemed as a systemic factor, as the age of a patient is a factor which cannot be modified.

Other risk factors which have been consistently found to be associated with increased risk of knee OA include gender, ethnicity and traumatic injury. Previous studies have found consistent evidence that women are at higher risk of knee OA then men (121, 122), and the risk of developing knee OA increases dramatically around the time of menopause (38). Differing patterns of knee OA prevalence have also been shown between racial and ethnic groups, with higher rates of knee OA found among Chinese, African-Americans and Japanese when compared

to Caucasians (123-125). One of the risk factors found to have the strongest associations with knee OA is a traumatic knee injury, such as meniscal, ligament, and joint capsule tears, joint dislocations, and intra-articular fractures. When Silverwood et al completed a meta-analysis using thirteen cohort studies to assess the association between a previous knee injury and knee OA they found a nearly three-fold increase in those developing knee OA who had had a previous knee injury when compared to those without an injury, with an OR of 2.83 (121).

Another systemic risk factor that has been established as influencing knee OA is family history. Previous studies have shown that if parents or siblings have OA then the patient is more likely to suffer from OA, for example a study by Spector and MacGregor in 2004 suggested that estimates of heritability of OA could be around 50% or more (126). These studies of heritability of OA have led to research examining the genetic contribution to the development of OA, and to date NCOA3 and ALDH1A2 have been identified as risk loci for OA (127).

The evidence of other factors which may increase the risk of knee OA is a little more inconclusive, for example there has been conflicting evidence on whether smoking affects the risk of knee OA. Some studies have demonstrated an increased risk between smoking and knee pain in OA (128), whereas other studies have reported a protective effect of smoking on OA status (129). The gold standard epidemiological approach for monitoring disease progression within knee OA patients is to monitor continuous knee joint measurements, however due to measurement error, any change observed may not be ‘organic’ change and thus true relationships between risk factors and disease progression maybe masked.

9.3 Statistical extensions

The following sections briefly describe extensions or further analysis which have been completed to assess risk factors associated with knee OA using the three statistical methods for assessing change as previously outlined. Only a few characteristics are explored during the further analysis, as the extensions performed in this chapter are completed as proof of concept and not an exhaustive analysis of the risk factors of knee OA.

9.3.1 RC index extension

The RC index gives the ability to group study participants according to their RC index score; those who have had a statically reliable increase, those who have had a statically reliable decrease and those who have experienced no reliable change or stable disease state. Currently there are no statistical extensions within the literature that would allow risk factors of progression to be considered at the same time as assessing change. But once an RC index score has been calculated it becomes possible to use these groupings to compare characteristics of participants between the groupings.

Although the RC index is able to identify study participants with a reliable increase in knee joint space width, increases in knee joint space width are biologically possible, they occur only in very rare circumstances. So the focus of exploring an extension to the RC index will use only those study participants who were identified as having a reliable decrease against those with no change in knee joint space width and as the disease progression of knee OA is slow, the results of the RC index assessing change across the 3-year SEKOIA study and across the full 8-years of the OAI data are used.

Among the 336 study participants who had a knee joint space width measurements at both baseline and year 3 in the placebo arm of the SEKOIA study, 89% were identified as having no reliable change in knee joint space width over the duration of the study, and 36 study participants, 11%, were identified as having a reliable decrease in knee joint space width over the study duration. The baseline characteristics of these two groups of study participants are compared in Table 9-1. Those study participants who were identified as having a reliable decrease in knee joint space over the duration of the SEKOIA study were significantly heavier at baseline than those with no change, 86kg compared to 81kg respectively. Although, at baseline, the BMI of those with a decrease in knee joint space width was slightly larger the difference did not reach statistical significance. The study participants in the two groups were similarly aged at baseline and had similar average heights. At baseline the knee joint space width of both groups were comparable, an average of 3.56mm in those with no change compared to 3.53mm in those 36 study participants identified as having reliable change over the 3-year SEKOIA study. Those identified as having a reliable decrease in knee joint space width ended the 3 year SEKOIA study with an average joint space that was significantly narrower than those without change, 1.79mm

compared to 3.30mm respectively. At baseline, the study participants in both groups had been diagnosed with knee OA for similar lengths of time and reported similar levels of knee symptoms. A higher percentage of study participants in those identified as having reliable disease progression reported not drinking alcohol at baseline, 61% compared to 48% in the stable group.

Table 9-1 Risk factors by RC index grouping in SEKOIA

	Stable (n=299)		**Reliable decrease (n=36)**		**p-value**
	Mean	**SD**	**Mean**	**SD**	
Age (years)	62.82	7.25	61.94	6.94	0.49[a]
Weight (kg)	80.12	15.74	86.02	17.37	0.04[a]
Height (cm)	164.76	9.24	165.72	10.47	0.56[a]
BMI (kg/m)	29.49	5.25	31.23	4.91	0.06[a]
Joint space width at baseline (mm)	3.56	0.84	3.53	0.91	0.84[a]
Joint space width at 3 years (mm)	3.30	0.94	1.79	0.88	<0.01[a]
Disease duration (months)	76.14	75.95	77.28	80.92	0.93[a]
VAS knee pain	53.92	23.19	58.33	23.06	0.28[a]
WOMAC Component score					
Physical function	679.58	376.62	733.86	399.76	0.42[a]
Stiffness	87.87	49.11	100.54	48.09	0.15[a]
Pain	204.90	107.00	223.75	113.01	0.32[a]
	n	**%**	**N**	**%**	
Kellgren and Lawrence score					
2	197	65.9	19	52.8	
3	102	34.1	17	47.2	0.12[b]
Alcohol					
No	144	48.2	22	61.1	
Yes	155	51.8	14	38.9	0.14[b]
Physical activity					
Low (occasional)	36	15.8	2	8.3	
Medium (regular)	159	69.7	16	66.7	
Intensive (competitive)	33	14.5	6	25.0	0.36[c]
Smoking status					
No	199	65.6	24	66.7	
Ex/Yes	100	33.4	12	33.3	0.99[b]
Gender					
Men	88	29.4	13	36.1	
Women	211	70.6	23	68.9	0.41[b]

All continuous variables were normally distributed; [a]p-value for t-test; [b]p-value for Chi-square; [c]p-value for Fisher's exact

It was possible to calculate an RC index score between baseline and the 8-year study visit for 1918 study participants in the OAI. Of those 1729, 90%, were identified as having no reliable change in knee joint space, while 9%, 178, study participants were identified as having a reliable decrease in knee joint space width. The baseline characteristics between these two groups of study participants were compared and the summary statistics are presented in Table 9-2. No significant difference was found in the average age of study participants or height at baseline of study participants. Nor was a significant difference found in the smoking status of study participants, with around half of study participants in both groups reporting they had never smoked. Around 40% of study participants identified with no change were men, similar to the proportion in the reliable change grouping. Those identified as having a reliable decrease in knee joint space width were on average statistical significantly heavier and had a higher BMI. Interestingly, on average those study participants in the reliable decrease RC index group had higher levels of physical activity at baseline compared to those in the stable group, 184.5 and 167.7 respectively. Those study participants in the reliable decrease RC index group had an average higher joint space width at baseline, 4.50mm compared to 4.23mm, and had a much narrower knee joint space width after 96 months, 2.06mm compared to 3.89mm. Both of these differences in joint space width reached statistical significance. At baseline, those who had a reliable decrease in knee joint space width had differing levels of pain and function compared to those with no change in knee joint space width. Interestingly those who had a decrease in knee joint space width had significantly lower levels of pain at baseline using the KOOS pain score, median score of 86 compared to 92. A significant difference was also found in the proportions of study participants K&L scores between the two RC index groupings, with those study participants identified as having a reliable decrease tending to have higher K&L scores. For example 24% of study participants in the reliable decrease group had a K&L score of 3 at baseline compared to 16% of study participants in the group with no change.

Table 9-2 Risk factors by RC index grouping in the OAI

	Stable (n=1729)		Reliable decrease (n=178)		p-value
	Mean	**SD**	**Mean**	**SD**	
Age (years)	60.10	8.76	59.75	8.03	0.61[a]
Weight (kg)	80.91	15.95	86.31	16.04	<0.01[a]
Height (cm)	168.24	9.12	168.56	9.39	0.66[a]
BMI (kg/m)	28.44	4.71	30.34	4.97	<0.01[a]
PASE* Score	167.7	80.9	184.5	90.2	0.01[a]
Joint space width at baseline (mm)	4.23	1.02	4.50	1.09	<0.01[a]
Joint space width at 96 months (mm)	3.89	1.23	2.06	1.09	<0.01[a]
	Median	**Interquartile range**	**Median**	**Interquartile range**	
KOOS Pain score	91.67	77.78 – 100.00	86.11	72.22 – 97.22	<0.01[b]
WOMAC Component score					
Physical function	3.00	0.00 – 10.00	5.00	0.00 – 16.00	<0.01[b]
Stiffness	1.00	0.00 – 2.00	1.50	0.00 – 3.00	0.08 [b]
Pain	1.00	0.00 – 3.00	2.00	0.00 – 4.00	<0.01[b]
	N	**%**	**n**	**%**	
Kellgren and Lawrence score					
0	414	24.0	27	15.4	
1	252	15.6	21	11.9	
2	767	44.4	84	47.7	
3	282	16.3	43	24.4	
4	12	0.7	1	0.6	0.01[c]
Alcohol					
No	288	16.7	37	20.9	
Yes	1434	83.3	140	79.1	0.16[c]
Smoking status					
Never	943	55.0	93	53.5	
Current	100	5.8	16	9.2	
Former	672	39.2	65	37.3	0.21[c]
Gender					
Men	693	40.1	79	44.4	
Women	1036	59.9	99	55.6	0.27[c]

*PASE – Physical activity scale for the elderly. [a]p-value for t-test; [b] p-value for Wilcoxon rank sum; [c]p-value for Chi-square

9.3.2 RC index extension summary

Weight was found to differ significantly between those identified as having a reliable decrease in knee joint space width and those having no change using the RC index in both SEKOIA and the OAI study, with study participants in the reliable decrease groups being on average heavier compared to those with no change. This is as expected since being overweight, and thus heavier

in weight, is a well-established risk factor for knee OA progression. The association was further confirmed in the OAI data as mean BMI was found to differ significantly between the two groupings, with those who had a reliable change having a higher BMI value than those with no change in knee joint space width.

Of interest, the mean joint space width of study participants with a reliable decrease in both SEKOIA and the OAI was significantly smaller at the end of the study than those study participants without any change. This result is as would be expected, as by calculating RC index scores the aim is to identify those individuals who have had a reliable change in knee joint space width.

There is little consensus in previous epidemiological studies about the effect alcohol consumption has on the risk of knee OA. Therefore it may be of interest that a significant association was found between reported alcohol consumption and RC index grouping in the SEKOIA study. A higher proportion of those study participants identified as having reliable change reported not drinking any alcohol on entry into the study than those with stable/no change in knee OA disease, however this difference did not reach statistical significance.

Despite age being a well-established risk factor for knee OA, no significant differences were found between the RC index groupings examined. This is likely because both studies had age eligibility criteria; to be eligible for the SEKOIA study, participants had to be aged 50 years or over, and, for the OAI study participants had to be between 45 and 79 years. Thus due to eligibility criteria a narrow age range existed within both datasets, possibly masking any relationship between age and disease progression.

Using the RC index scores to group study participants to compare characteristics allows for a crude assessment of characteristics that may be linked to progression. However it is not possible to determine a magnitude of change associated with the characteristics. For example it is possible to determine whether those study participants who were deemed to have no change using the RC index were on average younger compared to those study participants who had

been deemed to have a statistically reliable change. But it is not possible to quantify the magnitude of the relationship between change in knee joint space width and age.

Despite the weaknesses of assessing potential risk factors in this way, the advantage is that the statistical techniques used are very simple. Distributions of continuous variables would need to be assessed to ensure the correct statistical comparison tests are used, but other than that any characteristic could be compared between the groups. Although a simple method, comparing the groupings in this way could inform future research of potential risk factors to explore further.

9.3.3 Statistical modelling extensions

The statistical modelling techniques of LME and Bayesian hierarchical models are used to calculate an annual change in knee joint space width value for each study participants. Traditionally, when regression analysis is performed, regardless of whether in the frequentist or Bayesian framework, the results are used to make inferences at the population level, thus make statements about the 'average' study participant. Full use is not made of the ability to obtain individual level estimates of change, and as highlighted by Bartlett et al understanding of individuals trajectories and therefore variability of trajectories of change are necessary to progress therapeutic treatments (130). So a natural extension to both LME and Bayesian modelling used within this thesis is to use the individual change values in further analysis. Thus the individual value for annual change in knee joint space width can be used to gain a fuller understanding of characteristics that maybe associated with change in knee joint space, by using the individual change values as an outcome variable in further regression analysis.

When the annual change in knee joint space width values obtained during the LME and Bayesian hierarchical modelling were compared in chapter 8, the values for each individual were similar, but, as expected, the parameters were not identical and so the individual annual change in knee joint space for each study participant from both the statistical modelling methodologies are used in further analysis. The BLUPs obtained from LME modelling in SEKOIA and the OAI were both normally distributed, Figure 6-7 and Figure 6-17, as were the posterior estimates of the random effect, Figure 7-5 and Figure 7-10, so, to explore the associations between study participant

characteristics and change in knee joint space width linear regression was used in the software package STATA 14 (69).

Individual estimated annual changes in knee joint space width in all 559 study participants in SEKOIA that were obtained during the LME and Bayesian modelling process were used to investigate the magnitude of association between change and particular characteristics. The results of the univariate regression analysis are contained in Table 9-3. Joint space width at baseline, joint space width at year 3 study visits and K&L score were all found to have a statistically significant relationship with annual change in knee joint space width values from both LME and Bayesian models, and the direction and magnitude of the relationship was similar between both annual change values. For example, those study participants with a K&L grade of 3 at baseline had lost significantly more joint space width per year compared to those with a baseline K&L score of 2. The association between annual change in knee joint space width and WOMAC pain score reached significance using the LME BLUPs but the relationship was not significant when assessing the relationship using the Bayesian posterior estimates, however the magnitude of the relationship was similar. An increase of one point in the WOMAC pain component score was associated with a significant decrease in annual knee joint space width of 0.0001mm. Increasing age, increasing weight and increasing BMI were all associated with a decrease in annual knee joint space width. An increase in VAS knee pain score and WOMAC stiffness score was also associated with a decrease in knee joint space width, but regardless of the framework in which the annual change value was calculated, the associations did not reach statistical significance. Compared to those who reported not smoking at baseline, those who reported being either an ex or current smoker lost more joint space per year on average; using the annual change values from the LME models the difference was 0.0027mm per year or using the Bayesian posterior estimates of annual change the difference was 0.0007mm. At baseline those who reported they completed intensive levels of physical activity had greater annual loss of knee joint space width compared to those reporting only doing occasional activity, while it appeared that medium (regular) levels of physical activity was protective of loss in knee joint space width compared to those doing low level of physical activity.

Table 9-3 Predictors of annual change in SEKOIA

	LME annual change Beta 95% C.I	**Bayesian annual change Beta 95% C.I**
Age (years)	-0.0006 (-0.002,0.001)	-0.0004 (-0.002,0.001)
Weight (kg)	-0.0004 (-0.001,0.000)	-0.0004 (-0.001,0.000)
Height (cm)	<0.0001 (-0.001,0.001)	-0.0002 (-0.001,0.001)
BMI (kg/m^2)	-0.0016 (-0.003,0.000)	-0.0014 (-0.003,0.000)
Joint space width at baseline (mm)	0.0403 (0.031,0.050)	0.0176 (0.007,0.028)
Joint space width at 3 years (mm)	0.0872 (0.080,0.094)	0.0811 (0.072,0.090)
Disease duration (months)	<0.0001 (-0.000,0.000)	<0.001 (-0.000,0.000)
VAS knee pain	-0.0002 (-0.001,0.000)	-0.0001 (-0.000,0.000)
WOMAC Component score		
Physical function	<0.0001 (-0.000,0.000)	<0.001 (-0.000,0.000)
Stiffness	-0.0001 (-0.000,0.000)	-0.0001 (-0.000,0.000)
Pain	-0.0001 (-0.000,-0.000)	-0.0001 (-0.000,0.000)
Kellgren and Lawrence score		
2	-	-
3	-0.0512 (-0.068,-0.035)	-0.0349 (-0.052,-0.018)
Alcohol		
No	-	-
Yes	0.009 (-0.007,0.025)	0.0101 (-0.007,0.027)
Physical activity		
Low (occasional)	-	-
Medium (regular)	0.0013 (-0.025,0.028)	0.0022 (-0.025,0.029)
Intensive (competitive)	-0.0049 (-0.039,0.030)	-0.0069 (-0.042,0.028)
Smoking status		
No	-	-
Ex/Yes	-0.0027 (-0.020,0.014)	-0.0007 (-0.018,0.017)
Gender		
Men	-	-
Women	0.0036 (-0.014,0.022)	0.0100 (-0.008,0.028)

In the OAI data, the individual annual change in knee joint space calculated from the LME and Bayesian modelling were used in linear regression analyses to assess associations between various characteristics and change. The results from the univariate regression analysis is contained in Table 9-4. An increase in weight, height and BMI were all found to be significantly associated with an annual reduction in knee joint space width, using the change values from both statistical frameworks. An increase of 1 kg in weight was found to be associated with a reduction of 0.0007mm per year in knee joint space using results from both LME and Bayesian models, and an increase of one kg/m^2 was associated with a reduction of 0.0026mm and 0.0023mm per year in knee joint space width, using LME and Bayesian result respectively. Knee joint space width at baseline, knee joint space width measurements at the 96^{th} month study visits, KOOS pain scores and WOMAC component scores were all found to significantly predict annual change in knee joint space width. In both statistical frameworks, an increase in WOMAC pain score predicted an annual reduction of 0.0025mm using LME change and 0.0018mm using Bayesian change. In a similar pattern to the SEKOIA data, in the OAI data who reported either being an ex or current smoker at baseline had greater levels of reduction in annual knee joint space width compared to those who reported never smoking, but the difference did not reach statistical significance. The severity of radiographic knee OA at baseline was also found to be associated with annual change in knee joint space width, with study participants with worse K&L scores on entry having greater annual reductions in knee joint space width compared to those with a K&L score of 0. Increasing age was found to be associated with larger annual reductions in knee joint space width, with a reduction of 0.0004mm per additional year of age from the LME model, and 0.0002mm per additional year of age from the Bayesian model. Although the relationship between individual annual change in knee joint space width and age reached significance when using the BLUPs from LME models, it was not found to be significant when using the posterior estimates of individual change from Bayesian hierarchical models.

Table 9-4 Predictors of annual change in the OAI

	LME annual change	Bayesian annual change
	Beta 95% C.I	**Beta 95% C.I**
Age (years)	-0.0004 (-0.001,-0.000)	-0.0002 (-0.001,0.000)
Weight (kg)	-0.0007 (-0.001,-0.001)	-0.0007 (-0.001,-0.001)
Height (cm)	-0.0003 (-0.001,-0.000)	-0.0004 (-0.001,-0.000)
BMI (kg/m^2)	-0.0026 (-0.003,-0.002)	-0.0023 (-0.003,-0.002)
PASE* Score	-0.0001 (-0.0001, -0.0001)	-0.0001 (-0.0001, -0.0001)
Joint space width at baseline (mm)	0.0143 (0.012,0.016)	0.0068 (0.005,0.009)
Joint space width at 96 months (mm)	0.0427 (0.041,0.045)	0.0417 (0.040,0.044)
KOOS Pain score	0.0005 (0.000,0.001)	0.0004 (0.000,0.001)
WOMAC Component score		
Physical function	-0.0007 (-0.001,-0.000)	-0.0006 (-0.001,-0.000)
Stiffness	-0.0037 (-0.005,-0.002)	-0.0026 (-0.004,-0.001)
Pain	-0.0025 (-0.003,-0.002)	-0.0018 (-0.003,-0.001)
Kellgren and Lawrence score		
0	-	-
1	-0.0075 (-0.018,0.003)	-0.0076 (-0.018,0.003)
2	-0.016 (-0.024,-0.008)	-0.0158 (-0.024,-0.008)
3	-0.0507 (-0.059,-0.042)	-0.0415 (-0.050,-0.033)
4	-0.0283 (-0.040,-0.016)	-0.0126 (-0.024,-0.001)
Alcohol		
No	-	-
Yes	0.0091 (0.002,0.016)	0.0086 (0.002,0.015)
Smoking status		
Never	-	-
Ex/Current	-0.0047 (-0.010,0.001)	-0.0053 (-0.011,0.000)
Gender		
Men	-	-
Women	0.013 (0.007,0.019)	0.0134 (0.008,0.019)

**PASE* – Physical activity scale for the elderly*

9.3.4 Statistical modelling extension summary

When assessing predictors of annual change in knee joint space width in SEKOIA, using the BLUPs calculated following LME modelling and the posterior estimates from Bayesian hierarchical modelling, as the outcome variable in regression analysis, baseline joint space width, joint space width at the 3 year study visit and baseline K&L score were all found to be statistically significantly associated. Additionally when using individual annual change from LME model the relationship between WOMAC pain and annual change was found to also be significant, however the relationship did not reach significance using change estimated during Bayesian analysis. Although the relationships between age, weight and BMI, and annual change in knee joint space width did not reach statistical significance the associations were in the direction that would be expected from previous literature.

All the characteristics considered, bar smoking status, were found to be statistically significantly associated with annual change in knee joint space in the OAI when using individual change estimates from LME models. When using the posterior individual estimates of change from Bayesian analysis, although increasing age was associated with a reduction in annual knee joint space width the relationship did not reach significance. From the literature the direction of the relationship were as would be hypothesised, with increasing weight, BMI and K&L score associated with greater reduction in annual knee joint space width.

To date no studies have made use of individual study participant's annual change in knee joint space width rate to explore characteristics which affect disease progression. By using the individual level change, namely the BLUPs and the posterior estimates, the change values have already been smoothed and thus adjusted for measurement error. In addition to the change values already being smoothed, using these measurements has the added advantage that, rather than categorising the change, the analysis has higher levels of statistical power. Thus using the individual change values in this way is a very flexible approach to gaining a fuller understanding of factors that are associated with knee OA disease progression, as any characteristics, regardless of nature of the characteristics, can be related to estimates of annual change.

Assessing potential risk characteristics in this way has the added advantage that an estimate of the magnitude of the association can be obtained. Thus not only is it possible to determine if the relationship is significant, but it is also possible to quantify the magnitude of the effect that risk factor has on change in knee joint space width. This means that, clinically, the magnitude of effect of each characteristic can be explored to gain a better understanding of which characteristics may have the greatest effect on disease progression, and, as this extension to LME modelling can be performed using regression analysis, risk factors can also be explored while adjusting for established risk factors and other confounders.

In this section, the annual change estimates obtained after the LME modelling, BLUPs, and the posterior estimates obtained during Bayesian hierarchical modelling were used as outcome measures. The regression coefficients obtained during the univariate linear regression analysis to determine predictors of change were comparable regardless of the statistical framework the individual change parameter was obtained in. Interestingly, the regression coefficients explaining relationships using the Bayesian change values are often more conservative in their magnitude compared to the results using the LME change parameter.

9.3.5 Bayesian hierarchical modelling extensions

As described in the previous section, the individual annual change in knee joint space calculated during the Bayesian hierarchical modelling process could be used in further analysis to investigate risk factors for progression of knee OA. However rather than using these estimates as an outcome variable to explore relationship with characteristics that may affect progression it is possible to use Bayesian analysis to cluster the outcome data into groups containing similar patterns. This analytical approach is a nonparametric Bayesian process called the hierarchical Dirichlet process (DP) (131). This method was developed by Teh and colleagues in 2005, and is also known as DP mixture models.

One of the greatest advantages of completing analysis within a Bayesian framework is the ease at which different analysis components can be added together. So far, within this thesis, Bayesian hierarchical models have been used to account for lack of independence between

observations, i.e. to account for knee joint space width measurements being clustered within study participants. During the Bayesian process, individual posterior estimates of random effects are calculated, namely the individual estimates of average knee joint space width and annual change in knee joint space widths. It is likely that similar patterns of change exist between individuals, and Bayesian analysis can be used to assess any underlying latent groups, commonly known as 'clusters' (132).

To explore underlying clusters, finite mixture models are built into the hierarchical modelling process with each cluster being characterised by a different mixture model. Finite mixture models assume that there are *K* clusters within the data, with each cluster having its own unique estimate of the parameters though some components of the mixture models are shared between clusters. In other words, the posterior estimates of the clusters each occupy a different proportion of the same underlying population level distribution, and clustering of the data in this way is made possible by applying DP mixture models. The DP is a probability distribution, $\Theta \sim Dirichlet(\alpha_1, \alpha_2, \dots, \alpha_m)$, whose range is itself a set of probability distributions, and it is used within Bayesian analysis to determine how likely a random variable is to be distributed according to a particular distribution. The DP process can be used to cluster observations using multiple parameters that are contained within the Bayesian analysis, but, in the context of this thesis, the DP process was used to group study participants into two clusters using the random time effect. Thus the DP process was used in both the SEKOIA and the OAI data to investigate the hypothesis that there are two unique patterns in the annual change in knee joint space width of study participants.

To group observations into two clusters, the DP gives a value of 1 or 2 to a posterior estimate of the random effect of time, depending on the cluster to which the study participant has the highest probability of belonging. These DP cluster values are then averaged across all iterations to give an average value. Since the number of clusters was set to two in the analysis, the minimum value that the averaged DP cluster value could have been was 1 and the maximum value was 2. If an individual's DP cluster value was 1 it meant that in 100% of MCMC iterations the study participant's posterior estimate was contained within the distribution of cluster 1, and if the value was 2 it meant the study participant's posterior estimate was contained within cluster 2's distribution in 100% of the iterations. The average DP cluster values were used to

group study participants into the two clusters, and then the characteristics between the clusters were explored to ascertain if there were any statistically significant differences between the clusters or, in the context of this thesis, whether a different OA disease progression phenotype existed between the different groups. As average cluster allocation values ranged anywhere between 1 and 2, if a study participant had an average cluster allocation value greater than 1.5 then the study participants was most likely to belong to cluster 2, as the DP cluster value would have been 2 a higher proportion of the time. Visa versa, if the average cluster value was less than 1.5 then the study participant was contained within cluster 1.

The characteristics of those study participants allocated to different clusters using DP finite mixture models during the Bayesian hierarchical analysis in the SEKOIA data are contained in Table 9-5. Using the DP to group participants with similar annual change patterns, 526 study participants were grouped into a cluster, named cluster 1, and a much smaller number of 32 study participants into a different cluster, cluster 2. The average individual annual change in knee joint space width of those in cluster 1 was -0.039, and the mean individual annual change in knee joint space width in those study participants in cluster 2 was -0.505. Therefore the study participants had an annual reduction of knee joint space width of 0.505mm compared to the annual reduction in knee joint space width of 0.039mm in cluster 1. Thus it can be hypothesised that those in cluster 1 had slow or no progression in their knee OA disease when compared to those identified in cluster 2 whose knee OA progressed more rapidly when considering the annual change in knee joint space width. Consequently, cluster 1 was classified as the stable cluster and cluster 2 was classified as the progression cluster.

When considering the different baseline characteristics between the two clusters, those in the progression cluster were on average older and taller than those with stable knee OA disease state, but the difference was not statistically significant. Those identified in the progression cluster were significantly heavier than those with stable disease state, 87kg and 80kg respectively, and had a higher BMI, 31.5 kg/m^2 and 29.7kg/m^2 respectively. At baseline the progression cluster had a narrower mean knee joint space width than the stable cluster, and the difference between mean knee joint space at the 3 year study visit was significantly different, with those having disease progression having a mean joint space width of 1.36mm compared to those in the stable cluster who had a mean of 3.29mm. The study participants grouped into the

progression cluster reported having been diagnosed with knee OA for a longer period of time than those in the stable cluster, and had higher VAS knee pain scores. At baseline, a significantly higher percentage of the study participants within the progression cluster had more advanced radiographic knee OA according to their K&L score, with 56% of those in the progression cluster having a K&L score of 3 compared to 36% of those in the stable cluster. When comparing the two clusters, a greater percentage of study participants reported drinking alcohol at baseline, fewer participants reported doing intensive levels of physical activity and a greater percentage of study participants reported either being a current or ex-smoker in stable cluster, however none of these differences reached statistical significance.

Table 9-5 Characteristics of Bayesian latent cluster analysis from SEKOIA

	Stable cluster (n = 526)		Progression cluster (n = 32)		p-value
	Mean	**SD**	**Mean**	**SD**	
Age (years)	62.77	7.51	63.22	7.31	0.74[a]
Weight (kg)	80.45	15.82	87.09	18.94	0.02[a]
Height (cm)	164.55	9.46	165.88	12.13	0.45[a]
BMI (kg/m^2)	29.67	5.11	31.54	5.35	0.05[a]
Joint space width at baseline (mm)	3.52	0.82	3.27	0.95	0.09[a]
Joint space width at 3 years (mm)	3.29	0.92	1.36	0.66	<0.01[a]
Disease duration (months)	74.38	75.45	84.66	89.38	0.46[a]
VAS knee pain	53.65	22.33	55.53	23.99	0.65[a]
WOMAC Component score					
Physical function	694.57	380.47	711.16	381.73	0.81[a]
Stiffness	90.95	50.71	94.19	45.40	0.73[a]
Pain	211.72	108.53	206.44	100.27	0.79[a]
	n	**%**	**n**	**%**	
Kellgren and Lawrence score					
2	336	63.9	14	43.8	
3	190	36.1	18	56.2	0.02[b]
Alcohol					
No	264	50.2	20	62.5	
Yes	262	49.8	12	37.5	0.18[b]
Physical activity					
Low (occasional)	62	15.7	2	9.5	
Medium (regular)	277	70.1	14	66.7	
Intensive (competitive)	56	14.2	5	23.8	0.42[c]
Smoking status					
No	336	63.9	22	68.8	
Ex/Yes	190	36.1	10	31.2	0.58[b]
Gender					
Men	154	29.3	13	40.6	
Women	372	70.7	19	59.4	0.17[b]

All continuous variables were normally distributed; [a]p-value for t-test; [b]p-value for Chi-square; [c]p-value for Fisher's exact

When DP finite mixture models were used in conjunction with Bayesian hierarchical modelling in the 3689 study participants in the OAI, 87% of study participants were allocated were allocated into one cluster and the remaining 246 study participants were grouped into another cluster. The average annual change in knee joint space width, the random effect, calculated during the Bayesian hierarchical modelling process for those in cluster 1, containing 3223 study participants, was -0.031, and the mean annual change in joint space width in cluster 2,

containing 246 study participants, was -0.334. Therefore those study participants in cluster 2 lost 0.334mm per year of knee joint space width compared to those in cluster 1 who had a reduction of 0.031mm per year. So in a similar manner to the clusters identified in SEKOIA, those study participants in cluster 1 had little knee OA disease progression compared to those in cluster 2, and consequently cluster 1 was labelled as the stable cluster and cluster 2 was labelled as the progression cluster.

Table 9-6 contains a description of baseline characterises between the two clusters identified during the DP finite mixture modelling process. Those study participants in the progression cluster were found to be significantly heavier than those in the stable cluster, with a weight of 88kg compared to 82kg. The baseline height between the study participants in both clusters was almost identical. A statistically significant difference was found between baseline BMI, with those in the progression cluster having a mean BMI of 31kg/m^2 compared to those in stable cluster with a mean BMI of 30kg/m^2. The mean knee joint space width at baseline of those study participants in the progression cluster was 4.13mm, compared to the mean knee joint space width in the stable cluster of 3.97mm, so, at baseline, those in the progression cluster had larger knee joint space widths than those in the stable cluster but the difference did not reach statistical significance. By the end of the OAI study, 96 months later, those study participants in the progression cluster had significantly narrower mean joint space widths compared to those in the stable cluster, 1.96mm compared to 3.86mm respectively. Similar proportions of study participants in both clusters reported drinking alcohol at baseline, around 80% in each cluster, and there were similar proportions of men and women in each cluster. No difference was found in the proportions of study participants smoking at entry in the OAI. A statistically significant difference was found in the median WOMAC physical function, stiffness and pain component scores, with those study participants in the progression cluster having significantly worse median scores than those in the stable cluster, indicating that those study participants in the progression cluster had worse knee physical function, higher levels of knee stiffness and higher levels of pain at baseline compared to those in the stable cluster of study participants. A significant difference was also found in K&L scores between the two clusters, with those study participants in the progression cluster tending to have a higher K&L score at baseline than those in stable cluster; for example 33% of study participants in the progression cluster had a baseline K&L score of 3 compared to 25% in the stable cluster.

Table 9-6 Characteristics of Bayesian latent cluster analysis in the OAI

	Stable cluster (n=3223)		Progression cluster (n=246)		p-value
	Mean	**SD**	**Mean**	**SD**	
Age (years)	61.63	9.19	60.98	8.11	0.28[a]
Weight (kg)	82.05	16.11	88.14	15.59	<0.01[a]
Height (cm)	168.15	9.19	168.74	8.88	0.34[a]
BMI (kg/m^2)	28.92	4.81	30.94	4.64	<0.01[a]
Joint space width at baseline (mm)	3.97	1.35	4.13	1.26	0.07[a]
Joint space width at 96 months (mm)	3.86	1.26	1.96	1.32	<0.01[a]
	Median	**Interquartile range**	**Median**	**Interquartile range**	
KOOS Pain score	86.1	69.4 - 97.2	85.2	66.7 - 97.2	0.05[b]
WOMAC Component score					
Physical function	5	0.0 - 15.9	7	1.0 - 19.1	0.01[b]
Stiffness	1	0.0 - 3.0	2	0.0 - 3.0	0.03[b]
Pain	2	0.0 - 5.0	2	0.0 - 5.0	0.05[b]
	n	**%**	**n**	**%**	
Kellgren and Lawrence score					
0	520	16.3	25	10.2	
1	349	10.9	25	10.2	
2	1237	38.8	106	43.3	
3	803	25.2	80	32.7	
4	283	8.9	9	3.7	<0.01[c]
Alcohol					
No	609	18.9	43	17.7	
Yes	2590	80.3	200	82.3	0.61[d]
Smoking status					
Never	1707	53.6	122	50.8	
Current	193	6.1	18	7.5	
Former	1287	40.4	100	41.7	0.56[d]
Gender					
Men	1314	40.8	113	45.9	
Women	1909	59.2	133	54.1	0.11[d]

[a]p-value for t-test; [b] p-value for Wilcoxon rank sum;[c] p-value for Fisher's exact; [d]p-value for Chi-square

9.3.6 Bayesian hierarchical modelling extensions summary

The DP mixture modelling was used in conjunction with Bayesian hierarchical models to group study participants into two separate clusters according to their annual change in knee joint space width. In both the OAI and SEKOIA data a progression cluster was identified which contained

study participants with larger annual mean loss of knee joint space width than those in the stable cluster.

Those in the progression cluster of the SEKOIA data were on average significantly heavier, had larger BMI, narrower average knee joint space widths at the 3 year study visits and had higher K&L scores at baseline compared to those in the stable cluster. A similar pattern existed within the OAI data, with those in the progression clusters being significantly heavier, having higher BMI's, narrower joint space width measurements at the 96^{th} month study visits, higher K&L scores and worse WOMAC component scores.

Within this section, the DP mixture models were used to group study participants into two clusters. Using DP mixture models to cluster study participants is a statistical way of using the data to drive the cluster allocation, rather than using an arbitrary cut-point. This is a strength of DP mixture modelling as it means that the clusters are created without any need for a subjective element to the creation of the clusters, but the negative aspect of the cluster creation being data driven is that the cluster may not have any clinical relevance. Not only may the clusters have no clinical relevance, but during the cluster allocation process at each MCMC iteration a study participant is given DP cluster value depending on the cluster to which the observation is most likely to belong. These DP cluster values only give information on the cluster and give no information about the cluster, thus further analysis is required to understand the differences between the clusters.

In the context of this thesis, the DP mixture models were used to group the study participants from SEKOIA and the OAI into 2 clusters. Not only does this technique allow for data driven cluster allocation but the methodology can also been used to determine the appropriate number of clusters. Thus during the DP mixture modelling process, the number of clusters would reflect the number of different distinct patterns that exist within the data, ensuring that cluster allocation occurs using the number of distinct patterns within the data. So the novel application of the DP mixture models in conjunction with Bayesian hierarchical models to assess longitudinal change in knee joint space width may allow for new phenotypes to be discovered which have not been previously identified.

9.4 Comparison of extensions

Each of the three statistical extensions explored within this chapter aims to investigate potential risk factors for change in knee joint space width after accounting for measurement error. Comparing the RC index score groupings gives the ability to compare differences in baseline characteristics between those identified as having reliable knee OA disease progression over the study duration. Using individual study participant change, obtained during statistical modelling, as an outcome variable gives the ability to determine which characteristics predict annual change. While Bayesian analysis can be used to cluster study participants can be identified who have similar patterns of annual change in knee joint space width, and the characteristics of these clusters can then be compared. A comparison of the risk factors that were explored within each extension, and a summary of whether the characteristic was found to have a significant association is contained in Table 9-7 and Table 9-8.

The only characteristic that was consistently found to be have a statistically significant relationship with knee OA progression across all the three statistical extensions explored in SEKOIA was knee joint space at year 3. As expected, those identified as having reliable OA progression using the RC index and those identified in a cluster with greater average change in annual joint space width using Bayesian clustering had significantly narrower knee joint space widths at year 3. While a statistically significant association was also found between an annual decrease in knee joint space width and narrower end of study knee joint space width measurements when assess using individual annual change as an outcome in regression analysis.

In SEKOIA, baseline weight of a study participants was found to be significantly associated with OA progression in the RC index and the Bayesian clustering analysis, with those identified as having OA progression being heavier on average at baseline. Baseline height and age of those study participants in SEKOIA were not found to have a significant relationship with OA progression regardless of the statistical methodology used. A higher percentage of study participants identified as having reliable OA progression using the RC index reported not drinking alcohol compared to those having a no change in knee joint space width, however no relationship between alcohol and knee OA progression was found using either of the other methodologies. BMI was only found to be significantly different between study participants in

the two clusters identified using the DP finite mixture models in the Bayesian analysis, with those in the cluster who on average had greater levels of annual knee joint space loss having a higher average BMI. A relationship between OA progression and BMI was not found using any of the other methodologies, but the mean BMI of study participants in the placebo arm on entry to SEKOIA was already 29.8kg/m^2 and so the study participants were already predominantly overweight. The remaining characteristics were not found to have a statistically significant association with progression of knee OA, regardless of the statistical framework used to explore associations.

Table 9-7 Comparison of explored OA risk factors for progression in SEKOIA

	RCI groupings	Statistical model extension		Bayesian cluster analysis
		LME annual change	Bayesian annual change	
Age (years)	X	X	X	X
Weight (kg)	✓	X	X	✓
Height (cm)	X	X	X	X
BMI (kg/m)	X	X	X	✓
Joint space width at baseline (mm)	X	✓	✓	X
Joint space width at 3 years (mm)	✓	✓	✓	✓
Disease duration (months)	X	X	X	X
VAS knee pain	X	X	X	X
WOMAC Component score				
Physical function	X	X	X	X
Stiffness	X	X	X	X
Pain	X	X	X	X
Kellgren and Lawrence score	X	✓	✓	✓
Alcohol	X	X	X	X
Physical activity	X	X	X	X
Smoking status	X	X	X	X
Gender	X	X	X	X

Table 9-8 contains a summary of the relationships between OA progression and study participants baseline characteristics when using the three statistical extensions explored within this chapter in the OAI data. Weight, BMI, knee joint space width at the 96th month study visit, KOOS pain score, WOMAC physical function score, WOMAC pain score and baseline K&L score were all found to have a statistically significant relationship with knee OA progression across the three methodological extensions. As expected increasing weight was found to be predict

significantly greater reductions in annual change in knee joint space using individual change as an outcome in regression analysis, and on average those identified as having a reliable change in knee joint space using the RC index were heavier. The smoking status of study participants in the OAI was not found to have a significant relationship with OA progression, and only the additional analysis using individual annual change as an outcome in regression analysis found gender, alcohol and height to be significantly associated with OA progression.

Table 9-8 Comparison of explored OA risk factors for progression in the OAI

	RCI groupings	Statistical model extension		Bayesian cluster analysis
		LME annual change	Bayesian annual change	
Age (years)	x	✓	x	x
Weight (kg)	✓	✓	✓	✓
Height (cm)	x	✓	✓	X
BMI (kg/m^2)	✓	✓	✓	✓
Joint space width at baseline (mm)	✓	✓	✓	X
Joint space width at 96 months (mm)	✓	✓	✓	✓
KOOS Pain score	✓	✓	✓	✓
WOMAC Component score				
Physical function	✓	✓	✓	✓
Stiffness	X	✓	✓	✓
Pain	✓	✓	✓	✓
Kellgren and Lawrence score	✓	✓	✓	✓
Alcohol	X	✓	✓	x
Smoking status	X	X	X	x
Gender	X	✓	✓	x

The statistical modelling extension using individual annual change in knee joint space width as an outcome is the only methodology that allows for the magnitude of the relationship between change in knee joint space width and particular characteristics to be quantified, and allows for potential predictors of change in knee joint space width to be examined. But the RC index and Bayesian clustering extensions both provide a statistical methodology to group study participants. The RC index group those study participants together who were identified as having reliable change in knee joint space width and the Bayesian DP mixture model group those study participants together who have similar patterns of change in knee joint space width. When applying Bayesian clustering to the OAI and SEKOIA data, one of the clusters always had greater annual reduction in knee joint space width and thus was classified as the OA progression cluster. Consequently, the study participants identified within the OA progression group using the RC

index and Bayesian analysis were compared to determine if the different statistical methodologies were identifying the same group of study participants.

RC index scores in SEKOIA identified 36 study participants who had had a reliable change in knee joint space width between baseline and the 3rd year of the study, and 32 study participants were identified as being in the OA progression cluster using Bayesian DP finite mixture models. Of the 32 contained within the Bayesian OA progression cluster, 6 were not able to be compared with the RC index score because missing knee joint space widths meaning it was not possible to calculate RC index scores. In those study participants in which a comparison would be made, 64% of those identified as having a reliable decrease in knee joint space width were also identified as being in the progression cluster Bayesian analysis, Table 9-9. Thirteen study participants were identified as having a reliable change in knee joint space width that were not grouped into the Bayesian OA progression cluster, and 3 study participants were grouped into the Bayesian OA progression cluster that did not have an RC index score which indicated a reliable change in knee joint space width.

Table 9-9 Comparison between RC index and Bayesian latent grouping in SEKOIA

	RC index grouping	
Bayesian latent grouping	Stable	Decrease
Stable cluster	296	13
Progression cluster	3	23

In those study participants in the OAI data in whom it was possible to calculate both an RC index score and a Bayesian cluster allocation, 178 study participants were identified as having a reliable decrease in knee joint space width using the RC index, and 125 study participants were grouped within the Bayesian OA progression cluster. Of those study participants identified as having reliable knee OA disease progression, 65% were also identified in the OA progression cluster using Bayesian analysis, Table 9-10. Only 9 study participants within the Bayesian OA progression cluster were not identified by the RC index analysis, whereas an addition 62 were

identified as having disease progression using the RC index who were not identified within the Bayesian clusters. A further 121 study participants were identified as being in the OA progression cluster in the Bayesian analysis, however because of limitations in the RC index methodology a score for these study participants was not able to be calculated due to missing data and so comparisons cannot be made in these study participants.

Table 9-10 Comparison between RC index and Bayesian latent grouping in the OAI

	RC index grouping	
Bayesian latent grouping	Stable	Decrease
Stable cluster	1720	62
Progression cluster	9	116

Interestingly, when comparing the RC index and Bayesian cluster groups similar proportions of study participants were identified using both methodologies in both the SEKOIA and the OAI. In SEKOIA 64% and 65% in the OAI were identified in the clusters from both statistical methodologies, which is of note as both datasets contain study participants at different points along the knee OA disease pathway.

Chapter 10: Discussion

10.1 Overview of chapter

This chapter provides an overview of the main findings, and a discussion of the strengths and limitations of the work completed for this thesis and highlights areas for future work.

10.2 Main findings

This work has explored monitoring knee OA progression while accounting for the complex issue of measurement error within continuous knee joint space width measurements. Measurement error is very difficult to quantify due to the different situations in which it can arise.

The majority of the epidemiological research reviewed for this thesis used simple statistical techniques, such as t-tests, and only one of the studies reviewed mentioned the issue of measurement error. None of the studies reviewed developed or implemented statistical techniques that would account for measurement error, and so the review highlighted the need for statistical techniques that would enable identification of real change in knee joint space width measurements over time.

Data from the placebo arm of the SEKOIA study and from the open access dataset the OAI were used for the novel application of the RC index, LME modelling from the frequentist paradigm and Bayesian hierarchical modelling to monitor knee OA diseases progression, while accounting for the presence of measurement error.

10.2.1 RC Index

The first aim of the thesis was to assess the effectiveness of the RC index as a novel approach to estimating OA progression accounting for the presence of measurement error, through longitudinal assessment of continuous knee joint space width measurements. The principle behind the RC index is to be able to determine which study participants have had a statistically significantly reliable change in knee joint space width across the differing time frames of interest. Initially developed by Jacobson and Truax (76), there are currently many different variations of the RC index. All variations follow the same fundamental expression, but each variation differs in the assumptions made of the data and thus they differ in how the fundamental elements of the RC index are calculated. The assumption made of the data used within this thesis was that variance of the joint space width measurements was different between each study visit, thus allowing for different distributions of knee joint space width measurements at each time point. Making this assumption meant the variation of the RC index that was applied was the formula proposed by Christensen and Mendoza (79).

The crude change between study visits for each study participant was also calculated. When the RC index scores were compared to the crude change, the proportions of study participants, in both SEKOIA and the OAI, identified as having a reliable change in knee joint space width dramatically reduced. Measurement error presents itself within data as variation, and the more variable a measurement under consideration the greater the chance of it appearing that change has occurred. The RC index accounts for measurement error as the calculation uses the SDs and variation from each study visit to calculate an estimate of the standard errors of the measurements. The more variable a measurement is, the larger the measurements SD, the less study participants who would be identified as having had a reliable change. This was demonstrated by the application of the RC index within the SEKOIA and the OAI studies. The SEKOIA data had less variability in knee joint space widths than the OAI, and so, as would be expected on average the proportion of study participants identified as having a statistically reliable decrease in joint space width was lower in the OAI than in SEKOIA.

The RC index has its merits, but there have also been debate and criticism of the technique. The major criticism of the technique is the lack of consensus as to which RC index formula should be

used. Despite this the RC index is a simple formula, and requires no specialist statistical software to be able to calculate the individual scores.

The two most popular statistical techniques and metrics that are currently used within musculoskeletal research to identify change are the SRM and SEM. However, unlike the RC index, both of these metrics only give information at the population level. Therefore one of the merits of using the RC index is that assessment of change at the individual level is possible. Additionally the estimate of the standard error of the measurements, that is calculated during the process of obtaining the RC index scores, can be used to quantify the joint space width change above which change could be considered statistically reliable. This enables future change measurement, within the same population, to be assessed using the cut-point of change obtained during the RC index calculation. So, in a clinical setting, a clinician would be able to use the cut-point to determine if the change in knee joint space width observed between clinic visits in a particular patient can be considered as reliable change or largely attributable to error within the knee joint space width measurements.

In a research environment, once the RC groupings have been identified, further statistical analysis can be completed which enables characteristics of study participants who have reliable OA progression to be identified. Completing such research may enable the identification of characteristics associated with OA progression that have previously been masked through measurement error.

10.2.2 LME modelling

The second aim of this thesis was to assess the novel application of the established frequentist technique LME modelling to model longitudinal change in knee joint space width, while accounting for any measurement error within the continuous observations. One of the fundamental reasons for modelling the knee joint space width measurements using such a sophisticated modelling technique is that all the observations are used within the same analysis, thus accounting for the repeated nature of the data within individuals. LME modelling allows for random intercept and regression coefficients within the models, and thus individual

trajectories of change in knee joint space width can be obtained during the modelling process. The individual estimates of change, the BLUPs, are calculated post-hoc, as, during the modelling process, all knee joint space width observations are used to calculate average population level change. So, the way in which the LME modelling process accounts for measurement error is by 'smoothing', with extremely large and small values being 'smoothed' towards the population average. Thus LME models use the strength of the study population estimates to smooth individual estimates to account for the presence of measurement error.

LME models offer a very flexible framework in terms of the data structure. Unlike other statistical methods for longitudinal data, LME models are able to model data with unequal intervals between measurement occasion, and data which have differing numbers of observations for each participant. Although very flexible in relation to data structure, a major assumption of LME modelling is that, unlike other frequentist multilevel models, the relationship between the outcome and the predictor variable must be linear. There is little evidence in previous epidemiological research about the trajectory of knee joint space width across the lifecourse, and so in the context of this thesis it was assumed that trajectories of change were linear, and this assumption was then tested both informally and formally.

To date, multilevel regression analysis has been used in musculoskeletal research to make inferences at the population level while accounting for possible confounders. So, using the individual annual change values as an outcome in linear regression analysis is a novel use of the BLUPs obtained during LME modelling. Completing such analysis allowed for predictors of change to be assessed, and the magnitude of the relationship between annual change in knee joint space width and risk factors to be quantified. However it should be noted that in this two stage process, although during the first the estimates of annual change are 'smoothed' to take account of errors within the knee joint space width measurements, errors will still exist within the estimates and thus these errors will be carried forward into the subsequent modelling so the resulting estimates will have standard errors that are too small.

LME modelling, although a more complex statistical technique than the RC index, is an established methodology for handling longitudinal data and thus statistical packages have built-

in commands to handle such analysis. In a research setting, LME models can be easily applied using specialist statistical software, but this methodology is less useful in a clinical setting as a clinician will not perform statistical modelling during patient consultations. Although LME modelling could not be directly applied within a clinical environment to aid with patient management, the novel ability to analysis these longitudinal measurements within a research environment may lead to new determinates of change being discovered and when these results are disseminated this is turn may inform clinical about risk factors for progression.

10.2.3 Bayesian hierarchical modelling

None of the previous epidemiological studies reviewed during this thesis used Bayesian analysis to monitor change in knee joint space width, and so an aim of the thesis was to assess the novel application of Bayesian hierarchical modelling methodology to longitudinal knee joint space width observations in the presence of measurement error. In a similar vein to LME models, one of the major advantages of using the sophisticated statistical method of Bayesian modelling is that all observations within the longitudinal data can be used within the same analysis. So, regardless of whether study participants had missing knee joint space width measurements and differing intervals between the study visits, account is taken of all these challenges of longitudinal data during the Bayesian modelling. Another advantage of the way in which Bayesian hierarchical analysis models the data is account can be taken of different relationships between the clustering units, which in the context of the thesis are study participants. It can be assumed that study participants are completely identical, thus trajectories of change would be exactly the same between study participants, independent, for which trajectories would be completely different and share no similarities, or exchangeable, for which study participants have unique but similar trajectories of change in knee joint space width. The assumption of exchangeability was made to model knee joint space widths over time in this thesis, as, under this assumption, strength can be borrowed from other study participants leading to smaller credible intervals of estimates.

The Bayesian hierarchical modelling framework is not only flexible to the data structure, but is also flexible to the model structure, so it is possible to model and monitor random effects for both the intercept and the slope for each study participants. These parameters can be

monitored, and averaged across all MCMC iterations to allow for the calculation of individual posterior estimates for annual change in knee joint space width and intercept effects. By using the exchangeable assumption, the posterior estimates are 'smoothed', and therefore account for measurement error in the knee joint space measurements.

The principle behind Bayesian analysis is that the analysis is based on a combination of prior knowledge and likelihood, though this principle can be seen as both an advantage and a disadvantage of Bayesian analysis. Informative priors can be used when explicit information is known about the data being analysed, whereas vague priors are used when little is known and it is desired that the priors have little influence and the analysis is driven by the data. Even using vague, non-informative priors, some level of information is still required to be specified for the analysis to be completed, thus a disadvantage of using Bayesian analysis and stating prior information is that different statisticians will apply different priors, and so there may never be an agreed 'objective' prior. Vague priors were used in the analysis completed in this thesis, and to ensure that the priors did not unduly affect the results, two MCMC were used within the analysis with two different sets of initial values.

Despite careful selection and checking of prior assumptions, the results of Bayesian analysis rely on model convergence, as, if convergence has not occurred, any inferences made from the Bayesian analysis will be incorrect. There are a number of diagnostic tools that have been developed to assess convergence, but currently there is no gold standard method for monitoring and confirming whether convergence has indeed occurred. To ensure the posterior estimates obtained in this thesis were as reliable as possible, early iterations of the MCMC simulations were discarded and both informal, trace plots, and formal, BGR diagnostic, were used to assess convergence of the hierarchical models.

In the same way as the BLUPs obtained during LME modelling, the individual posterior estimates of change can be used as an outcome variable to explore factors which may predict OA progression. This two stage analysis approach is a novel approach to assessing risk factors for change in knee joint space width which has not been previous discussed within the literature. Although using Bayesian modelling to 'smooth' the annual change estimates to errors within the

knee joint space width measurements, it should be noted that the annual change values would not be completely free of error and caution should be used when completing subsequent modelling to assess factors associated with progression.

Additionally the Bayesian framework allows for the use of DP mixture models, which enables groups of study participants with similar patterns of annual change in knee joint space width to be grouped together. This is a novel application of Bayesian analysis, as, to date, DP mixture models have not been previously used to identify phenotypes of OA progression.

A further consideration of Bayesian hierarchical models is that the analysis is by far the most complex of the three methodologies explored. Specialist software has been developed which has led to Bayesian analysis being more widely applied, but the computer power and time required to complete analysis is still a consideration. Despite the freely available specialist software, the complexity and the time required to run of Bayesian analysis means that this methodology would not be of direct use within a clinical environment. However application of DP mixture methods in a research environment may enable researchers to identify previously unknown risk factors of progression, and using DP mixture models could identify previously unknown phenotypes of OA progression. If confirmed these newly identified risk factors could inform clinicians of potential behaviour changes to improve patient management.

10.3 Epidemiological implications

Knee OA is a complex disease, not only affecting the cartilage within the joint but also many of the surrounding tissues. The complexity of the disease is not confined to the structures affected by disease progression, but also in the wide variety of signs and symptoms with which patients present. The most widely reported symptom of knee OA is pain, and this is what usually drives a patient to seek treatment. There are many different commonly-used and validated assessments for monitoring pain, and pain is often used as an outcome in epidemiological studies that focus on behaviour change. However, the only approved endpoint in clinical trials assessing the potential of DMOADs is joint space width. In previous epidemiological studies that monitored disease progression using joint space width, no mention was made of the issue of

measurement error. Therefore the statistical modelling methodologies explored within the work for this thesis may help identify new DMOADs. Thus, the clinical implications of applying these novel modelling methods could ultimately lead to new treatments being identified.

When used in a research environment, the implements of either of the statistical modelling techniques explored within this thesis could lead to the discovery of new and novel treatments for patients. But the RC index methodology has the potential to be used directly in a clinical setting, as the RC index formula is simple, requires no specialist statistical software and is quick to calculate. The RC index threshold could be calculated for the patient population within the clinical setting of interest, and if a clinician were presented with a patient from the same population then the calculated threshold could be used to aid the clinician to monitor whether the change in joint space observed in a patient is reliable change.

A number of risk factors for knee OA have been previously established, such as weight and age, although evidence of other factors that may increase the progression of knee OA is a little more inconclusive, such as smoking status. The novel application of the statistical methods explored in this thesis to monitor longitudinal knee joint space width measurements while accounting for the effect of measurement error, may lead to the identification of previously unknown risk factors or confirmatory evidence for risk factors with conflicting evidence.

10.4 Strength and limitations

This thesis assessed the novel application of established statistical techniques to musculoskeletal longitudinal data. To date no studies have been identified that have applied the RC index or Bayesian hierarchical modelling to data on knee OA disease progression. Despite a handful of studies being identified that used LME models in the frequentist framework to assess change, their authors focused on population level change. Thus it is believed this is the first time that individual study participants BLUPs have been used to assess change over time in knee joint space widths.

A comparison of the three statistical methods was made in chapter 8, along with a discussion of the strengths and limitations of each method. However there are limitations to this thesis study. Although all three statistical methods explored within this project appear to be able to monitor change in the presence of measurement error, none of the methods explored were able to quantify measurement error. This also means the 'true' value of each knee joint space width measurement is unknown, and thus there is no 'gold' standard against which to compare. Consequently a major limitation of this study is that it was not possible to compare the results obtained using statistical methods to adjust for measurement error, to observations that are 'true' measurements and contain no measurement error.

The longitudinal studies, SEKOIA and the OAI, used within this thesis are both 'prospective' studies, as the observations were made at the time of the study visit (133) rather that observations gathered about previous time points. Although the statistical methodologies are similar regardless of the type of longitudinal data, whether 'retrospective' or 'prospective', the type of data contained within this project can be regarded as a strength. Prospective data collection often tends to be prone to less measurement error (134), as there is no recall error and data collection is usually governed by protocols. Hence the studies used in the analyses contained within this thesis are large datasets that were rigorously collected, following strict protocols and thus measurement error has already been reduced to as small a magnitude as possible. Not only was the SEKOIA study data obtained under strict trial conditions, but on recruitment study participants had to have established knee OA and so the variability in knee joint space width due to disease status was reduced in these data. In contrast, there was no restriction due to knee OA disease state in the OAI dataset, and, on recruitment, study participants covered the entire spectrum of disease progression from onset to joint failure, allowing for further assessment of how the identified statistical methodologies performed across the knee OA lifecourse.

10.5 Further work

The focus of this thesis has been on developing statistical methodology to monitor change in continuous knee joint space width obtained from radiographs. From a musculoskeletal research perspective, future work could include evaluating the methodologies using knee joint space

widths obtained from other methods, such as MRI scans. Doing this would allow for further evaluation of how well the proposed statistical methods are able to identify change in knee joint space width over time.

In Chapter 6 BLUPS were obtained from LME modelling and in Chapter 7 posterior estimates from Bayesian modelling were obtained, both giving an estimate of annual change in joint space width. These estimates of change were used in Chapter 9 as outcomes in further analysis, but, although not stored during analysis for this thesis, both estimates have standard errors. Running the statistical analysis and storing the corresponding standard error would enable a reliable change to be calculated in a similar way to the RC index.

A Bland-Altman plot was used to compare the estimates of annual change in SEKOIA obtained from LME modelling and Bayesian modelling in Chapter 8. On inspection there was a clear line on the plot, Figure 8-1, and this line was found to be made up of those estimates from study participants who only had a baseline joint space width measurement. From looking at this line of estimates there appears to be no systematic difference, i.e. the differences between the two statistical method estimates are positive and negative, therefore no one method either over or under estimates annual change compared to the other. Further work could involve simulation or restriction of data used within the modelling to explore and understand these differences.

The work in Chapter 9 explored potential extensions to the three statistical methodologies being used within this thesis. Knee OA is a complex disease, and previous research has been able to identify several risk factors, but to be able to treat and prevent the knee OA disease a full and complete understanding of disease progression of risk factors is required. The analysis contained within chapter 9 only explored a handful of characteristics, so future work following this thesis would be to explore characteristics which have previously had conflicting evidence or new characteristics which have not been previously identified as being associated with knee OA disease progression.

Additionally, the DP mixture model extension for the hierarchical analysis in a Bayesian framework was explored using only 2 clusters. However this was an arbitrary number chosen to explore the technique. The DP mixture modelling process allows for an unknown number of mixture components to be modelled, and thus the number of clusters to be inferred from the data can vary. It would be interesting to explore this further, as it would allow for the issue of whether a variety of phenotypes of radiographic progression exists.

From a statistical perspective future work would include applying the identified methods to other types of longitudinal data to further evaluate how well the identified methods are able to handle measurement error, as different types of data will contain differing magnitudes of measurement error. For example, do the statistical methods explored perform just as well within longitudinal data that has greater levels of measurement error than in knee joint space width measurements?

10.6 Conclusion

In conclusion, this research applies previously established statistical techniques to the novel application of monitoring longitudinal knee joint space width measurements while accounting for the presence of measurement error within the observations. Application of the RC index dramatically reduced the proportion of study participants identified as having a change in knee joint space widths when compared to crude differences. The RC index score is used to classify those study participants with either stable disease state or reliable change in knee joint space width, while the application of statistical modelling in both a frequentist and Bayesian framework enabled individual annual change in knee joint space width values to be obtained. These estimates were used to investigate factors that may predict annual change, and using the change values in this manner also allows for the quantification of the effect each risk factors has on annual change.

If replicated in further studies, these methods could prove valuable in monitoring of disease progression and in identification of as yet undiscovered risk factors for the progression of knee OA. The RC index could be used to calculate threshold of change which could be directly applied

within a patient population to aid clinicians in monitoring progression of knee OA. In a research environment, if the research in focused on assessing change over time then the simpler LME modelling technique could be implemented to 'smooth' data for measurement error and obtain individual estimates of change. Yet, more complex Bayesian methodology can be used to investigate phenotypes for disease progression. Additionally, there has only been limited application of the statistical methods in this thesis in other clinical areas, such as the application of the reliable change index to monitor recovery following concussion (135), but wider application of these statistical methods in other clinical areas may aid in monitoring of other conditions.

Appendices

Appendix 1. List of studies included in literature review

Cromer MS, Bourne RM, Fransen M, Fulton R, Wang SC. Responsiveness of quantitative cartilage measures over one year in knee osteoarthritis: comparison of radiography and MRI assessments. Journal of Magnetic Resonance Imaging. 2014 Jan 1;39(1):103-9.

Benichou OD, Hunter DJ, Nelson DR, Guermazi A, Eckstein F, Kwoh K, Myers SL, Wirth W, Duryea J. One-year change in radiographic joint space width in patients with unilateral joint space narrowing: Data from the osteoarthritis initiative. Arthritis care & research. 2010 Jul 1;62(7):924-31.

Boegård TL, Rudling O, Petersson IF, Jonsson K. Joint space width of the tibiofemoral and of the patellofemoral joint in chronic knee pain with or without radiographic osteoarthritis: a 2-year follow-up. Osteoarthritis and cartilage. 2003 May 31;11(5):370-6.

Botha-Scheepers S, Dougados M, Ravaud P, Le Graverand MP, Watt I, Breedveld FC, Kloppenburg M. Effect of medial tibial plateau alignment on serial radiographs on the capacity to predict progression of knee osteoarthritis. Osteoarthritis and cartilage. 2008 Feb 29;16(2):272-6.

Bowes MA, Vincent GR, Wolstenholme CB, Conaghan PG. A novel method for bone area measurement provides new insights into osteoarthritis and its progression. Annals of the rheumatic diseases. 2015 Mar 1;74(3):519-25.

Bruyere O, Cooper C, Pavelka K, Rabenda V, Buckinx F, Beaudart C, Reginster JY. Changes in structure and symptoms in knee osteoarthritis and prediction of future knee replacement over 8 years. Calcified tissue international. 2013 Dec 1;93(6):502-7.

Bruyere O, Pavelka K, Rovati LC, Deroisy R, Olejarova M, Gatterova J, Giacovelli G, Reginster JY. Glucosamine sulfate reduces osteoarthritis progression in postmenopausal women with knee osteoarthritis: evidence from two 3-year studies. Menopause. 2004 Mar 1;11(2):138-43.

Bruyère O, Richy F, Reginster JY. Three year joint space narrowing predicts long term incidence of knee surgery in patients with osteoarthritis: an eight year prospective follow up study. Annals of the rheumatic diseases. 2005 Dec 1;64(12):1727-30.

Christgau S, Henrotin Y, Tanko LB, Rovati LC, Collette J, Bruyère O, Deroisy R, Reginster J. Osteoarthritic patients with high cartilage turnover show increased responsiveness to the cartilage protecting effects of glucosamine sulphate. Clinical and experimental rheumatology. 2004 Jan 1;22(1):36.

Cicuttini F, Hankin J, Jones G, Wluka A. Comparison of conventional standing knee radiographs and magnetic resonance imaging in assessing progression of tibiofemoral joint osteoarthritis. Osteoarthritis and cartilage. 2005 Aug 31;13(8):722-7.

Cline GA, Meyer JM, Stevens R, Buckland-Wright C, Peterfy C, Beary JF. Comparison of fixed flexion, fluoroscopic semi-flexed and MTP radiographic methods for obtaining the minimum medial joint space width of the knee in longitudinal osteoarthritis trials. Osteoarthritis and cartilage. 2006 Dec 31;14:32-6.

Conrozier T, Mathieu P, Piperno M, Favret H, Colson F, Vignon M, Conrozier S, Vignon E. Selection of knee radiographs for trials of structure-modifying drugs in patients with knee osteoarthritis: A prospective, longitudinal study of lyon schuss knee radiographs with the definition of adequate alignment of the medial tibial plateau. Arthritis & Rheumatology. 2005 May 1;52(5):1411-7.

Cotofana S, Buck R, Dreher D, Wirth W, Roemer F, Duryea J, Nevitt M, Eckstein F. Longitudinal (One-Year) Change in Cartilage Thickness in Knees With Early Knee Osteoarthritis: A Within-Person Between-Knee Comparison. Arthritis care & research. 2014 Apr 1;66(4):636-41.

Deberg MA, Labasse AH, Collette J, Seidel L, Reginster JY, Henrotin YE. One-year increase of Coll 2-1, a new marker of type II collagen degradation, in urine is highly predictive of radiological OA progression. Osteoarthritis and cartilage. 2005 Dec 31;13(12):1059-65.

Dieppe P, Cushnaghan J, Young P, Kirwan J. Prediction of the progression of joint space narrowing in osteoarthritis of the knee by bone scintigraphy. Annals of the rheumatic diseases. 1993 Aug 1;52(8):557-63.

Dieppe PA, Cushnaghan J, Shepstone L. The Bristol 'OA500'study: progression of osteoarthritis (OA) over 3 years and the relationship between clinical and radiographic changes at the knee joint. Osteoarthritis and Cartilage. 1997 Mar 1;5(2):87-97.

Eaton CB, Sayeed M, Ameernaz S, Roberts MB, Maynard JD, Driban JB, McAlindon TE. Sex differences in the association of skin advanced glycation endproducts with knee osteoarthritis progression. Arthritis research & therapy. 2017 Feb 17;19(1):36.

Eckstein F, Boudreau R, Wang Z, Hannon MJ, Duryea J, Wirth W, Cotofana S, Guermazi A, Roemer F, Nevitt M, John MR. Comparison of radiographic joint space width and magnetic resonance imaging for prediction of knee replacement: A longitudinal case-control study from the Osteoarthritis Initiative. European radiology. 2016 Jun 1;26(6):1942-51.

Edwards MH, Parsons C, Bruyère O, Dop FP, Chapurlat R, Roemer FW, Guermazi A, Zaim S, Genant H, Reginster JY, Dennison EM. High Kellgren-Lawrence grade and bone marrow lesions predict worsening rates of radiographic joint space narrowing; The SEKOIA Study. The Journal of rheumatology. 2016 Mar 1;43(3):657-65.

Fukui N, Yamane S, Ishida S, Tanaka K, Masuda R, Tanaka N, Katsuragawa Y, Fukui S. Relationship between radiographic changes and symptoms or physical examination findings in subjects with symptomatic medial knee osteoarthritis: a three-year prospective study. BMC musculoskeletal disorders. 2010 Nov 24;11(1):269.

Gandy SJ, Dieppe PA, Keen MC, Maciewicz RA, Watt I, Waterton JC. No loss of cartilage volume over three years in patients with knee osteoarthritis as assessed by magnetic resonance imaging. Osteoarthritis and cartilage. 2002 Dec 1;10(12):929-37.

Gensburger D, Roux JP, Arlot M, Sornay-Rendu E, Ravaud P, Chapurlat R. Influence of blinding sequence of radiographs on the reproducibility and sensitivity to change of joint space width measurement in knee osteoarthritis. Arthritis care & research. 2010 Dec 1;62(12):1699-705.

Hunter DJ, Beavers DP, Eckstein F, Guermazi A, Loeser RF, Nicklas BJ, Mihalko SL, Miller GD, Lyles M, DeVita P, Legault C. The Intensive Diet and Exercise for Arthritis (IDEA) trial: 18-month radiographic and MRI outcomes. Osteoarthritis and cartilage. 2015 Jul 31;23(7):1090-8.

Jubb RW, Piva S, Beinat L, Dacre J, Gishen P. A one-year, randomised, placebo (saline) controlled clinical trial of 500-730 kDa sodium hyaluronate (Hyalgan) on the radiological change in osteoarthritis of the knee. International journal of clinical practice. 2003;57(6):467-74.

Karsdal MA, Bihlet A, Byrjalsen I, Alexandersen P, Ladel C, Michaels M, Andersen JR, Riis BJ, Kraus V, Bay-Jensen AC, Christiansen C. OA phenotypes, rather than disease stage, drive structural progression–identification of structural progressors from 2 phase III randomized clinical studies with symptomatic knee OA. Osteoarthritis and Cartilage. 2015 Apr 30;23(4):550-8.

Kraus VB, Feng S, Wang S, White S, Ainslie M, Graverand MP, Brett A, Eckstein F, Hunter DJ, Lane NE, Taljanovic MS. Subchondral bone trabecular integrity predicts and changes concurrently with radiographic and magnetic resonance imaging–determined knee osteoarthritis progression. Arthritis & Rheumatology. 2013 Jul 1;65(7):1812-21.

Lanyon P, Jones A, Doherty M. Assessing progression of patellofemoral osteoarthritis: a comparison between two radiographic methods. Annals of the rheumatic diseases. 1996 Dec 1;55(12):875-9.

Lapane KL, Yang S, Driban JB, Liu SH, Dubé CE, McAlindon TE, Eaton CB. Effects of prescription nonsteroidal antiinflammatory drugs on symptoms and disease progression among patients with knee osteoarthritis. Arthritis & rheumatology. 2015 Mar 1;67(3):724-32.

Le Graverand MH, Buck RJ, Wyman BT, Vignon E, Mazzuca SA, Brandt KD, Piperno M, Charles HC, Hudelmaier M, Hunter DJ, Jackson C. Change in regional cartilage morphology and joint space width in osteoarthritis participants versus healthy controls: a multicentre study using 3.0 Tesla MRI and Lyon–Schuss radiography. Annals of the rheumatic diseases. 2010 Jan 1;69(01):155-62.

Le Graverand MH, Vignon EP, Brandt KD, Mazzuca SA, Piperno M, Buck R, Charles HC, Hunter DJ, Jackson CG, Kraus VB, Link TM. 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. Annals of the rheumatic diseases. 2008 Nov 1;67(11):1562-6.

le Graverand MP, Clemmer RS, Redifer P, Brunell RM, Hayes CW, Brandt KD, Abramson SB, Manning PT, Miller CG, Vignon E. A 2-year randomised, double-blind, placebo-controlled, multicentre study of oral selective iNOS inhibitor, cindunistat (SD-6010), in patients with symptomatic osteoarthritis of the knee. Annals of the rheumatic diseases. 2013 Feb 1;72(2):187-95.

Lu B, Ahmad O, Zhang FF, Driban JB, Duryea J, Lapane KL, McAlindon T, Eaton CB. Soft drink intake and progression of radiographic knee osteoarthritis: data from the osteoarthritis initiative. BMJ open. 2013 Jul 1;3(7):e002993.

Marsh M, Souza RB, Wyman BT, Le Graverand MP, Subburaj K, Link TM, Majumdar S. Differences between X-ray and MRI-determined knee cartilage thickness in weight-bearing and non-weight-bearing conditions. Osteoarthritis and cartilage. 2013 Dec 31;21(12):1876-85.

Mazzuca SA, Brandt KD, Katz BP, Lane KA, Buckwalter KA. Comparison of quantitative and semiquantitative indicators of joint space narrowing in subjects with knee osteoarthritis. Annals of the rheumatic diseases. 2006 Jan 1;65(1):64-8.

Mazzuca SA, Brandt KD, Lane KA, Katz BP. Knee pain reduces joint space width in conventional standing anteroposterior radiographs of osteoarthritic knees. Arthritis & Rheumatology. 2002 May 1;46(5):1223-7.

Mazzuca SA, Le Graverand MP, Vignon E, Hunter DJ, Jackson CG, Kraus VB, Link TM, Schnitzer TJ, Vaz A, Charles HC. Performance of a non-fluoroscopically assisted substitute for the Lyon schuss knee radiograph: quality and reproducibility of positioning and sensitivity to joint space narrowing in osteoarthritic knees. Osteoarthritis and cartilage. 2008 Dec 31;16(12):1555-9.

Messent EA, Ward RJ, Tonkin CJ, Buckland-Wright C. Tibial cancellous bone changes in patients with knee osteoarthritis. A short-term longitudinal study using Fractal Signature Analysis. Osteoarthritis and cartilage. 2005 Jun 30;13(6):463-70.

Michel BA, Stucki G, Frey D, De Vathaire F, Vignon E, Bruehlmann P, Uebelhart D. Chondroitins 4 and 6 sulfate in osteoarthritis of the knee: a randomized, controlled trial. Arthritis & Rheumatology. 2005 Mar 1;52(3):779-86.

Nevitt MC, Peterfy C, Guermazi A, Felson DT, Duryea J, Woodworth T, Chen H, Kwoh K, Harris TB. Longitudinal performance evaluation and validation of fixed-flexion radiography of the knee for detection of joint space loss. Arthritis & Rheumatology. 2007 May 1;56(5):1512-20.

Oak SR, Ghodadra A, Winalski CS, Miniaci A, Jones MH. Radiographic joint space width is correlated with 4-year clinical outcomes in patients with knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthritis and Cartilage. 2013 Sep 30;21(9):1185-90.

Pavelka K, Gatterova J, Altman RD. Radiographic progression of knee osteoarthritis in a Czech cohort. Clinical and experimental rheumatology. 2000 Jul 1;18(4):473-8.

Pavelka K, Gatterova J, Gollerova V, Urbanova Z, Sedlackova M, Altman RD. A 5-year randomized controlled, double-blind study of glycosaminoglycan polysulphuric acid complex (Rumalon®) as a structure modifying therapy in osteoarthritis of the hip and knee. Osteoarthritis and Cartilage. 2000 Sep 1;8(5):335-42.

Pavelká K, Gatterová J, Olejarová M, Machacek S, Giacovelli G, Rovati LC. Glucosamine sulfate use and delay of progression of knee osteoarthritis: a 3-year, randomized, placebo-controlled, double-blind study. Archives of internal medicine. 2002 Oct 14;162(18):2113-23.

Piperno M, Le Graverand MP, Conrozier T, Bochu M, Mathieu P, Vignon E. Quantificative evaluation of joint space width in femorotibal osteoarthritis: comparison of three radiographic views. Osteoarthritis and Cartilage. 1998 Jul 1;6(4):252-9.

Ravaud P, Giraudeau B, Auleley GR, Edouard-Noel R, Dougados M, Chastang C. Assessing smallest detectable change over time in continuous structural outcome measures: application to radiological change in knee osteoarthritis. Journal of clinical epidemiology. 1999 Dec 31;52(12):1225-30.

Ravaud PH, Giraudeau B, Auleley GR, Chastang C, Poiraudeau S, Ayral X, Dougados M. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. The Journal of rheumatology. 1996 Oct;23(10):1756-64.

Reginster JY, Badurski J, Bellamy N, Bensen W, Chapurlat R, Chevalier X, Christiansen C, Genant H, Navarro F, Nasonov E, Sambrook PN. Efficacy and safety of strontium ranelate in the treatment of knee osteoarthritis: results of a double-blind, randomised placebo-controlled trial. Annals of the rheumatic diseases. 2012 Nov 1:annrheumdis-2012.

Ritter SY, Collins J, Krastins B, Sarracino D, Lopez M, Losina E, Aliprantis AO. Mass spectrometry assays of plasma biomarkers to predict radiographic progression of knee osteoarthritis. Arthritis research & therapy. 2014 Oct 7;16(5):456.

Sawitzke AD, Shi H, Finco MF, Dunlop DD, Bingham CO, Harris CL, Singer NG, Bradley JD, Silver D, Jackson CG, Lane NE. The effect of glucosamine and/or chondroitin sulfate on the progression of knee osteoarthritis: a report from the glucosamine/chondroitin arthritis intervention trial. Arthritis & Rheumatology. 2008 Oct 1;58(10):3183-91.

Soto-Hermida A, Fernández-Moreno M, Oreiro N, Fernández-López C, Pértega S, Cortés-Pereira E, Rego-Pérez I, Blanco FJ. Mitochondrial DNA (mtDNA) haplogroups influence the progression of knee osteoarthritis. Data from the Osteoarthritis Initiative (OAI). PloS one. 2014 Nov 12;9(11):e112735.

Tezcan ME, Goker B, Lidtke R, Block JA. Long-term effects of lateral wedge orthotics on hip and ankle joint space widths. Gait & posture. 2017 Jan 31;51:36-40.

Tindall EA, Sharp JT, Burr A, Katz TK, Wallemark CB, Verburg K, Lefkowith JB. A 12-month, multicenter, prospective, open-label trial of radiographic analysis of disease progression in osteoarthritis of the knee or hip in patients receiving celecoxib. Clinical therapeutics. 2002 Dec 1;24(12):2051-63.

Tourville TW, Johnson RJ, Slauterbeck JR, Naud S, Beynnon BD. Assessment of early tibiofemoral joint space width changes after anterior cruciate ligament injury and reconstruction: a matched case-control study. The American Journal of sports medicine. 2013 Apr;41(4):769-78.

Uebelhart D, Malaise M, Marcolongo R, DeVathaire F, Piperno M, Mailleux E, Fioravanti A, Matoso L, Vignon E. Intermittent treatment of knee osteoarthritis with oral chondroitin sulfate: a one-year, randomized, double-blind, multicenter study versus placebo. Osteoarthritis and cartilage. 2004 Apr 30;12(4):269-76.

Vignon E, Brandt KD, Mercier C, Hochberg M, Hunter D, Mazzuca S, Powell K, Wyman B, Le Graverand MP. Alignment of the medial tibial plateau affects the rate of joint space narrowing in the osteoarthritic knee. Osteoarthritis and cartilage. 2010 Nov 30;18(11):1436-40.

Vignon E, Piperno M, Graverand L, Hellio MP, Mazzuca SA, Brandt KD, Mathieu P, Favret H, Vignon M, Merle-Vincent F, Conrozier T. Measurement of radiographic joint space width in the tibiofemoral compartment of the osteoarthritic knee: comparison of standing anteroposterior and Lyon schuss views. Arthritis & Rheumatology. 2003 Feb 1;48(2):378-84.

Vilım V, Olejarova M, Macháček S, Gatterova J, Kraus VB, Pavelka K. Serum levels of cartilage oligomeric matrix protein (COMP) correlate with radiographic progression of knee osteoarthritis. Osteoarthritis and cartilage. 2002 Sep 30;10(9):707-13.

Wilkinson CE, Carr AJ, Doherty M. Does increasing the grades of the knee osteoarthritis line drawing atlas alter its clinimetric properties?. Annals of the rheumatic diseases. 2005 Oct 1;64(10):1467-73.

Wirth W, Duryea J, Le Graverand MP, John MR, Nevitt M, Buck RJ, Eckstein F, OA Initiative Investigators Group. Direct comparison of fixed flexion, radiography and MRI in knee osteoarthritis: responsiveness data from the Osteoarthritis Initiative. Osteoarthritis and Cartilage. 2013 Jan 31;21(1):117-25.

Zuiderbaan HA, Khamaisy S, Thein R, Nawabi DH, Pearle AD. Congruence and joint space width alterations of the medial compartment following lateral unicompartmental knee arthroplasty. Bone Joint J. 2015 Jan 1;97(1):50-5.

Appendix 2. Comparison of radiographic imaging protocol

Image acquisition element	SEKOIA	OAI
Radiographic view taken	Weight bearing fixed flexion postero-anterior view	Weight bearing fixed flexion
Study participant positioning	Positioned using a positioning frame (Synaflexer)	Positioned using a positioning frame (Synaflexer)
Restriction on timing of radiographic image	No mention of any restrictions	No mention of any restrictions
Physical activity proceeding radiographic image	No mention of any restrictions	No mention of any restrictions
Reading of radiographic images	Read in pairs using baseline image	Read viewing all time points simultaneously but blinded to the correct order
Number of centres for image acquisition	98 centres across 18 countries	4 centre across America
Centralised reading	Yes	Yes
Number of readings	Read once	Read once
Types of readers	Radiology technicians	Trained readers, passing a two-stage certification

List of References

1. Murray CJ, Richards MA, Newton JN, Fenton KA, Anderson HR, Atkinson C, et al. UK health performance: findings of the Global Burden of Disease Study 2010. The lancet. 2013;381(9871):997-1020.
2. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum. 1986;29(8):1039-49.
3. Lane NE, Brandt K, Hawker G, Peeva E, Schreyer E, Tsuji W, et al. OARSI-FDA initiative: defining the disease state of osteoarthritis. Osteoarthritis Cartilage. 2011;19(5):478-82.
4. Peat G, Croft P, Hay E. Clinical assessment of the osteoarthritis patient. Best Practice & Research Clinical Rheumatology. 2001;15(4):527-44.
5. Parry E, Ogollah R, Peat G. Significant pain variability in persons with, or at high risk of, knee osteoarthritis: preliminary investigation based on secondary analysis of cohort data. BMC Musculoskelet Disord. 2017;18:80.
6. Braun HJ, Gold GE. Diagnosis of Osteoarthritis: Imaging. Bone. 2012;51(2):278-88.
7. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957;16(4):494-502.
8. Altman RD, Hochberg M, Murphy WA, Jr., Wolfe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. Osteoarthritis Cartilage. 1995;3 Suppl A:3-70.
9. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage. 2007;15 Suppl A:A1-56.
10. 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(5):995-1000.
11. Dennison E, Cooper C. Osteoarthritis: epidemiology and classification. Rheumatology New York Mosby. 2003:1981-84.
12. Thomas E, Peat G, Croft P. Defining and mapping the person with osteoarthritis for population studies and public health. Rheumatology. 2014;53(2):338-45.
13. Van der Pas S, Castell MV, Cooper C, Denkinger M, Dennison EM, Edwards MH, et al. European project on osteoarthritis: design of a six-cohort study on the personal and societal burden of osteoarthritis in an older European population. BMC Musculoskelet Disord. 2013;14.
14. Parsons C, Clynes M, Syddall H, Jagannath D, Litwic A, van der Pas S, et al. How well do radiographic, clinical and self-reported diagnoses of knee osteoarthritis agree? Findings from the Hertfordshire cohort study. SpringerPlus. 2015;4(1):1-5.
15. Felson DT, McAlindon TE, Anderson JJ, Weissman BW, Aliabadi P, Evans S, et al. Defining radiographic osteoarthritis for the whole knee. Osteoarthritis Cartilage. 1997;5(4):241-50.
16. Culvenor AG, Engen CN, Oiestad BE, Engebretsen L, Risberg MA. Defining the presence of radiographic knee osteoarthritis: a comparison between the Kellgren and Lawrence system and OARSI atlas criteria. Knee Surg Sports Traumatol Arthrosc. 2015;23(12):3532-9.
17. 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(6):1513-7.
18. Yu D, Peat G, Bedson J, Jordan KP. Annual consultation incidence of osteoarthritis estimated from population-based health care data in England. Rheumatology. 2015;54(11):2051-60.
19. Litwic A, Edwards M, Dennison E, Cooper C. Epidemiology and Burden of Osteoarthritis. Br Med Bull. 2013;105:185-99.

20. Neogi T, Zhang Y. Epidemiology of osteoarthritis. Rheum Dis Clin North Am. 2013;39(1):1-19.
21. Silman AJ, Hochberg MC. Epidemiology of the rheumatic diseases: Oxford University Press; 2001.
22. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis. 2014.
23. Curtis L, Netten A. Unit costs of health and social care: University of Kent; 2005.
24. Chen A, Gupte C, Akhtar K, Smith P, Cobb J. The global economic cost of osteoarthritis: how the UK compares. Arthritis. 2012;2012.
25. Conaghan PG, Hunter DJ, Maillefert JF, Reichmann WM, Losina E. Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group. Osteoarthritis Cartilage. 2011;19(5):606-10.
26. Ravaud P, Auleley G-R, Chastang C, Rousselin B, Paolozzi L, Amor B, et al. Knee joint space width measurement: an experimental study of the influence of radiographic procedure and joint positioning. Rheumatology. 1996;35(8):761-6.
27. Hellio Le Graverand MP, Mazzuca S, Duryea J, Brett A. Radiographic-Based Grading Methods and Radiographic Measurement of Joint Space Width in Osteoarthritis. Radiol Clin North Am. 2009;47(4):567-79.
28. Buckland-Wright JC, Macfarlane DG, Lynch JA, Jasani MK, Bradshaw CR. Joint space width measures cartilage thickness in osteoarthritis of the knee: high resolution plain film and double contrast macroradiographic investigation. Ann Rheum Dis. 1995;54(4):263-8.
29. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, et al. Radiological progression of hip osteoarthritis: definition, risk factors and correlations with clinical status. Ann Rheum Dis. 1996;55(6):356-62.
30. Lequesne M. Quantitative measurements of joint space during progression of osteoarthritis: chondrometry. Osteoarthritic disorders. 1995;30:427-44.
31. Kirwan J, Cushnaghan J, Dacre J, McAlindon T, Dieppe P, Rogers J. Progression of joint space narrowing in knee osteoarthritis. Arthritis Rheum. 1992;35(9):S134-S.
32. Deep K, Norris M, Smart C, Senior C. Radiographic measurement of joint space height in non-osteoarthritic tibiofemoral joints. A comparison of weight-bearing extension and 30 degrees flexion views. J Bone Joint Surg Br. 2003;85(7):980-2.
33. Ravaud P, Giraudeau B, Auleley G-R, Drape J-L, Rousselin B, Paolozzi L, et al. Variability in knee radiographing: implication for definition of radiological progression in medial knee osteoarthritis. Ann Rheum Dis. 1998;57(10):624-9.
34. Benichou OD, Hunter DJ, Nelson DR, Guermazi A, Eckstein F, Kwoh K, et al. One-Year Change in Radiographic Joint Space Width in Patients With Unilateral Joint Space Narrowing: Data From The Osteoarthritis Initiative. Arthritis Care Res (Hoboken). 2010;62(7):924-31.
35. Coggon D, Reading I, Croft P, McLaren M, Barrett D, Cooper C. Knee osteoarthritis and obesity. Int J Obes Relat Metab Disord. 2001;25(5).
36. 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(10):1500-5.
37. Heidari B. Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part I. Caspian Journal of Internal Medicine. 2011;2(2):205-12.
38. Srikanth VK, Fryer JL, Zhai G, Winzenberg TM, Hosmer D, Jones G. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis Cartilage. 2005;13(9):769-81.

39. Uitterlinden AG, Burger H, Huang Q, Odding E, Van Duijn CM, Hofman A, et al. Vitamin D receptor genotype is associated with radiographic osteoarthritis at the knee. J Clin Invest. 1997;100(2):259-63.
40. Wirth W, Buck R, Nevitt M, Le Graverand MPH, Benichou O, Dreher D, et al. MRI-based extended ordered values more efficiently differentiate cartilage loss in knees with and without joint space narrowing than region-specific approaches using MRI or radiography - data from the OA initiative. Osteoarthritis Cartilage. 2011;19(6):689-99.
41. Vilim V, Olejarova M, Machacek S, Gatterova J, Kraus VB, Pavelka K. Serum levels of cartilage oligomeric matrix protein (COMP) correlate with radiographic progression of knee osteoarthritis. Osteoarthritis Cartilage. 2002;10(9):707-13.
42. Hunter DJ, Beavers DP, Eckstein F, Guermazi A, Loeser RF, Nicklas BJ, et al. The Intensive Diet and Exercise for Arthritis (IDEA) trial: 18-month radiographic and MRI outcomes. Osteoarthritis Cartilage. 2015;23(7):1090-8.
43. Bowes MA, Vincent GR, Wolstenholme CB, Conaghan PG. A novel method for bone area measurement provides new insights into osteoarthritis and its progression. Ann Rheum Dis. 2015;74(3):519-25.
44. Conrozier T, Mathieu P, Piperno M, Favret H, Colson F, Vignon M, et al. Selection of knee radiographs for trials of structure-modifying drugs in patients with knee osteoarthritis: A prospective, longitudinal study of Lyon schuss knee radiographs with the definition of adequate alignment of the medial tibial plateau. Arthritis Rheum. 2005;52(5):1411-7.
45. Bruyere O, Cooper C, Pavelka K, Rabenda V, Buckinx F, Beaudart C, et al. Changes in structure and symptoms in knee osteoarthritis and prediction of future knee replacement over 8 years. Calcif Tissue Int. 2013;93(6):502-7.
46. Fukui N, Yamane S, Ishida S, Tanaka K, Masuda R, Tanaka N, et al. Relationship between radiographic changes and symptoms or physical examination findings in subjects with symptomatic medial knee osteoarthritis: A three-year prospective study. BMC Musculoskelet Disord. 2010;11 (no pagination)(269).
47. Pavelka K, Gatterova J, Altman RD. Radiographic progression of knee osteoarthritis in a Czech cohort. Clin Exp Rheumatol. 2000;18(4):473-7.
48. Eckstein F, Boudreau R, Wang Z, Hannon MJ, Duryea J, Wirth W, et al. Comparison of radiographic joint space width and magnetic resonance imaging for prediction of knee replacement: A longitudinal case-control study from the Osteoarthritis Initiative. Eur Radiol. 2016;26(6):1942-51.
49. Hellio le Graverand MP, Clemmer RS, Redifer P, Brunell RM, Hayes CW, Brandt KD, et al. A 2-year randomised, double-blind, placebo-controlled, multicentre study of oral selective iNOS inhibitor, cindunistat (SD-6010), in patients with symptomatic osteoarthritis of the knee. Ann Rheum Dis.72(2):187-95.
50. Eaton CB, Sayeed M, Ameernaz S, Roberts MB, Maynard JD, Driban JB, et al. Sex differences in the association of skin advanced glycation endproducts with knee osteoarthritis progression. Arthritis Res Ther. 2017;19(1):36.
51. Tourville TW, Johnson RJ, Slauterbeck JR, Naud S, Beynnon BD. Assessment of early tibiofemoral joint space width changes after anterior cruciate ligament injury and reconstruction: a matched case-control study. The American journal of sports medicine. 2013;41(4):769-78.
52. Lapane KL, Yang S, Driban JB, Liu SH, Dube CE, McAlindon TE, et al. Effects of prescription nonsteroidal antiinflammatory drugs on symptoms and disease progression among patients with knee osteoarthritis. Arthritis and Rheumatology. 2015;67(3):724-32.
53. Reichmann WM, Maillefert JF, Hunter DJ, Katz JN, Conaghan PG, Losina E. Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: A systematic review. Osteoarthritis Cartilage. 2011;19(5):550-6.

54.Spector T. Measuring joint space in knee osteoarthritis: position or precision? J Rheumatol. 1995;22(5):807-8.
55.Buckland-Wright JC. Quantitative radiography of osteoarthritis. Ann Rheum Dis. 1994;53(4):268-75.
56.Ravaud P, Dougados M. Radiographic assessment in osteoarthritis. The Journal of rheumatology. 1997;24(4):786-91.
57.Waterton JC, Solloway S, Foster JE, Keen MC, Gandy S, Middleton BJ, et al. Diurnal variation in the femoral articular cartilage of the knee in young adult humans. Magn Reson Med. 2000;43(1):126-32.
58.Coleman JL, Widmyer MR, Leddy HA, Utturkar GM, Spritzer CE, Moorman CT, 3rd, et al. Diurnal variations in articular cartilage thickness and strain in the human knee. J Biomech. 2013;46(3):541-7.
59.Ravaud P, Giraudeau B, Auleley GR, Edouard-Noel R, Dougados M, Chastang C. Assessing smallest detectable change over time in continuous structural outcome measures: Application to radiological change in knee osteoarthritis. J Clin Epidemiol. 1999;52(12):1225-30.
60.Gustafson P. Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments: CRC Press; 2003.
61.Carroll RJ, Ruppert D, Stefanski LA, Buonaccorsi J. Measurement error in nonlinear models. Metrika. 1997;45(3):182-3.
62.Felson D, Niu J, Sack B, Aliabadi P, McCullough C, Nevitt MC. Progression of osteoarthritis as a state of inertia. Ann Rheum Dis. 2013;72(6):924-9.
63.Eckstein F, Boudreau RM, Wang Z, Hannon MJ, Wirth W, Cotofana S, et al. Trajectory of Cartilage Loss within Four Years of Knee Replacement: A Nested Case-Control Study from the Osteoarthritis Initiative. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society. 2014;22(10):1542-9.
64.Reginster J-Y, Badurski J, Bellamy N, Bensen W, Chapurlat R, Chevalier X, et al. Efficacy and safety of strontium ranelate in the treatment of knee osteoarthritis: results of a double-blind, randomised placebo-controlled trial. Ann Rheum Dis. 2013;72(2):179-86.
65.Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken). 2011;63(S11):S240-S52.
66.Cooper C, Reginster JY, Chapurlat R, Christiansen C, Genant H, Bellamy N, et al. Efficacy and safety of oral strontium ranelate for the treatment of knee osteoarthritis: rationale and design of randomised, double-blind, placebo-controlled trial. Curr Med Res Opin. 2012;28(2):231-9.
67.Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833-40.
68.Nevitt M. The Osteoarthritis Initiative: Protocol for the Cohort Study [Available from: https://oai.epi-ucsf.org/datarelease/docs/StudyDesignProtocol.pdf.
69.2015 S. Stata Statistical Software. Texas2015.
70.Washburn RA, Smith KW, Jette AM, Janney CA. The physical activity scale for the elderly (PASE): Development and evaluation. J Clin Epidemiol.46(2):153-62.
71.Peterfy C, Li J, Zaim S, Duryea J, Lynch J, Miaux Y, et al. Comparison of fixed-flexion positioning with fluoroscopic semi-flexed positioning for quantifying radiographic joint-space width in the knee: test-retest reproducibility. Skeletal Radiol. 2003;32(3):128-32.

72.Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health and Quality of Life Outcomes. 2003;1:64-.
73.Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220-33.
74.Theiler R, Stucki G, Schütz R, Hofer H, Seifert B, Tyndall A, et al. Parametric and non-parametric measures in the assessment of knee and hip osteoarthritis: interobserver reliability and correlation with radiology. Osteoarthritis Cartilage. 1996;4(1):35-42.
75.Buckler JM. Variations in height throughout the day. Arch Dis Child. 1978;53(9):762-.
76.Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59(1):12-9.
77.Ferguson RJ, Robinson AB, Splaine M. Use of the Reliable Change Index to Evaluate Clinical Significance in SF-36 Outcomes. Qual Life Res. 2002;11(6):509-16.
78.Hinton-Bayre AD. Deriving reliable change statistics from test-retest normative data: comparison of models and mathematical expressions. Arch Clin Neuropsychol. 2010;25(3):244-56.
79.Christensen L, Mendoza J. A method of assessing change in a single subject: An alteration of the RC index. Behav Ther. 1986;17(3):305-8.
80.Maassen GH. The standard error in the Jacobson and Truax Reliable Change Index: the classical approach to the assessment of reliable change. J Int Neuropsychol Soc. 2004;10(6):888-93.
81.Kendall PC. Clinical significance. J Consult Clin Psychol. 1999;67(3):283-4.
82.Ogles BM, Lunnen KM, Bonesteel K. Clinical significance: History, application, and current practice. Clin Psychol Rev. 2001;21(3):421-46.
83.Nevitt MC, Peterfy C, Guermazi A, Felson DT, Duryea J, Woodworth T, et al. Longitudinal performance evaluation and validation of fixed-flexion radiography of the knee for detection of joint space loss. Arthritis Rheum. 2007;56(5):1512-20.
84.Vignon E, Brandt KD, Mercier C, Hochberg M, Hunter D, Mazzuca S, et al. Alignment of the medial tibial plateau affects the rate of joint space narrowing in the osteoarthritic knee. Osteoarthritis Cartilage. 2010;18(11):1436-40.
85.Reginster JY, Badurski J, Bellamy N, Bensen W, Chapurlat R, Chevalier X, et al. Efficacy and safety of strontium ranelate in the treatment of knee osteoarthritis: Results of a double-blind, randomised placebo-controlled trial. Ann Rheum Dis. 2013;72(2):179-86.
86.Locascio JJ, Atri A. An overview of longitudinal data analysis methods for neurological research. Dement Geriatr Cogn Dis Extra. 2011;1(1):330-57.
87.Xanthakis V, Sullivan LM, Vasan RS. Multilevel modeling versus cross-sectional analysis for assessing the longitudinal tracking of cardiovascular risk factors over time. Stat Med. 2013.
88.Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
89.Henderson CR. Best linear unbiased estimation and prediction under a selection model. Biometrics. 1975;31(2):423-47.
90.Twisk JWR. Longitudinal data analysis. A comparison between generalized estimating equations and random coefficient analysis. Eur J Epidemiol. 2004;19(8):769-76.
91.West BT, Welch KB, Galecki AT. Linear mixed models: a practical guide using statistical software: CRC Press; 2014.
92.Heidari B. Knee osteoarthritis diagnosis, treatment and associated factors of progression: part II. Caspian Journal of Internal Medicine. 2011;2(3):249-55.
93.Stevens JP. Applied multivariate statistics for the social sciences: Routledge; 2012.
94.Altman DG. Practical statistics for medical research: CRC press; 1990.
95.Raudenbush SW, Bryk AS. Hierarchical linear models: Applications and data analysis methods: Sage; 2002.
96.Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 1986;42(1):121-30.

97.Diggle P. Analysis of longitudinal data: Oxford University Press; 2002.
98.StataCorp L. Stata Multilevel mixed-effects Reference manual.
99.Bayes M, Price M. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. Philosophical Transactions. 1763;53:370-418.
100.Geyer C. Introduction to Markov Chain Monte Carlo. Handbook of Markov Chain Monte Carlo. 2011:3-48.
101.Gelman A. Bayesian data analysis. 2nd ed. Boca Raton, Fla.: Chapman & Hall/CRC; 2004. xxv, 668p. p.
102.Gilks WR, Richardson S, Spiegelhalter DJ. Introducing markov chain monte carlo. Markov chain Monte Carlo in practice. 1996;1:19.
103.Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and computing. 2000;10(4):325-37.
104.RStudio: Integrated Development for R. Boston, MA2015.
105.Kass RE, Wasserman L. The selection of prior distributions by formal rules. Journal of the American Statistical Association. 1996;91(435):1343-70.
106.Lunn D. The BUGS book : a practical introduction to Bayesian analysis. Boca Raton ; London: CRC Press; 2013. xvii, 381 p. p.
107.Jeffreys H. An Invariant Form for the Prior Probability in Estimation Problems. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. 1946;186(1007):453-61.
108.Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. Journal of computational and graphical statistics. 1998;7(4):434-55.
109.Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association. 1996;91(434):883-904.
110.Little RJ. Modeling the drop-out mechanism in repeated-measures studies. Journal of the American Statistical Association. 1995;90(431):1112-21.
111.Rubin DB. Inference and missing data. Biometrika. 1976:581-92.
112.Cro S. Relevant Accessible Sensitivity Analysis for Clinical Trials with Missing Data [Doctorial Thesis]: London School of Hygiene & Tropical Medicine; 2017.
113.Emrani PS, Katz JN, Kessler CL, Reichmann WM, Wright EA, McAlindon TE, et al. Joint space narrowing and Kellgren-Lawrence progression in knee osteoarthritis: an analytic literature synthesis. Osteoarthritis Cartilage. 2008;16(8):873-82.
114.Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stat Med. 2005;24(15):2401-28.
115.Shor B, Bafumi J, Keele L, Park D. A Bayesian multilevel modeling approach to time-series cross-sectional data. Political Analysis. 2007:165-81.
116.Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. Journal of Statistical software. 2005;12(3):1-16.
117.Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.
118.Grotle M, Hagen KB, Natvig B, Dahl FA, Kvien TK. Obesity and osteoarthritis in knee, hip and/or hand: an epidemiological study in the general population with 10 years follow-up. BMC Musculoskelet Disord. 2008;9:132.
119.Messier SP, Loeser RF, Miller GD, Morgan TM, Rejeski WJ, Sevick MA, et al. Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis: the Arthritis, Diet, and Activity Promotion Trial. Arthritis & Rheumatology. 2004;50(5):1501-10.
120.Garstang SV, Stitik TP. Osteoarthritis: epidemiology, risk factors, and pathophysiology. Am J Phys Med Rehabil. 2006;85(11 Suppl):S2-11; quiz S2-4.

121.Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan JL, Protheroe J, Jordan KP. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. Osteoarthritis Cartilage.23(4):507-15.
122.Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis Rheum. 1987;30(8):914-8.
123.Zhang Y, Xu L, Nevitt MC, Aliabadi P, Yu W, Qin M, et al. Comparison of the prevalence of knee osteoarthritis between the elderly Chinese population in Beijing and whites in the United States: The Beijing Osteoarthritis Study. Arthritis Rheum. 2001;44(9):2065-71.
124.Jordan JM, Helmick CG, Renner JB, Luta G, Dragomir AD, Woodard J, et al. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. J Rheumatol. 2007;34(1):172-80.
125.Muraki S, Oka H, Akune T, Mabuchi A, En-yo Y, Yoshida M, et al. Prevalence of radiographic knee osteoarthritis and its association with knee pain in the elderly of Japanese population-based cohorts: the ROAD study. Osteoarthritis Cartilage. 2009;17(9):1137-43.
126.Spector TD, MacGregor AJ. Risk factors for osteoarthritis: genetics. Osteoarthritis Cartilage. 2004;12:39-44.
127.Loughlin J. Genetic contribution to osteoarthritis development: current state of evidence. Curr Opin Rheumatol. 2015;27(3):284-8.
128.Felson DT, Zhang Y. Smoking and osteoarthritis: a review of the evidence and its implications. Osteoarthritis Cartilage. 2015;23(3):331-3.
129.Anderson JJ, Felson DT. Factors associated with osteoarthritis of the knee in the first national Health and Nutrition Examination Survey (HANES I). Evidence for an association with overweight, race, and physical demands of work. Am J Epidemiol. 1988;128(1):179-89.
130.Bartlett SJ, Ling SM, Mayo NE, Scott SC, Bingham CO, 3rd. Identifying common trajectories of joint space narrowing over two years in knee osteoarthritis. Arthritis Care Res (Hoboken). 2011;63(12):1722-8.
131.Teh YW, Jordan MI, Beal MJ, Blei DM, editors. Sharing clusters among related groups: Hierarchical Dirichlet processes. Adv Neural Inf Process Syst; 2005.
132.Gershman SJ, Blei DM. A tutorial on Bayesian nonparametric models. J Math Psychol. 2012;56(1):1-12.
133.Goldstein H. The design and analysis of longitudinal studies: their role in the measurement of change: Academic Press; 1979.
134.White E, Hunt JR, Casso D. Exposure measurement in cohort studies: the challenges of prospective data collection. Epidemiol Rev. 1998;20(1):43-56.
135.Hinton-Bayre AD, Geffen GM, Geffen LB, McFarland KA, Frijs P. Concussion in Contact Sports: Reliable Change Indices of Impairment and Recovery. J Clin Exp Neuropsychol. 1999;21(1):70-86.