Long-term Progression of Structural Joint Damage in Early Rheumatoid Arthritis

Lewis Carpenter

School of Life and Medical Sciences

University of Hertfordshire

September 2016

A thesis presented for the degree of

Doctor of Philosophy

Dedicated to my late

Grandmother Catherine Mckeen

Diagnosed with rheumatoid arthritis in her 30s,

she spent the best part of her life living with the disease

While I wasn't fortunate enough to meet her,

stories of her kind nature and gentle soul, despite the difficulties she faced,

continue to inspire me to this day

Abstract

Rheumatoid Arthritis (RA) is a chronic auto-immune disease that causes inflammation in the joints. Left uncontrolled, this prolonged inflammation can lead to pain and structural damage, resulting in erosions to the bones and total breakdown of the surrounding cartilage. Structural joint damage, measured by plain radiographs, is an important outcome measure of RA. It provides an objective marker of disease activity to assess any improvements or failures of treatments in controlling for the disease. Increased long-term joint damage has been linked with increased functional disability and decreased quality of life for RA patients. While a range of studies have looked at radiographic outcomes from observational data, they tend to be restricted to historical cohorts, with little long-term data on how radiographic progression may have changed in line with changes in clinical management. Additionally, these studies have not used the appropriate statistical methods to account for non-normal data distributions and within-patient variation over time.

As a result, the main aim of this thesis is to investigate the long-term progression of structural joint damage in patients with early RA. The specific objectives were to; (1) investigate the current evidence base to identify common methods in measuring and analysing radiographic outcomes, (2) assess what statistical methods are most appropriate in modelling long-term radiographic data, (3) use these models to understand the natural progression of radiographic damage using data from two UK inception cohorts, and finally, (4) expand these models to investigate the long-term relationship of radiographic damage with two important clinical outcomes; disease activity and functional disability. The analysis is based on longitudinal data from two UK prospective, multi-centre, early RA observational cohorts. These cohorts represent two distinct eras in the management and treatment of RA, making them invaluable for investigating how key RA outcomes have progressed in clinical practice over time.

Using multi-level count models, precise rates of radiographic progression for both cohorts are presented. The models look at how seropositive RA and increased disease activity are related to increased radiographic progression, and what impact this has on functional disability. The results show that rates of radiological damage have declined dramatically in recent years. Possible attributable factors to these declines include both milder disease and more effective treatment strategies.

Analysis of the earlier cohort (1986-2001) shows how seropositive RA and increased disease activity lead to clinically meaningful increases in radiological damage. Conversely, their impact on patients in the more recent cohort (2002-2011) suggest that their effect on radiographic progression is reduced, where increases in radiological damage were not larger than clinically meaningful thresholds. This has large implications on the debate around the use of biologic therapies in patients with less severe RA. However more data is sorely needed, particularly long-term radiographic data from those patients on biologics treatments, before any definitive conclusions can be made.

The possible impact of these declines on functional disability appears to be relatively small. The analysis shows that radiographic damage is more strongly associated with functional disability in later disease, but there is little evidence to indicate that declines in radiographic damage has lead to large improvements in long-term functional disability. These findings are explored within the framework of a dual-pathway model, which suggests that functional disability is caused by two distinct mechanisms, either structural joint damage, or through increased pain. Research so far has predominantly focused on pharmacological treatments in reducing inflammation. More research is needed to explore the role of psychosocial factors and pain perception in order to create a more holistic treatment programme for RA patients.

Acknowledgements

I am indebted to all my supervisors, without whom this thesis would not have been possible. I thank both Professor Adam Young and Dr David Wellsted for providing the funding to allow me to undertake the work carried out within this thesis. Since starting at the university, David has been instrumental in facilitating my pursuit into medical statistics, and I thank Adam for helping formulate the initial research questions and ideas. Adam has continued to show complete faith in my ability and has been a fantastic mentor from inception through to completion. Dr Kirsten Rennie and Dr Sam Norton have been absolutely invaluable in helping me shape my research, improve my understanding of scientific methods and of course in developing my statistical knowledge to undertake much of the analysis performed in this thesis. It has been a pleasure to work with all my supervisors so closely and I look forward to continuing this working relationship, as I look to progress my career in research.

Professor David L Scott has also been very supportive of my work and I thank him for providing funds so that my work can be directly involved in the TITRATE project currently underway. I am honoured to be part of such an exciting, expansive and innovative research project, involving some of the most prestigious professors in the field of rheumatology.

I of course also want to extend my thanks to the wider ERAS/ERAN team, who have been exceptionally helpful in reviewing much of my work and have provided tremendous support for me in terms of publications and presentations at conferences. Of specific note, I owe a huge amount of gratitude to both Dr Keeran Jayakumar and Daniel McWilliams, who painstakingly scored the thousands of x-rays in the cohorts to provide me with the data to conduct the analysis in this thesis. Without their tremendous patience and hard work this thesis would not have been possible. I would also like to pay particular thanks to Dr Elena Nikiphorou, who has been there for me as a colleague and a great friend throughout.

There have been a number of colleagues at the Centre for Clinical and Health Services Research that have been extremely supportive and have provided great advice throughout, and it is only appropriate that I also thank everyone involved in the ERAS and ERAN cohorts; from the rheumatologists, nurses, administrative team and allied health professionals, right through to the patients themselves. Without them, none of this work would be possible.

Finally a massive, loving thanks to my parents, brothers and sister, and last but by no means least, my partner Gabriella. Their immense support has been vital in the undertaking and completion of this thesis and am forever grateful.

Contents

List of Figures

List of Tables

Abbreviations

ACPA	**A**nti-**C**itrullinated **P**eptide **A**ntibodies
AIC	**A**kaike **I**nformation **C**riterion
AKA	**A**nti-**K**eratin **A**ntibodies
ANA	**A**nti-**N**uclear **A**ntibodies
Anti-CCP	**Anti-C**yclic **C**itrullinated **P**eptides
APC	**A**ntigenic **P**resenting **C**ells
APF	**A**nti-**P**erinuclear **F**actor
ARA	**A**merican **R**heumatism **A**ssociation
AS	**A**nkylosing **S**pondylitis
BIC	**B**ayesian **I**nformation **C**riterion
BSR	**B**ritish **S**ociety for **R**heumatology
BSR-BR	**B**ritish **S**ociety for **R**heumatology-**B**iologics **R**egister
CDAI	**C**linical **D**isease **A**ctivity **I**ndex
CI	**C**onfidence **I**nterval
CRP	**C R**eactive **P**rotein
CRF	**C**ase **R**eport **F**orm
DAI	**D**isease **A**ctivity **I**ndex
DANBIO	**DAN**ish Registry for **BIO**logic Therapies in Rheumatology
DAS	**D**isease **A**ctivity **S**core
DJS	**D**estruction of the **J**oint **S**urface
DMARD	**D**isease **M**odifying **A**nti-**R**heumatic **D**rugs
DZ	**DiZ**ygotic
EAC	**E**arly **A**rthritis **C**ohort
ERAN	**E**arly **R**heumatoid **A**rthritis **N**etwork
ERAS	**E**arly **R**heumatoid **A**rthritis **S**tudy

ESR	**E**rythrocyte **S**edimentation **R**ate
EULAR	**EU**ropean **L**eague **A**gainst **R**heumatology
FIML	**F**ull-**I**nformation **M**aximum **L**ikelihood
FLS	**F**ibroblast-**L**ike **S**ynoviocytes
GLM	**G**eneralised **L**inear **M**odels
GEE	**G**eneralised **E**stimating **E**quation
GMM	**G**rowth **M**ixture **M**odels
HAQ	**H**ealth **A**ssessment **Q**uestionnaire
Hb	**H**emoglobin
HR-pQCT	**H**igh **R**esolution-perpheral **Q**uantitative **C**omputed **T**omography
HRT	**H**ormone **R**eplacement **T**herapy
ICC	**I**ntraclass **C**orrelation **C**oefficient
IFN-γ	**I**nter**F**ero**N**-γ
IMR	**I**ncident **M**ortality **R**atio
IRR	**I**ncidence **R**ate **R**atio
IQR	**I**nter-**Q**uartile **R**ange
JSN	**J**oint **S**pace **N**arrowing
LOWESS	**LO**cally **WE**ighted **S**scatterplot **S**moothing
MAR	**M**issing **A**t **R**andom
MCID	**M**inimal **C**linically **I**mportant **D**ifference
MCP	**M**eta**C**arpo**P**halangeal
MHC	**M**ajor **H**istocompatibility **C**omplex
MICE	**M**ultiple **I**mputation **C**hained **E**quations
MMP-3	**M**atrix **M**etallo**P**roteinase **P**-3
MRI	**M**agnetic **R**esonance **I**maging
MTP	**M**eta**T**arso**P**halangeal
MZ	**M**ono**Z**ygotic
NB	**N**egative **B**inomial
NHS	**N**ational **H**ealth **S**ervice
NICE	**N**ational **I**nstitute for **C**linical **E**xcellence
NIHR	**N**ational **I**nstitute for **H**ealth **R**esearch
NSAID	**N**on-**S**teroidal **A**nti-**I**nflammatory **D**rugs
NOAR	**NO**rfolk **A**rthritis **R**egister

NRAS	**N**ational **R**heumatoid **A**rthritis **S**ociety
OR	**O**dds **R**atio
OMERACT	**O**utcome **ME**asures for **R**heumatoid **A**rthritis **C**linical **T**rials
OA	**O**steo**A**rthritis
OCP	**O**ral **C**ontraceptive **P**ill
p-ANCA	perinuclear-**A**nti**N**eutrophil **C**ytoplasmic **A**ntibodies
PGA	**P**atient **G**lobal **A**ssessment
PIP	**P**roximal **I**nter**P**halangeal
PsA	**Ps**oriatic **A**rthritis
RANKL	**r**ANK **L**igand
RA	**R**heumatoid **A**rthritis
RCT	**R**andomised **C**ontrolled **T**rials
RF	**R**heumatoid **F**actor
RR	**R**ate **R**atio
SDAI	**S**implified **D**isease **A**ctivity **I**ndex
SDD	**S**mallest **D**etectable **D**ifference
SD	**S**tandard **D**eviation
SE	**S**hared **E**pitope
SENS	**S**implified **E**rosion **N**arrowing **S**core
SES	**S**hort **E**rosion **S**cale
SF-36	**S**hort **F**orm-36
SJC	**S**wollen **J**oint **C**ount
SLE	**S**ystemic **L**upus **E**rythematosus
SMR	**S**tandardised **M**ortality **R**atio
SvdH	**S**harp/**v**an **d**er **H**eijde
TJC	**T**ender **J**oint **C**ount
TNF-α	**T**umour **N**ecrosis **F**actor-α
TVC	**T**ime **V**arying **C**ovariate
T2T	**T**reat-**To**-**T**arget

Chapter 1

Introduction

1.1 Aims and objectives of the thesis

The overall aim of this thesis is to investigate the long-term progression of radiographic joint damage in early Rheumatoid Arthritis (RA) using longitudinal data from two UK longitudinal patient cohorts. The specific objectives are to; (1) investigate the current evidence base to identify common methods in measuring and analysing radiographic outcomes, (2) assess what statistical methods are most appropriate in modelling long-term radiographic data, (3) use these models to understand the natural progression of radiographic damage using data from two UK inception cohorts, and finally, (4) expand these models to investigate the long-term relationship with two important clinical outcomes; disease activity and functional disability. The thesis is broken down into distinct chapters, which set out to explore and address these specific aims. These aims are explored in specific chapters as follows:

1.2 Outline of the thesis

Chapter 2 provides an overview of the epidemiology, aetiology, and pathophysiology of RA, as well as how the clinical management developed, and how radiographic scoring methods are used in the UK over the last three decades. The chapter will explore how Randomised Controlled Trials (RCT) and observational studies have utilised radiographic outcomes, and how the study designs differ in what research questions they are able to answer. Of particular interest to this research is how observational studies have evaluated radiographic outcomes, and what

the current prevailing theories are that involve long-term radiographic progression. To this end, a comprehensive systematic review is conducted in Chapter 3 to evaluate all published studies looking at observational data on radiographic damage in early RA. Through meta-analytic techniques and narrative synthesis, the progression, and predictive factors, of long-term radiographic damage are explored and summarised.

Chapter 4 and Chapter 5 detail the methodological aspects of this thesis. Chapter 4 introduces the Early RA Study (ERAS) and Early RA Network (ERAN) datasets, and the radiographic data contained within. Important aspects of the data, such as inter-reader reliability between the two readers and missing data are explored. Finally the chapter looks at the other key outcome measures collected as part of the ERAS and ERAN cohorts that were also examined alongside radiographic outcomes. Chapter 5 investigates the methods commonly used to analyse radiographic outcomes, and introduces count regression models. It explains why these methods are most suited to the analysis of radiographic outcomes.

Chapter 6 explores the progression of radiographic damage over the first 5 years of disease in both the ERAS and ERAN cohorts and establishes if, using the modelling techniques outlined in Chapter 5, there have been any changes in the natural progression of radiographic damage. This chapter also assesses whether the association of seropositive RA with radiographic progression has been altered as a result of any secular changes. Due to the non-randomised nature of observational data, it is difficult to explore any direct treatment cause and effects. However, key differences in both the clinical management, and indeed any differences in the patient populations at first presentation will be explored to ascertain the possible reasons for any secular differences in long-term progression.

The final aim is then addressed as part of Chapter 7. The models developed in Chapter 6 are extended to provide a more in-depth evaluation of the relationship between radiographic progression with two key clinical RA outcomes; disease activity and functional disability. Disease activity has been known to influence the extent of radiographic progression. Increased radiographic progression is commonly considered to be the main driver of increased functional disability, particularly in later disease. However the precise nature of how these relationships develop over the course of the disease is not known. Currently no study has investigated this complex relationship longitudinally, using statistical methods appropriate to the characteristics of the radiographic data. This chapter investigates the extent of radiographic progression in

those patients with 'moderate' disease. Under current UK guidelines, effective, but more expensive treatments are reserved only for those patients with very severe disease. However, there is a concern that those patients that fall just short of the threshold for these expensive treatments suffer from similar levels of disability and structural joint damage.

Chapter 8 summarises the main findings from each chapter and discusses the implications of these findings in relation to the current evidence base. This includes the impact that these results may have on the treatment and management of RA. This chapter will also summarise the strengths and limitations of using the ERAS and ERAN cohort data, as well as the methods used to conduct the analyses. Finally, it will examine potential directions for future research and what more is needed to improve the evidence base in this area of RA research.

Chapter 2

General Introduction

2.1 Rheumatoid Arthritis

2.1.1 Introduction

Rheumatoid Arthritis (RA) is a common and chronic inflammatory arthropathy[1]. It is an autoimmune disease, characterised by persistent synovitis, systemic inflammation and the presence of auto-antibodies, such as Rheumatoid Factor (RF), and Anti-Citrullinated Peptide Antibodies (ACPA), namely Anti-Cyclic Citrullinated Peptides (anti-CCP)[2]. RA is a heterogeneous disease, ranging from mild and remitting forms, to highly active and disabling. The long-term prognosis of RA is poor, with 80% of affected patients becoming disabled after 20 years, and an average reduction of between 3-18 years in life expectancy[3]. RA typically occurs in the small joints of the hands, wrists, feet and the knees and symmetry is common. Symptoms include stiffness (usually worse in the morning), tenderness, pain, swelling and deformities of the affected joints. Over time, if the disease is not adequately controlled, RA can have a large impact on a patient's quality of life. Routine daily tasks, such as the ability to button shirts, tie shoelaces or washing hair can become severely impaired. It has a profound impact on the patients ability to work, and a report by the National Rheumatoid Arthritis Society (NRAS) estimated that RA as a disorder results in a loss of nearly £8 billion in the UK due to work disability[4].

Since no medical tests are pathognomonic of RA, a range of classification criteria were developed to assist with the diagnosis of RA in the medical community[5]. This was first brought about in

Criterion	Definition
1) Morning Stiffness	Morning stiffness in and around the joints, lasting at least 1 hour before maximal improvement
2) Arthritis of 3 or more joint areas	At least 3 joint areas simultaneously have had soft tissue swelling or fluid (not bony overgrowth alone) observed by a physician. The 14 possible areas are right or left PIP, MCP, wrist, elbow, knee, ankle, and MTP joints
3) Arthritis of hand joints	At least 1 area swollen (as defined above) in a wrist, MCP, or PIP joint
4) Symmetric arthritis	Simultaneous involvement of the same joint areas (as defined in 2) on both sides of the body (bilateral involvement of PIPs, MCPs, or MTPs is acceptable without absolute symmetry)
5) Rheumatoid nodules	Subcutaneous nodules, over bony prominences, or extensor surfaces, or in juxtaarticular regions, observed by a physician
6) Serum rheumatoid	Demonstration of abnormal amounts of serum rheumatoid factor by any method for which the result has been positive in <5% of normal control subjects
7) Radiographic changes	Radiographic changes typical of rheumatoid arthritis on posteroanterior hand and wrist radiographs, which must include erosions or unequivocal bony decalcification localized in or most marked adjacent to the involved joints (osteoarthritis changes alone do not qualify)

TABLE 2.1: 1987 American College of Rheumatology Criteria for rheumatoid arthritis classification

1958, when the American Rheumatism Association (ARA) developed a list of 11 criteria in an attempt to define RA[6]. This was later revised in 1987 to improve the specificity and simplicity of the classification criteria[7], with a major change being to remove the hierarchy of certainty around diagnosis, which ranged from 'possible' to 'classic', as this was largely unhelpful[5]. Table 2.1 outlines the 7 major criteria for classifying RA from the 1987 ARA criteria, ranging from the type of stiffness experienced, location of the affected joints, appearance of Rheumatoid nodules, high levels of RF in the serum, and radiographic changes. While the latest set of criteria from the American College of Rheumatology (ACR; renamed from the aforementioned ARA) has been successful at distinguishing RA from other generalised rheumatic diseases, such as osteoarthritis (OA), Systemic lupus erythematosus (SLE) and fibromyalgia[5], there are still debates about its effectiveness in classifying patients with recent-onset polyarthritis[8]. Nevertheless, current ACR classification criteria defines a patient's presenting with at least 1 joint with definite synovitis, which cannot be explained by another disease, with RA if they score 6 out of 10 based on

the number of joints involved, serology, acute phase reactants and duration of symptoms. As research improves our understanding of how genetics and laboratory markers are involved in the aetiology of the disease, absolute definitions of RA that can be used in clinical practice will become more attainable[5].

2.1.2 Epidemiology

Rheumatoid Arthritis is the most common form of chronic inflammatory arthritis[5], with estimated prevalence rates in the Northern European countries and Northern America estimated at between 0.5-1.1%, with an estimated annual incidence rate of between 20 and 50 cases per 100,000 population[9]. There does appear to be geographical differences in the prevalence of RA, with Southern European and developing countries reporting relatively lower rates of between 0.1-0.7%[10–17]. Interestingly, RA is thought to be a 'modern disease', with no records of symptoms or any definite description of RA before 1800 [5]. This has led researchers to hypothesise that urbanisation or industrialisation is likely to be a cause of RA[18], although the exact causes are still unknown.

When studying secular trends in RA incidence over time, studies have shown evidence of a decrease in incidence rates[5, 19, 20], particularly in women[5]. However, improvements in healthcare and living standards continue to increase average life expectancy, resulting in an increase in the number of elderly patients living with RA. So while the incidence of RA may be declining, the prevalence of the disease is increasing[21], and likely to increase further in the coming years, particularly at the higher (55+ years) age groups[22].

Patients with RA have higher premature mortality compared to the general population[23, 24] primarily from cardiovascular disease, viral infections and cancer[25]. A recent meta-analysis conducted by Dadoun et al. in 2013[25] assessed whether the excess risk in mortality amongst RA patients had changed over time. The review found that the Incident Mortality Rate (IMR) decreased from 4.7/100 person-years before 1970 to 2.0/100 person-years after 1983. In contrast, the Standardised Mortality Ratio (SMR) did not change over time, with an estimated pooled SMR of 2.01 from eight studies. They concluded that excess mortality had decreased amongst RA patients over the last 5 decades, but at a slower rate compared to the general population. This finding was substantiated further in the Norfolk Arthritis Register (NOAR), where the SMR for all-cause mortality over the first 7 years in RA patients had not changed from 1990-2004[24]. However, the advent of newer more effective treatments, such as methotrexate, was

shown to reduce mortality[26]. This suggests that better treatments were having an effect on reducing mortality rates.

2.1.3 Aetiology

While the exact causes of RA is unknown, it is widely accepted that it is a combination of both environmental and genetic factors[9, 27, 28]. However, to date no complete hypothesis that incorporates both these elements has been formulated[29]. The HLA-DR shared epitope (SE) has been consistently linked with the susceptibility of RA[30], with more than 80% of RA patients carrying the HLA-DRB1*04 epitope (often referred to as the 'Rheumatoid Epitope') [31]. It is thought to be a significant factor in both the onset and the severity of the disease over the long-term[5]. Quantifying the extent to which genetic factors are involved in the onset of RA has proved challenging for a number of reasons; including strong gene-environment interactions, accounting for secular changes in the disease, and the specificity of identified gene alleles.

Early studies indicated a low concordance rate of around 15% between Monozygotic (MZ) twins[17], leading to the belief that environmental factors play a significant role[30]. However, MacGregor et al. argued that there was a need to account for changes in population prevalence when quantifying the heritability of RA, and subsequently estimated that RA had a higher heritability rate of approximately 60%[32], a finding validated through the use of two large twin cohorts from the UK and Finland. The expression of the HLA-DR SE alleles in other non-RA diseases has also questioned its specificity to RA, along with the absence of the HLA-DR SE allele in approximately 20% of RA patients[30]. Consequently, it has been suggested that the stronger link between HLA-DR SE in established RA indicates that its presence in patients with RA is indicative of more severe and aggressive forms of the disease, rather than a precursor of disease onset[5].

A range of environmental factors have been hypothesised to be involved with the onset of RA[5, 33], and smoking is thought to be a fundamental environmental risk for the development of RA[28]. In 1996, Silman et al.[34] conducted a study in twins to investigate the effect that smoking has on the susceptibility of RA, where an interview questionnaire was sent out to 79 identical Monozygotic (MZ) and 71 same-sex Dizygotic twins (DZ), both with and without RA. The number of discordant pairs, that is one twin with RA that smoked and one twin without RA that didn't smoke, was low due to both twins being exposed to similar environmental factors. Nevertheless, the analysis indicated a strong association between RA and smoking (Odds Ratio

(OR) of 12 (95% Confidence Intervals (CI) 1.78-513) and 2.5 (95% CI 0.92-7.87) for MZ and DZ twins respectively, and a OR of 3.9 (95% CI 1.64-10.5) for both MZ and DZ combined).

Subsequently, it has become clear that smoking is strongly associated with a specific sub-type of RA referred to as seropositive RA[28, 35]. The presence of Rheumatoid Factor (RF) or Anti-citrullinated Peptide antibodies (ACPA) have been shown to be a significant risk factor in the development of RA[36, 37], along with being a significant predictor of increased erosive damage[38]. The presence of ACPAs, like anti-citrullinated protein antibodies (anti-CCP), have been found to be superior in predicting the risk of developing RA compared to RF antibodies[36, 39], with research showing that production of ACPA in the synovium can occur potentially years before symptom onset, this provides strong evidence for its role as a biomarker in RA[40]. Interestingly, research has indicated that the presence of the HLA-DRB SE is only associated with ACPA positive RA, leading to speculation that the production of ACPA is mediated by the presence of the HLA-DR SE[28, 41, 42]. While direct evidence for this causal relationship is lacking, it does highlight the need to include both genetic and ACPA information when looking at these subsets of RA patients[28, 41]. Further still, the relationship between smoking and ACPA positive RA suggests a interaction between the involvement of HLA-DR SE, smoking and ACPA positive RA[29]. It is not clear how these specific genetic and environmental factors lead to the pathophysiology seen in ACPA positive RA, but recent evidence suggests those who have the HLA-DR SE and who also smoke have a 21-fold increase in relative risk (RR) compared to those without the gene who do not smoke, significantly higher than either factor in isolation[29].

Other environmental factors, such as infectious agents that trigger the auto-immune response, have also been studied as the possible cause of RA[43]. However, no single organism has been found in the synovial fluid or tissue to date, and there is no evidence of incidence clusters of RA coinciding with spikes in either recorded bacterial infections or viruses in cohort data[5, 43]. Although, case studies have been documented that show RA can be caused by infections such as the parvovirus and rubella[43]. These are limited to single cases only and therefore are lacking in their generalisability.

The prevalence of RA is much higher in post-menopausal woman when compared to pre-menopausal women[44], leading to speculation that hormonal factors could play a role in the onset of RA. The incidence of RA in women who take an Oral Contraceptive Pill (OCP) was shown to be around half that compared to women who had never taken one[45], along with

reduced susceptibility during pregnancy[46]. What is not clear is whether there is a direct causal link between OCP use and onset of RA, or whether OCP use is a surrogate marker of other lifestyle choices in women who choose to take the OCP[43], e.g. smoking. Furthermore, research has not shown any association between women during their menopause and their use of Hormone Replacement Therapy (HRT) at the onset of RA[43], which would be expected to decrease the risk.

2.1.4 Pathophysiology

Figure 2.1 shows a cross-section of a synovial joint. Part (A) indicates a normal joint unaffected by RA, whilst part (B) shows the physiological changes of a joint affected by RA. Around the joint is the synovium (or synovial membrane) that encapsulates the articular cartilage between the two bones. Its two main purposes are to provide structure as well as nutrients to the joint. The inner layer (intima) of the synovium is lined with fibroblast-like synoviocytes (FLS); cells that produce unique proteins crucial for maintaining joint lubrication. This lubrication comes in the form of synovial fluid, which allows the joint to move freely.

The immune response is the body's natural defence against pathogens that enter the body. It triggers an inflammatory response with the sole purpose of isolating and removing the pathogen from the body. Initially, phagocytes, such as macrophages, will begin to consume the pathogens in an attempt to contain it. These phagocytes send messenger cells called cytokines (e.g. Interleukin and Tumour Necrosis Factor-α(TNF-α)) to warn the other immune cells of the invading cells. T-Cells and B-Cells are major types of lymphocytes involved in the immuno-inflammatory response. B-cells provide 'humoral immunity', that is the B-cells react to specific antigens by producing antibodies that bind to them. Conversely, T-cells provide 'cell mediated immunity', which in turn provides several functions to the immune response. This includes the production of helper T-cells to mediate B-cells and macrophages, suppressor T-cells to reduce the immune response and cytotoxic T-cells, which signal infected cells to perform apoptosis (self-destruct).

While this inflammation response is vital in maintaining good health, in auto-immune diseases such as RA, the prolonged inflammation at the site of the joints leads to the destruction of healthy cells. The cause of this activation and reason that the response is localised to the joint regions is currently unknown[47]. However, T-Cells, B-Cells and pro-inflammatory cytokines are known to play a key role in the pathophysiology of RA[27]. The first stages of RA involve

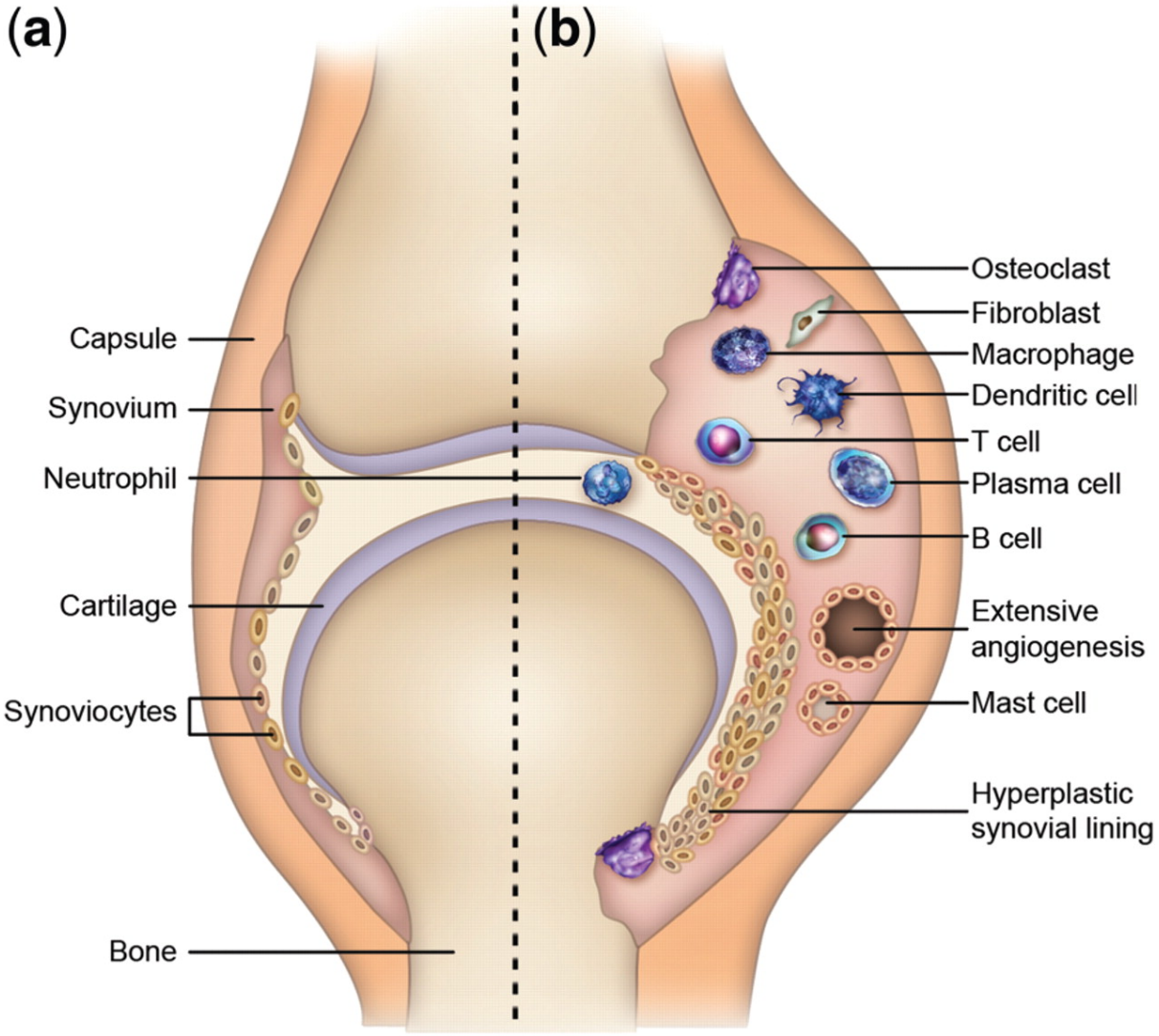


FIGURE 2.1: Cross-section of a (a) a normal joint and (b) a joint affected by Rheumatoid Arthritis

the activation of the immune response through Antigenic Presenting Cells (APCs) in the synovium[27]. APCs, including dendritic cells, macrophages and activated B cells, process unknown antigens into peptides that are inserted into the groove of HLA-DR4, which is located on the surface of the APCs. Attached to these APCs are class II major histocompatibility complex (MHC) proteins, which are unique to specific B-cells. T cells with the specific T-cell receptors are then activated by engaging with this trimolecular complex. Predominately it is CD4+ memory cells (helper T-Cells) that infiltrate the synovial membrane[47], leading to the release of the cytokine Interleukin-2 (IL-2), which in turn leads to the expansion of T-cells and expression of surface molecules like CD69, TNF-α and rANK ligand (RANKL)[27]. Production of soluble mediators IL-17 and interferon-γ (IFN-γ), stimulates macrophages to produce large numbers of pro-inflammatory cytokines IL-1, IL-6 and TNF-α[48]. The expression of these molecules

and pro-inflammatory cytokines in the synovial fluid results in the formation of osteoclasts; the primary mediators of bone destruction, fibroblasts and synoviocytes.

While the role of T cells in the pathogenesis of RA is well established, the precise role of B Cells is not as well characterised [48]. It is thought that they could have a number of potentially critical roles; including presenting as APCs to activate helper T-cells, but also in the production of RF. B-cells with RF specificity may migrate to the synovium, creating a 'self-perpetuating' cycle of increased B-cell activation and amplification of RF production, thereby prolonging and exacerbating the inflammation at the joint site[48]. The role of ACPA in the inflammation response is also of great interest, as it has been shown to be more specific to RA than RF antibodies[41]. There is growing evidence that ACPA positive RA may represent a distinct sub-class of RA[41]. Gaining a better understanding of the pathophysiology of RA has lead to the development of more effective medications that are able to target specific cells in the immuno-inflammatory response in order to modify the disease, such as anti-TNF biologics.

2.1.5 Treatment

Pharmacological treatment of RA can be broadly split into two principals; control of the underlying symptoms caused by RA (e.g. pain) through the use of non-steroidal anti-inflammatory drugs (NSAIDs), or modification of the inflammation process through the use of Disease Modifying Anti-rheumatic drugs (DMARDs)[47]. The treatment of RA has seen paradigmatic changes over the last 15-years[49]. In the 1980's, RA was treated in a pyramidal approach, where symptoms were managed, and drugs were increased in both dose and number as the disease progressed[50]. Given the high toxicity of steroid treatments and early DMARDs, such as Gold and ciclosporin, the aim of this treatment method was to reduce the burden of the disease while limiting the side-effects of these relatively potent medications[2]. With the emergence of new generation DMARDs in the late 1990s, particularly sulphasalsine and methotrexate, data from RCTs proved that early and more intensive treatment led to significantly better clinical and radiological outcomes[51–53]. This then led to the realisation that RA had a 'Window of Opportunity' in the early stages, when treatment was most effective [54]. From the late 2000s, the use of Biologic DMARDs, which are highly specific in targeting specific pro-inflammatory cytokines, has proved to be more effective than single DMARD therapy, showing increased efficacy when used in conjunction with methotrexate[55, 56]. However, given the relatively high

cost per patient of the biologic DMARD therapies, their use in the UK is currently restricted to only those patients with severe disease.

Treat-2-Target

Currently, RA is treated using a Treat-To-Target (T2T) approach. In 2010, findings from an international task force were published that provided recommendations for achieving optimal therapeutic outcomes in RA[49, 57]. Drawing parallels with other chronic illnesses, such as hypertension and diabetes, it was highlighted how the use of treatment targets using predefined biomarkers/markers of disease activity could greatly reduce the risk of organ damage. With regards to the recommendations set out, the committee concluded that remission, defined as the 'absence of signs and symptoms of significant inflammatory disease activity', should be the primary goal of all RA patients[49, 57]. However, it should be noted that this recommendation had low levels of support due to the common use of low disease activity as the primary target in clinical trials, rather than remission. As such, it was acknowledged that while remission should remain the ultimate goal, there might be instances where low disease activity is a viable alternative. Recent advances in the use of DMARDs such as methotrexate, particularly in combination with other DMARD therapies, has resulted in remission, or at the very least low disease activity becoming a realistic goal in the majority of RA cases[58]. This therefore became the target that all clinicians aimed to achieve with their patients. However, defining remission or indeed low disease activity, has proven to be no easy feat. In cases where the disease is particularly severe, and where the patient is not responding to conventional DMARD therapies, the clinician may prescribe biologic DMARDs. These are potentially more effective, but ultimately more expensive DMARDs, which are used to help attain remission or low disease activity in those patients with persistent high RA.

Clinical definitions of remission

Clinical definitions of remission in RA began with the ACR definition in 1981[59]. While clinically useful, its use in the setting of clinical trials was restricted by its difficulty to implement quickly and reliably[60]. The formulation of core sets of outcomes set out by the World Health Organisation in 1994[61] and the ACR in 1993[62] proved instrumental in quantifying disease states in a easy, quick and standardised manner. Based on these core outcomes, the ACR developed a standardised tool of assessing response to therapies through the use of the ACR20[63].

This defined improvement as a 20% reduction in the core outcome measures set out previously (swollen joint count, tender joint count, patient and physician global assessments, pain, disability, and an acute-phase reactant), which could be used across clinical trials as a standardised primary outcome measure.

In conjunction with the ACR tool, the Disease Activity Score (DAS) was also developed as a means of quantifying the severity of the disease using a similar set of outcomes as those described by the ACR. Much like the ACR definition, the aim of the DAS was to enable effective therapeutic decisions in clinical practice, as well as help standardise results across clinical trials[64]. Van der Heijde and 5 other rheumatologists used 6 year data collected on 113 early RA patients to develop the first DAS index. Factor analysis was conducted on a variety of laboratory and clinical markers to establish which factors could be used to devise an index that was easy to implement in the clinical setting, while minimising any loss of critical information. Five factors were established for the score and consisted of the Ritchie index (a measure of Tender Joint Counts (TJC)), a 44-count for Swollen Joint Count (SJC), Erythrocyte Sedimentation Rate (ESR; a marker of acute phase inflammation) and general health. These separate measures, which could be obtained relatively easily in routine clinical care, were then calculated into a single DAS index, using a pre-defined formula, to ascertain overall disease levels[64].

The DAS was later modified by Prevoo et al.[65] to the DAS28, which included a 28-joint count for both swollen and tender joints, rather than the Ritchie Index for tender joints and the 44-joint count for swollen joints, both of which took longer to implement[66]. Furthermore, the formula to calculate the overall score was adjusted, where the level of ESR had a larger impact on the overall score for the DAS28 compared to the original DAS[66]. Because of these subtle, but key differences, it is important that these two measures of disease activity are not used interchangeably, as direct comparisons between patient groups assessed with the two scores will invariably lead to bias[66]. That being said, formulas have been devised to transform the DAS-44 to the DAS28[67]. While the DAS28 has proved to be popular as a primary outcome, modifications have been attempted to simplify the score, such as the development of the Simplified Disease Activity Index (SDAI)[68]. This score has been shown to be more sensitive to changes in disease activity, and a validated means of quantifying overall disease activity[69].

Although strictly considered as a continuous measure of disease activity, ranging from 0-9.4,

categorisation of the DAS28 has been used widely as a means of stratifying patients into remission, low, moderate and high disease activity groups[67, 70]. When compared to the ACR remission criteria, a DAS28 score of <2.6 has been validated as a suitable cut-point in defining remission[70]. The definition of high disease activity as a DAS28 score of >5.1 is also the basis of the National Institute for Clinical Excellence (NICE) guidelines on the use of biologic DMARDs in the UK[71]. Currently, NICE restricts the prescription of biologic DMARDs to those patients that have a DAS28 score of >5.1 on two separate occasions, who have also failed to respond to two conventional synthetic DMARDS, one of which must be methotrexate. Recent data is questioning whether this threshold is set too high, and whether patients with moderate DAS28 (i.e. just below the threshold for biologics) are adequately controlled on conventional DMARDs alone[72–74].

2.2 Radiographic damage

RA is characterised by persistent synovitis, and over time this leads to structural damage to both the bone and surrounding cartilage of the affected joint. Erosive damage has been shown to occur in early RA, and while it is a heterogeneous disease, the majority of patients show signs of erosive arthritis after 3 years[75–77]. Radiological damage is crucial in understanding the severity of the disease, and has been used extensively as a primary end-point in RCTs[53, 55, 56, 78–83]. A recent report by a European League Against Rheumatology (EULAR) task force highlighted the importance of imaging techniques in the diagnosis, clinical management and detection of joint inflammation in RA[84]. It is significantly associated with increased functional disability[38, 85–87] and shown to significantly increase the risk of orthopaedic surgery in later disease[88]. By taking radiographs of a patient's hands and feet, it can provide the clinician with an objective and accurate snap-shot of how the disease is affecting the joints at any one time. Collected over time, these can be used to document the ability of treatments in reducing the erosive damage over the course of the disease[89].

It is theorised that radiological damage encompasses two main components of structural joint damage; that is erosive damage of the boney structures and narrowing of the joint space[90]. While a combination of both these components as a total score is typically the focus in RCTs and observational studies alike[91, 92], there is evidence that both components represent related, but ultimately distinct biological mechanisms in the disease process. Erosive joint destruction is the product of invading synovial osteoclasts[93] and joint space narrowing (JSN) is largely

the involvement of cartilage damage due to metalloproteinases, which are mediated by pro-inflammatory cytokines[94]. Understanding the precise differences in disease mechanisms could have important implications on therapeutic management, since additional therapies could be targeted specifically at the inhibition of osteoclasts to reduce further erosions in patients with established disease[93].

2.2.1 Measuring Radiographic Damage

Various methods have been designed and validated as a means of quantifying the extent of radiographic joint damage[95]. The aim of the scoring methods is to evaluate the small joints of the hands, wrists and feet (with the exception of the early Sharp scoring method[96], which only scored the small joints of the hands and wrists) and rate the joint with respect to the severity of the erosions and JSN. While some scoring methods provide just one score for each joint that encompasses the severity of both the erosions and JSN, other scoring methods provide two scores for each joint, indicating the severity of erosions and JSN individually. More often than not, the separate erosion and JSN score are combined to provide an overall total score. Figure 2.2 shows a plain x-ray of two fingers from an RA patient. RA can be seen to be affecting the third and fourth Proximal InterPhalangeal (PIP) joints. The red arrows in the figure highlight erosive damage to the bones, while the blue arrow indicates reduction in JSN. The white arrow shows evidence of soft tissue swelling.

Sharp scoring method

The Sharp scoring method was first published in 1971[96] and scored 29 areas in the hands and wrists for erosions and 27 areas for the JSN. Each area was given a score ranging from 0 to 5 for the severity of the erosions, and 0 to 4 for the severity of the JSN. 0 indicated a normal joint, whereas 4 indicated severe erosions and complete reduction of JSN (ankylosis). The scores could therefore range from 0 to 290 for the erosion score, and 0 to 216 for the JSN score. This was later modified in 1985[97], which reduced the number of assessed areas for the erosion score from 29 to 17 and the number of areas assessed for JSN from 27 to 18. This resulted in a reduced total score ranging from 0 to 314, with a separate maximum score of 170 and 144 for the erosion and JSN score respectively.

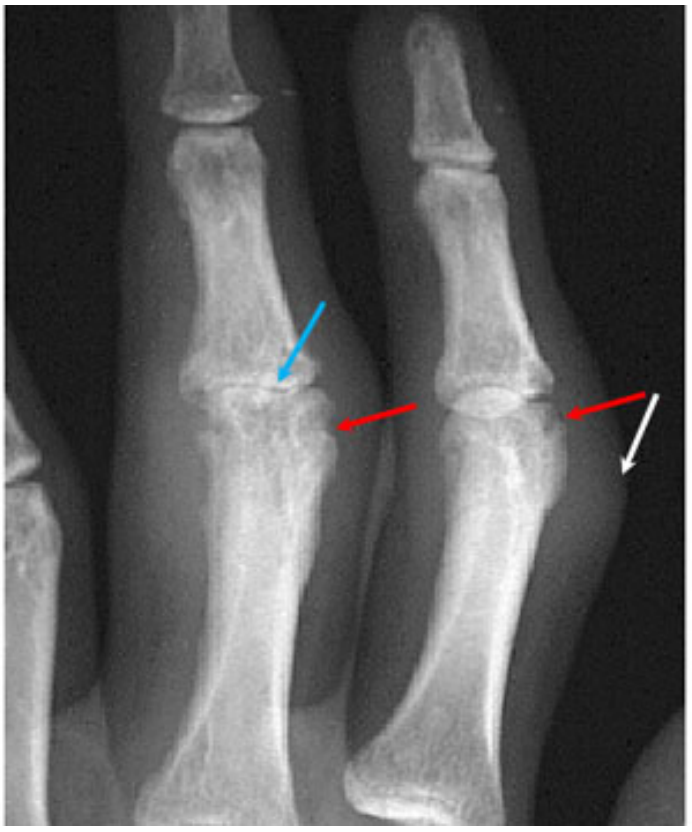


FIGURE 2.2: X-ray of two Proximal InterPhalangeal (PIP) joints affected by Rheumatoid Arthritis. The white arrow indicates soft tissue swelling, the red arrows indicate erosive damage and the blue arrow indicates joint space narrowing.

There was a further modification of the Sharp score by Fries et al.[98], which aimed to incorporate both the size and count of each joint, as well as an indication of the 'global' severity of the affected joint. It accomplished this using a weighted score, however the additional time needed to conduct this score was not outweighed by any significant improvements in sensitivity or reliability[95], and therefore is not widely adopted.

The original Sharp method, including its early modifications, was the only scoring method to not include the joints of the feet; namely the metatarsophalangeal and interphalangeal joints. The importance of assessing the radiological damage in the small joints of the feet was highlighted by van der Heijde in 1992[99], which demonstrated how radiographic damage was more common in the feet than in the hands during the first 3 years in patients with early RA. This was further substantiated by Plant et al.[100] in 1994, who demonstrated that radiographic scores in the feet where significantly correlated with later progression[100]. As such, it is perhaps no surprise that the final modification of the Sharp score that does include the joints of the feet is the most widely used scoring method, being particularly popular in RCTs over recent decades. Developed in partnership with Désirée van der Heijde[101], and commonly referred to as the Sharp/van der

Heijde (SvdH) scoring method, this modification of the Sharp score rates radiographic damage based on the severity of the erosions in 32 joints in the hands and 12 joints in the feet, and the severity of JSN in 30 joints in the hands and 12 joints in the feet. Each joint is rated from 0-5 for both erosions and JSN (however a score of 0-10 for erosions in the joints of the feet was used) giving a maximum score of 280 for the erosion score and 168 for the JSN score. These scores are combined to give a total SvdH score ranging from 0 to 448 (See Appendix D).

In 1999 attempts were made by van der Heijde[102] to simplify the SvdH method by condensing the score into a Simplified Erosion Narrowing Score, or SENS. Rather than incorporating a grading of the severity of the joint, each joint is merely scored 1 if it has presence of erosions, 1 if it has presence of JSN or 0 if neither are present. Each joint can therefore have a score ranging from 0 to 2. The same joints that are assessed using the SvdH method are also assessed using the SENS, therefore a total score can range from 0 to 86. Sample data using patients from a cohort with established RA has indicated a high level of agreement between the SENS method and the SvdH method (84%, k=0.565)[103]. It should be noted however, that the small sample of 25 patients and the omission of Bland and Altman plots to investigate how agreement varies over the range of the scores (with higher variation likely towards the higher end of the scale) makes full interpretation of these results difficult. The criticisms outlined were discussed in a study by Klarenbeek et al.[104] that indicated, through the use of cumulative probability plots, how the sensitivity of the SENS was reduced at the higher end of the scale. As such, they recommended that the advantages of the SENS with respect to speed and ease of use are not outweighed by the reduction in sensitivity, and therefore the SvdH method remains the scoring method of choice in RCT and cohort studies[104].

Larsen scoring method

The Larsen scoring method was first developed by Avri Larsen in 1974[105]. Like the Sharp scoring methods, it measures the severity of both the erosions and JSN but includes both components in one score instead. A total of 30 joints of the hands and 12 joints of the feet are assessed using the Larsen method. The wrist is considered 1 unit and the score is multiplied by 5. Each joint is scored from 0 to 5, where 0 indicates a normal joint, 1 indicates slight abnormalities, 2 indicates definite early abnormalities, 3 indicates medium destructive abnormalities, 4 indicates severe definite abnormalities and 5 indicates mutilating abnormalities. As with the Sharp score,

there have been various modifications of the Larsen score, however the method most often cited is the 1977 version[106], which has a total score ranging from 0-250 (See Appendix D).

In 1995, a significant modification was made to the Larsen score in an attempt to make it more suitable for use in long-term studies[107]. The wrist was divided into 4 sections, rather than being treated as 1 area, and the thumb and first metatarsophalangeal (MTP) joint were omitted. The severity of erosions was graded based on size of the erosion on the joint, rather than subjective classifications of 'severe'. As a result this Larsen score ranged from 0 to 160. Variations on this theme were also developed further by Scott et al[108] and Rau and Herborn[109]. While the modification detailed by Scott et al. restored the original 0 to 250 range of the Larsen score, it changed the focus of the grading to more clinical definitions of erosive and JSN damage. Likewise, Rau et al. looked at employing a concept of 'destruction of the joint surface' (DJS), which rated the proportion of the joint surface that was affected by erosive damage. This ranged from <25%, 26-50%, 51-75% and >75% for grade 2, 3, 4 and 5 respectively. The Rau and Herborn method retained the same 0 to 160 score range as the same joints were assessed.

Other scoring methods

While the Larsen and more recently the Sharp (particularly the SvdH) scoring methods are typically the most commonly used radiographic outcomes in observational studies and RCTs, there are other scoring methods that attempt to quantify the extent of radiographic damage in RA patients. The Genant method was developed in 1983 by Genant et al.[110] and considered erosive damage at 16 sites in the hand and six in the feet, and JSN in 11 sites in the hand and six at the feet. Each site is graded from 0 to 4, however this method requires a standard reference set of radiographs for comparison. This was modified in 1998 by Genant et al.[111] to include 0.5 increments for the 0 to 4 gradings, and then again by Kaye et al.[112] that details two ways of implementing the score; a similar but more detailed version of the original Genant scoring method, and a simplified method. Both methods grade 21 joints in the hands and wrists, however the more detailed approach grades joints from 0 to 4 for erosions and zero to five for JSN. The total score for the detailed version ranges from 0 to 168 and 0 to 210 for erosions and JSN respectively. The simplified method combines both erosions and JSN into one score from between 0 to 4, but also includes a grade P for post-operative joints, and a grade X for joints that cannot be evaluated. The simplified version ranges from 0 to 168, which is calculated based

on the sum of the scores divided by the number of evaluated joints (i.e. not those with grade X).

Other methods include the Ratingen score, a method derived from the Larsen scoring method by Rau et al.[113]. This method can be seen as a natural extension from the Rau and Herborn modified Larsen score that focuses on the proportion of the DJS on twenty joints in the hand, four sites in the wrist and ten joints in the feet. The method restricts scoring to only those joints with definite changes of erosion and joint destruction and grades the joints as 1 for one or more joints with <20% DJS, 2 for a DJS of 21-40%, 3 for a DJS of 41-60%, four for a DJS of 61-80% and finally five for a DJS of >80%. The score ranges from 0 to 190.

Finally, the Short Erosion Scale (SES) was proposed by Wolfe et al.[114] as a means of determining the minimum number of joints needed to gain a suitable estimate of the 'global' radiographic damage of a RA patient. To do this, the authors used Rasch analysis on the Larsen score to identify the minimum number of joints without compromising on specificity or accuracy of the score in quantifying radiological damage. The analysis concluded that only twelve joints, three of the four wrist regions and metacarpophalangeal joints (MCP) 2, 3 and 5 of the hand. Each joint is graded from 0 to 5 using the criteria set out by Larsen in their 1995 modification of the Larsen score[107]. However, there is little evidence of its use in the literature, making any assessments about its reliability difficult.

Use of other imaging techniques

So far, the scoring methods outlined have been restricted to images displayed using plain radiographs from X-ray. The EULAR task force[84] highlights how the use of other imaging techniques, such as magnetic resonance imaging (MRI) and ultrasound, can also prove useful in the diagnosis of RA, as well as predicting long-term outcomes. MRI, and to a lesser extent ultrasound, were highlighted as particularly useful in measuring clinical features that cannot be detected with Computerised Radiography (CR) alone, such as bone marrow oedema and synovitis, both of which have been found to be associated with increased erosions in later disease [115–117]. However, its widespread use in clinical practice is limited due to its relatively high cost compared to plain radiographs.

2.2.2 Common issues with radiography

The use of radiography to measure disease is popular in both observational and clinical trials as it provides an objective means of quantifying the extent of structural damage caused by RA[118, 119]. However, there are a range of issues that need to be considered to ensure that it remains a 'gold standard' measure of disease[118]. Firstly, there are technical considerations with both the quality of the x-ray film, and positioning of the joint to ensure that the radiographic score can be appropriately applied. Variation in aspects such as joint rotation and film exposure has been shown to have a marked effect on the interpretation of the erosive and JSN damage of joints[95]. The posteroanterior view for both hand and foot radiographs has been shown to be the most superior when compared to other angles[120], while under- and over-penetrated films can have a marked effect on the loss of erosions[95].

For longitudinal studies assessing multiple radiographs for one patient over time, the sequence in which the radiographs are read has been shown to have a marked impact on the measurement[121]. Studies have shown that when the radiographs are read in chronological order, they are more sensitive to changes over time, particularly in studies with a long follow-up[121]. While this approach reduces the potential within-subject random error, it does introduce bias in that the reader may be expecting the erosive damage to progress over time[95], thereby artificially inflating the rate of progression. Alternatively, blinding the reader to the sequence of the radiographs (referred to as 'paired' reading), or to both the patient and the sequence (referred to as 'single' reading), has been shown to reduce bias, however the level of measurement error is likely to increase[102]. It is not clear which method is more 'desirable'[119], but in the context of longitudinal analyses it is clear that chronological order in order to minimise within-reader variability is of more importance[121].

The presence of measurement error can greatly increase the 'noise' during the analysis of radiographic data, which in turn leads to reduced precision and biased estimations. When looking at differences in radiographic damage between groups, or change in scores over time within patients, it is important to be able to quantify this measurement error and distinguish between real clinical change and random variation due to the scoring method used. Indeed, this was identified as a key objective of the Outcomes Measures in Rheumatology (OMERACT) RA Imaging Module[122]. Often referred to as the Minimal Clinically Important Difference (MCID), clinical panels would investigate subsets of radiographs and decide the change in units of a particular scoring method that demonstrated a clinically meaningful change[123]. Alongside opinion based

methods for defining this threshold, statistical methods have also been used as a means of defining the level of measurement error in any one study. The Smallest Detectable Difference (SDD) was calculated by Lassere et al.[124], by calculating the level of agreement between two scorers. Using Bland and Altman agreement analysis, the standard deviation (SD) of the differences between two scorers was used to define the level of random measurement error present. Anything below this SD represented random variation, whereas anything above would reflect actual changes in radiographic damage. Using this statistical definition, Bruynesteyn et al. could investigate how the clinical consensus of the MCID compared with the statistical quantification of measurement error[123]. The study found that the estimated MCID and calculated SSD were very similar for the SvdH method, however the MCID for the Larsen method was much smaller than its SDD, perhaps indicative of the Larsen method being less sensitive to change in scores. Since its publication, a meta-analysis was conducted by Navarro-Compan et al.[125], which reaffirmed the use of the SDD in quantifying the measurement error was appropriate. However, these studies mostly assume a continuous normal distribution, whereas radiographic damage scores have been shown to be highly skewed[126]. It is therefore likely that most estimates of the MCID are likely to be too low at low values of the scale, and too high at high values of the scale.

The concept of measurement error also became considerably pertinent with respect to negative progression[127]. It was often assumed that erosions and JSN could only worsen over time, and this was reflected in the development of both the Sharp and Larsen scoring methods, which do not directly allow for negative progression[95, 128]. Despite this, there have been several documented clinical cases where erosive healing had occurred[128], particularly in those patients in sustained remission with no signs of inflammation[129]. As such, the emergence of erosive healing has become an accepted occurrence in clinical practice[130], although still debated amongst clinical circles. Importantly, it is key to establish that any negative progression is the reflection of true erosive healing, and not down to measurement error alone[127]. As such, it is seen as integral that at least two scorers are used in RCTs and observational trials that include radiographic outcomes to ensure that any random variation in the score can be appropriately accounted for[131, 132]. It is possible that in the era of new biologics and increased efficacy of conventional DMARDs where erosive healing is now evident, new radiographic scoring methods that account for this are needed.

2.3 Randomised controlled trials and observational studies

The choice of an appropriate study design for a predetermined hypothesis is paramount. Rather than which design is *better*, the consideration is which study design is most suitable to facilitate the collection of strong, robust evidence for a specific research question[133].

The current paradigm in medicine is that clinical practice should be guided and developed through the use of high quality evidence generated from trials and studies. In the context of treatment efficacy, RCTs are heralded as the 'gold standard' in medical research due to their rigorous use of methodological techniques that aim to minimise bias and reduce the effect of human error[134]. In the absence of counterfactuals, RCTs aim to test the efficacy of a treatment by comparing it to a group of patients, similar clinically and demographically, who do not have the treatment. Bias from confounding effects is reduced through the use of random allocation of patients in groups, and systematic bias is eliminated through the use of blinding both the patients and researchers involved as to what treatment the patient group is taking, either the drug or a placebo, as well as the statistician analysing the data. The aim is to create a vacuum, whereby any potential associations between the treatment and unknown factors are minimised.

RCTs provide a solid framework to test the efficacy of a drug over a relatively short time-scale. However, common criticisms of RCTs are the fact that they generally have relatively short follow-ups, usually no more than 1 year, and that they only include very specific sub-groups of patients, not generally representative of the patient population as a whole. In order to investigate the natural, long-term progression of a disease, long-term observational studies are needed[133]. The pathway from drug development to implementation includes observational studies at phase IV, with the aim to determine long-term safety and efficacy. Furthermore, they are useful for determining long-term prognosis where random allocation of patients is no longer necessary, since the goal can be centred on describing and predicting, rather than estimating treatment efficacy.

Observational cohorts follow-up a group of patients with a similar characteristic, such as a diagnosis of a specific disease, and the patients are either enrolled at the time of diagnosis (inception) or anytime during their disease (prospective). The aim of observational cohorts is to examine patterns over time from a wide range of patients, covering a wide spectrum of the disease. This ranges from mild to severe, and patients are treated according to published guidelines at the

time. As a result, the data collated represents a more 'real-world' sample of patients, and therefore the natural, treated progression of the disease can be analysed. Observational cohorts can also be invaluable to health commissioners and physicians alike, since important data regarding the cost and management of patients can be monitored and analysed. This is particularly important when new therapies or treatments are introduced, where long-term data on a large group of patients is needed to monitor adverse effects and clinical management.

Randomised control trials with radiographic outcomes

RCTs have been instrumental in demonstrating the effect of different treatment strategies on the progression of radiographic damage. While direct comparisons between trials is difficult, since each study recruits specific and unique patient populations[135], there has been breakthroughs in which treatment strategies provide the best radiographic outcomes, particularly in patients with most severe disease.

RCT for Conventional DMARD therapies

The reduction of radiographic damage in RCTs was first reported in a placebo-controlled trial looking at intramuscular gold in the 1970s[136], followed by a number of trials in the 80s and 90s comparing sulphasalazine and methotrexate with other conventional synthetic DMARDs, such as hydroxychloroquine[137], auranofin[138] and azathioprine[139]. All these early trials indicated an increased efficacy of both sulphasalazine and methotrexate in reducing radiographic progression over the comparison DMARD, thereby cementing their use as the anchor DMARDs of choice in routine care (as detailed in the latest National Institute for Clinical Excellence (NICE) guidelines[71]).

Following the evidence that single DMARD therapies, particularly sulphasalazine and methotrexate, were superior to placebo and other conventional synthetic DMARDs, a range of RCTs were conducted to investigate whether there were any additional benefits of combination DMARD therapies on retarding radiographic progression. In 1997, the COBRA (Combinatietherapie Bij Reumatoide Artritis) trial[140] investigated the efficacy of combined sulphasalazine, methotrexate and prednisolone with the use of sulphasalazine alone. They found significant reductions in disease activity over the 56-week trial period, along with reductions in the total SvdH score at the 28, 56 and 80 week periods, for those patients on the COBRA treatment strategy (that

is, the combination of sulphasalazine and methotrexate with concomitant prednisolone) when compared to the patients on sulphasalazine only. Patients indicated similar baseline levels of radiographic damage, and the difference between the groups was greater for the erosions score, although it should be noted that the analysis was restricted to univariate rank sum testing, a statistical method more prone to confounding bias.

Since the publication of the main trial, Landewè et al.[141] conducted a further analysis on the COBRA trial using 5 year follow-up data on the same patients. The longer follow-up and more robust statistical techniques (Generalised Estimating Equation (GEE) modelling) allowed for more precise estimation of the radiographic progression in these patient groups. Although, it is important to note that patients were not randomised to their treatment arms beyond the first year of the trial. From 148 patients completing follow-up measures, the combination group was found to have significantly lower annual progression compared to the sulphasalazine only group, with an annual unit increase of 5.6 vs. 8.6 SvdH units per year (p=0.03). Interestingly, unlike the original analysis for the trial, this paper indicated similar reductions in both erosions scores and JSN scores between the groups, once again highlighting the importance of using appropriate analysis techniques when analysing radiographic data. The likelihood of patients switching to methotrexate during the additional 4 years of follow-up was significant, showing that the use of methotrexate after the initial 1 year trial period increased, while the use of sulphasalazine decreased. Despite these changes between the trial arms over the subsequent 4 year follow-up, adjusted analysis indicated that the changes in DMARD did not significantly alter the estimates for the two groups. The results of this additional follow-up study provided evidence that the early use of intensive DMARD therapy with the combination of prednisolone during the so-called 'window of opportunity'[54], has an effect on reducing radiographic progression over the longer term.

The COBRA study was instrumental in demonstrating the efficacy of combination DMARD therapy in early RA. However, the use of concomitant prednisolone within the combination arm meant that it was not possible to establish whether the reduction in radiographic progression rates was due to the combination of methotrexate and sulphasalazine, the use of predinisolone, or the combination of all three. In 1999, Dougados et al.[142] conducted a three arm trial to investigate the efficacy of methotrexate, sulphasalazine and the combination of the two on disease progression over 1 year. While significant declines in DAS were seen for the combination group, a lack of difference between the groups in the ACR or EULAR response criteria, along with increased toxicity in the combination group questioned the use of combination therapy

in routine care. Nevertheless, there was evidence of decreased radiographic progression in the combination group compared to both monotherapy arms. As with the COBRA trial, further follow-up was also collected to investigate the impact of initial combination therapy on long-term outcomes, including radiography[143]. Interestingly, this study found no perceived benefits of early combination therapy on long-term outcomes, including disease activity, functional disability and radiographic progression. While factors such as differences in levels of disease severity and differences in treatment following the initial 1 year follow-up between the COBRA trial and the trial conducted by Dougados et al. could explain the differences in results, it is likely that the early use of corticosteroids in the COBRA study led to the significant reductions in radiographic progression over the first 5 years in their trial.

Given that the monotherapy groups in both the COBRA trial[140] and the one conducted by Dougados et al.[142] did not include additional prednisolone, it was also unclear whether monotherapy with concomitant prednisolone could be as effective as combination DMARD therapies, such as the COBRA strategy. The Finnish RA combination therapy trial (FIN-RACo) trial[81] assessed the use of triple combination DMARD therapy (methotrexate, sulphasalazine, hydroxychloroquine and prednisolone) against a single DMARD (initially sulphasalazine) with out without prednisolone. A total of 199 patients were recruited into the trial, with 87 randomised to the combination arm and 91 in the single treatment arm. Using the Larsen scoring method, radiographic damage in the combination arm increased from a median of 2 [Interquartile range (IQR) 0-4] to 4 [IQR 0-14], compared to an increase in the single therapy arm from a median of 2 [IQR 0-8] to 12 [IQR 4-20]; indicating a statistically significant increase in progression for the single therapy group over the first 2 years (p=0.002). Furthermore, they found that the number of new erosions was decreased in the combination arm, and interestingly that those patients who received prednisolone in the single treatment arm indicated higher rates of radiographic progression over the 2-year follow-up. Although, it is likely that this increase is a reflection of higher disease severity (thus the need for additional prednisolone along with single DMARD), rather than the inability of prednisolone to reduce radiographic progression when used concomitantly with single DMARDs. The data from the FIN-RACo trial was also supplemented with additional follow-up data, which included a further two follow-ups at 5 and 11 years[144]. The study found that when adjustment for baseline Larsen score was included, there was a statistically significant difference between the combination and single arm treatment groups. The combination arm indicated a much smaller mean change from baseline to 11 years of 17 Larsen units [95% CI 12-26], compared to a mean change of 27 Larsen units [95%CI

22-33] in the single treatment arm. It was concluded that intensive therapies using combination DMARDs with concomitant prednisolone at the early stages, with a focus on remission as the target, leads to tight disease control along with significant reductions in radiological damage over the long-term.

The role of prednisolone in retarding radiographic progression over the short-term was further supported by trials conducted by Kirwan[145] and Svensson[146], which both demonstrated reductions in radiographic progression over the first 2 years. This was further supplemented by a systematic review published in Cochrane, which evaluated 15 RCTs looking at the effect of glucocorticoid steroids in combination with traditional DMARD therapies. The review concluded that all but one study indicated a benefit in favour of glucocorticoids, with a pooled standardised mean difference of 0.40 in reduction of the erosion score over the first 2 years[147].

As is often the case in medicine, standardising treatment based on the best evidence at the time is difficult. Often different clinics will operate in different ways with regards to choice and, more importantly, dose of a particular drug[148]. The CAMERA trial (computer assisted management in early RA) investigated whether intensive treatment of methotrexate mediated by a computer decision based tool could improve RA outcomes over 2 years when compared to conventional treatment[148]. They found that higher doses of methotrexate optimised earlier in the computer-assisted arm led to improvements in disease activity after 2 years, however there was no difference in radiographic progression between the two arms. This is likely owing to the very little progression in both patient groups over the first 2 years. The authors do note that of those with radiographic damage, the progression was higher in the conventional arm compared to the computer-assisted arm, suggestive of some beneficial effect of high dose methotrexate in reducing long-term radiographic progression. With respect to radiographic damage, this trial may suggest that methotrexate dose is less important in inhibiting radiographic progression, although more long-term evidence is needed.

RCT for biologic DMARD therapies - Anti-TNF

The introduction of biologic therapies brought about the potential for a new line of RA treatment with the potential of being more effective than conventional synthetic DMARDs, particularly methotrexate. One of the first trials published in the New England Journal of Medicine by Bathon et al.[78] compared the use of etanercept (belonging to a class of anti-TNF-α inhibitor biologics) at both 10mg and 25mg doses against methotrexate. A total of 632 patients were

randomised into the three treatment groups, with measures of disease activity and radiographic damage reported over 12 months. The trial found that overall disease activity was reduced in the 20mg etanercept group when compared to the methotrexate only group, along with greater reductions in mean changes of the total SvdH and erosion score. The total SvdH and erosion score between the methotrexate and etanercept at 10mg group were similar over the 12 months and the rate of JSN progression was similar for all three groups.

The efficacy of etanercept and methotrexate in combination was also highlighted in the combination of methotrexate and etanercept in active early rheumatoid arthritis (COMET) trial[149]. This 2-arm RCT investigated 542 early RA patients recruited from 22 countries. Two hundred and sixty-three were randomised to the methotrexate only arm, while 265 were randomised to a combined treatment of etanercept and methotrexate. The trial found that those patients randomised to the combination arm were more likely to achieve remission as defined using the DAS28 score of <2.6 over the whole 12 month trial period. Radiographic assessment using the SvdH score also indicated a lower rate of radiographic progression in the combination group, progressing just 0.27 units over the 12 months, compared to 2.44 units in the methotrexate only arm. This was largely driven by increases in erosions, rather than JSN. Furthermore, post-hoc analysis of the COMET data[150] examined the impact of very early treatment in both the combination and monotherapy patient sub-groups. In both trial arms patients were stratified into very early treatment (<4 months,) or early treatment (>4 months).

In the combination arm, very early treatment resulted in better disease outcomes in 12 months, with a higher proportion of patients in the very early group achieving low disease activity and DAS28 remission (69.8% vs. 47.8%). In contrast, the very early treatment group within the methotrexate only arm indicated similar proportions of patients in the low disease activity and DAS28 remission (34.7% vs. 31.8%), although significantly different statistically. However, what is particularly interesting is the disparity between radiographic outcomes and disease severity in this trial. The proportion of patients with an annual increase of <0.5 SvdH was similar in the very early and early groups within the combination arm.

Although, there was a significantly high proportion of patients treated very early in the methotrexate arm with low radiographic progression compared to the early treatment group (73.9% vs. 50% respectively). No direct comparisons were made in the data analysis. The proportion of patients in the very early methotrexate group showed similar levels of radiological progression

when compared to both the very early and early groups within combination arm. For radiographic progression outcomes, there was evidence that very early methotrexate use could be as effective in retarding radiographic progression in the short-term as combination therapy with biologic DMARDs. This evidence is limited by two factors; often, the precise timing of DMARD initiation in RCTs is not recorded, and definitions of early RA can vary from 1 to 3 years from symptom onset. Therefore, there is currently no other evidence that very early DMARD use has such a profound effect on short-term, and indeed long-term, radiographic progression.

Other anti-TNF biologics have also shown similar results to entanercept. The PREMIER trial conducted by Breedveld et al.[151] demonstrated the efficacy of adalimumab in a multi-centre RCT of 799 early RA patients. They found that the early use of combination adalimumab and methotrexate was superior to both drugs when used alone, with a greater proportion of patients achieving the ACR90 criteria and significantly reduced radiographic progression over the first 2 years. Using the SvdH scoring method, the mean change from baseline to year 2 was 10.4, 5.5 and 1.9 for the methotrexate, adalimumab and combination therapy arms respectively. However, the authors do note that the methotrexate group had higher radiographic damage at baseline, specifically a higher erosion score, partly explained by a marginally longer disease duration compared to the other two arms.

The ASPIRE trial conducted in 2004[152] demonstrated the superiority of infliximab when used in combination with methotrexate, again indicating better disease control over 12 months along with a greater reductions in radiographic joint damage. In a follow-up study conducted by Smolen et al.[153], detailed analysis of the ASPIRE data was conducted to investigate potential prognostic markers of poor disease control and increased radiographic progression. While the relative efficacy of methotrexate is highlighted in the report, particularly when the cost-to-benefit ratio is considered, there are those patients that continue to progress radiographically with poor disease outcomes. This analysis identified this sub-group in the methotrexate only treatment arm, showing relatively higher levels of acute phase markers and swollen joint counts. In contrast, patients receiving methotrexate in combination with infliximab had consistently low radiographic progression, independent of increased disease markers.

RCT for biologic DMARD therapies - Rituximab, Tocilizumab and Infliximab

Along with the anti-TNF biologics outlined in trials so far, a host of other biologics, which target other cells involved in the inflammation cascade, have also been authorised for the use in

the treatment in RA. The IMAGE trial[154] demonstrates the improved efficacy of rituximab, a biologic which targets CD20+B cells, when used in combination with methotrexate compared to methotrexate alone. Those patients randomised to rituximab at 2x1000mg dose had significantly better disease outcomes by 12 months, along with significant reductions in radiographic progression. Methotrexate demonstrated a mean change in total SvdH score of 0.7 and 0.38 from baseline to 24-weeks and from 24-weeks to 52-weeks respectively. This compared to a change in total SvdH score of 0.33 and 0.14 in the rituximab 2x1000mg group at baseline to 24-weeks and 24-weeks to 52-weeks respectively.

As with the other trials, the total SvdH score was largely driven by changes in the erosion score, rather than JSN. Finally the efficacy of tocilizumab, a humanised anti-IL-6 receptor, was investigated in the SAMURAI trial[155]. Unlike previous trials that compared the biologic monotherapy to methotrexate monotherapy, the SAMURAI trial included patients on a range of different conventional DMARD treatment strategies in the control arm. Having said that, patients in the control arm were predominately on methotrexate monotherapy, however some patients also received non-methotrexate DMARDs either in monotherapy or in combination with methotrexate. The trial assessed disease severity and radiographic progression over 12 months in a total of 265 patients. RA duration was typically longer compared to the previous trials, with patients having RA for up to 5 years, as opposed to 2 or 3 years. Nevertheless, the trial demonstrated the superiority of tocilizumab monotherapy in reducing disease severity and radiographic damage over the first year when compared to conventional DMARD therapy.

RCT comparing combination DMARDs with biologic DMARDs

With the exception of the SAMURAI trial, the sub-sample of RCTs outlined so far have demonstrated the efficacy of biologics either alone or in combination with methotrexate against only methotrexate monotherapy. The apparent superiority of biologics over conventional DMARD therapies may therefore be unsurprising, as the COBRA[148] and FIN-RACo[81] have demonstrated that early use of conventional synthetic DMARDS in combination is more effective than DMARD monotherapy[83]. To this end, data from the Swefot trial[156] investigated the use of add-on combination DMARDs (sulaphasalazine and hydroxychloroquine) compared to the use of add-on biologic therapies (infliximab) on those patients receiving initial methotrexate monotherapy who fail to respond after the first 3 to 4 months. The initial 12-month report indicated that patients randomised to receive biologic add-on therapy had a higher proportion

achieving the EULAR defined 'good-response' criteria compared to those patients receiving combination conventional DMARDs. However, the report of the same trial data after 24 months showed that the number of patients achieving EULAR defined 'good-response' at 24 months was similar in the two trial arms[83]. Although, there was evidence of reduced radiological progression in the patients treated with combination methotrexate and infliximab.

The results from this extended report indicate that conventional DMARDs used in combination early on may provide a suitable and cost-effective therapy option in the short-term, compared to the use of add-on combination biologics. What is unclear is whether the radiographic benefit of add-on biologic therapy in this patient sub-group that fail to respond to methotrexate monotherapy is significant in the long-term, as this could still highlight the need for the use of biologics in all but the most severe cases of RA, where radiographic retardation is difficult.

What is the optimal treatment for early RA?

The small sub-sample of RCTs described have shown that, on the whole, early intensive combination DMARD therapy produces better long-term radiologic outcomes in early RA when compared to delayed, monotherapy DMARD treatments. This has been demonstrated using a range of different combination treatments; with combination methotrexate and sulphasalazine, including low-dose corticosteroids in the COBRA trial[148], and triple therapy with methotrexate, sulphasalazine and hydroxychloroquine, including low dose corticosteroids in the FIN-RACo trial[81]. The use of biologic therapies, largely anti-TNF-α, have shown superiority to methotrexate in monotherapy, and the potential for greater radiographic retardation when compared to combination DMARDs in the Swefot trial[83].

The increase in treatment options for early RA has made it difficult to ascertain what the optimal treatment strategy should be in early RA[56]. As such, the BeST trial aimed to evaluate four different treatment strategies in order to shed some light on which would lead to the best outcomes in the long-term. The first group (sequential) received methotrexate monotherapy, with other conventional DMARDs being used in sequence if response was insufficient. The second group (add-on) were assigned to step-up combination DMARDs, with methotrexate monotherapy being used as first line, and other conventional DMARDs were added if response was insufficient. The third group (combination DMARD) were assigned to combination DMARD therapy, with a combination of both methotrexate and sulphasalazine, along with low-dose predinsolone being used as first line. Finally, the fourth treatment group (combination biologics) were assigned

to combination biologic therapy, with methotrexate and infliximab being used as first line. In all groups, treatment was escalated over the duration of the trial from increased methotrexate dose, switching of conventional DMARDs used, and included the use of combination DMARDs with biologic therapies if response was lacking.

A total of 508 patients were enrolled in the study and completed the first year follow-up. Patients in the combination DMARD and combination biologics groups had a higher proportion of patients achieving remission by 12 months compared to the sequential and add-on group. The proportion of patients achieving remission in the combination DMARD group was similar to the combination biologic group (71% vs. 74%). A similar pattern was seen for functional disability and radiographic progression, where combination therapy with either prednisone or infliximab was superior in controlling disease when compared to monotherapy. As with the Swefot trial[83], the results of the BeST trial do indicate that early combination DMARD therapy with concomitant prednisolone could be as effective as combination biologic therapy in reducing radiographic progression. Van der Broek et al.[157] argue that the common theme in all four trial arms was the implementation of T2T principles, whereby adjustments were made continuously when response was not being achieved. It could therefore be argued that early combination therapy has the advantage of faster and more sustained remission in the first few years of the disease while ensuring long-term sustainment of remission is achieved by continued monitoring and tailored adjustments for more intensive therapies made when needed. In the case of the BeST trial this appeared to result in sustained long-term remission in all treatment groups over the 7 year follow-up.

Observational studies of early RA with radiographic outcomes

Early records of RA cohorts date back to the 1950s and 60s and were primarily population based cohorts with the aim of estimating the prevalence of RA[158]. The first hospital based RA cohort was developed in Bath, which recruited 100 early RA patients from between 1957 and 1963. The cohort collected data for over 40-years and looked at aspects of RA surrounding disability and loss of functional capacity[159]. A similar cohort was developed in Middlesex named the RAPS study, which also looked at serial x-rays of the hands of feet, providing some of the first evidence about the high rate of structural joint damage in the first 5 years of RA[160]. Following these relatively small cohort studies, a host of large inception cohorts were developed with the aim of collecting higher numbers of patients with similar lengths in follow-up.

The Early RA study (ERAS) was formed in 1986 and aimed to collect data on 1000 early RA patients, and follow those patients for at least 5 years[161, 162]. Recruited across 9 different centres in the UK from the years 1986 to 2001, a total of 1465 patients were recruited with standard clinical, radiological, laboratory and genetic data collected at baseline, six-months, then yearly thereafter for up to 25 years; thereby exceeding its target for both patient numbers and length of follow-up. Data collected included a range of demographic, clinical and laboratory markers, the most fundamental being disease activity (DAS-44), functional disability (Health Assessment Questionnaire (HAQ)) radiographic data (recorded using both the Larsen and SvdH scoring methods), health survey data (Short-Form 36), genetics, surgical interventions, co-morbidities and mortalities. All patients recruited were treated according to published guidelines at the time.

The successor to ERAS, the Early RA Network (ERAN) had similar aims, design and clinical assessments[163]. The 21st century saw a number of key developments in the management of RA, largely due to the publication of national and international guidelines, increased activity from patient support groups and greater access to care. The NHS as a whole was more focused on clinical governance to advance the practice of evidence-based medicine, and governing bodies such as NICE relied on high quality evidence to guide clinical management and evaluate current healthcare provisions within the National Health Service (NHS). This led to higher demand for high quality clinical data, such as those provided from observational cohorts. As a result, ERAN, a natural extension to ERAS, was developed. It maintained many similarities with ERAS in its core aims and design. The broad aims of the cohort were to contribute to good clinical practice and facilitate in the decisions made by clinical governance. Importantly, ERAN documented the clinical management of patients during the era in which biologic therapies were introduced.

During the same time as the ERAS and ERAN cohorts, the Norfolk Arthritis Registry (NOAR) was established to investigate the incidence of polyarthritis and RA in rural areas of East Anglia[164]. It was one of the first studies to update UK incidence and prevalence rates on RA, which led to the finding that age-specific prevalence was decreasing in women, but increasing in men[5]. Patients were recruited from 1989 to 1994 with data recorded for up to 15 years, and from 2000 to 2008 with data collected for up to 2 years. As with the ERAS and ERAN cohorts, demographic, clinical and laboratory markers were collected. Unlike ERAS/ERAN, patients with undifferentiated RA were also included in the cohort, which allowed the cohort

to compare outcomes in other inflammatory arthritis sub-groups, indicating that patients with polyarthritis experienced similar levels of worsening disability and mortality as RA[165–167].

Possibly one of the most important registries in recent years for RA is the development of the British Society for Rheumatology Biologics Register (BSR-BR)[168]. The register was set up to monitor all patients who were newly started on biologic DMARD therapies for safety outcomes. While including a comparison cohort of patients on conventional synthetic DMARDs, their primary aim is to detect any increases in risk of lymphomas owing to the use of biologic DMARDs. To date, the register has data on several thousand RA patients on biologic DMARDs, and is thought to represent approximately 80% of patients starting on biologics since 2002. Their primary aim is to detect any increases in risk of lymphomas from the use of biologic DMARDs, over and above the risk from RA itself. While no increased risk has been found for lymphoma, but did indicate increased risk of non-melanoma skin cancers and infections[169].

Outside the UK, many cohorts have also been developed. The Leiden Early Arthritic Cohort (EAC), started in 1993, recruited patients with early inflammatory arthritis from GP referrals[170]. The aim was to detect and treat inflammatory disorders early in the disease course, particularly early RA. Outcome measures collected included clinical measures, laboratory measures, radiographic measures and genetics. By 2003, 1,600 patients with early inflammatory arthritis were recruited into the cohort. Also from the Netherlands, the Nijmegen cohort was set up to monitor patients with early RA[171]. By 2005, a total of 525 patients were enrolled into the cohort with over 10 years follow-up[172]. Of particular note is also the Danish Registry for Biologic Therapies in Rheumatology, commonly referred to as the DANBIO registry, which was set up in 2000 with the aim of collecting a large amount of data on patients in rheumatology clinics. To date, it has collected data on over 10,000 patients with RA, Psoriatic Arthritis (PsA) and Ankylosing Spondylitis (AS)[173]. What makes this cohort particularly unique is the way in which it has integrated the collection of cohort data through the use of simple computer software, which can be easily implemented in routine clinical care. In the early stages, the registry was largely paper based; however in 2005, the introduction of DANBIO-online allowed for much more efficient data collection methods. The advantages of adopting an online only system were three-fold. Firstly, it allowed for more accurate data entry since the additional step of entering data from paper forms was eradicated. Secondly, it also included algorithms for checking missing data, which would prompt the user to fill in sections before submission. This was an effective means of reducing missing data, which is not possible using paper forms. Finally, the user-friendly interface and simple mode of data entry provided clinicians with a

less burdensome means of enrolling patients and entering follow-up data. In fact, the online database provided a means for clinicians to access individual patient data during consultations, allowing them to see how both routine clinical markers and study-specific markers of disease had changed over the course of the disease.

Other non-UK cohorts of note are the Lund cohort in Sweden[174], the Heinola and Jyväskylä cohort in Finland[175] and the ESPOIR cohort in France[176]. To assess the radiographic findings from these observational cohorts, the next chapter will conduct a systematic review on all published literature concerning the progression and predictive factors of radiographic joint damage.

2.4 Discussion

Rheumatoid arthritis is a disabling a chronic disease. It is multi-faceted and involves a number of key outcomes, including pain, disability, quality of life, inflammation, and structural damage. Understanding the aetiology and mechanisms of disease has helped the development of effective treatment strategies, including the development of drugs, the optimum time to administered these drugs, and the long-term management of the disease.

There remains a wide variety of outcome measures for quantifying the different aspects of RA, from biomarkers of disease, to scoring systems for establishing the extent of joint damage using plain x-rays. While these measures have their limitations, they have been important for research into RA, as they allow for tangible assessments of objective disease outcomes for the use in RCTs and observational studies.

RCTs have been instrumental in measuring the efficacy of different treatment strategies, and long-term extensions of these trials are important in determining the long-term impact of these early treatment decisions on the disease course. However, the limitations of RCTs are not to be ignored, and their inability to answer important research questions, such as the implementation of treatment strategies in routine clinical care and the progression of disease in patients with less severe disease, highlights the importance of long-term observational studies.

The emergence of large observational cohorts and registries has been key in understanding the natural course of RA, and how treatments are implemented in real-world clinics, using patients with a wide spectrum of disease severity and co-morbidities. It has allowed a greater

understanding of how critical changes in treatment practices have affected the progression of key RA outcomes, such as radiographic joint damage. The use of computerised systems and open-source software has proven to be of great assistance in the set up and maintenance of long-term observational trials, seemingly by-passing many barriers faced when trying to set up and run these often time-consuming and expensive studies. Implementation of these techniques not only allows for the opportunity to collect rich and useful data for research purposes, but can also help facilitate the clinical consultation with patients, whereby an accurate progression of any individual patients' disease can be easily illustrated and explained in real time.

Chapter 3

Systematic Review

3.1 Introduction

Towards the later sections of Chapter 2 a host of RCTs and observational studies that recorded radiographic outcomes were introduced. The main focus of the RCTs were to investigate the efficacy of different treatment regimes on radiographic progression, since restricted follow-up periods and strict patient inclusion criteria render RCTs inappropriate for prognostic modelling and investigation of long-term progression. In contrast, observational studies provided detailed accounts of the natural progression radiographic damage over time, including patients covering the wide spectrum of disease severity, treated in a natural clinical setting according to published guidelines at the time.

The specific aim of this chapter is to expand on the observational studies detailed in Chapter 2 and conduct a systematic review on all published observational studies looking at radiographic outcomes in early RA over the long-term. This will directly address the first aim outlined in Chapter 1; that is to investigate the current evidence base to identify common methods in measuring and reporting radiographic outcomes.

The first recorded systematic reviews on radiographic progression were published by Scott et al. in 2000[85], which was later updated in 2003[38]. It reviewed the rates and clinical predictors of radiographic progression in patients with RA. The review concluded that 39-73% of early RA patients develop one or more erosions in the first 5 years and radiographic joint damage progresses constantly over the first 20 years of disease[38]. Two systematic reviews on radiographic progression have been published since[86, 177]. Neither reported radiographic progression rates,

concentrating instead on specific predictors (function and disease activity) of radiological damage. To date, no review has used quantitative analysis techniques, such as meta-analysis, to investigate radiographic progression rates from the published literature.

Despite some evidence of erosive healing[128, 129], structural damage is largely considered irreversible[38, 178], and therefore clinicians need to identify patients at higher risk of severe radiographic damage to tailored treatment earlier on. Predictive modelling is a useful statistical method to identify all clinical factors that are associated with primary RA outcomes[178, 179]. Previous studies have provided contemporary accounts on the relationship between radiographic progression with functional disability[86] and disease activity[177]. Other factors, such as anti-CCP antibodies and genetic factors have not been fully reviewed.

The objectives of this systematic review are therefore to evaluate firstly, all published data on baseline and annual progression rates of radiographic damage from all longitudinal observational cohorts, and secondly, the association of standard clinical and laboratory parameters with long-term radiographic joint damage. Where appropriate, meta-analyses were conducted on the baseline and annual progression rates of radiographic joint damage scores, and their predictive markers.

3.2 Methods

A systematic review protocol was developed to ensure that objectives and aims were clearly outlined from the outset and submitted and approved by PROSPERO in February 2014 (Registration Number: CRD42014007589).

Identifying publications

Publications were identified by computerised searches of PubMed, Cochrane Library (including CENTRAL, CDSR, DARE, HTA) and Scopus, supplemented by lateral search techniques: checking reference lists, performing key word searches in Google Scholar and using the 'cited by' option in PubMed. All databases were searched from January 1st 1975 to February 31st 2014. The search strategy used a mixture of key words and MeSH terms on the title/abstract and full text as appropriate.

The following search strategy was used to search for all published literature in PubMed:

((‘arthritis, rheumatoid’[MeSH Terms] AND ((((radiographic[Title/Abstract] OR X-ray [Title/Abstract]) OR structural joint damage[Title/Abstract]) OR Larsen[Title/Abstract]) OR Sharp [Title/Abstract])) AND (((((((‘randomized controlled trials as topic’[MeSH Terms] OR ‘randomized controlled trials as topic’[MeSH Terms]) OR (controlled[All Fields] AND (‘clinical trials as topic’[MeSH Terms] OR (‘clinical’[All Fields] AND ‘trials’[All Fields] AND ‘topic’[All Fields]) OR ‘clinical trials as topic’[All Fields] OR ‘trial’[All Fields]))) OR (‘randomized controlled trial’[Publication Type] OR ‘randomized controlled trials as topic’[MeSH Terms] OR ‘randomised controlled trial’[All Fields] OR ‘randomized controlled trial’[All Fields])) OR (‘clinical trial’[Publication Type] OR ‘clinical trials as topic’[MeSH Terms] OR ‘clinical trial’[All Fields])) OR observational[All Fields]) OR (‘longitudinal studies’[MeSH Terms] OR (‘longitudinal’[All Fields] AND ‘studies’[All Fields]) OR ‘longitudinal studies’[All Fields] OR ‘prospective’[All Fields])) OR (‘epidemiologic studies’[MeSH Terms] OR (‘epidemiologic’[All Fields] AND ‘studies’[All Fields]) OR ‘epidemiologic studies’[All Fields] OR (‘epidemiological’[All Fields] AND ‘studies’[All Fields]) OR ‘epidemiological studies’[All Fields])) OR longitudinal[All Fields])).

Inclusion/exclusion criteria

The following inclusion criteria were used to select publications: (1) investigated the progression or predictive/prognostic markers of radiographic joint damage, (2) use of validated diagnosis criteria (.e.g European League Against Rheumatology (EULAR) and/or the American College of Rheumatology (ACR) criteria), (3) baseline assessments occurred no later than 3 years from symptom onset, (4) prospective cohort study design, (5) radiographic follow-up data available for at least 5 years for progression rates, and at least 3-years for predictive markers, (6) Larsen or Sharp van der Heijde methods (SvdH) to score radiographic damage as the primary outcome, and (7) only publications in English.

Publication screening

One reviewer (Lewis Carpenter) screened all titles and abstracts identified by the electronic search and applied the selection criteria to potentially relevant papers. A second reviewer (Elena Nikiphorou) independently screened the full text of 10% to compare against agreed inclusion criteria. Agreement was achieved in 97% and any disagreements were resolved through discussion. Figure 3.1 provides a flow diagram of all publications identified, screened and included in this review.

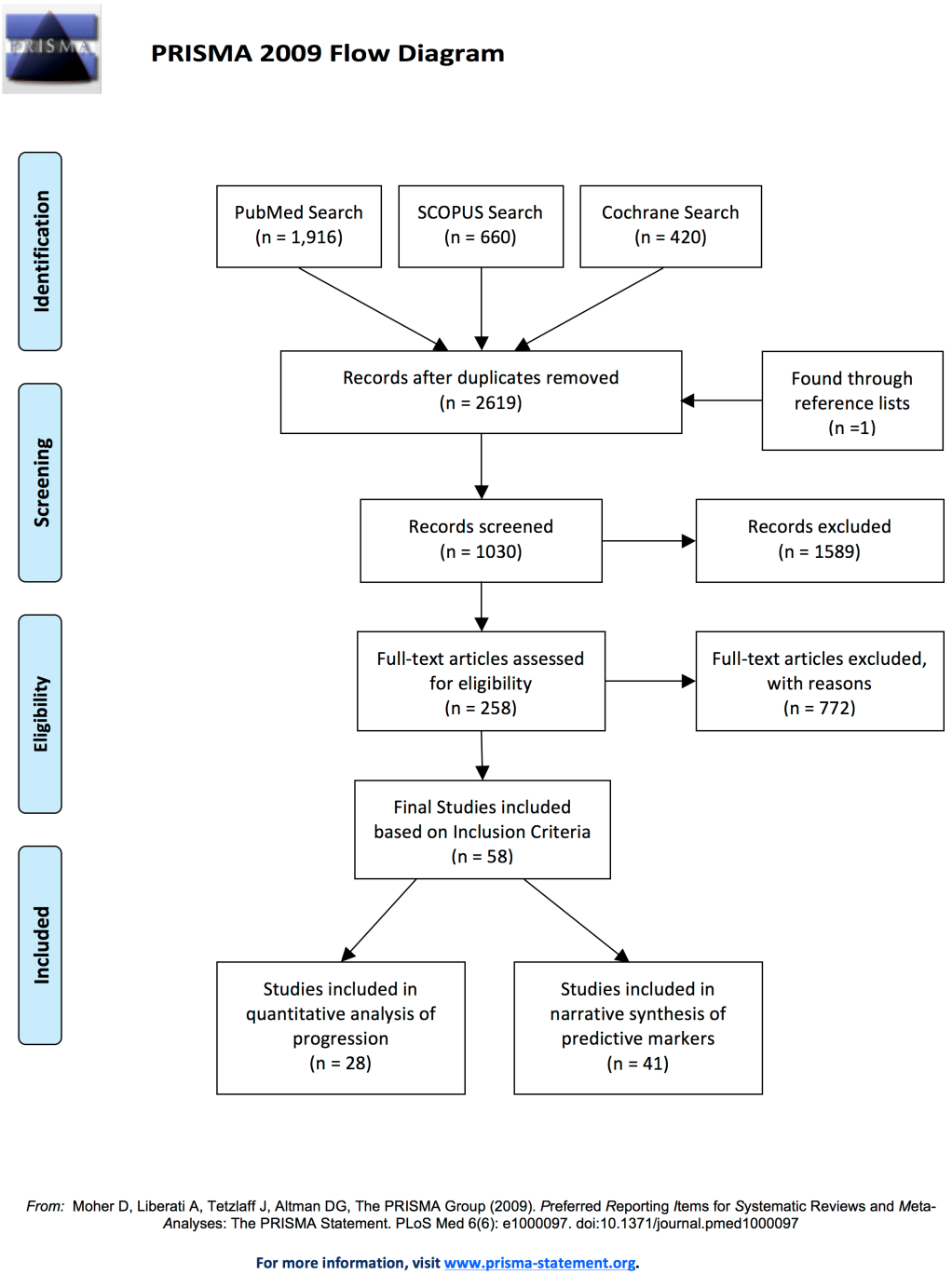


FIGURE 3.1: PRISMA Flow Diagram

Data extraction

Two reviewers (Lewis Carpenter and Rachel Sharpe) extracted data using a pre-designed form, piloted to ensure all data necessary for the analysis could be included: cohort study name,

country of study population, scoring method, patients numbers, recruitment years, length of follow-up, sex, mean age, baseline DAS and HAQ, DMARD use by 12 months, RF positivity, number, mean/median and standard deviation/interquartile range of radiographic scores at each follow-up, analysis method, and significant and non-significant predictors with the effect estimate and 95% confidence intervals (95%CI). Where the raw data was not published, authors were contacted to provide this data (n=21).

Quality Assessment

All studies were screened and rated based on the modified version of the Downs and Blacks checklist for non-randomised studies of health care interventions[180]. Since the studies involved were not examining effectiveness of health care interventions, all items on the checklist that related to comparative groups (e.g. randomisation and blinding procedures) were not used. One reviewer scored all studies using the amended checklist and a second reviewer independently scored 10% drawn at random. All discrepancies between the two reviewers were discussed and consensus achieved.

Analysis

Means and standard deviations of either the Larsen or Sharp score were recorded at each follow-up for each study. Where only a median scores were obtained, the median and range was converted into a mean score and standard deviation[181]. In order to estimate the annual rate of change, along with standard errors, a linear regression model was conducted with follow-up year as the independent variable. Baseline scores and annual rates of progression, with respective standard errors, were transformed into percentage of maximum damage dependent on which radiographic scoring method was used[182, 183]. Transformed scores were entered into random effects meta-analysis to calculate pooled effect estimates for both baseline radiographic scores and annual rate of change.

To assess the strength of predictive markers, the regression coefficients and odds ratios (OR), with their respective 95%CI, were collated. Unadjusted effect estimates were primarily sought, when not reported adjusted estimates were used. Random effects meta-analysis was used for all models because of the likely heterogeneity between studies.

Heterogeneity

The study entry criteria of this review included studies as homogenous as possible to allow appropriate meta-analysis. Heterogeneity between studies was predicted a priori, mainly due to differences in cohort start dates and scoring methods. The i-Squared statistic for each model was found to be consistently above 80%, and therefore random effects models were used throughout. To investigate possible sources of heterogeneity, scoring method and recruitment year were entered into meta-regression models and were the basis of two separate stratified analyses. As there were only ten studies for analysis, these were stratified into two recruitment periods, 1965-1989 and 1990-2000. Not only did this provide equal groupings for stratified analysis, but around 1990 onwards signified marked changes in clinical management of RA, including early and more intensive therapies and treat-to-target strategies.

Narrative synthesis of predictive factors

Every marker identified from each study, both significant and non-significant, was recorded and counted. Where possible, meta-analysis was used to assess the strength of predictive markers. However, for several predictive markers meta-analysis was not possible because too few studies conducted or reported analyses in such a way as to make pooling of effects possible and appropriate. Where a meta-analysis was not appropriate a narrative synthesis of the data was conducted.

3.3 Results

Meta-analysis of long-term radiographic progression

Of the 28 studies identified, ten studies provided the necessary data for meta-analysis[184–192] (See Table 3.1). Patients were recruited between 1965 and 2000, with follow-up ranging from 5-20 years. The number of patients included with baseline radiographic data ranged from 73-1121. Four studies used Larsen; six used the SvdH scoring method. Five recruited patients from 1965-1989 and five from 1990-2000.

Author	Cohort	Country	Scoring Method (Max)	Sample Size	Recruitment Year	Years Follow-up	% Female	Mean Age	% RF+	Radiographic Damage	
										Mean Baseline (SD)	Annual Rate (SE)
Post 1990											
Bridges Jr., et al.	CLEAR I	USA	Sharp (448)	357	2000	5	82.4	50	80.1	2.89 (7.65)	1.87 (0.70)
Tanaka, et al.	Japan Cohort	Japan	Sharp (448)	130	1995	10	69	54	54	5 (10.33)	3 (0.23)
Courvoisier, et al.	French Cohort	France	Sharp (448)	117	1993	10	80.3	50.4	78.6	5.8 (9)	3.08 (0.42)
Knevel, et al.	Leiden	Nether-lands	Sharp (448)	678	1993	7	67.4	56.6	57.9	8.74 (10.74)	4.34 (0.11)
Viatte, et al.	NOAR	UK	Larsen (200)	1446	1990	5	68	56	44	10.74 (13.89)	0.83 (0.61)
Pre 1990											
James et al.	ERAS	UK	Larsen (200)	1465	1986	9	66.4	55.3	62.7	4.32 (10.13)	2.44 (0.70)
Kuper, et al.	Nijmegen Cohort	Netherlands	Sharp (448)	126	1985	6	64	50	83	1 (16.17)	8 (0.10)
Kapetanovic et al.	Lund Cohort	Sweden	Larsen (200)	135	1985	20	62.8	52.1	83	8.13 (1.47)	3.4 (0.31)
Kaarela et al.	Hienola Cohort	Finland	Larsen (200)	103	1973	20	68	45	100	4.3 (6.8)	4.12 (0.45)
Knevel, et al.	Groningen cohort	Netherlands	Sharp (448)	261	1965	25	67.8	45.1	93.2	3 (56.5)	3.67 (0.50)

TABLE 3.1: Summary table of studies included with long-term radiographic data

3.3.1 Baseline radiographic score

The first analysis examined baseline radiographic score across the ten studies. The overall rate of damage at baseline was estimated at 2.02% (95%CI 1.37-2.67) of maximum damage. When stratified by scoring method, the sub-group pooled estimate for Larsen score was 3.41% (95%CI 1.80-5.01) of maximum damage (6.82 units), and the sub-group pooled estimate for the SvdH score was 1.20% (95%CI 0.60-1.80) of maximum damage (5.38 units). Studies recruiting patients between 1965-1989 had a sub-group pooled estimate of 2.01% (95%CI 1.14-2.89) of maximum damage, and those between 1990-2000 reported a sub-group pooled estimate of 2.03% (95%CI 1.05-3.01) of maximum damage (See Figure 3.2).

3.3.2 Annual rate of change

In the second analysis the overall annual rate of change was estimated at 1.08% (95%CI 0.72-1.44) of maximum damage. The sub-group pooled estimate for Larsen score was 1.38% (95%CI 1.80-5.01) of maximum damage, or 2.76 units per year, and for the SvdH score was 1.20% (95%CI 0.88-1.88) of maximum damage, or 4.03 units per year. When stratified by recruitment periods, 1965-1989 had a sub-group pooled estimate of 1.50% (95%CI 1.08-1.92) of maximum damage, and for 1990-2000 was 0.68% (95%CI 0.47-0.90) of maximum damage (See Figure 3.3).

Meta-Regression

Although the small sample size (n=10 studies) limits the power to conduct meta-regression models with appropriate number of covariates, a sensitivity analysis using meta-regression was

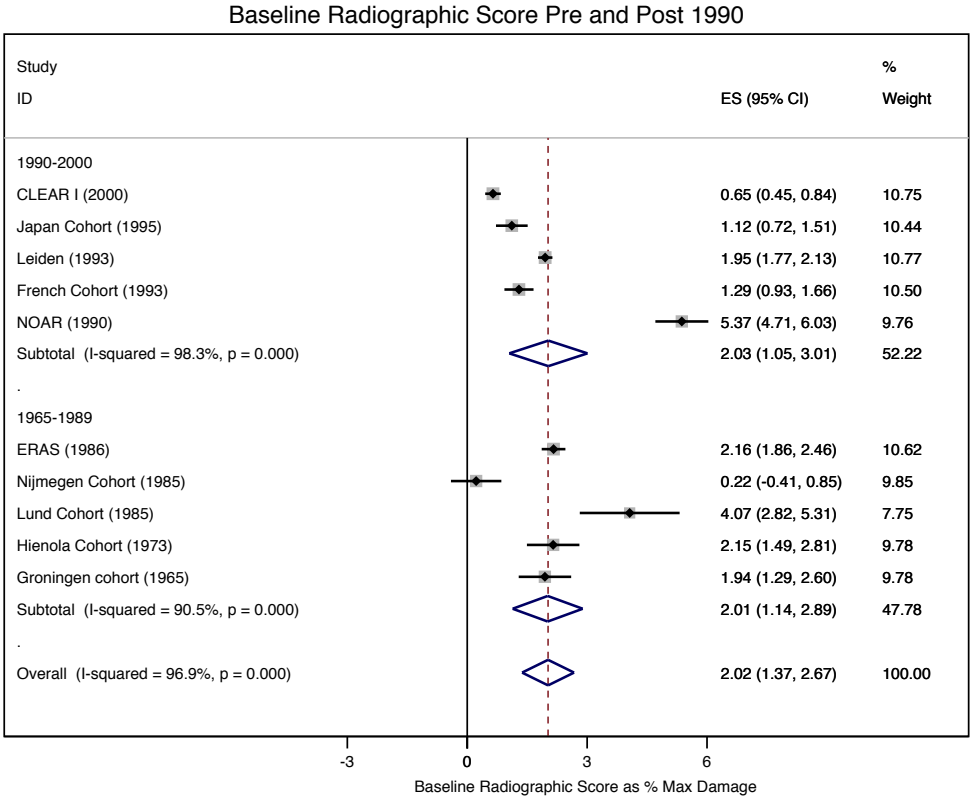


FIGURE 3.2: Forest plot of baseline radiographic scores stratified by recruitment periods

used to indicate whether differences in baseline and annual progression rates were significant while controlling for scoring method. It was also important to investigate possible factors that might influence the overall effect estimate given the high levels of heterogeneity between studies (i-squared score ranging from 90.5%-98.3%).

In line with visual inspection of stratified effect sizes, the difference of 0.02% for baseline progression rates was non-significant ($p>0.1$), while the difference in annual progression rate of 0.79% was significant ($p<0.05$). Of note is that differences between Larsen and SvdH scoring methods was not significantly different for annual progression rates ($p<0.01$), suggesting that relative increase in either scoring method was comparable.

3.3.3 Predictive markers of long-term radiographic damage

Forty-one published papers were identified that examined predictive markers of radiographic joint damage, representing 21 different cohort studies. Despite a number of papers being based

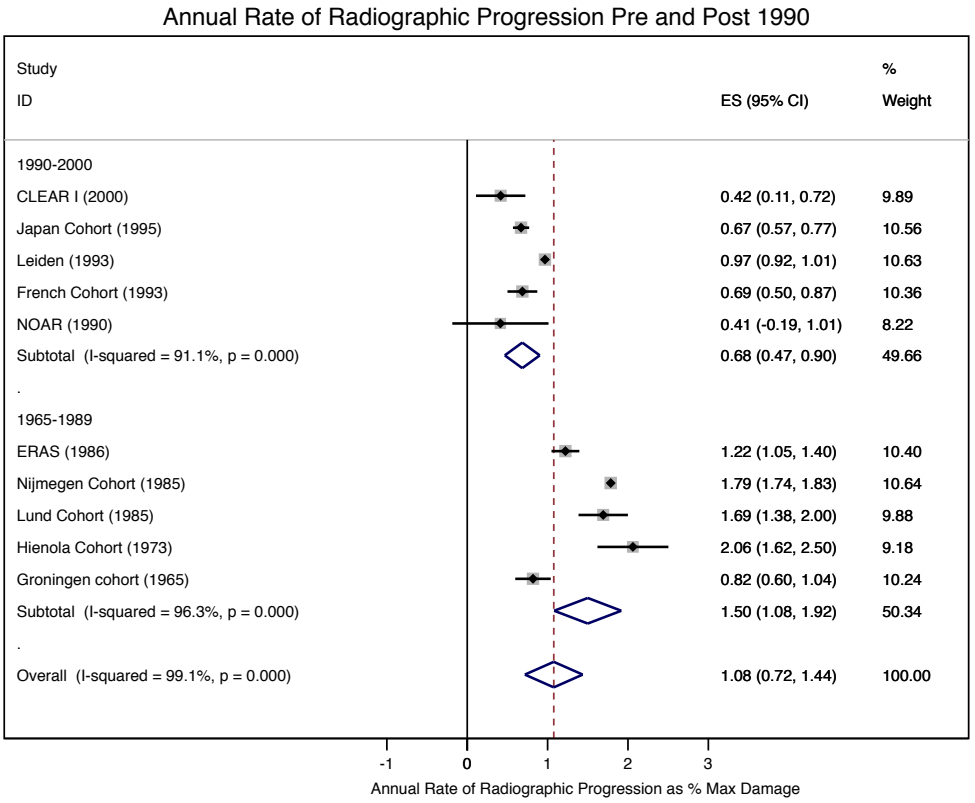


FIGURE 3.3: Forest plot of annual rates of change stratified by recruitment periods

on the same cohort data (See Table 3.2), the analysis techniques used were sufficiently different from each other to allow their inclusion in the analysis.

A total of 28 studies used the SvdH[117, 185, 190–215] and 13 used the Larsen scoring method[75, 77, 188, 216–225] as the primary outcome measure of radiographic joint damage. A total of 24 of the 41 studies examined radiographic damage at a single time point, while 17 investigated radiographic damage expressed as a change in score over two time points. A total of 13 studies transformed radiographic scores into binary variables and 27 treated the radiographic score as a continuous score. Only one study treated the radiographic score as an 'event' for use in a 'time-to-event' analysis[223]. In total, 12 different analysis methods were used (Table 3.2).

Acute phase Markers

One of the most reported covariates was ESR, followed by C-Reactive Protein (CRP), both key markers of the acute-phase response (See Figure 3.5). Of the 15 studies that included ESR, 13 studies reported it as a significant predictor, and of the 11 studies that included CRP 10

reported it as a significant predictor of radiographic joint damage. Some studies examined acute phase markers as continuous, others as a categorical predictors, using either pre defined cut-points, or quartiles. Although there was sufficient data to conduct a meta-analysis, these differences made formal meta-analysis inappropriate and direct comparison between the effect estimates unfeasible.

Courvoisier et al.[185] reported that increased ESR indicated over a three-fold increase risk of a radiological damage score above the median at 10 years. Similar effect estimates were seen in other studies using similar analysis techniques, where an odds ratio (OR) of 2.7 (CIs not reported) was reported by Fex et al.[216], an OR of 2.9 (95% CI 1.01-5.88) reported by Tanaka et al.[192], and an Incidence Rate Ratio (IRR) of 2.0 (95% CI 1.4-3.0) reported by Bukhari et al.[194]. Using linear regression techniques, Lindqvist et al.[221] reported an average increase of 0.42 (95% CI 0.62-1.04) units of the Larsen Score for every one-unit increase in CRP. Similarly, Mustila et al.[222] reported that only ESR was significantly associated with radiographic joint damage at 12, 36, 60 and 84-months in univariate analysis, whereas RF was only significant at 36 months, and perinuclear Antineutrophil Cytoplasmic Antibodies (pANCA), Antikeratin Antibodies (AKA), Antiperinuclear Factor (APF) and Age were not associated at any time. Antinculear Antibodies (ANA) were not investigated in this study.

Anti-Cyclic Protein Antibodies (ACPA) and Rheumatoid Factor (RF)

The presence of ACPA, largely anti-CCP, was reported in 16 studies, with 14 reporting significant associations. Using linear regression, Lindqvist et al.[221] reported that patients positive for anti-CCP had on average an increase of 37 units on the Larsen score compared to anti-CCP negative patients over 10 years, while Nyhäll-Wählin et al.[207] reported an increase of 14.74 over 5 years. Anti-CCP positive patients were also reported to have between a 2.3 and 9.3 fold increase in risk of rapid radiological progression[201, 210].

The evidence for the predictive role of RF was reported in 21 studies, with 12 reporting statistical significance. Four studies investigating radiographic progression based on low or high radiographic damage groups indicated that patients positive for RF were between 1.8 and 2.8 times more likely to have high rates of long-term radiographic joint damage[77, 192, 194, 210].

To assess the relative strength of anti-CCP and RF, all studies reporting OR and their relative 95%CIs were entered into a random effects meta-analysis. Five of the 13 studies reporting Anti-CCP, and ten of the 21 studies reporting RF were included in the meta-analysis. Reasons for exclusion included insufficient data, lack of data on measures of variation, or did not calculate ORs. The overall pooled effect estimate for anti-CCP was 2.49 (95%CI 1.96-3.15) and for RF was 2.07 (95%CI 1.61-2.65) (See Figure 3.4). While this does suggest a moderate difference between the two markers, with Anti-CCP proving to be more strongly associated, the overlap of the 95%CIs would indicate this difference is non-significant. All five studies included in the meta-analysis for anti-CCP showed an increased risk, with only one reporting a non-significant result, which was also the only adjusted effect estimate included[218]. All but two of the studies included in the RF analysis reported an increased risk[205, 218].

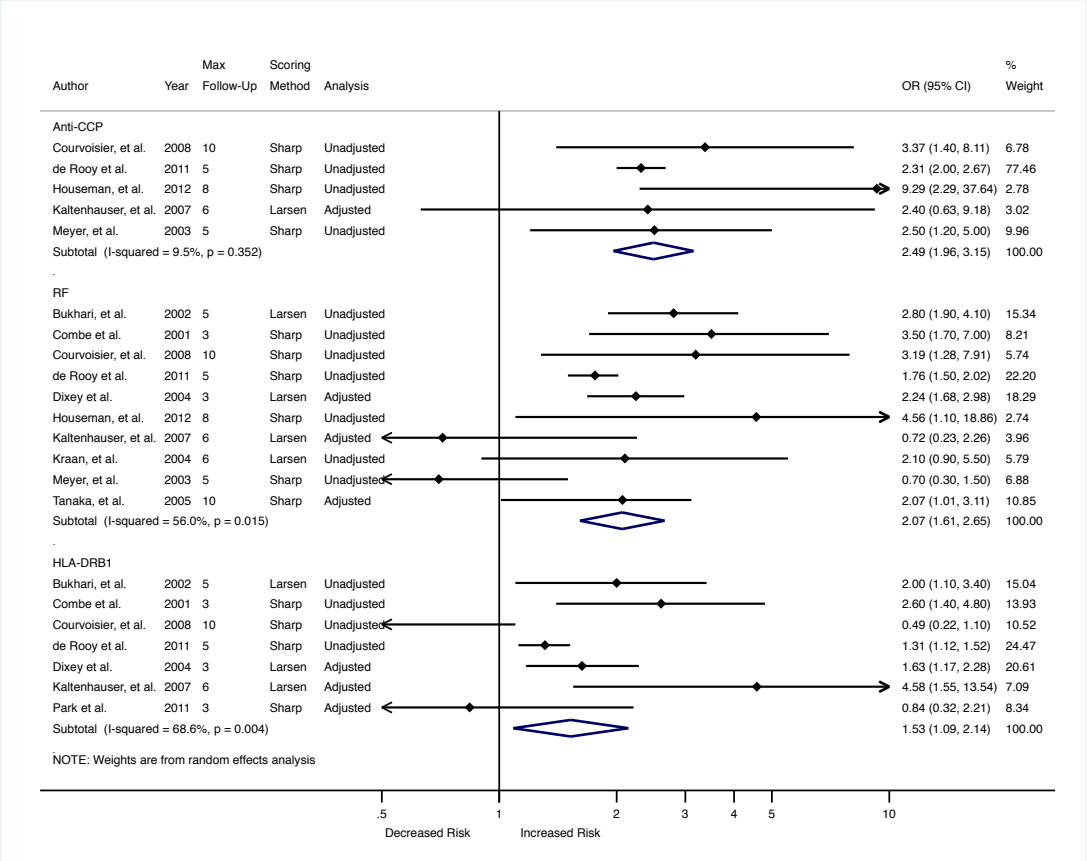


FIGURE 3.4: Forest Plot of RF, Anti-CCP and HLA-DRB1

Genetic factors

Of the 16 studies investigating the influence of genetic factors on radiographic progression, 12 found a statistically significant association. Four studies used follow-up data of at least 5 years,

while 12 were restricted to between 3 and 4 years follow-up. ORs for the presence of HLA-DRB1-SE ranged between 1.31 and 2.6[194, 195, 210]. Constantin et al. demonstrated in two studies that presence of HLA-DRB1 was associated with increased radiographic progression over 4 years[196, 197].

Seven of the 16 studies provided sufficient data for meta-analysis. A random effects model indicated an overall pooled estimate of 1.53 (95%CI 1.09-2.14) (see Figure 3.4). Two of the seven studies included reported a decreased risk[185, 208].

Other factors

The evidence for age and female sex as predictors of radiographic joint damage were limited, with only 4/12 and 4/15 studies reporting significant results respectively. The reported effect sizes of both age and sex was low, with age indicating a 1.14[210] and 1.2[194] increase in risk, and female sex indicating a 25% reduction in risk[210]. The paucity of studies in this review investigating swollen joint counts, tender joint counts, DAS, Matrix MetalloProteinase-3 (MMP-3) and functional disability makes it impossible to draw any conclusions about their impact on radiographic damage. Due to the lack of data provided by studies investigating these predictive markers, it was not possible to conduct a meta-analysis for these markers.

3.3.4 Quality Assessment

All studies included in this review were assessed for study quality through the use of the Downs and Blacks Quality Assessment Checklist[180] (See Table 3.7 and Figure 3.6). Generally the studies were of good quality, all reporting clear aims and objectives, outcome measures and recruiting representative patients. However, only 3 studies (6%) reported on missing data, and only 7 (15%) reported on any losses to follow-up over time. The use of appropriate statistical methods was variable, particularly in the 3-5 year follow-up predictive studies. Only 13 (27%) were using appropriate statistical methods, many of which used step-wise regression models, that is the process of systematically removing covariates that do not satisfy a pre-defined level of statistical significance, that can omit important confounders despite being statistically non-significant.

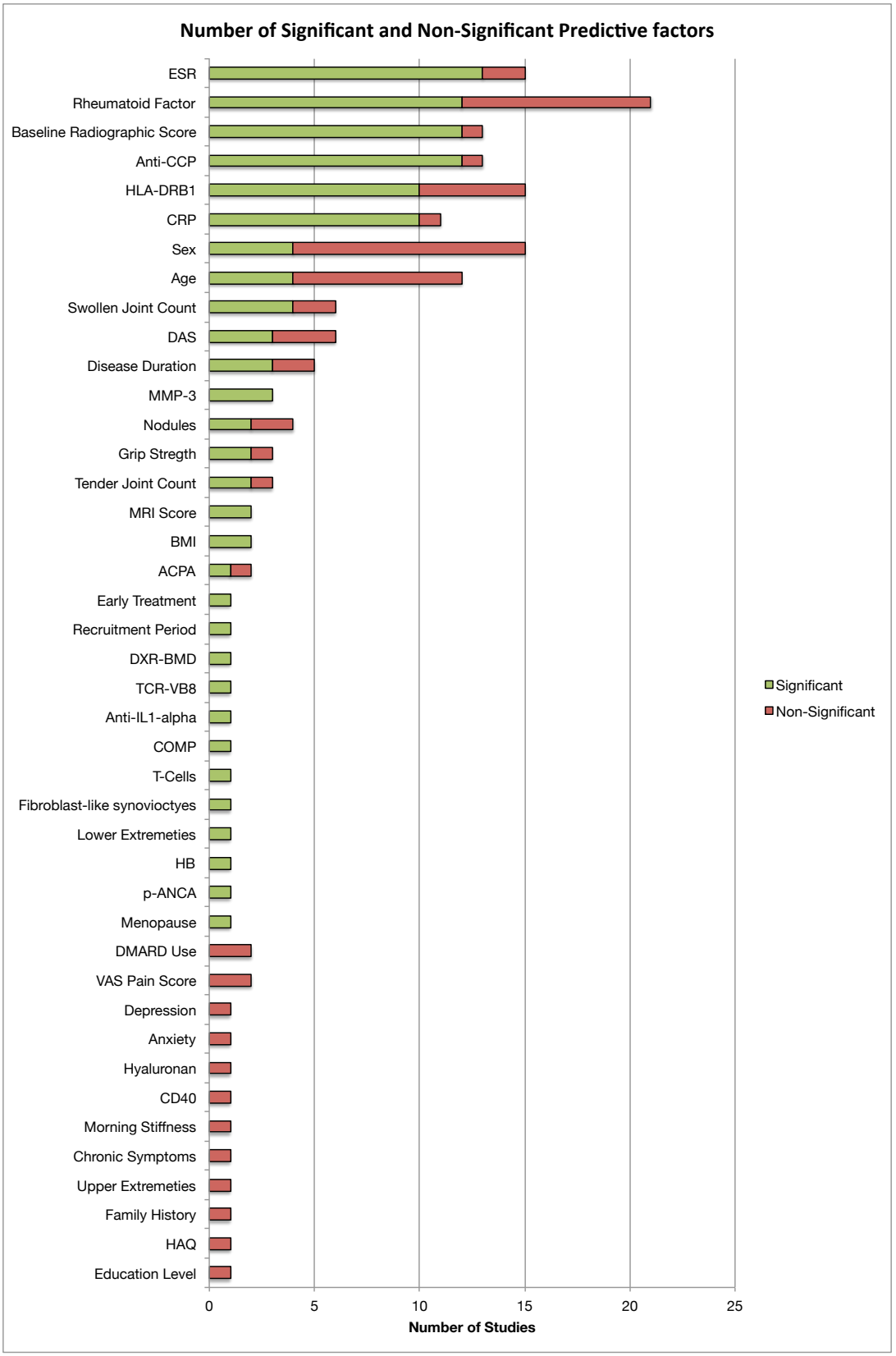


FIGURE 3.5: Histogram of Significant Factors identified and Number that were Significant

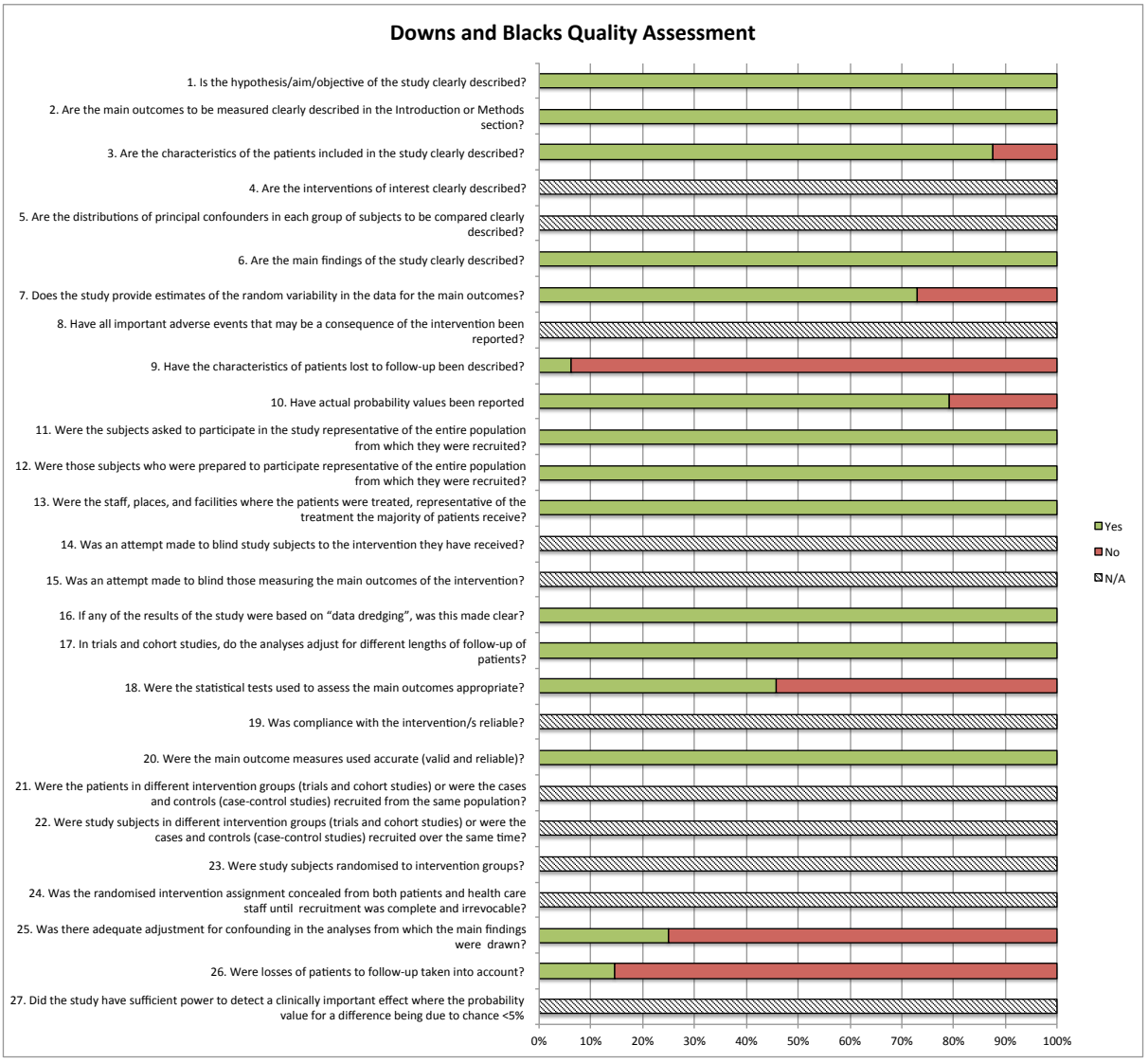


FIGURE 3.6: Figure of each item of the Downs and Blacks Quality Assessment

3.4 Discussion

Progression of radiographic damage

This review is the first to use meta-analysis techniques to provide more accurate estimates of overall radiographic damage in patients with early RA, both at presentation and over a 20 year period. Using data from 10 studies, the overall radiographic damage rate at presentation was 2.02% of maximum damage, and the overall annual rate of progression was 1.08% of maximum damage.

Previous reports[38] have estimated the total annual radiographic progression rate to be 1.9% of maximum damage, with the Larsen score progressing 3.8 units per year (2.5% of maximum

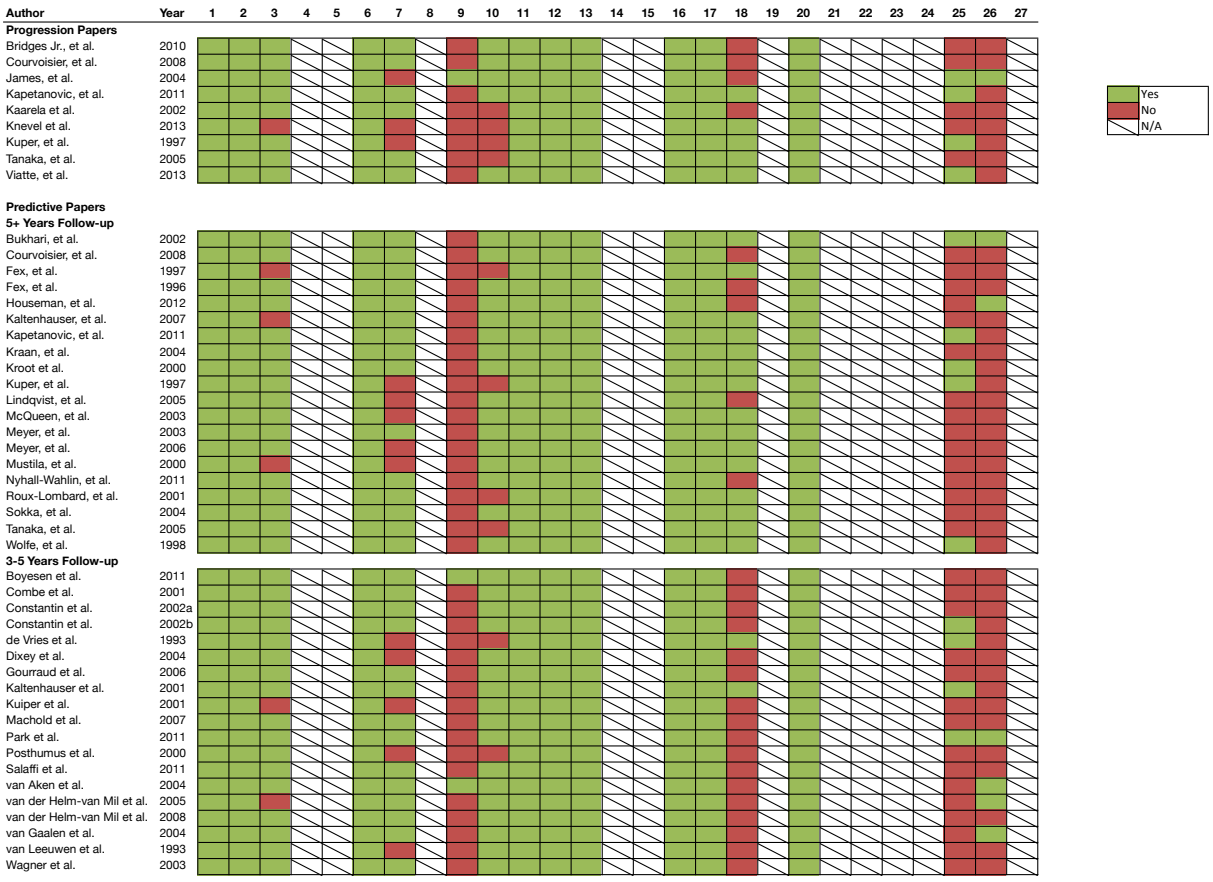


FIGURE 3.7: Table of each study assessed by the Downs and Blacks Quality Assessment

damage) and the SvdH score progressing 4.3 units per year (1.3% of maximum damage) over the first 15 years. This review provides relatively similar rates, at an overall rate of 1.08% (95% CI 0.72-1.44) of maximum damage, with Larsen scores progressing 2.76 units per year (1.38% of maximum damage), and the SvdH score progression at 4.03 units per year (1.20% of maximum damage) over the first 20 years of disease. The differences in rates between this review, and that conducted by Scott et al.[38] , are more than likely to be multifaceted. Firstly, meta-analytical techniques to calculate pooled effect estimates will produce different rates compared to methods relying only on averages. The former is a more robust method since studies with higher patient samples are given a higher weighting, thereby reducing the influence of less precise estimates derived from smaller studies, along with providing an estimate of precision (95%CI). Secondly, inclusion criteria for this review were different, i.e. only observational cohort studies recruiting patients with early RA. This ensured a more homogenous study sample, since patients included in RCTs are highly selected and likely to have higher levels of disease activity and higher rates of radiographic progression[183, 208], whereas this review describes data based on patients from 'true-to-life' clinical settings.

By stratifying studies by recruitment year, the annual rate of progression in studies recruiting between 1990-2000 was more than half the rate reported for those recruiting between 1965-1989. Baseline radiographic damage however, remained similar across the recruitment periods. The reduction in radiographic progression from 1965-2000 is concordant with data published by Finckh et al.[226], which indicated decreased radiographic progression rates from 1970 to 1990, and Sokka et al.[224], who found decreased 5-year radiographic progression rates across three cohorts spanning 1983-1985, 1988-1989 and 1995-1996. Finkch et al. argue that this is likely to be a consequence of increasingly more intensive therapies, since the effect over time was diminished once DMARD use was controlled for in their model[226]. More recent data from RCTs have demonstrated the impact of modern treatment on radiographic progression, with combination synthetic DMARDs and biologics proving to be effective in slowing the progression of radiographic joint damage[227], especially when used early, during the 'Window of Opportunity'[182]. The differences found in the two recruitment periods examined in this review also coincides with large changes in clinical management, with increased use of early, more intensive treatment in the 1990s, with methotrexate the anchor DMARD, as either monotherapy or in combination therapies[228]. Indeed, Pincus et al.[229] reported that not only radiographic related outcomes have improved from 1985 to 2000, but there have also been significant improvements in joint scores, functional capacity and mortality outcomes. The extent to which this is directly attributable to advances in treatment strategies is however difficult to determine, due to non-randomised study designs[224, 229].

While both the Larsen and SvdH score showed similar levels of annual progression when expressed as a proportion of maximum damage, it is worth noting the difference between the scoring methods at baseline. it is likely that these differences in baseline scores is due to differences in the way the scoring methods quantify both erosions and JSN, and ultimately the differences in the maximum value of both scoring methods.

Of interest is the paucity of observational cohort studies in more recent years, which is likely to be multifactorial. The advent of biologics resulted firstly in many more clinical trials of erosive damage and secondly led to the development of national biologic registers. In addition, fewer cohorts may report radiographic assessments.

Predictive factors of radiographic damage

The predictive factors found in this review are in agreement with review paper by Scott et al.[38], which found increased levels of acute phase markers and RF positivity to be the most consistent predictive markers. This review also found strong evidence for the association between anti-CCP positivity and long-term radiological damage.

In support of the predictive use of acute phase reactants, Navarro-Compán et al.[177] recently published a systematic review specifically on the relationship between radiographic joint damage and Disease Activity Indices (DAI), such as the Disease Activity Score (DAS). They found that, while DAIs are clinically useful, the individual components of the DAI's, particularly SJC and acute phase markers, were much more predictive of radiographic joint damage compared to DAI score itself. Interestingly in the current review, the extent to which measures of function and disease activity were used as covariates was relatively rare, with the majority of studies opting to examine individual components of disease activity measures, rather than the composite measure itself.

To the author's knowledge, this is the first review to summarise evidence of associations of anti-CCP and genetic factors with radiographic progression in long-term cohort studies. De Rooy et al.[210] found HLA-DRB1 shared epitopes to increase the risk of radiographic joint damage at 5 years, although this study did not include anti-CCP in the models. More recent studies[230, 231] have highlighted the importance on the dependence of RA related genetic markers on anti-CCP for associations with radiographic progression. Kaltenhauser et al.[218], reported that a combination of both anti-CCP and DRB1*04 SE as a compound marker was significantly associated with increased radiographic damage at 4 years and yielded the highest specificity, although Kroot et al.(26), found anti-CCP to be significantly associated, but not HLA-DRB4 in multivariate analysis. This evidence is suggestive of an association between SE-positive alleles and anti-CCP antibodies, however the exact pathogenetic mechanisms are unclear[218]. Further study of specific HLA-DRB1 haplotypes may determine a prognostic role[231], but at the present time, in the context of current clinical practice, the assessment of genetic markers does not provide much, if any, additional prognostic value, which is already available in an ordinary setting.

A number of studies included in the current review[205, 217, 218] found RF was not a significant predictor in the presence of anti-CCP, suggesting that anti-CCP may be a superior marker of long-term radiographic damage. This was further evidenced by our meta-analysis, which found

Anti-CCP to be more highly associated with increased risk of radiographic damage. Although, it should be noted that differences in specific RF antibodies and titre levels could explain variations between studies.

Given the large heterogeneity of the methods and analysis techniques used, it was not possible to conduct a formal meta-analysis that would allow a direct aggregation of these results for predictive markers. The most likely explanation for contrasting predictive factors is differences in study design[232]. When investigating novel markers in the absence of multivariate methods, the importance of well-established factors, such as seropositivity and acute phase reactants, are not appropriately accounted for[177]. In these cases, it is likely that the effect of the single marker has been masked, or over-exaggerated in the absence of these other, already established factors. Novel protein markers, such as MMP-3[201, 223], have been found from this review to have potentially strong associations with radiographic joint damage, but more evidence is needed, particularly with large patient samples using appropriate multivariate modelling techniques.

Strengths and limitations

This review is the first to apply meta-analytical techniques to investigate radiographic progression rates from a large pool of long-term observational cohort studies of patients with early RA. The methodologies used are novel and provide interesting new insights into the general progression of radiographic joint damage, and how they have changed over time. This review has shown evidence for the importance of multivariate analysis, and highlighted the potential strength of anti-CCP antibodies over RF for predicting both erosive damage and progression in medium to long-term observational cohort studies. It is unfortunate that while a large number of studies examining predictive markers have been conducted, the variability between studies in analytical methods made pooling the results of these studies inappropriate.

A further limitation is that it was not possible to stratify patients based on important disease markers, such as seropositivity, when modelling radiographic progression rates. The review on predictive markers presented in this paper highlights the difference in radiographic progression in patients with anti-CCP positivity. Producing separate radiographic progression rates for both seropositive and seronegative patients would have been more ideal, as it would allow the investigation of whether similar patterns that were found in the current review are also observed in these patient sub-groups.

It was also not possible to obtain substantive information on the impact of treatment. While the investigation into recruitment years provides a surrogate marker of changes in treatment practices, it was not possible to model the effect of treatment directly. Nevertheless, given the cohorts represent a 'true-to-life' clinical setting, those studies that recruit patients during the same years are likely to be following published guidelines on treatment regimes, and would therefore be largely similar.

Conclusions

In conclusion, the progression of radiographic damage has more than halved since 1990, with advances in treatment likely to be the cause. There is good evidence that anti-CCP positivity is a more consistent and a stronger marker of severe radiographic damage over RF positivity in the long-term, and increased markers of acute phase reactants have continued to be strongly associated with radiographic damage. In the context of clinical management, other novel antibodies need further study, and genetic data at present does not prove to be a cost-effective marker of disease that provides additional prognostic value independent of anti-CCP positivity.

3.5 Concluding remarks

The aim of this chapter was to review the current evidence base and identify common methods in measuring and reporting radiographic outcomes. Much of this chapter has provided the background for the primary analyses to be conducted in Chapters 6 and 7. The findings from the meta-analysis of progression rates indicate that radiographic progression has significantly fallen post 1990. This will be of interest when the progression of radiographic damage is modelled in both the ERAS and ERAN cohorts, representing roughly the same transition in time periods. The review has also highlighted the prominent role that seropositive RA has on increased radiographic progression, therefore the analysis in Chapter 6 will also explore seropositive and seronegative sub-groups to ascertain what impact this has on the long-term progression of radiographic progression in both cohorts.

Critically, this review has demonstrated that the statistical methods used to explore long-term radiographic data have been varied and often inappropriate. The next chapter of the thesis will therefore look at the methodological aspects of the thesis. Chapter 4 will introduce the ERAS and ERAN datasets in detail and explore the radiographic data contained within. Then Chapter

5 will look at the various statistical methods used in the studies examined in this systematic review and ascertain which models are most suitable for analysing longitudinal radiographic data collected in observational studies.

Author	Year	Sample	Cohort	Scoring Method (Max)	Max Follow-up	Data Used for Analysis	Data Type	Analysis Used	Multivariate
5+ Years Follow-up									
Bukhari, et al.	2002	439	NOAR	Larsen (200)	5 Years	Single time point	Continuous	Negative Binomial Regression	Yes
Courvoisier, et al.	2008	117	French Cohort	Sharp (448)	10 Years	Single time point	Binary	Logistic Regression	Yes
de Rooy et al.	2011	676	Leiden Cohort	Sharp (448)	5 Years	Change in score	Continuous	Linear Regression	Yes
Fex, et al.	1997	113	Lund Cohort	Larsen (200)	5 Years	Single time point	Continuous	Linear Regression	Yes
Fex, et al.	1996	113	Lund Cohort	Larsen (200)	5 Years	Change in score	Binary	Logistic Regression	Yes
Houseman, et al.	2012	58	Portsmouth Cohort	Sharp (448)	8 Years	Change in score	Binary	Logistic Regression	Yes
Kaltenhauser, et al.	2007	93	Leipzig Cohort	Larsen (200)	6 Years	Single time points	Binary	Logistic Regression	Yes
Kapetanovic, et al.	2011	183	Lund Cohort	Larsen (200)	10 Years	Single time point	Continuous	Linear Regression	Yes
Kraan, et al.	2004	36	Netherland/Australia Study	Larsen (200)	6 Years	Change in score	Binary	Logistic Regression	Yes
Kroot et al.	2000	273	Nijmegen Cohort	Sharp (448)	6 Years	Single time point	Continuous	Linear Regression	Yes
Kuper, et al.	1997	157	Nijmegen Cohort	Sharp (448)	6 Years	Single time point	Continuous	Linear Regression	Yes
Lindqvist, et al.	2005	157	Lund Cohort	Larsen (200)	10 Years	Single time points	Continuous	Linear Regression	Yes
McQueen, et al.	2003	31	Auckland Cohort	Sharp (448)	6 Years	Single time point	Continuous	Linear Regression	Yes
Meyer, et al.	2003	156	French Cohort	Sharp (448)	5 Years	Single time points	Binary	Logistic Regression	No
Meyer, et al.	2006	99	French Cohort	Sharp (448)	5 Years	Single time point	Binary	Logistic Regression	No
Mustila, et al.	2000	82	Helsinki Cohort	Larsen (210)	7 Years	Single time points	Binary	Logistic Regression	Yes
Nyhall-Wahlin, et al.	2011	191	BARFOT	Sharp (448)	5 Years	Single time points	Continuous	Linear Regression	Yes
Roux-Lombard, et al.	2001	24	Lund Cohort	Larsen (200)	5 Years	Change in score	Survival	Cox Regression Model	Yes
Sokka, et al.	2004	197	Jyväskylä Cohort	Larsen (100)	5 Years	Change in score	Continuous	Quantile Regression	Yes
Tanaka, et al.	2005	114	Japan Cohort	Sharp (448)	10 Years	Change in score	Binary	Logistic Regression	Yes
Welsing et al.	2004	185	Nijmegen Cohort	Sharp (448)	9 Years	Score at each time point	Continuous	Auto-regressive GEE	Yes
Wolfe, et al.	1998	256	Wichita Cohort II	Sharp (448)	19 Years	Change in score	Continuous	Linear Regression	Yes
3-5 Years Follow-up									
Boyesen et al.	2010	55	Diakonhjemmet Cohort	Sharp (280)	3 Years	Change in score	Continuous - Square Root	Linear Regression	Yes
Combe et al.	2001	172	French Cohort	Sharp (448)	3 Years	Single time point	Binary	Logistic Regression	Yes
Constantin et al.	2002a	96	Rangueil Cohort	Sharp (448)	4 Years	Change in score	Continuous	ANOVA	No
Constantin et al.	2002b	96	Rangueil Cohort	Sharp (448)	4 Years	Single time point	Continuous	ANOVA	No
de Vries et al.	1993	111	Nijmegen Cohort	Sharp (448)	3 Years	Change in score	Continuous - Square Root	ANOVA	Yes
Dixey et al.	2004	866	ERAS Cohort	Larsen (200)	3 Years	Single time point	Binary	Logistic Regression	Yes
Gourraud et al.	2006	144	Rangueil Midi-Pyrénées Cohort	Sharp (448)	4 Years	Single time points	Continuous	Rank Sum Tests	No
Kaltenhauser et al.	2001	48	Leipzig Cohort	Larsen (200)	4 Years	Change in score at each time point	Continuous	Mixed Effects Linear Regression	Yes
Kuiper et al.	2001	332	Nijmegen Cohort	Sharp (448)	3 Years	Single time point	Continuous - Square Root	Linear Regression	Yes
Machold et al.	2007	55	Austrian Early Arthritis Cohort	Larsen (168)	3 Years	Change in score	Binary	Logistic Regression	Yes
Park et al.	2010	184	CPR Cohort	Sharp (406)	3 Years	Change in score at each time point	Continuous	Cluster Analysis	Yes
Posthumus et al.	2000	33	Groningen Cohort	Sharp (448)	3 Years	Change in score	Continuous	Correlation Analysis	No
Salaffi et al.	2011	48	Italian Cohort	Sharp (448)	3 Years	Change in score	Continuous	Linear Regression	Yes
van Aken et al.	2004	153	Leiden Cohort	Sharp (448)	4 Years	Change in score	Continuous	Rank Sum Tests	No
van der Helm-van Mil et al.	2005	324	Leiden Cohort	Sharp (448)	4 Years	Single time points	Continuous	Chi-Square	No
van der Helm-van Mil et al.	2008	488	Leiden Cohort	Sharp (448)	3 Years	Single time point	Continuous	Linear Regression	Yes
van Gaalen et al.	2004	268	Leiden Cohort	Sharp (448)	4 Years	Single time point	Continuous	Linear Regression	Yes
van Leeuwen et al.	1993	110	Groningen/Nijmegen Cohort	Sharp (448)	3 Years	Single time points	Continuous	ANOVA	Yes
Wagner et al.	2003	77	Leipzig Cohort	Larsen (200)	4 Years	Single time point	Binary	Logistic Regression	Yes

TABLE 3.2: Summary table of studies included with predictive factors of long-term radiographic progression

Chapter 4

Methodology I - Data

4.1 Introduction

The aim of the following two chapters of the thesis is to detail the methodology behind the primary analyses conducted in Chapters 6 and 7. This chapter marks the first part of this methodology section, and provides an in depth look at the observational cohorts, which collected the radiographic data used to conduct the analyses. The second part of the methodology section detailed in Chapter 5 looks at the statistical models that will be used to analyse this data appropriately.

The chapter will begin by providing an overview of both the observational cohorts that provided the primary data for this thesis; ERAS and ERAN. The main focus of this chapter will be on the radiographic data collected by both these cohorts, including the methods used to score this data. Specific properties of the radiographic data will then be explored, as understanding the data in detail will inform the statistical methods used, and is crucial in interpreting radiographic outcomes. Importantly, since each cohort had the radiographs read by individual readers, the level of agreement will be assessed by Bland and Altman plots and measures of inter-reader reliability. This will enable the data from both cohorts to be combined for subsequent analyses. Finally, the chapter will look at the extent of missing data for the radiographic outcomes and discuss what limitations and implications this has on the statistical methods used.

4.2 ERAS and ERAN

Both ERAS and ERAN were briefly introduced in Chapter 2, outlining some of the more general aspects of the observational studies. While conceptually identical, there were a few methodological differences between them. By study close, ERAS had recruited 1,465 and ERAN 1,236. While ERAS recruited patients at first presentation to 9 rheumatology outpatient clinics across the UK, ERAN recruited from 23 centres, two of which were shared between the two cohorts. These geographical locations were specifically selected to represent as much of the UK population as possible, including rural, urban and ethnically diverse populations. The maximum follow-up for ERAS was 25 years (median 10 years), although many centres opted to stop follow-up at 10 years in line with the studies original aim. As ERAN began much later, maximum follow-up is 10 years (6 years). Patients were enrolled into the ERAS if they satisfied the 1987 ACR criteria for RA, had a disease duration of <2 years, and no previous treatment with DMARDs, whereas patients in ERAN needed to satisfy the 1987 ACR criteria, had a disease duration of <3 years and some patient had received some treatment prior to first visit.

The data collected on these patients were a myriad of demographic, standard clinical and laboratory measures; including age, sex, disease activity measured using the DAS 44-joint count method in ERAS[64] and the DAS28 method in ERAN[65], functional disability measured using the HAQ[233], biomarkers, such as acute phase markers (ESR), RF status and haemoglobin (HB), and co-morbidities. Alongside these, more specialised outcome measures not often seen in observational and clinical trials in early RA were also collated, such as work disability, psychological wellbeing (Hospital Anxiety and Depression Scale (HADS)[234] and Short-Form 36 (SF-36)[235]) and orthopaedic surgical outcomes. Both cohorts also collected yearly radiographs on all patients, which were later assessed using both the Larsen scoring method[106] and SvdH method[101] in ERAS, and the SvdH method in ERAN. The combination of both ERAS and ERAN provides a powerful database in which data from 2,701 patients with maximum follow-up of 25 years is available. It documents the changes in clinical management since the 1980s with the increased used of methotrexate, to the adoption of earlier and more intensive therapies in the 1990s, and finally the beginnings of the biologic era in RA in the early 2000s.

	ERAS (1986-2001)	ERAN (2002-2013)	**Total**
Demographics			
Age (Mean (SD))	55.3 (14.6)	57.1 (14.0)	56.1 (14.4)
Female (%)	66	68	67
Clinical Markers			
RF+ (%)	63	60	62
Baseline HAQ (Median (IQR))	1.00 (1.25)	1.00 (1.13)	1.00 (1.13)
Baseline ESR (Mean (SD))	42.2 (28.8)	30.3 (24)	37.2 (27.5)
Baseline SJC* (Mean (SD))	15.5 (9.46)	5.94 (5.74)	- (-)
Baseline TJC* (Mean (SD))	11.7 (8.67)	7.25 (6.89)	- (-)
Baseline PGA* (Mean (SD))	44 (26.4)	43.4 (25.6)	- (-)
Baseline Low HB (%)	41	28	35
Months to First Visit (Median (IQR))	6 (7)	6 (9)	6 (8)
Observations	1465	1236	2701

Note: *SJC, TJC and PGA measured differently in ERAS and ERAN

TABLE 4.1: Summary Statistics for the ERAS and ERAN cohorts

4.3 Baseline differences between the cohorts

ERAS and ERAN reflect different eras of RA management. As such, improvements in healthcare nationwide alongside improvements in daily living have led to some notable changes in the characteristics of patients presenting in both cohorts. Some of these differences can be seen in Table 4.1. Patients in ERAN tended to present slightly older, reflecting the increasingly ageing population in the UK. There is some evidence that patients in ERAN were presenting with less severe forms of RA, with slightly lower proportions of RF positive patients, and marginally lower ESR scores at baseline. However, direct comparisons using the DAS is made impossible due to differences in the methods used to measure SJC and TJC. Although, of particular note is the marked decrease in the proportion of patients in ERAN presenting with low HB compared to ERAS. This was defined as a HB count of <12 for men and <13 for women.

Treatment strategies used for patients in both ERAS and ERAN reflect the common practices of RA management in the UK at the time. Patients in ERAS were typically treated with conventional DMARDs in combination with steroids. DMARDs were commonly prescribed in sequential monotherapy, with those with more severe disease treated using step-up combination therapy. The 1990s saw two major advances in the treatment of RA[236]. The first was the adoption of a pyramidal treatment strategy, whereby more aggressive treatments were prescribed earlier, and the second was the increased use of methotrexate as the first line DMARD of choice.

These changes in UK treatment guidelines were reflected towards the later stages of ERAS and the early stages of ERAN.

Figure 4.1 indicates similar levels of patients on monotherapy treatment, however there is a marked increase in the use of DMARD add-on and the use of combination DMARD in both double and triple therapy. ERAN also documents the first cases of early RA patients treated with biologic DMARDs in the UK.

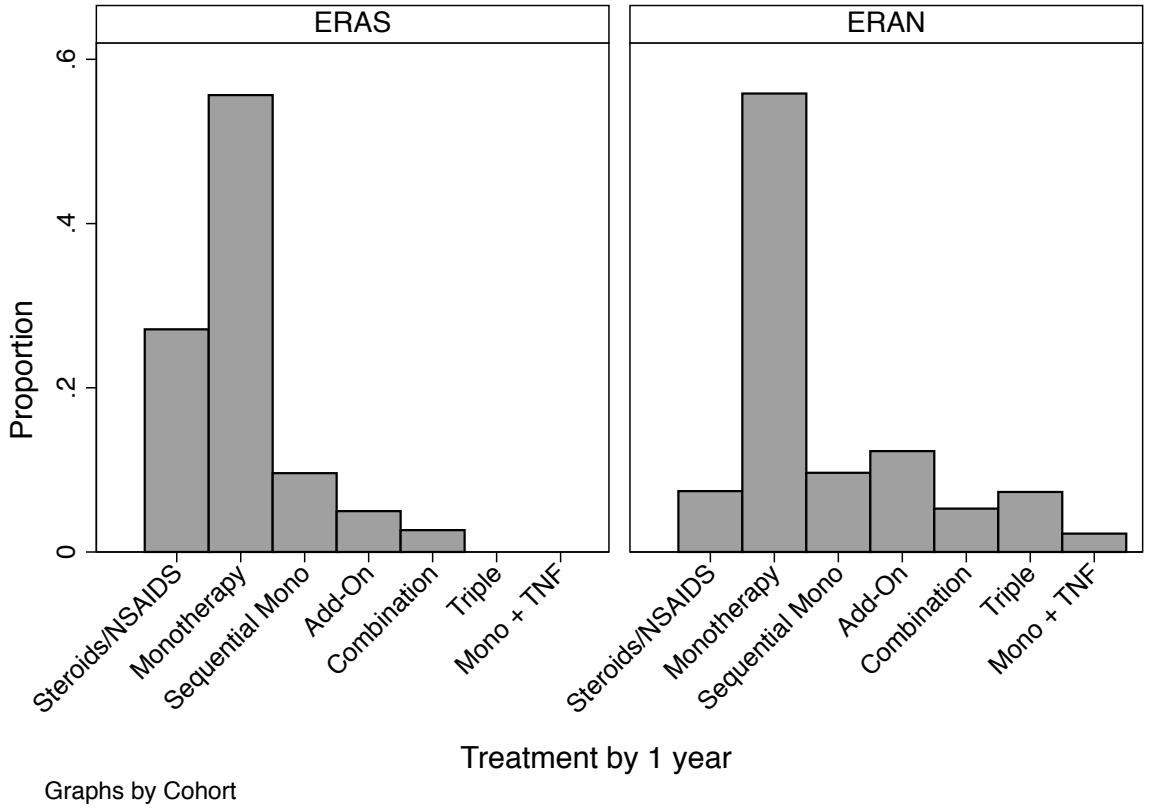


FIGURE 4.1: Prescription of DMARDs in ERAS and ERAN over the first 12 months

Figure 4.2 then highlights the dramatic increase in use of methotrexate as the first DMARD of choice in ERAN. In ERAS, the most popular DMARD of choice was sulphasalazine, followed by methotrexate and intramuscular gold.

4.4 Radiographic data

Radiographic data was recorded in ERAS using the SvdH and Larsen scoring method, while ERAN data was scored using only the SvdH method. These methods were explained in detail

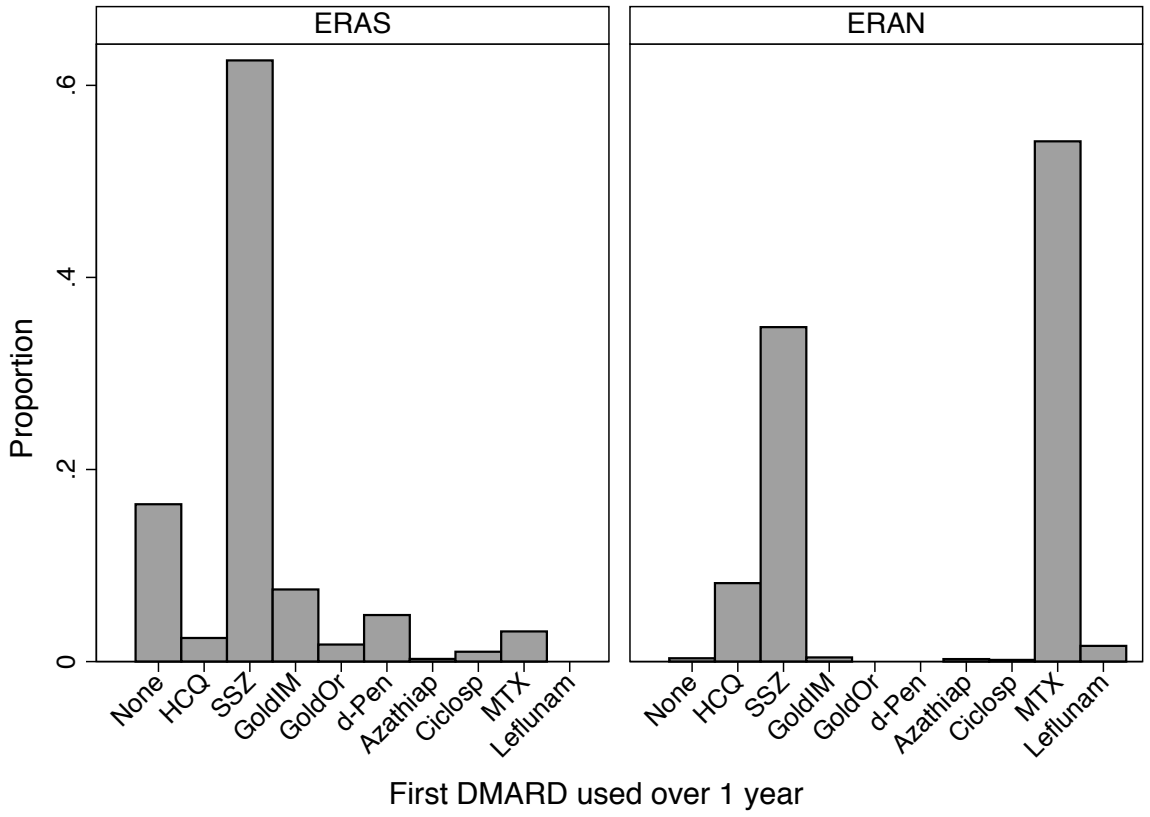


FIGURE 4.2: First DMARD of choice in both ERAS and ERAN over the first 12 months

in Chapter 2, but briefly the Larsen score evaluates 42 joints in the hands and feet, and ranges from 0 to 200. The score incorporates both the prevalence and severity of the erosions and JSN in one total score. The SvdH score evaluates the same number of joints, and assess the JSN (0-168) and erosion score (0-280) as separate scores. The total score ranges from 0 to 448.

For ERAS, a total of 1,234 patients had SvdH data and 1,157 patients had Larsen data at any time between baseline and 9 years follow-up, while for ERAN, a total of 447 had SvdH data at any time from baseline to 9 years follow-up. A box plot, indicating the total SvdH over the first 9 years for both ERAS and ERAN, is depicted in Figure 4.3 and shows the total SvdH score increasing at a linear rate over time in ERAS. In contrast, the total SvdH score in ERAN does not indicate a similar linear progression over time, with the median score increasing at a much slower rate.

The Larsen score over the first 9 years for the ERAS cohort is depicted in Figure 4.4. Much like the total SvdH score for ERAS, the Larsen score shows a similar trend over time, with the median Larsen score increasing at a fairly linear rate. Comparisons between Larsen and SvdH

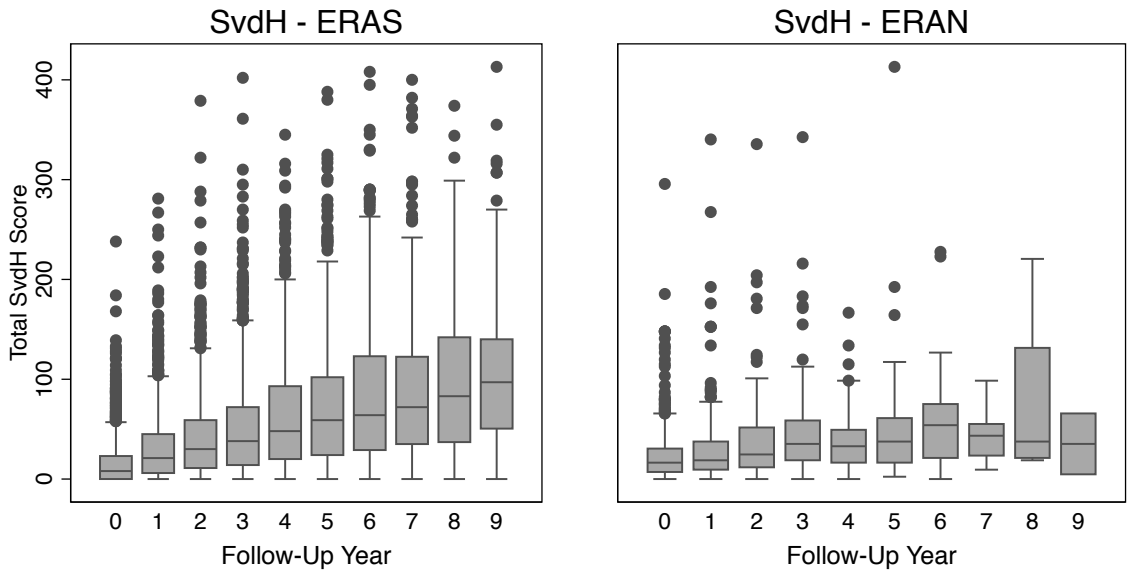


FIGURE 4.3: Box plot of Total SvdH score for ERAS and ERAN over first 9 years follow-up

in the ERAS cohort are difficult given the difference in the way erosions and JSN are scored by the methods, as well as differences in the maximum scores (Larsen = 200, SvdH = 448).

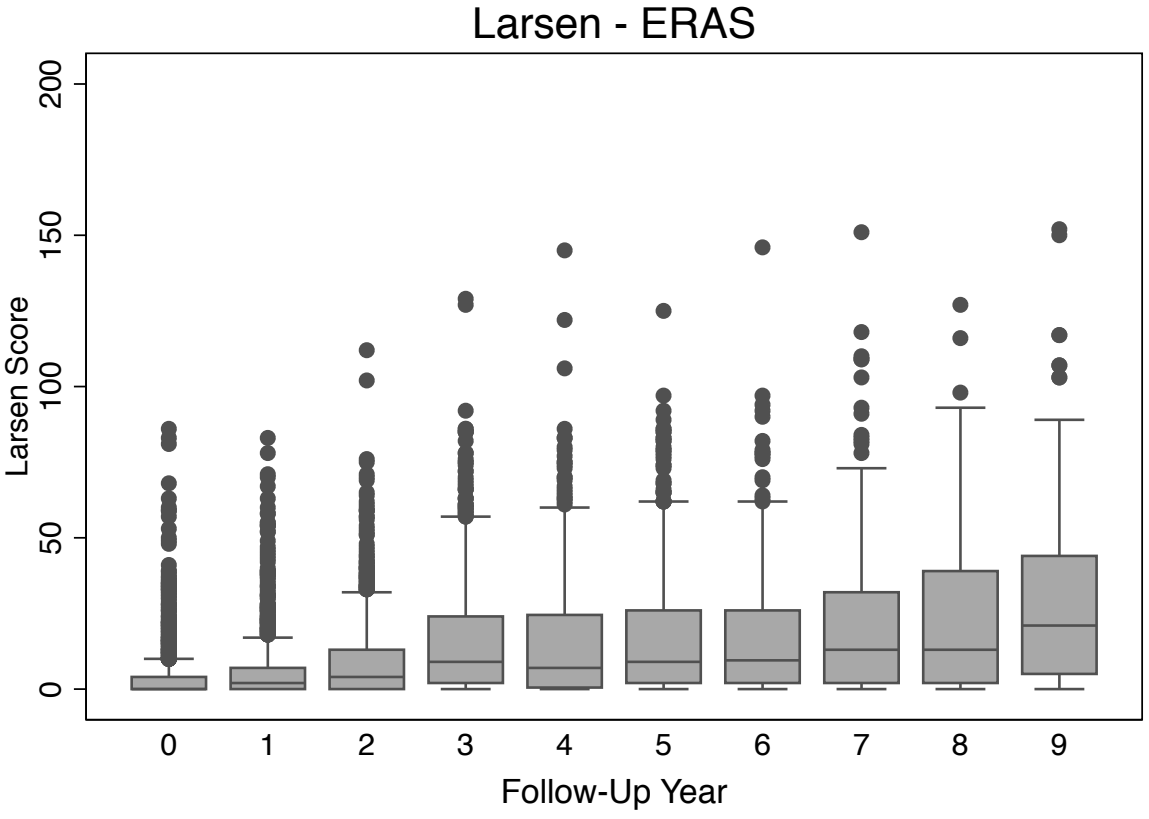


FIGURE 4.4: Box plot of Larsen score for ERAS and ERAN over first 9 years follow-up

Investigation of both the mean and median scores indicate large differences between the two across all follow-up measures. Figure 4.5 highlights not only the differences in the progression of both ERAS and ERAN seen in Figure 4.4 above, but also that the mean score is consistently estimated to be higher than the median score. Investigation of the Standard Deviation (SD) also highlights the large variation around the summary estimate, where the SD is often higher or equal to the mean estimate.

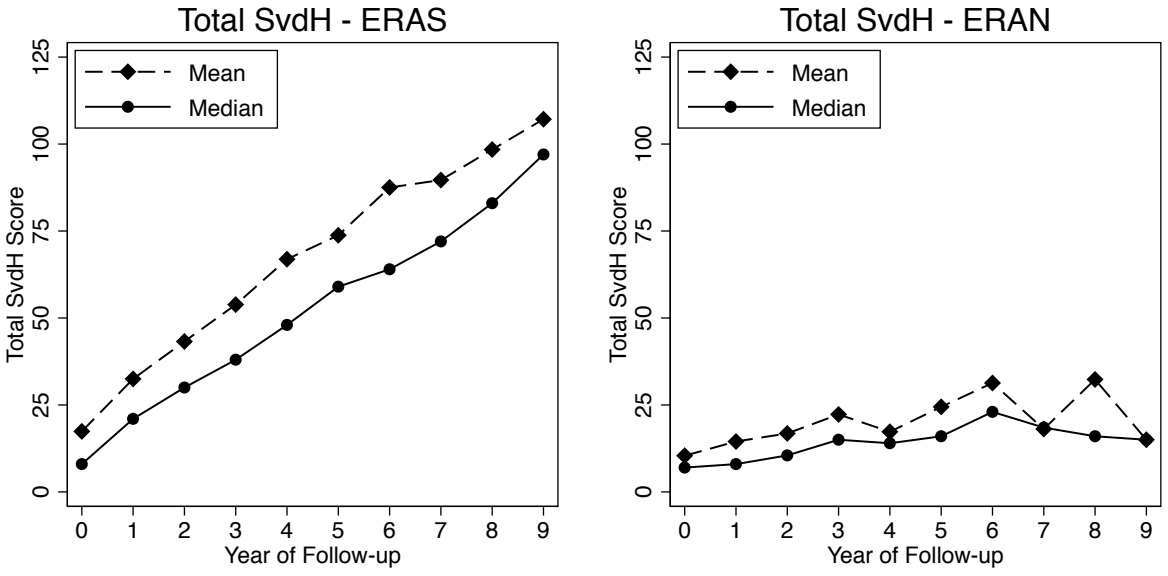


FIGURE 4.5: The mean and median score over the first 9 years for the Total SvdH score in ERAS and ERAN

These differences in summary estimates, along with the high SD relative to the mean, are suggestive of highly non-normal data distributions. Understanding the properties of the data, including its underlying data distribution, is pivotal in the interpretation of any summary statistics, as well as guiding the decision about which statistical methods to use to model the data. As such, the next section will look to explore the data distribution of radiographic outcomes in more detail.

4.4.1 Data distribution

The histogram in Figure 4.6 depicts the data distribution for the baseline SvdH scores for patients in the ERAS cohort. The solid red line indicates the mean score at 17.4 (SD 24.7). This has a variance of 608.4, nearly 35 times higher than the mean. The histogram highlights that the majority of the data lies on the left, with many patients presenting with no radiographic damage at baseline. The distribution then shows a large tail on the right, highlighting the high

positive skew of the data (also driven by the large range of scores seen on the SvdH scoring method). This high positive skew is denoted by a large positive skewness value of 2.9, confirming the asymmetry of the data distribution shape. The distribution also estimates a high kurtosis score of 15.3, highlighting the impact of the extreme values across the range of the score. The median is shown on the figure as a dashed red line and is reported at 8 (IQR 23), again showing the difference between these two summary estimates.

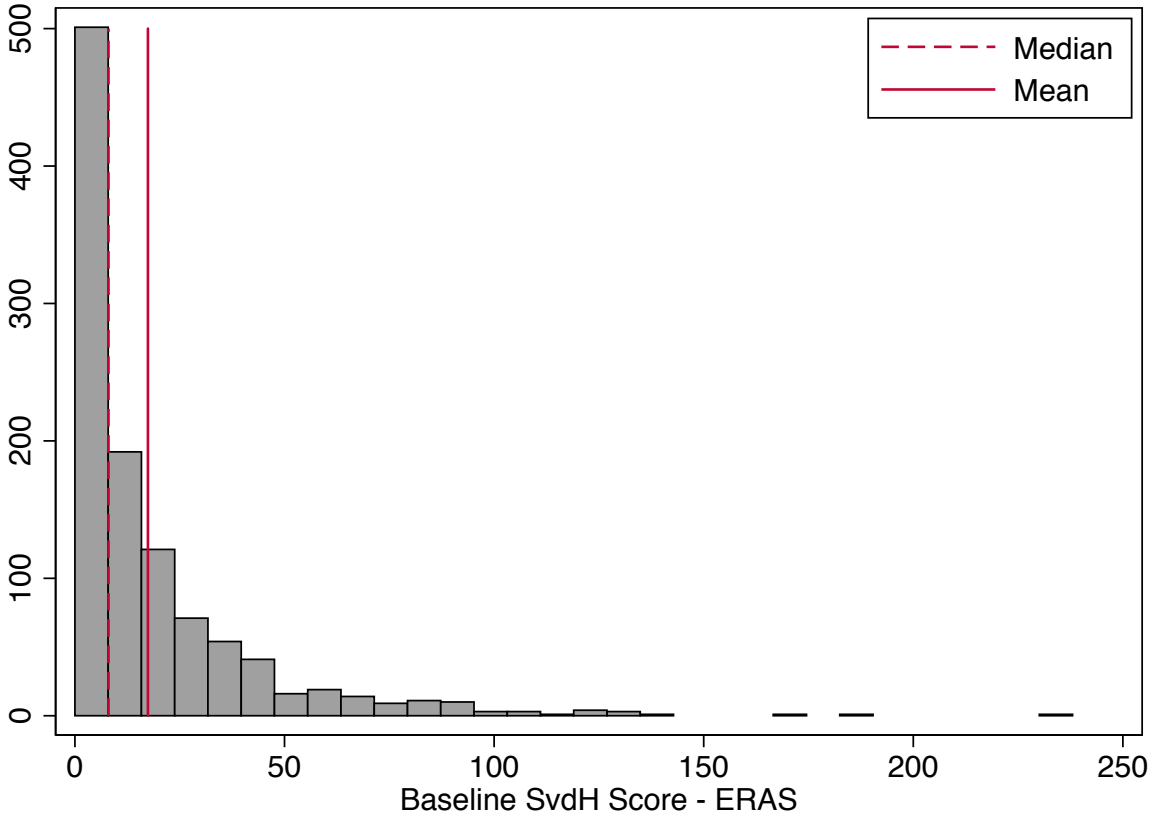


FIGURE 4.6: Histogram of Baseline Larsen Scores from ERAS. Solid red line indicates the mean, the dashed red line indicates the median

The fact that the median score lies closer to the left hand side, where the majority of the data lies, has led to many studies advocating the use of the median, rather than the mean, when summarising radiographic data[183, 237]. While the data is not shown, similar data distributions are seen for the Larsen score in ERAS, and the SvdH scores in ERAN. These data distributions are also consistent across the different follow-up measures.

4.4.2 Agreement analysis

For the SvdH score, one reader scored the x-rays in ERAS (Keeran Jayakumar) and one reader scored the x-rays in the ERAN cohort (Daniel McWilliams). To ensure good internal validity of the score between both cohorts, it is important to test the level of agreement between the two readers. An agreement analysis was conducted whereby the scorer from ERAN (DM) scored 40 radiographs from the ERAS cohort. A sub-sample of 20 patients from ERAS were chosen at random, and both their radiograph at 1 year and 5 years were assessed. Due to missing data from either the erosion or JSN score as a result of unreadable radiographs, only 25 total scores were analysed.

To assess levels of agreement for these radiographs Bland and Altman plots were used[238]. These plots graph the mean score between the two readers against the difference between the two readers with an incorporated trend line to highlight changes over the scale of the score[239]. This was calculated for the total SvdH score, JSN score and erosion score.

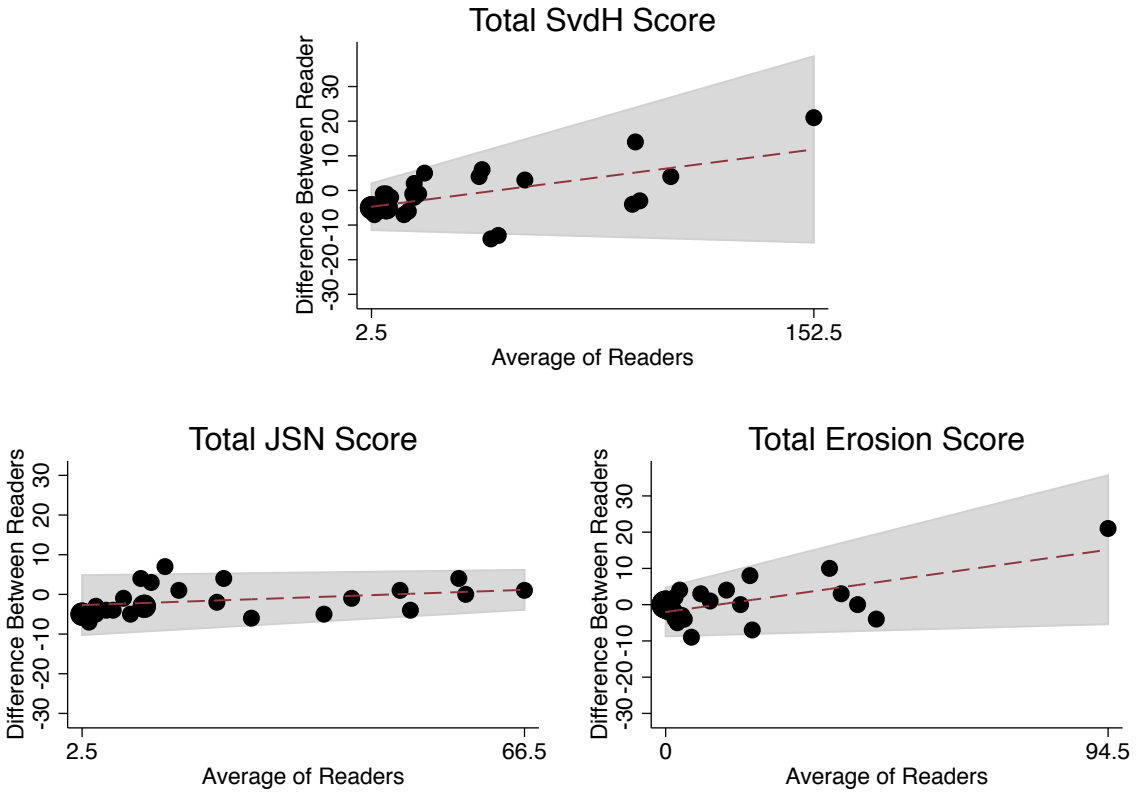


FIGURE 4.7: Bland and Altman plots for the total SvdH, JSN and erosion score. The shaded grey area indicates the 95% confidence intervals, while the dashed red line shows the mean difference between the readers

The Bland and Altman plots for the total SvdH, JSN and erosion score are given in Figure 4.7. For the total SvdH score and erosion score, there is a small increase in the difference as the average score increases. This suggests that agreement between the two scorers decreased at the higher ranges of the radiographic score. For the total JSN score however, the level of agreement was uniform across the range of the radiographic score.

A common method of measuring the agreement between two continuous measures is to calculate the Intra-class Correlation Coefficient (ICC). A Pearson correlation coefficient, which is typically used to summarise the correlation between two variables, would not be appropriate in the context of agreement. A situation where two readers consistently score the same radiograph with a difference of 5 units would indicate a near perfect Pearson correlation, but far from perfect agreement[240]. The Pearson correlation is insufficient because it does not estimate both the within-subject (intra-observer), and the between subject (inter-observer) reliability. In contrast the ICC estimates both the within and between subject variability, and can therefore provide a meaningful coefficient to indicate the level of agreement between two readers. Much like the Pearson coefficient, a value of 0.3, 0.5 and 0.8 indicate a low, moderate and high effect size [241].

For the erosion, JSN and total score, the ICC was estimated at 0.95, 0.98 and 0.98 respectively. This represents a very high level of agreement, indicating that any error between the two scorers is likely due to measurement error alone, rather than systematic bias between the two scorers.

4.4.3 Missing Data

Missing data is inherent in both RCTs and observational cohorts[242]. This can occur for a variety of reasons; for RCTs, this is likely to be due to patient drop-outs over the follow-up. In the context of an observational cohort missing data can occur due to missed clinical visits, patients moving away, being referred to another hospital or death. Specific to radiographic data, missing data can also be caused by missing x-ray records, lack of radiographs at follow-up visits and unreadable radiographs.

The proportion of missing data for the Larsen score in the ERAS cohort was 27%, 35%, 48%, 72% and 74% for baseline and years 3, 5, 7 and 9 respectively. For the SvdH score, the proportion of patients with missing data in ERAS was 26%, 37%, 52%, 73% and 89% and in ERAN was 58%, 79%, 82%, 99% and 99% for baseline and years 3, 5, 7 and 9 respectively. The extent of

missing data for the SvdH score and the Larsen score in ERAS is depicted in Figure 4.8, while the extent of missing data for the SvdH score in ERAN is depicted in Figure 4.9. This is shown for baseline and follow-up years 1 to 9. Radiographs were collected from all 9 recruiting centres in ERAS and all 23 recruiting centres in ERAN, however radiographs were only scored from 6/23 centres recruiting patients to ERAN. Investigation of those patients recruited from those 6 centres versus the entire cohort indicated no significant difference in demographic or baseline clinical characteristics.

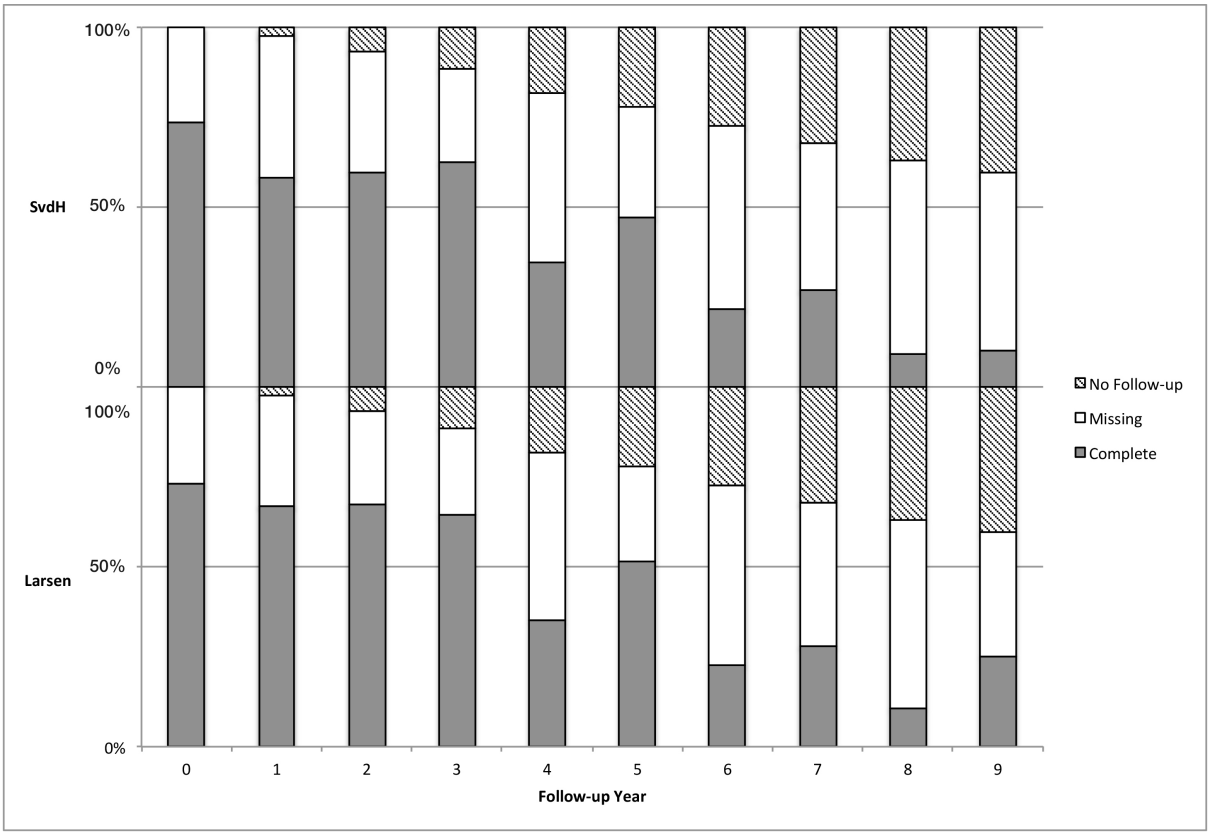


FIGURE 4.8: Missing data for the total SvdH and the Larsen score in ERAS over the first 9 years follow-up

4.5 Discussion

This chapter has provided an in depth overview of the ERAS and ERAN cohorts, along with the radiographic data collected. It has shown how radiographic outcomes typically produce very skewed data distributions, largely due to a large number of patients with early RA experiencing no radiographic damage. This results in a high number of zero scores. The effect of this high positively skewed distribution is that there is a large discrepancy between the mean and median

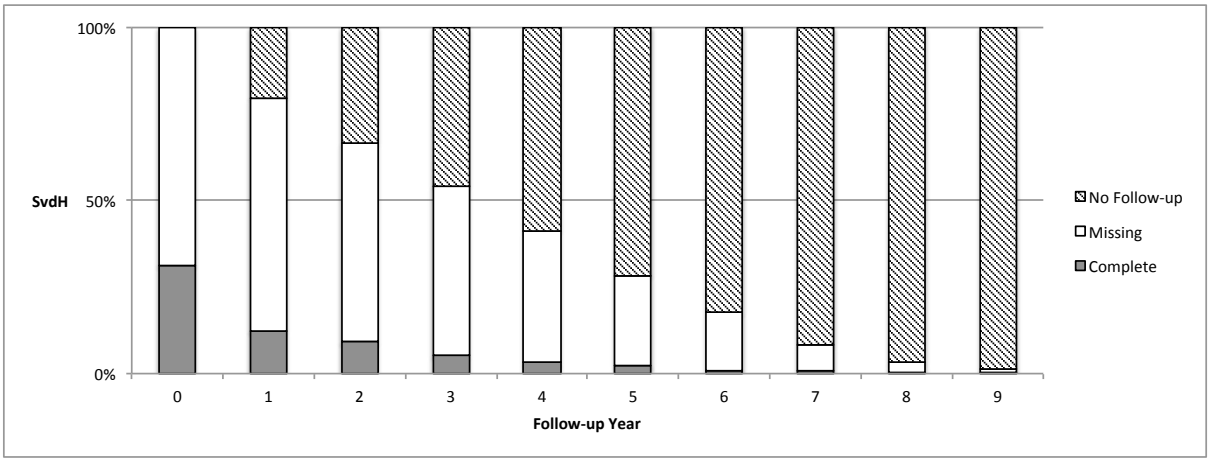


FIGURE 4.9: Missing data for the total SvdH Score in ERAN over the first 9 years follow-up

summary estimates, along with very large values for the variance, often 20 to 30 times larger than the mean. Therefore, the median rather than the mean provides a more accurate summary estimate of the data.

Investigation of the SvdH data from both ERAS and ERAN indicate a tendency for a linear increase over the first 9 years in ERAS, with a much slower progression in ERAN. While the agreement analysis conducted does indicate that combining the data from both the ERAS and ERAN cohort is appropriate, the large amount of missing data from the ERAN cohort does restrict the statistical power to evaluate trends beyond 5 years in the ERAN data.

The implications of the data distribution on which statistical methods should be used to model radiographic data will be explored in detail in the second part of the methodology section in Chapter 5. It will evaluate common statistical methods used in the literature and look at which model is best placed for accounting for the unique data distribution created from radiographic outcomes. Finally, it will look at the impact of the missing data in longitudinal data and evaluate methods for handling this missing data to reduce the bias that can be introduced.

Chapter 5

Methodology II - Statistical Models

5.1 Introduction

Radiographic data produced from scoring methods such as the Larsen and SvdH methods are well known for producing skewed data distributions[126, 243, 244]. This was highlighted in the first part of this methodology section (Chapter 4), which explored the radiographic data from the ERAS and ERAN cohorts. Radiographic scores are more accurately defined as a semi-continuous weighted count score, where each affected joint is counted, and then weighted by the severity of the erosions or JSN present. This produces skewed distributions with high frequencies of zero scores, due to a large number of patients in early RA presenting with no observable radiographic damage. The non-normality of these radiographic outcomes results in a wide variety of statistical methods being used in the literature, as was evidence in the systematic review conducted in Chapter 3. Often the methods used are inappropriate as they do not account for the unique properties of the radiographic data and violate fundamental assumptions.

In the context of secondary analyses, large heterogeneity makes it difficult to compare and contrast results. Furthermore, while there rarely is a 'right' approach to statistical analyses, there are inherent advantages and disadvantages to the different analysis methods used. Understanding these is imperative, and choosing the most appropriate method to analyse data ensures more precise and less biased estimates. This will not only allow for a more coherent understanding of the underlying effect, but general agreement of the most suitable analysis methods will also reduce the heterogeneity between studies to allow for more effective secondary analyses.

The objective of this chapter is to address the second aim of the thesis and determine which statistical methods are best suited to modelling longitudinal radiographic data. To begin, the chapter will first explore some of the methods used from previous studies identified from the systematic review conducted in Chapter 3. Using the information gained from Chapter 4 on the specific properties of radiographic data from the ERAS and ERAN cohorts, the chapter will begin by looking at the general linear models and why they are not suitable for use on radiographic data due to the non-normal data distributions they create. The chapter will then look at Generalised Linear Models (GLM) and the methods that can be used to account for this non-normality, including logistic regression and two main types of count regression; Poisson and Negative Binomial (NB). The chapter will focus specifically on count regression methods and their suitability to modelling radiographic data. The chapter will then look at methods for modelling longitudinal data, and introduce the concept of multi-level models to analyse data over time. Finally, the issue of multivariate regression techniques and missing data will be examined. The ultimate aim of this chapter is to provide a statistical framework, from which the primary analysis conducted in Chapters 6 and 7 can be based on.

5.2 General Linear Models

Regression analysis is commonly used in the medical literature to develop prognostic models, whereby the researcher wishes to be able to predict a value of a dependent variable based on the value of a/several independent variable(s). General linear models is one form of regression analysis, which is used when the dependent variable being modelled is a continuous outcome. If only one independent variable is specified in the model, it is referred to as univariate regression, and if more than one independent variable is used it is referred to as multivariate regression. For a general linear model to be deemed suitable it must satisfy a range of pre-defined assumptions concerning the data. The general linear model assumes that: 1) the dependent variable is a continuous outcome, 2) the relationship between the dependent and independent variable(s) is linear, 3) each observation is independent, 4) the variance is homoscedastic (i.e. variance is constant at each value of the dependent variable) and , 5) the residuals (predicted - observed) are normally distributed.

A simple linear regression with one independent variable (univariate) can be denoted using the formula given in Equation (5.11), where y_i is the dependent variable of the ith observation,

α denotes the intercept, β denotes the slope, x denotes the independent variable of the ith observation and ε denotes the error term.

$$y_i = \alpha + \beta x_i + \varepsilon \tag{5.1}$$

This equation can be extended to multivariate linear regression through the formula given in Equation (5.2), where β_1, β_2, and β_x is the 1st, 2nd and xth covariate in the model.

$$y_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_x x_i + \varepsilon \tag{5.2}$$

A linear regression will estimate a coefficient for each covariate in the model, which is interpreted as the change in the dependent variable given a 1-unit increase in the independent variable.

While it is not a requirement for the outcome to be normally distributed for linear regression, data that follows a normal distribution is favourable, since skewed data distributions and extreme outliers can have a large impact on the normality of the residuals. Ensuring the residuals are normally distributed is one of the major assumptions, as it can have a large impact on the calculation of confidence intervals, and therefore the reliability of the significance test. In cases where data are highly skewed, and the residuals are found to be highly non-normal, other regression techniques should be used, which assume some other data distributions that are non-normal.

5.2.1 Linear regression for radiographic outcomes

As was seen in Chapter 4, radiographic outcomes produce highly positively skewed data distributions. It is therefore likely that linear regression techniques will not be suitable to this data. This is demonstrated in Figure 5.1, which shows the observed frequency of the baseline Larsen and SvdH scores (black line with black crosses), along with the predicted probability density from a general linear model (red line).

As expected, the general linear model fails to adequately model for the large number of zero scores, and the predicted probabilities tend to be over estimated across the range of the score. Furthermore, examination of the variance and the distribution of the residuals indicates that the model violates two key assumptions; homoscedascity of the variance and normally distributed

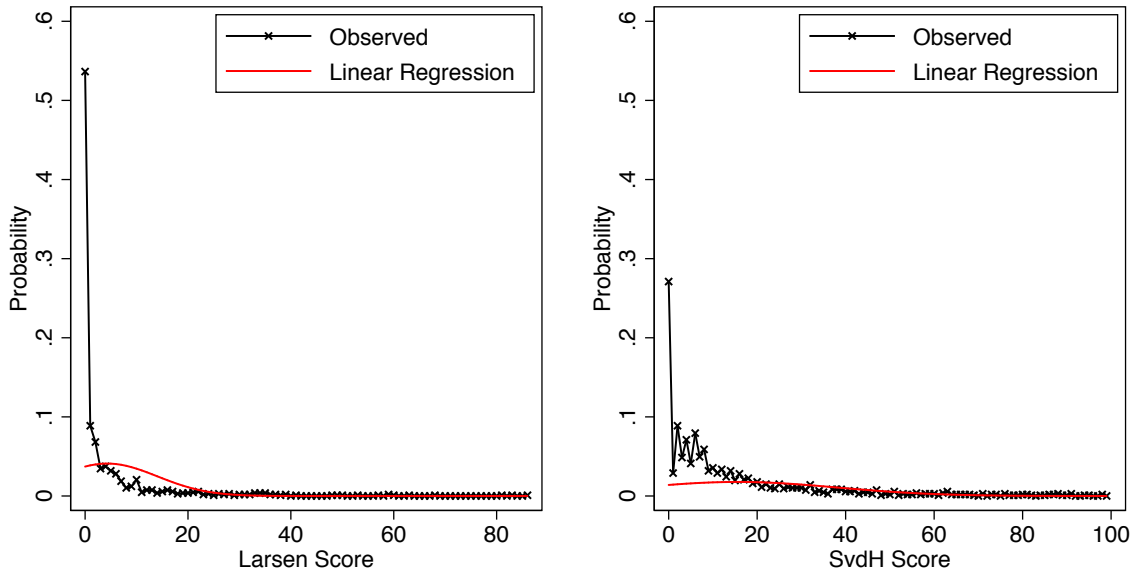


FIGURE 5.1: Observed Score vs. Predicted Probabilities from Linear Regression for the Larsen and SvdH Score in ERAS

residuals. Figure 5.2 depicts the heteroscedastic nature of radiographic outcomes, with the variance narrower at the low end, but increased at the high end, while Figure 5.3 shows the non-normally distributed residuals when applying linear regression to both the baseline Larsen and SvdH scores.

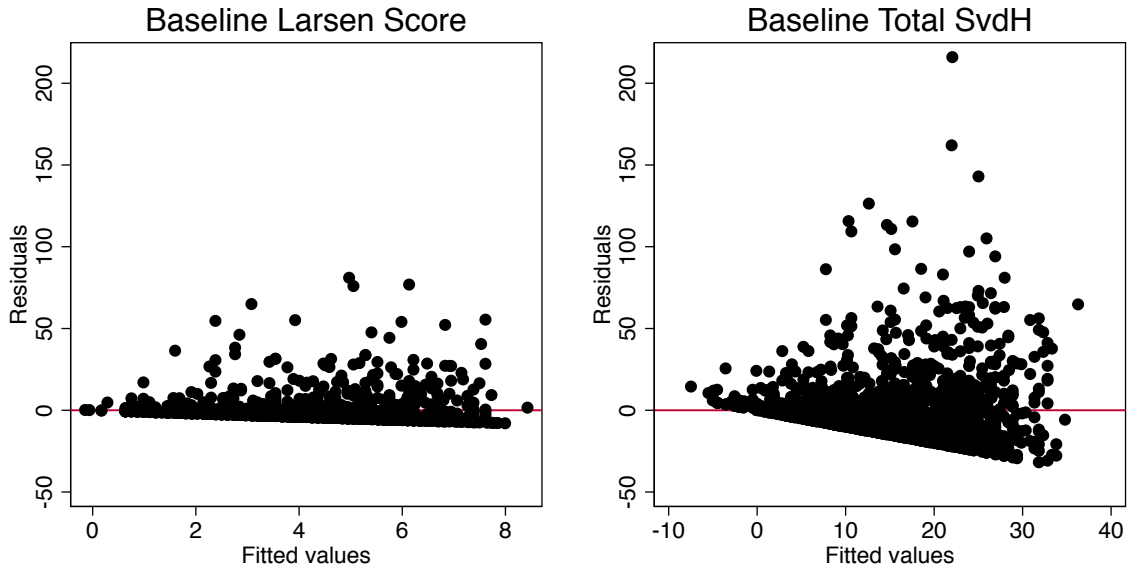


FIGURE 5.2: Residual vs. Fitted Plot to investigate Homoscedasticity for the Larsen and SvdH Score in ERAS

Despite its unsuitability, linear regression models are widely used in studies investigating the relationship between radiographic damage and functional disability[245]. The systematic review

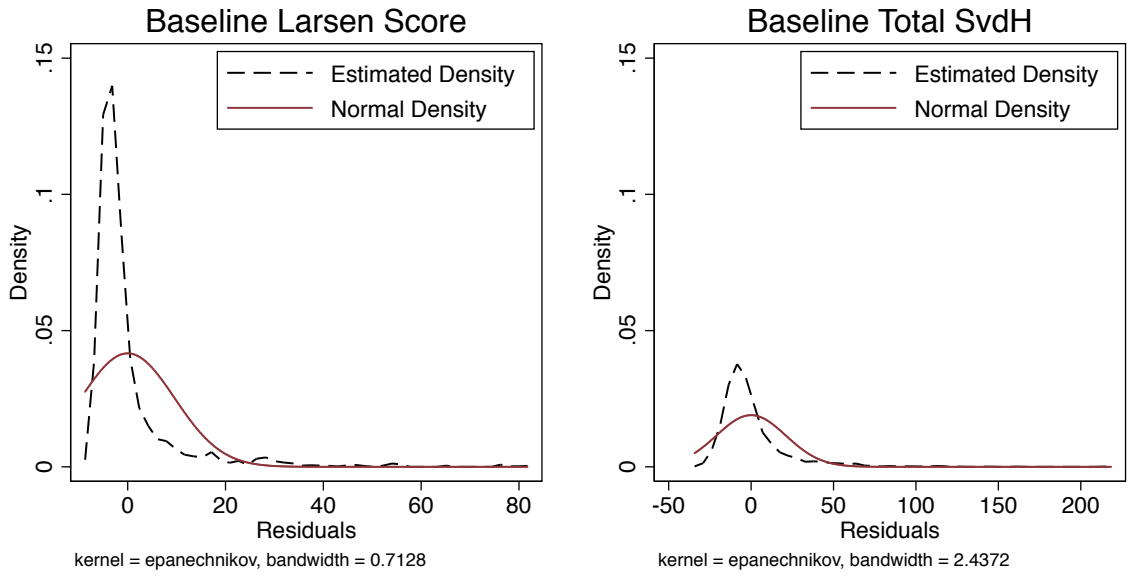


FIGURE 5.3: Distribution of the Residuals vs. Normal Density plot for the Larsen and SvdH Score in ERAS

conducted in Chapter 3 identified a host of published papers where linear regression was used to predict the level of radiographic damage based on data from clinical markers.

5.2.2 Data transformation

A common method for dealing with non-normal data is to transform it using functions that are known to reduce any positively or negatively skewed data. Common methods include squared, cubic, natural log, or square-root transformation. The aim is to transform the data so that it is normally distributed, and then apply general linear models. Once the model estimates have been estimated from the transformed data, the coefficients are then back-transformed so that they can be interpreted based on the original scale. The systematic review conducted in Chapter 3 found that square-root transformation was the most popular method of transformation, with log+1 transformation also being used.

However, the main issue with radiographic data is the preponderance of zero scores. Any method of transformation will only shift this large frequency count to a different value. This can be clearly illustrated in Figure 5.4, which shows the baseline SvdH scores being transformed using square, cubic, square-root, and natural log + 1 transformation methods. In all instances, the data still exhibits a degree of non-normality as a result of the large frequency of zero scores. In the case of the natural log +1 transformation, there is evidence of a mixture of distributions, with a binary zero, non-zero score, then log-normal if damage is present.

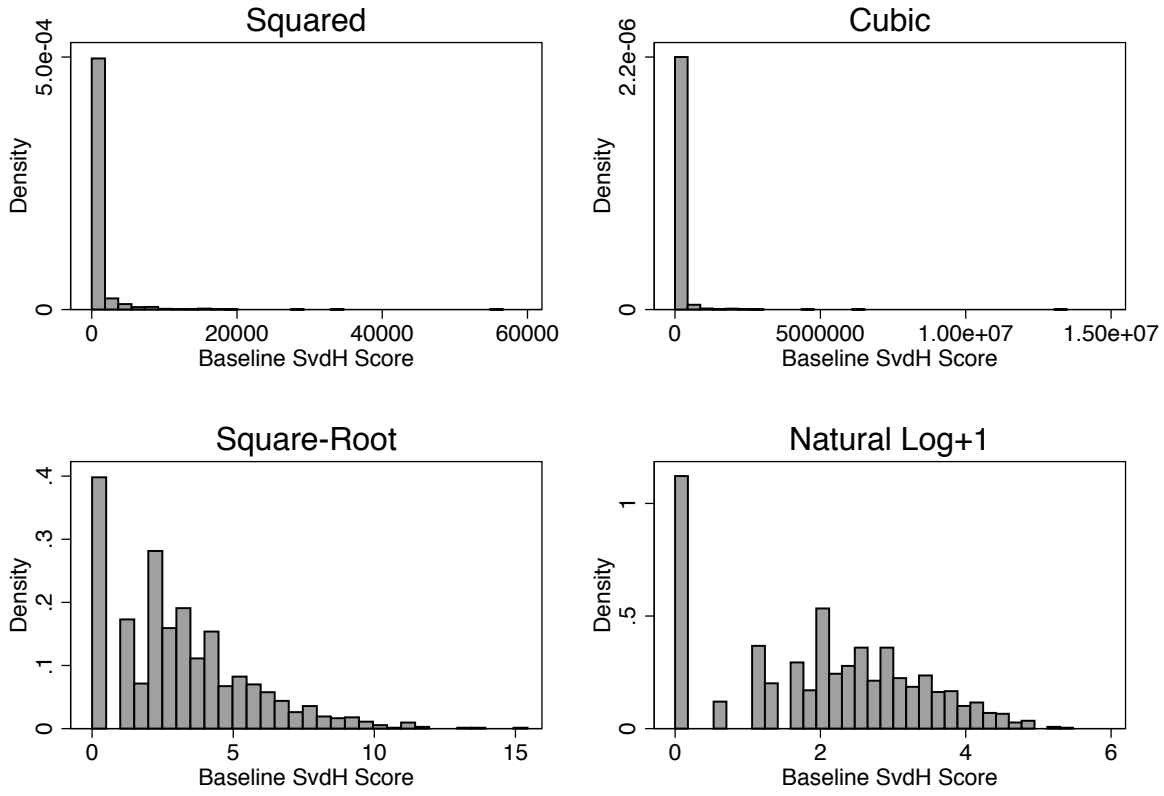


FIGURE 5.4: Common methods of transformation on Baseline SvdH Scores in ERAS

Figure 5.5 indicates how the use of log+1 transformed data still does not enable the general linear model to accurately predict the skewed distribution of the radiographic outcomes. In both the models using the baseline Larsen and total SvdH score, the number of zero scores was underestimated and the density of the low scores was greatly overestimated.

5.3 Generalised Linear Models (GLM) and non-normal distributions

This chapter has so far demonstrated that general linear models, even with data transformation, are not suitable in modelling radiographic outcomes. To address the issue of non-normally distributed data that is produced by radiographic outcomes, it is therefore imperative that non-normal models that assume the correct data distributions are used.

Generalised Linear Models (GLM) are extended versions of the general linear model that allow for the specification of non-normal distributions. Many different data distributions can be

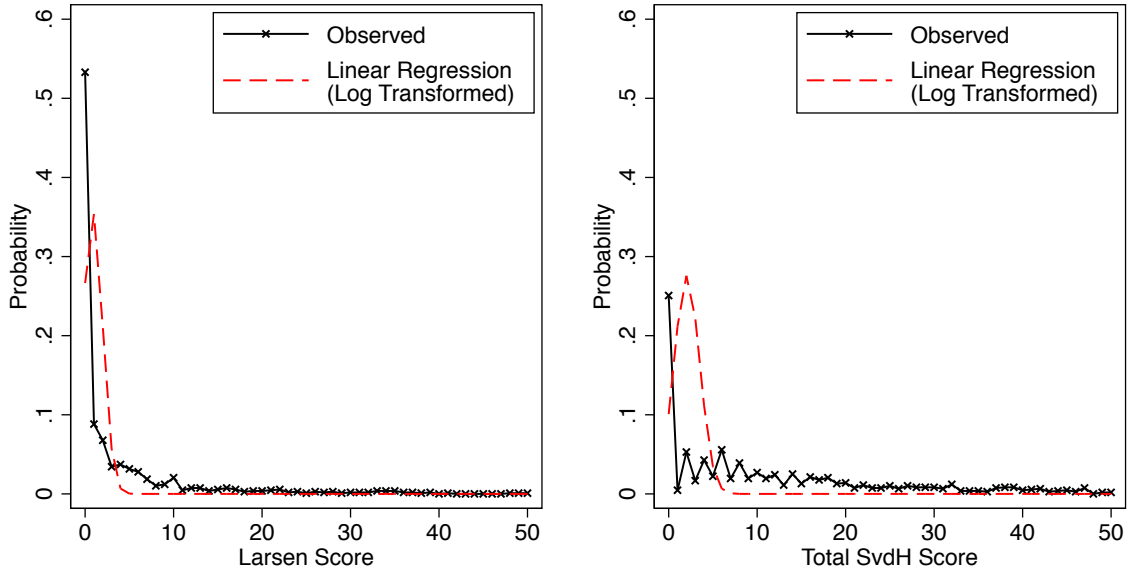


FIGURE 5.5: Observed Score vs. Predicted Probabilities from Linear Regression (Log Transformed) for the Larsen and SvdH Score in ERAS

specified, including Bernoulli, Poisson, Gamma, Binomial and multinomial. This section will focus on the application of the Bernoulli, Poisson and Gamma distributions, which are used to conduct logistic, Poisson and Negative Binomial (NB) regression. It will look at the application of these non-normal modelling techniques in the context of radiographic outcomes.

5.3.1 Logistic Regression

Logistic regression is the analysis of binary data. Whereas general linear models apply a normal, Gaussian distribution, a logistic regression applies a Bernoulli distribution to account for the dichotomous nature of the dependent variable. The data is modelled using the log of the dependent variable, and the resulting coefficient is interpreted as the anti-log (exponentiated) of the log score to create a ratio. Whereas the coefficient from the general linear model is an estimate of the mean change in the dependent variable for every 1-unit increase in the independent variable, logistic regression estimates the change in probability for every 1-unit increase in the independent variable, expressed as an Odds Ratio (OR).

The equation for a simple logistic regression is given in Equation (5.3), where $ln\left(\frac{P}{P-1}\right)$ is the log of the OR, and P is the probability of belonging in either category. The OR is defined as the probability of being in the category divided by the probability of not being in the category. As with the linear regression, α is the intercept parameter, β is the slope parameter, x is the independent variable and ε is the residual.

$$ln\left(\frac{P}{P-1}\right) = \alpha + \beta x + \varepsilon \tag{5.3}$$

The logistic regression model is estimated based on the log of the OR, as this allows any value of the model parameters to be estimated, while confining the log of the OR between 0 and 1. The log OR can then be exponentiated to provide the OR.

5.3.1.1 Logistic regression for radiographic outcomes

With radiographic data, categorisation is typically dichotomised into 2 groups consisting of a 'low/slow' group and a 'high/fast' group. These groups can be based on the score at the final follow-up visit, or based on a change in score over time. However, defining the cut-off is often arbitrary and various different methods are used. These include using the median value, a value clinically agreed to show clinically meaningful change[122, 246, 247], or a value determined through statistics, such as the Smallest Detectable Change (SDC)[124]. The SDC attempts to identify a threshold whereby any increase above this threshold reflects a true, meaningful increase in radiographic damage, rather than a change that might be expected due to measurement error alone[248]. For example, if the SDC is found to be 4 Larsen units, then any changes of less than 4 units may be due to measurement error alone, whereas any change in score greater than 4 units is likely to indicate true progression. This is particularly important in studies where the radiographer has scored the x-rays in a random order, and the chances of negative change, i.e. erosive healing, is more likely[89, 249, 250]. Further still, the method of estimating the SDC is based on the assumption of normality, with homoscedastic errors. As was demonstrated in Figure 5.3, the residuals are heteroscedastic, with the variance increasing at the high end of the radiographic score.

The ability to use statistical methods to quantify measurement error has been investigated, but with varying degrees of success[248, 251]. While the estimation of SDC and clinically meaningful change is very similar for the Sharp score, there is less consensus between these thresholds for the Larsen score, making the decision about what threshold to use more difficult[251]. Deciding how to categorise the data and which method to use becomes challenging and can introduce systematic bias if not defined *a-priori* to the analysis. Furthermore, if the goal is to facilitate the use of meta-analysis, this becomes difficult if separate studies are using different cut-points, with varying definitions of 'high/fast' progression.

5.3.2 Count Data

Count data is the number of times an event or outcome of interest occurs. In a clinical context this could be number of the hospital visits, the number of sick days off work or the number of times a particular prescription has been administered. In the case of count variables, very often the underlying distribution is positively skewed, that is the mass of the data distribution is at the lower end of the scale, with a low mean score, and a variance equal to, or greater than, the mean. Since the count cannot be negative, the variability of scores above the mean will typically be higher than those below the mean, meaning that the SD as a measure of variation is also inappropriate.

When modelling count data, it is therefore imperative that appropriate distributions are specificed in order to obtain relevant, unbiased and appropriate estimates. A variety of count distributions exist, the two most popular of which are the Poisson and NB. Both the Poisson and NB distributions can also be extended to account for increased or decreased zero counts, through the addition of either a zero-inflated or zero-truncated parameter.

The following section will introduce the Poisson and NB distributions and explain how their application in GLM makes them ideal for analysing count outcomes.

5.3.3 Poisson distribution

The Poisson distribution is outlined in Equation (5.4), where P is the probability, x is the number of events, e is Euler's constant, μ is the average number of events and $x!$ is the factorial of the total number of events.

$$P(x) = \frac{e^{-\mu}\mu^{x}}{x!} \tag{5.4}$$

The mathematical term μ is used in the Poisson distribution to describe both the mean and the variance of the distribution. The Poisson distribution assumes these are both the same, referred to as equidispersion. As the value of μ increases (i.e. the number of counts or events increases) the distribution curve shifts further to the right (See Figure 5.6). When the estimates of μ are >10, the Poisson distribution is approximately symmetric with a similar cumulative density function to the normal distribution. The difference however, is that the normal distribution allows for non-integer values, where as the Poisson does not.

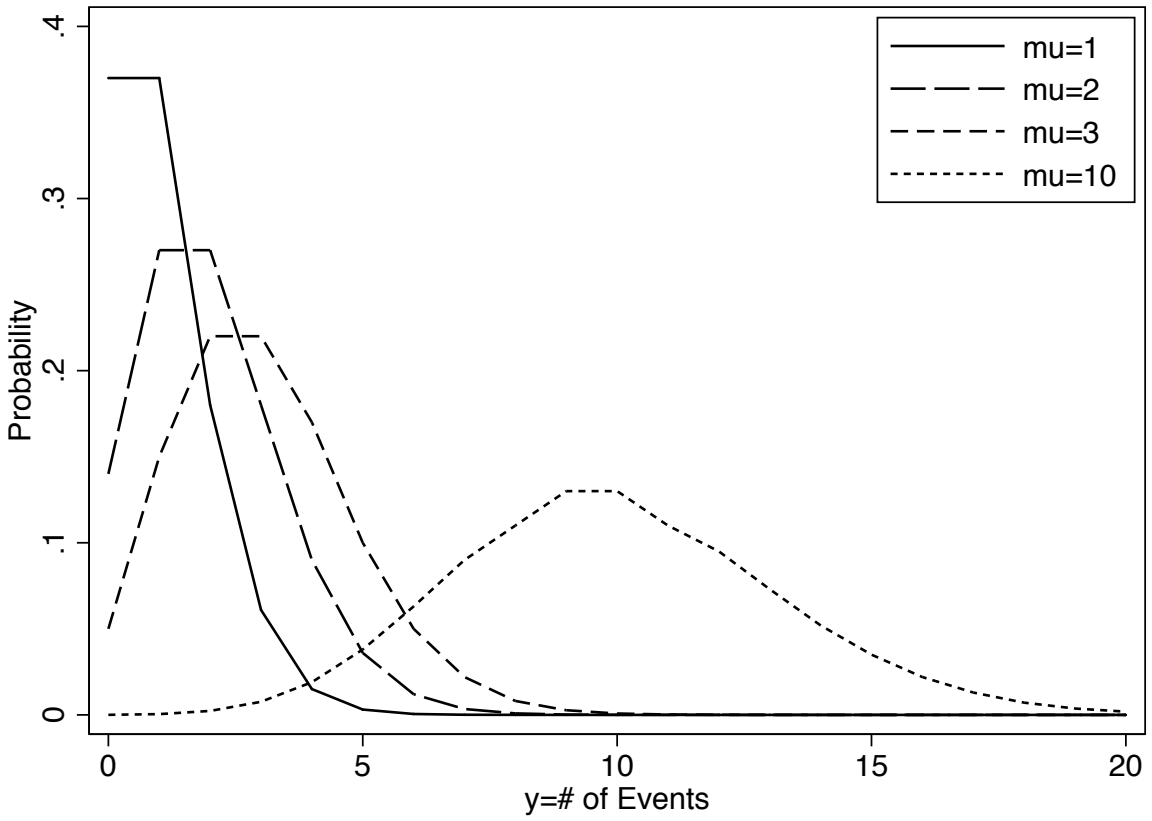


FIGURE 5.6: Poisson Probability Distributions

When the Poisson distribution is specified as part of a GLM, each observation i is drawn from this Poisson distribution with a mean of μ. Since the count of any observation cannot be less than 0, the linear estimate of the mean is based on log-linear scale. This is shown in Equation (5.5), where μ is the expected mean, x is the independent variable from the ith observation and β is the slope.

$$\mu_i = e\Big\{x_i^{'}\beta\Big\} \tag{5.5}$$

The predicted effect estimate is the exponentiation of the predicted log count, and is interpreted as a Rate Ratio (RR). That is, the ratio of the expected count divided by the expected count when the covariate increases by 1 unit, where δ is the change in score. This is depicted in Equation (5.6).

$$\frac{E(y \mid \mathbf{x}, x_k + \delta)}{E(y \mid \mathbf{x}, x_k)} = e^{\beta_k \delta} \tag{5.6}$$

5.3.4 Negative binomial (NB) distribution

While the Poisson distribution assumes a equidispersion, the NB distribution includes an additional parameter that accounts for data where the variance is larger than the mean, commonly referred to as *overdispersion*. In general, overdispersion is where there is greater variability than is expected given the chosen statistical test used. In the context of Poisson regression, there is not an additional parameter whereby the variance can be defined independently of the mean, so the Poisson regression is unable to model instances where the variance of the outcome is much larger than the mean of the outcome.

The mean structure of the NB is identical to that seen in the Poisson, and therefore the estimated mean by both models can be similar. However, it is the standard errors that will change if the data is found to have overdispersion. The Poisson will have spuriously large z-scores and therefore small p-values where the variance is larger than the mean, whereas the NB, which accounts for this overdispersion, will provide accurate estimates of the standard errors, and unbiased p-values (**N.B.** It is worth noting that the NB regression can account for underdispersion, but this is not discussed, since it does not occur in radiographic outcomes).

$$Pr\left(y_i|\mathbf{x}_i\delta_i\right) = \frac{e^{-\tilde{\mu}_i}\tilde{\mu}_i^{yi}}{y_i!} \tag{5.7}$$

Equation (5.7) shows the equation for the probability distribution of the NB, where δ is the additional overdispersion parameter, which is drawn from a gamma distribution.

5.3.5 Excess zero scores and zero-inflated models

GLM incorporating a Poisson distribution is a useful way of modelling count data where the mean and the variance are equal. In cases where the count data has a variance of greater than the mean, the NB distribution is useful at modelling this overdispersion and estimating more precise standard errors. This overdispersion is typically caused by high frequency of zero scores. However, the probability of the zero score occurring in the data may be varied between observations. As a result, there are instances where modelling the probability of the zero score occurring is also needed.

Any zero score in count data can be thought to occur in one of two ways. To illustrate this, an example using the number of fish caught by fishermen will be used. Each observation is a

fisherman, and the count data is the number of fish caught. They all went on a camping trip near a river, however, not all of them made it to the river to fish. There were those who went on the trip and had time to fish in the river, and those that never made it to the river to fish. For those fishermen who made it to the river, they could have caught any number of fish. For those that made it to the river, if no fish were caught then their zero score would be considered a 'sampling zero', since the zero score occurred as a product of the random sampling distribution. However, for those who never made it to the river , they would also record a zero score, but their probability of catching any fish is very different to those who went to the river. As a result, it would be useful to record these zero scores differently, as 'structural zeros', since the structure of the data means the probability of achieving a non-zero count is different within the group.

A zero-inflated function is a useful way of incorporating this difference in probability of experiencing the count outcome (i.e. accounting for sampling zeros). This can be applied to either the Poisson or NB functions. A Poisson or NB zero-inflated models the data over two-components. The first component for a Poisson zero-inflated model is depicted in Equation (5.8) an shows the first binary process of determining structural zeros. The second component is depicted in (5.9), and shows the count model part for non-zero scores using Poisson distribution introduced in Equation (5.4), for when x is greater than or equal to 1.

$$\Pr(y_j = 0) = \pi + (1 - \pi)e^{-\lambda} \tag{5.8}$$

$$\Pr(y_j = x_i) = (1 - \pi)\frac{\lambda^{x_i}e^{-\lambda}}{h_i!}, \qquad x_i \geq 1 \tag{5.9}$$

Understanding whether to use a zero-inflated parameter in either a Poisson or NB model is dependent on the data structure, and whether the zero scores occur purely as a product of the sampling distribution, or whether some zero scores could occur due to the fact that the observation had no opportunity to experience the event.

5.3.5.1 Count models for radiographic data

The use of count data regression models, such as Poisson and NB regression, has been shown in previous studies to effectively account for these skewed data distributions observed by radiographic data[252–254]. In the systematic review conducted in Chapter 3, two of the studies

identified used count models to look at predictive markers of radiographic damage[191, 194]. Rather than treating the radiographic score as a continuous outcome, it is treated as a count outcome, weighted by the severity of the erosion of JSN.

The predicted probability functions of a GLM with a Poisson (blue line) and NB (green line) distribution specification, applied to both the baseline Larsen and total SvdH score, are depicted in Figure 5.7.

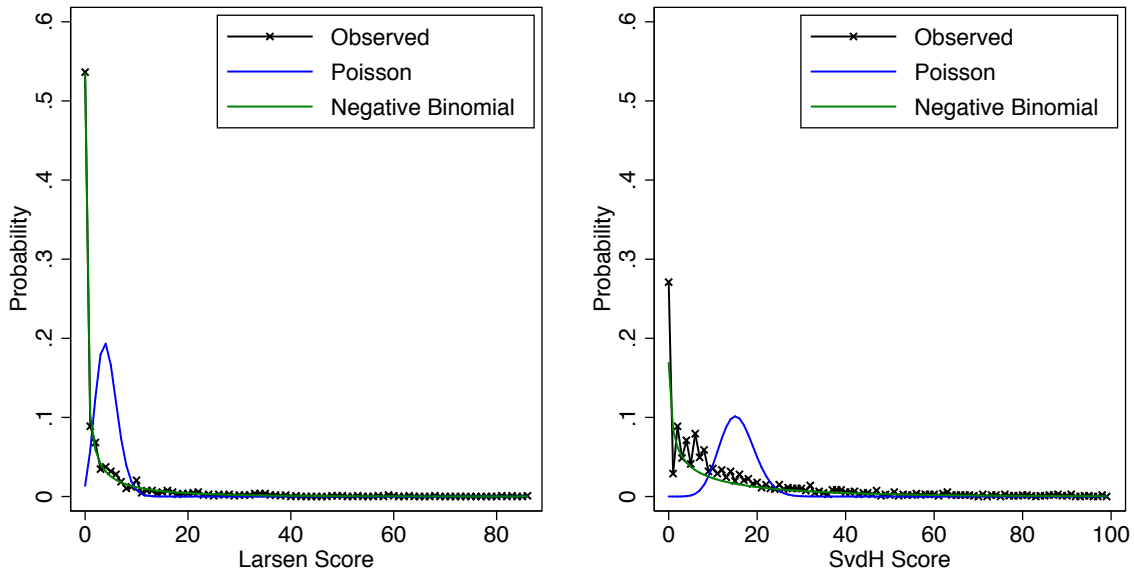


FIGURE 5.7: Observed vs. Predicted Probabilities from PR and NB for the Larsen and SvdH Score in ERAS

The Poisson (blue line) does not fit the observed data well, mainly due to the excess zero scores. It also greatly overestimates the scores at the low end of the scale. However, the NB (green line) does a very good job predicting the observed counts, appropriately accounting for the overdispersion that is present in the data. In the case of the SvdH score, it would appear that the NB does underestimate the zero count.

The predicted probability functions for the zero-inflated Poisson and NB models are shown in Figure 5.8. The zero-inflated Poisson (dashed blue line) is better for modelling the high level of zero counts compared to the Poisson, but still overestimates the scores at the low-end. The zero-inflated NB (dashed green line) is similar to the standard NB for the baseline Larsen and total SvdH score.

The probability density functions have provided a good means of looking at how well each distribution can accurately model radiographic outcomes. Model fit can also be presented statistically through the use of the Akaike Information Criterion (AIC) and Bayesian Information

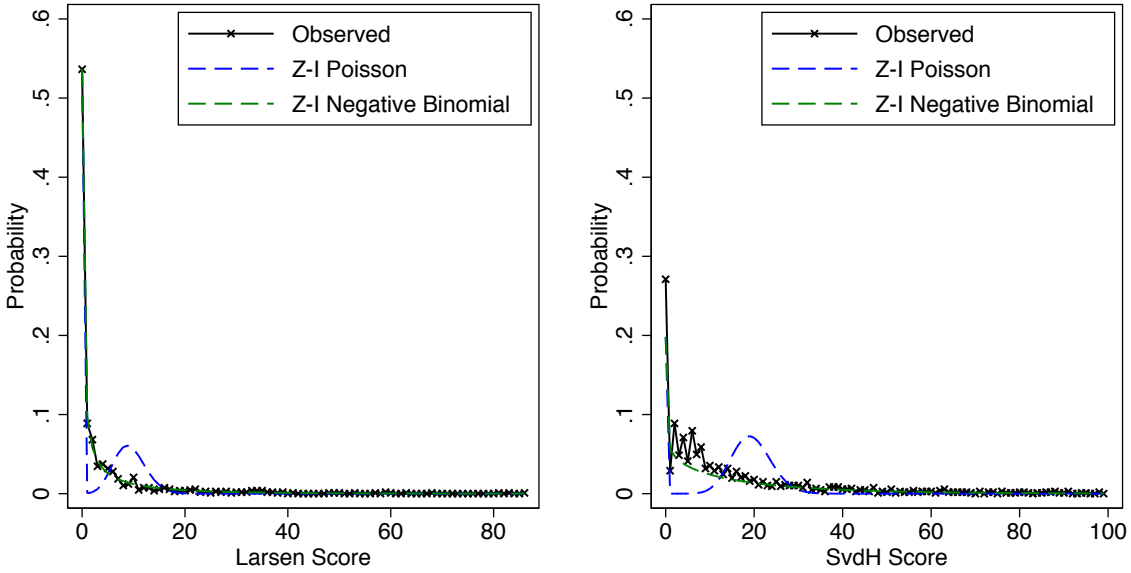


FIGURE 5.8: Observed vs. Predicted Probabilities from Zero-inflated PR and NB for the Larsen and SvdH Score in ERAS

Criterion (BIC) fit statistics. The lower the value of the AIC or BIC, the better the model fit. Table 5.1 indicates that the NB regression models provided the best AIC and BIC values, whereas the Poisson provided the worst model fit. For reference, the linear regression model fit statistics were also supplied.

	Linear	Poisson	NB	ZI-Poisson	ZI-NB
Larsen Score					
Observations	1070	1070	1070	1070	1070
AIC	7904.8	14502.8	4537.5	8899.9	4538.8
BIC	7909.8	14507.8	4547.4	8909.8	4553.8
Total SvdH Score					
Observations	1077	1077	1077	1077	1077
AIC	9961.8	31891.9	7978.3	22286.3	7873.0
BIC	9966.8	31896.9	7988.2	22296.2	7888.0

AIC = Akaike information criterion, BIC = Bayesian information criterion

TABLE 5.1: Model fit statistics of linear and count regression for the total SvdH score in ERAS

In the case of the baseline Larsen score, the NB and zero-inflated NB were very similar, however the zero-inflated NB was found to provide a small improvement in model fit over the standard NB for the baseline total SvdH.

Estimation of the proportion of zero counts estimated by each count model is provided in Table 5.2.

	Larsen Zero Counts	SvdH Zero Counts
Observed	0.54	0.25
PR	0.01	0.00
ZI-PR	0.54	0.25
NBR	0.53	0.20
ZI-NBR	0.53	0.20

TABLE 5.2: Predicted Zero Counts for baseline Larsen and SvdH scores in ERAS

The zero-inflated Poisson, NB and and zero-inflated NB all provide a reasonable estimate of the zero counts for both the baseline Larsen and total SvdH. Interestingly the zero-inflated Poisson provides the closest prediction to the observed count for both the Larsen and SvdH. However, the substantially larger AIC and BIC highlights how in all other aspects, the zero-inflated Poisson fails to adequately model the data.

To illustrate the impact of model fit on the model estimates, a series of count models will be applied to the baseline Larsen score and SvdH score (from the ERAS cohort only) with a set number of covariates of interest; sex, age at disease onset, baseline HAQ, baseline DAS, baseline HB and RF positivity. As before, the following count models will be estimated: Poisson, zero-inflated Poisson, NB and zero-inflated NB. The interpretation and prognostic abilities will not be assessed at this point, and the model is purely being used for illustrative purposes in order to assess how the goodness of fit impacts on the model estimates when covariates are introduced.

Table 5.3 and Table 5.4 provide the model estimates, including their standard errors, for each of the four count models. Investigation of the Pearson Chi^2, AIC and BIC once again indicates that both the NB and the zero-inflated NB have consistently lower test statistics compared to the Poisson and zero-inflated Poisson.

The Poisson typically under estimates the standard errors, thereby increasing the Z-score and increasing the chances of reporting a significant estimate. Given that the fit statistic is so high compared to the NB or zero-inflated NB, it can be deduced that these estimates for the standard errors are biased, and therefore the chances of a Type-I error is greater.

Finally, the difference in the Rate Ratios (RR) between the standard and zero-inflated versions of both Poisson and NB indicates that additional parameterisation of the excess zero scores as structural zero scores has a marked effect on the estimation of the effect estimate.

The connotations of this in the context of radiographic damage is that there could be two types of RA patients with respect to zero radiographic scores. There are those patients that have an erosive form of RA, but do not experience any damage (sampling zero), and those with a form of non-erosive RA, and therefore do not have radiographic damage as they do not have the capacity to experience the event (structural zero).

	Poisson	ZI-Poisson	NB	ZI-NB
Baseline Larsen Score				
Female	0.95	0.87	0.86	0.72
	(0.03)	(0.03)	(0.12)	(0.11)
Age at Onset	1.03	1.01	1.03	1.02
	(0.00)	(0.00)	(0.00)	(0.01)
Baseline HAQ Score	0.97	0.94	1.00	1.04
	(0.02)	(0.02)	(0.11)	(0.12)
Baseline DAS Score	1.17	1.05	1.21	1.08
	(0.02)	(0.02)	(0.07)	(0.08)
Baseline HB	0.91	0.86	0.89	0.80
	(0.00)	(0.01)	(0.02)	(0.04)
RF Positive	1.27	1.16	1.39	1.36
	(0.04)	(0.04)	(0.20)	(0.19)
Chi2	12049.1	398.5	440.5	65.3
AIC	13081	8254	4405	4401
BIC	13111	8294	4440	4446

Exponentiated coefficients (IRR); Standard errors in parentheses

TABLE 5.3: Model Estimates from Count Models on Larsen Scores in ERAS

Finally, the importance of using the NB model to account for the overdispersion between the mean and the variance can be graphically depicted after running these multivariate models. Both the mean and variance were estimated for each patient based on modelling the Larsen and SvdH score. The mean values were then grouped into twenty groups and their associated variance were depicted in Figure 5.9. It can be seen how for both the baseline Larsen and total SvdH score, the NB model (blue line) is much better at accounting for the high variance at the higher mean values when compared to the Poisson model (red line).

5.4 Longitudinal data

Longitudinal data is the process of collecting data on participants at multiple time-points, sometimes referred to as repeated measures. The time interval between each time-point can

	PR	ZI-PR	NBR	ZI-NBR
Baseline SvdH Score				
Female	1.47	1.16	1.44	1.16
	(0.02)	(0.02)	(0.13)	(0.10)
Age at Onset	1.04	1.02	1.04	1.02
	(0.00)	(0.00)	(0.00)	(0.00)
Baseline HAQ Score	0.94	0.98	0.92	0.97
	(0.01)	(0.01)	(0.07)	(0.06)
Baseline DAS Score	1.08	1.01	1.10	1.01
	(0.01)	(0.01)	(0.04)	(0.04)
Baseline HB	1.01	0.93	1.01	0.91
	(0.00)	(0.00)	(0.01)	(0.02)
RF Positive	1.10	1.04	1.12	1.07
	(0.02)	(0.02)	(0.10)	(0.08)
Chi2	164603.8	1624.7	3810.1	94.4
AIC	27220	20016	7763	7666
BIC	27250	20056	7798	7711

Exponentiated coefficients (IRR); Standard errors in parentheses

TABLE 5.4: Model Estimates from Count Models on SvdH Scores in ERAS

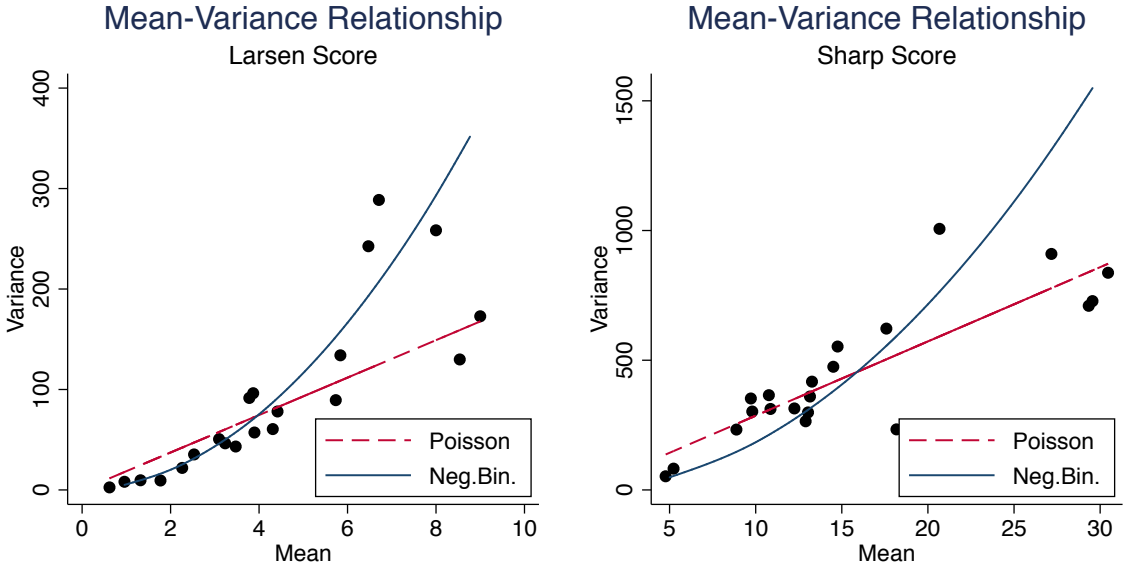


FIGURE 5.9: Mean–Variance Relationship for Baseline Larsen and SvdH Scores in ERAS

either be at fixed intervals, such as every year, or unequal. It is not suitable to use cross-sectional regression techniques, such as those described above, since these models assume that each observation is independent. In the context of longitudinal data, it is almost certain that a participant's score at one time point is highly correlated to their previous or indeed subsequent scores[255]. To appropriately handle longitudinal data analysis it is crucial to estimate the likely

intraclass correlation between their time points. There are two main methods in which this can be appropriately modelled, Generalised Estimating Equations (GEE) and multi-level modelling. This section will explore both methods, and explain why the use of multi-level models are the most suitable for modelling longitudinal radiographic outcomes.

5.4.1 Generalised Estimating Equations (GEE)

GEE can be regarded as an extension of GLM, which were defined earlier in this chapter. They allow the estimation of parameters from GLMs, while incorporating a correlation structure between observations. GEE can be depicted in the Equation (5.10), where μ_{ij} is the mean for the ith subject at the j time point, β is the regression coefficient and V is the variance.

$$U(\beta) = \sum_{i=1}^{N} \frac{\partial \mu_{ij}}{\partial \beta_k} V_i^{-1} \{Y_i - \mu_i(\beta)\} \tag{5.10}$$

Unlike the other regression methods detailed so far, GEE does not rely on the maximum likelihood estimation (MLE) method of estimating the model parameter. Instead, GEE is described as a semi-parametric method and does not require specification of the data distribution[256]. This approach is often referred to as a quasilikelihood function and looks at estimating the population-averaged estimates[256]. As with GLM, categorical and count based distributions can be specified in any GEE model.

5.4.2 Multi-level modelling

Multi-level modelling, sometimes referred to as mixed-effects or hierarchal models, is a method that allows the incorporation of multiple levels of data. The general linear model and GLMs, described so far, assume a single data structure. However, data can often be structured over multiple levels. They are therefore seen as an extension of GLMs. As with the GEE, observations within each level of the data are likely to be related to one another, and therefore it is crucial that the model accounts for the correlation within these levels of data. Classic examples of multi-level data structures include patients in a study all being recruited from a particular hospital, or children being assessed in different classes across different schools. In the later example, each child is clustered within a class, which is clustered within a school, thereby creating 3 'levels' of data; child, class and school. In the context of longitudinal data, the additional 'level'

of data is time, with each patient having multiple data points. The advantage of multi-level models over GEE is that level-specific estimates can be obtained, whereas GEE only provides population-averaged estimates. In longitudinal data, multi-level models allow for the estimation of individual change over time, as well as the variance between subjects.

The equation for simple linear regression is outlined in Equation (5.11).

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon \tag{5.11}$$

Where β_0 = intercept and β_1 = the slope, adjusted by covariate x_i in the model, and e = residual error.

The equation for the multi-level linear regression has some subtle, but important differences as seen in Equation (5.12). The addition of the $_j$ denotes that each parameter will be estimated for $_j$ time point. This is referred to as Level 1 of the multi-level model.

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij} \tag{5.12}$$

Level 2 of the multi-level model is the calculation of the intercept and slope parameter given the group level mean. Equation (5.13) refers to the constant (intercept) and Equation (5.14) refers to the subject-specific regression coefficient (slope), where γ_{00} indicates the overall constant, γ_{01} and γ_{10} indicates the regression coefficient between the dependent variable and the level 1 predictor, and the dependent variable and the level 2 predictor respectively. Finally, $u_0 j$ and $u_1 j$ indicates the error component of the constant and regression coefficient respectively (i.e. the deviation from the overall intercept/slope and the group level intercept/slope).

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \tag{5.13}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{5.14}$$

5.4.3 GEE vs. Multi-level

Both GEE and multi-level modelling appropriately handle the correlated data seen in longitudinal studies, and as such, the decision of which approach to use for any particular study should be based on the research questions asked.

While GEE benefits from not having to specify the specific data distribution, it is restricted to only providing population average effects, and not the group level effects that may be of interest. Secondly, since it does not rely on MLE it cannot handle missing data through ML techniques, and therefore in cases where data are largely missing, it will be restricting the data to only those patients with complete data.

In contrast, multi-level models are able to specifically estimate the group level effects, and its use of MLE enables it to handle missing data in the analysis. As a result however, this does require the correct specification of the data distribution to ensure the mean and associated standard errors are appropriately modelled.

Given that this particular thesis is concerned with radiographic data, which does suffer from missing data, and is also interested in quantifying the group level effect (that is the effect over time), the multi-level approach was chosen as the preferred method. The next section will explore its application to radiographic outcomes, and assess how multi-level negative binomial models accurately fit the data, and the ways in which it handles missing data.

5.4.3.1 Multi-level count models for radiographic outcomes

As was observed in the cross-sectional data of baseline radiographic scores, the skewed distribution of the radiographic score produces different values for the mean and median estimate. Figure 5.10 demonstrates how the radiographic score can be much higher at any point in time when looking at the mean compared to the median score over time of the Larsen and SvdH data in ERAS.

To illustrate the need for using multi-level models, the total SvdH scores for patients in the ERAS are plotted in Figure 5.11 along with a Locally Weighted Scatterplot Smoothing (LOWESS) of this data. The plot highlights a large variation in the total SvdH score at baseline (intercept) and different rates of progression over the 9 years follow-up (slopes). A multi-level model will

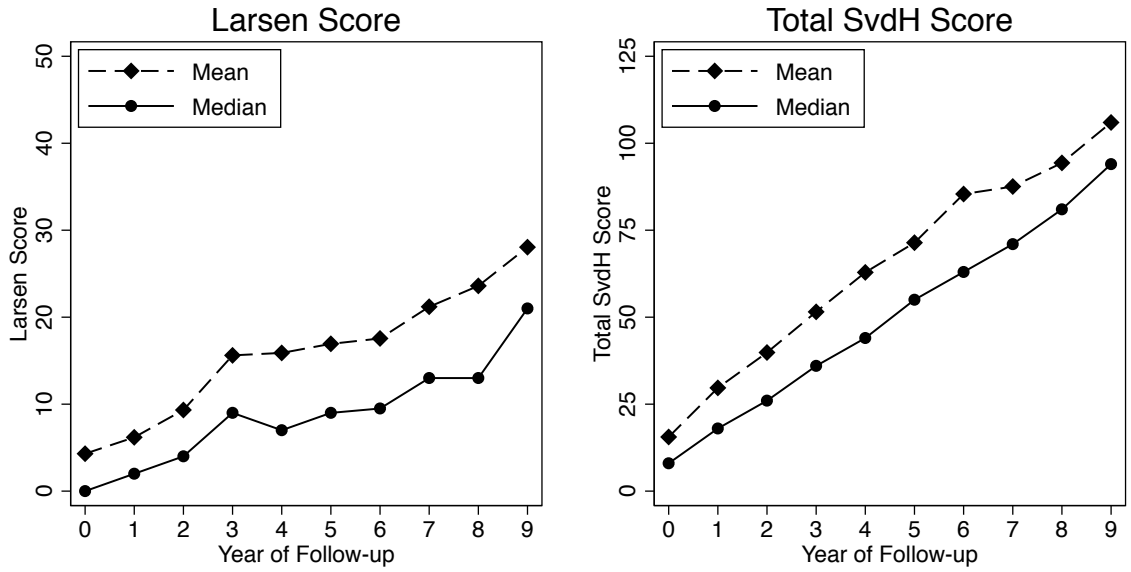


FIGURE 5.10: Mean and median Larsen and SvdH scores over the first 9 years for ERAS patients

allow for the intercepts and slopes to vary over individuals, thereby accounting for the additional variance within, as well as between individuals.

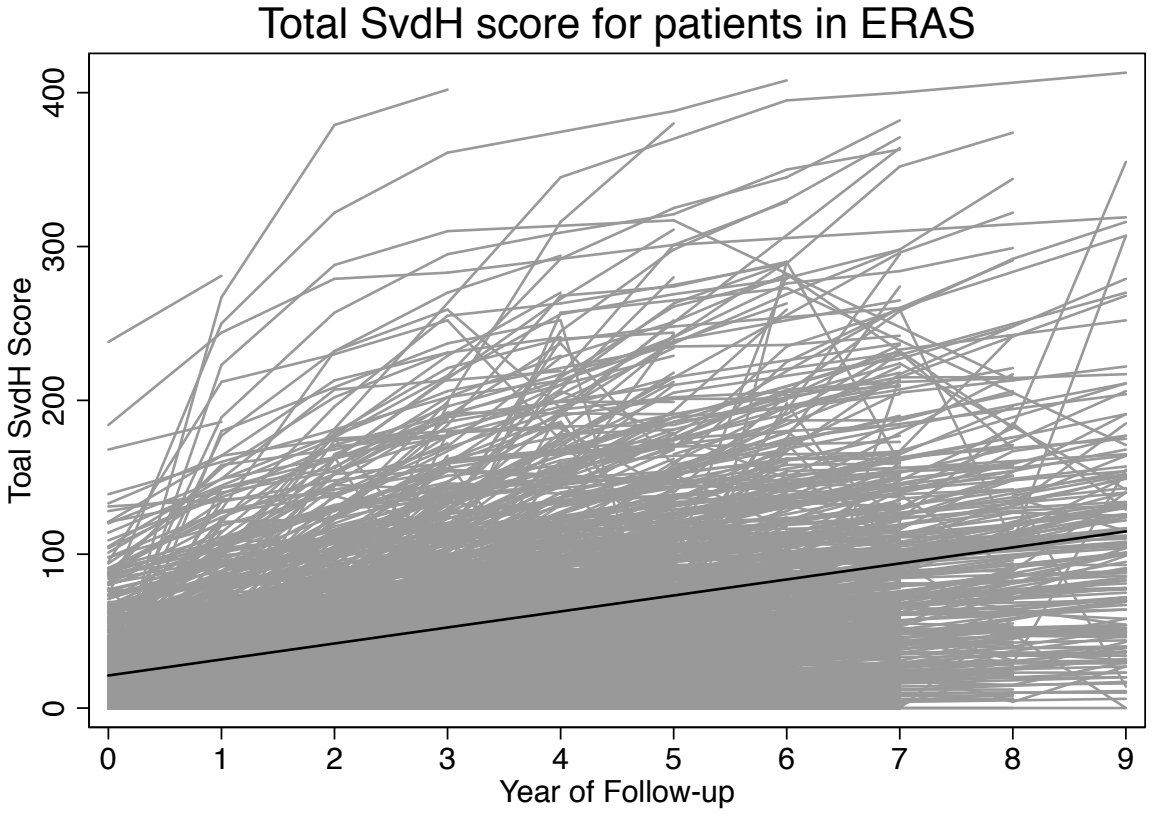


FIGURE 5.11: The total SvdH score (grey plot) and LOWESS smoother (black plot) over the first 9 years for patients in ERAS

As with the cross-sectional data earlier in the chapter, comparisons between the Poisson and NB models can also be compared in the multi-level context. To understand how the multi-level Poisson and NB regression methods differ in modelling change over time, the estimated sample means for the Larsen and SvdH score from the ERAS dataset have been calculated using each of these models. The estimates are given in Table 5.5 and graphed in Figure 5.12.

	Larsen		Total SvdH	
	Poisson	NB	Poisson	NB
Follow-up Year	1.20	1.28	1.20	1.27
	(0.002)	(0.006)	(0.001)	(0.005)
Chi2	16199.02	2436.88	44233.33	4003.53
AIC	48773	38381	74669	51394
BIC	48793	38408	74689	51421

Exponentiated coefficients - Rate Ratios (RR); Standard errors in parentheses

TABLE 5.5: Estimated yearly progression rates for the Larsen and total SvdH Score using multi-level Poisson and NB regression in ERAS

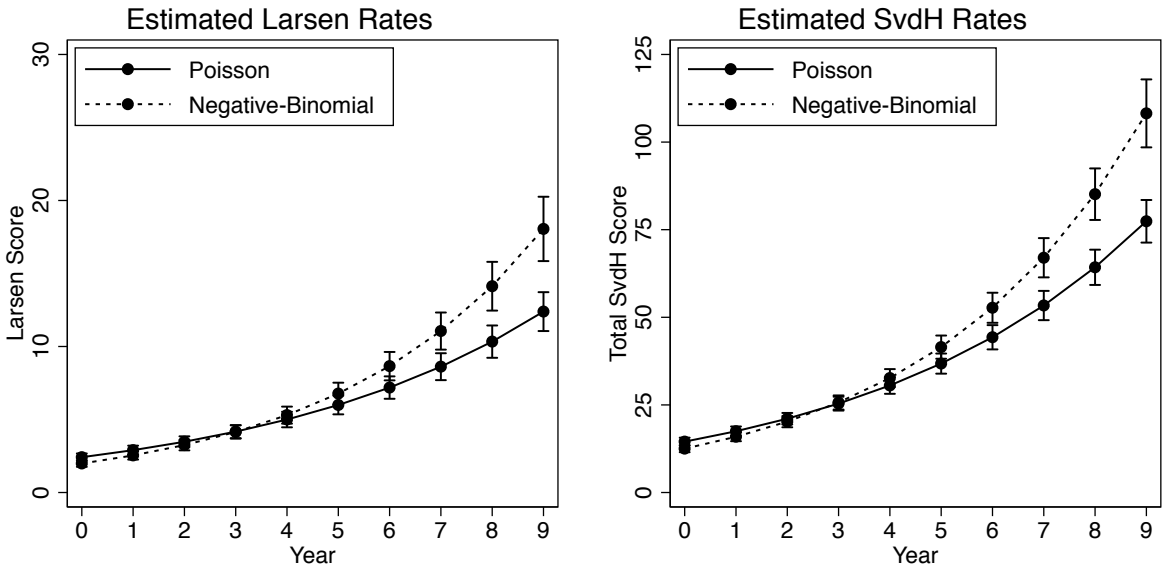


FIGURE 5.12: Estimated radiographic rates over first 9 years for the Poisson and NB multi-level models for the Larsen and SvdH Score in ERAS

The estimated coefficients and standard errors for the follow-up year covariate (slope) for both the Larsen score and the total SvdH score are shown in Table 5.5. For both radiographic scores, the coefficient and the associated standard errors are smaller for the Poisson compared to the NB. Given the lower Chi^2 test, AIC and BIC, it can be concluded that the multi-level NB provides better model fit, and therefore a less biased estimate of the yearly progression.ftb

The differences in yearly progression are also highlighted in Figure 5.12, which illustrates how the NB has accelerated progression from year 5 and onwards.

5.4.3.2 Multi-level models and missing data

Chapter 4 outlined the extent of missing data for both radiographic outcomes in ERAS and ERAN. The application of general linear models and non-linear GLMs for cross-sectional data will exclude patients case-wise, meaning that a case, or patient, with any missing data will result in the complete omission of that case in the analysis. This is commonly referred to as 'complete case analysis'. Complete case analysis makes various assumptions about the characteristics of those cases that have been omitted. In order to assume that the complete case analysis is not biased by the fact that it only includes patients with complete data, it assumes that all those cases with missing data are missing at random. That is to say that their characteristics are similar to those cases included in the complete case analysis, and that if their data were complete, and they were included in the analysis, similar effect estimates would be obtained. Whether this assumption is valid can be difficult to ascertain. Understanding the nature of the missing data is the first step, and understanding whether the data is indeed *Missing At Random (MAR)*, or whether there could be a reason for the missing data, can help understand whether there is any bias in a complete case analysis.

Once the the data is assumed to be MAR, various methods can be applied to impute this missing data to reduce the bias inherent in complete case analysis. Various methods have been devised, such as mean imputation; however these methods have been shown to bias results further than complete case analysis, as they artificially accentuate any mean effect by reducing the variance around the mean point estimate. In contrast, multiple imputation chained equations (MICE) have been recommended as a more appropriate way of reducing the bias from missing data. Using data from the outcome of interest, as well as covariates included in the model, MICE imputes data using chained equations to provide a range of imputed values that could represent the missing value. These imputations are then combined under *Rubins Rule*[257] to estimate one specific value for estimation in the analysis. Simulation data has shown the use of MICE to provide more robust and less biased effect estimates, namely standard errors[242].

With longitudinal data, missing data is even more likely to occur, and attrition is likely to increase over the follow-up period. Unlike GLM, which as we discussed will omit cases with

missing data list wise, multi-level modelling has the advantage of using full-information maximum likelihood (FIML) estimation. For example, if a patient only has radiographic data at baseline and year 3, that data will be used to estimate the rate over the full follow-up period, thereby maximising all information for the model. While the methods are very different to MICE, the assumptions are similar and studies have shown it to yield similar results[258]. However, it must be noted that MICE is able to impute values for missing predictor variables, as well as the dependent variable, whereas multi-levelling methods will only apply FIML estimation for the dependent variable. Multiple Imputation methods in multi-level modelling has received little attention, however methods are being explored, and the added benefit of using these more complicated models where missing data for the independent variables are low is not clear[259].

5.5 Multivariate models

In epidemiology, choosing the covariates to include in a predictive model is challenging[260]. There is a variety of different methods available to the researcher when deciding which variables to include, and which to exclude. Some of these methods rely on prior theory and knowledge generated from previous studies, while other methods look at developing statistical algorithms that aim to include only those variables that, for example, satisfy a pre-defined level of statistical significance.

In 2009, Walter and Tiemeier[260] looked at current practices involving variable selection in epidemiological studies. They found the most common methods used were either based on prior knowledge, automated techniques, such as stepwise selection, or a combination of the two. This was directly in line with the systematic review on predictive factors of radiographic progression in Chapter 3. Selecting variables based on previous research remains the most encouraged method of model building amongst statistical circles[260], however Walter and Tiemeier argue that more referencing of the previous research is needed so that the rationale behind its selection can be discussed and examined.

In contrast, the use of statistical methods to determine which covariates to include, such as stepwise selection, is highly debated in epidemiological research[260–262]. Stepwise selection looks at systematically including each variable in the data based on a pre-defined statistical criterion. The most popular of which is the statistical significance level, whereby only those

variables below a certain significance level are included. While other statistical criteria, such as the information criterion, are much more favoured, they are rarely used. This is likely due to the way in which most statistical software packages apply stepwise selection, where the default is commonly based on the significance test.

5.6 Discussion

This chapter looked at addressing the second aim of the thesis; to provide a strong statistical framework for the primary analyses conducted in the subsequent chapters. There is seldom a 'one size fits all' approach to statistical analysis. Often there are multiple methods available to researchers to summarise the data, and provide evidence on a given hypothesis. However, this chapter has provided a comprehensive overview of the modelling methods typically used to look at radiographic data, which were identified from the systematic review conducted in Chapter 3. This chapter has highlighted why the skewed data distribution of radiographic outcomes renders general linear models inappropriate, even following data transformation techniques. While categorisation of the radiographic score to conduct logistic regression does overcome the issues with the skewed distribution, other assumptions about how the data should be categorised have also been shown to introduce bias. Although this chapter has demonstrated the inherent bias with adopting these modelling techniques to radiographic outcomes, the interpretation and conclusions drawn from the literature which have applied these methods are unlikely to change completely. However, the magnitude of the effect estimates, and ultimately the precision of the standard errors, and therefore the confidence intervals and associated level of significance could be over estimated.

An in-depth assessment of count regression models was undertaken, along with their suitability in modelling radiographic outcomes. The high preponderance of zero scores, along with large overdispersion, led to the conclusion that NB regression models were able to provide the most precise effect estimates with the best model fit. This was in comparison to other count models, such as Poisson regression. The chapter also explored the use of zero-inflated count models and the impact this has on the theoretical conception of zero scores in radiographic damage. The use of zero-inflated parameters would suggest that forms of non-erosive RA exist, and therefore the estimates would be adjusted based on the probability that some patients will not experience any damage throughout the course of the disease. However, while there are patients that do not exhibit radiographic damage, it is unclear whether they have a specific form of the

disease that would mean they would never develop radiographic damage. A recent study[263], using primary cross-sectional data and secondary data from a systematic review, found that the prevalence of non-erosive RA was rare, only occurring in around 2% of established RA patients. The justification for the use of zero-inflated NB over NB regression is therefore not valid from a theoretical or statistical stand-point.

With longitudinal data, it is important that appropriate methods are used to account for the correlated data structure. This chapter looked at two popular methods, GEE and multi-level modelling. Given the interest of cluster-specific estimates, and the use of FMIL estimators, the multi-level procedure was favoured and this section went on to explore how the multi-level NB regression continued to provide best model fit to longitudinal data compared to multi-level Poisson regression methods.

This chapter also highlighted the importance of accounting for missing data in a study, particularly in longitudinal observational studies where missing data is likely to occur over time. The role of MICE in reducing the bias of complete case analysis was discussed, as well as the inherent benefit of multi-level models, given their use of FMIL estimators. Finally, the chapter looked at the importance of multivariate methods, as well as the correct methods for model building. While automated procedures, such as stepwise selection, remain popular, they are typically chosen based solely on their level of significance. This can result in two main problems. Firstly, it can omit variables of interest that while non-significant, are important to control for in order to produce unbiased estimates for the variables of interest. Secondly, they are often based on arbitrary thresholds of significance, which can lead to selection bias and overestimation of the standard errors and overall model fit[264]. This results in capitalisation on chance findings, where effect estimates are larger in the sample than might be found in the true population.

In summary, radiographic damage provides complex data that, in the context of longitudinal analysis, is increasingly more complex to model. This section has shown the appropriateness of count models, specifically NB regression in both the cross-sectional and longitudinal context. These models will be used throughout the thesis in order to establish the progression of radiographic damage in both the ERAS and ERAN cohort, as well as its association with disease activity and functional disability over time.

Chapter 6

Secular Declines in Radiographic Damage

6.1 Introduction

Published literature has suggested that the incidence of Rheumatoid Arthritis (RA) has declined over the last three decades[5, 19, 20, 265–270]. This corresponds with reports of declines in disease activity[271, 272], functional disability[226, 273], orthopaedic surgery [274] and radiographic progression[224, 226, 275].

While the causal nature of this decline is not entirely clear, it is hypothesised that these declines in disease severity are related to widespread changes in treatment strategies during the 1990s[229]. Data from RCTs has demonstrated that early initiation of conventional synthetic DMARDs can significantly improve patient outcomes, particularly the increased use of methotrexate in combination with other DMARDs[53, 142, 276, 277], and indeed biologic DMARDs[56, 149, 151, 152].

Radiographic joint damage is often used in RCTs as a primary outcome, and has been shown to be strongly related to levels of functional disability[86] and disease activity[177]. Although commonly expressed as a global score[278], radiographic joint damage comprises of two main components; erosions and joint space narrowing (JSN). While related, they are thought to be the result of two distinct pathophysiological mechanisms[279, 280]. Possible causes of erosive joint destruction is the product of invading synovium into the boney structures of the joints, and increased osteoclast activity[93]. Likewise, JSN has been hypothesised to reflect cartilage damage

as a result of metalloproteinases, which are upregulated by pro-inflammatory cytokines[94]. JSN is common to a range of pathologies, including osteoarthritis (OA), and is a common comorbid condition in people with RA[281]. Despite this, much of the focus of longitudinal data concerning radiographic damage has reported the combination of these two processes as one composite score[279, 280], for example using the radiographic scoring methods of Ratingen or Larsen, that lack the ability to distinguish progression of erosions and JSN as separate domains[224, 226, 275].

Further still, seropositive status has been strongly associated with worse radiographic progression[38, 282]. However, to date no study has looked at whether the relative strength of this association has changed given the wider demographic changes seen in many other aspects of RA, including disease severity. It might be hypothesised that radiographic measures of RA will show significant changes given recent declines in disease activity, but whether previously demonstrated risk factors for progression continue to be influential remains unclear.

The aim of this chapter is to present the first analysis looking at the progression of radiographic damage in both the ERAS and ERAN cohorts. This chapter will address the third aim of the thesis and look at the natural progression of radiographic damage in early RA. Furthermore, this chapter will look to extend the findings of the meta-analysis conducted in Chapter 3 and examine whether the progression of radiographic damage has changed between ERAS and ERAN. Using radiographic data from the SvdH, the total score, as well as the separate JSN and erosion score will be investigated. The review also highlighted how seropositive RA remains one of the strongest predictors of increased damage, so the association of seropositive RA on long-term radiographic progression within each cohort will also be examined.

6.2 Methods

6.2.1 Sample

The analysis focuses on a sub-sample of patients from both the ERAS and ERAN cohorts with SvdH data over the first 5 years follow-up. The sub-sample consists of a total of 1,662 with SvdH data. Of these, 1,216 were from ERAS, while 446 were from ERAN. Due to the availability of data from the ERAN cohort, the analysis was restricted to just the first 5 years. This was to allow direct comparisons between the cohorts, and although data up to 9 years was available,

the substantial amount of missing data would have resulted in imprecise estimates during years 5 to 9.

6.2.2 Statistical analysis

To assess differences in the use of first-line conventional DMARDs between the two cohorts, the cumulative incidence of time to first DMARD within the first 12 months from first outpatient appointment was estimated. This was estimated for any DMARD use, as well as separate estimates for the two most commonly used first-line DMARDS; methotrexate and sulphazalasine.

As was established in Chapter 5, a multi-level NB regression will be used to model the progression of radiographic damage in both cohorts. Cohort membership (either ERAS or ERAN) was the main covariate of interest. To assess the impact of missing data and the potential selection bias that may occur as a result, a sensitivity analysis based on the availability of radiographs will investigate the baseline characteristics of those with and without radiographic data. Furthermore, as detailed in Chapter 5, the use of multi-level models with FIML estimators ensures the use of all available data under the missing at random assumption, so that all patients with data are included.

To model the progression of radiographic damage, time was defined as years from enrolment and was included as a continuous variable. The maximum number of years follow-up was 5 years. Seropositive RA was the secondary covariate of interest and entered as a main effect, along with a three-way interaction term with cohort and time to allow for progression rates to be estimated separately by seropositivity status for each cohort. Sex, age, DAS28, HAQ, low Hb (<12/13), months from symptom onset to first rheumatology visit, steroid use prior to first assessment and DMARD use within first 12 months were all entered into the model to control for any potential confounding effects. To allow direct comparisons of the DAS score between the two cohorts, the DAS-44 score was transformed to the DAS28 score using the formula outlined by van Gestel et al.[67].

To aid interpretation, the results from the models were also expressed as an absolute change in the SvdH score using the estimated mean SvdH, along with 95% Confidence Intervals [95% CI]. This allowed for a more direct interpretation of the effect that each factor had in terms of absolute difference in SvdH units, the percentage of maximum possible damage, and annual

progression greater than the minimum clinically important difference of 5 units[283]. These models were estimated separately for the total SvdH score, JSN and erosion score.

6.3 Results

The demographic and baseline clinical characteristics of both ERAS and ERAN patients, including only those with radiographic data, are shown in Table 6.1. Reasons for missing radiographic data included loss of records, unreadable radiographs and loss to follow-up. Patients from ERAS were marginally younger at presentation and had higher DAS28, ESR, HAQ and more likely to be anaemic at baseline. Patient's characteristics with recorded radiographic data were similar to the total number of patients in their respective cohort.

Differences in treatment strategies between the two Cohorts

For all DMARDs, ERAS reported a 12-month cumulative incidence of 71.6% [95%CI 69.2-73.8] and for ERAN 95.3% [95%CI 93.9-96.4] (See Figure 6.1). The 12-month cumulative incidence of sulfasalazine use was higher in ERAS (55% [95%CI 52.4-57.5]) than ERAN (33.1% [95%CI 30.4-35.8]), while methotrexate use was substantially lower in ERAS (1.4% [95%CI 0.9-2.1]) compared to ERAN (52.1% [95%CI 49.2-55.0]).

6.3.1 Radiographic progression rates of ERAS and ERAN

For the Mixed Effects Negative-Binomial Regression (MENBR) analysis, a total of 1,508 patients contributing 5,430 observations (mean observations per patient = 3.6) were included. Overall, the ERAN cohort exhibited a 41% lower total SvdH score at baseline compared to ERAS (Exponentiated coefficient 0.59 [95%CI 0.50-0.70], $p<0.001$), along with a 65% slower annual rate of progression over the first 5 years (Exponentiated coefficient 0.35 [95%CI 0.24-0.47], $p<0.001$) (See Figure 6.2). The differences in absolute and relative scores for both cohorts are shown in Table 6.2. When expressed as a proportion of maximum possible damage, the estimated values indicated an increase of 1.5% [95%CI 1.4-1.7] per year for ERAS and 0.6% [95%CI 0.4-0.7] per year for ERAN. The total proportion of patients who had annual progression estimated to be greater than the MCID (>5 SvdH units) was 74% for ERAS and 27% for ERAN.

	ERAS				ERAN				**ERAS+ERAN**			
	Total	With SvdH	Without SvdH	Missing (%)	Total	With SvdH	Without SvdH	Missing (%)	**Total**	**With SvdH**	**Without SvdH**	**Missing (%)**
Demographics												
Year of Recruitment	1986-2001	1986-2001	1986-2001	0 (0)	2002-2013	2002-2013	2002-2013	0 (0)	1986-2013	1986-2013	1986-2013	0 (0)
Age at Onset (Mean (SD))	55.3 (14.6)	54.9 (14.5)	57.4 (14.9)	0 (0)	57.1 (14)	58 (13.5)	56.6 (14.3)	0 (0)	56.1 (14.4)	55.7 (14.3)	56.8 (14.4)	0 (0)
Female (%)	66	66	69	0 (0)	68	65	70	0 (0)	67	65	70	0 (0)
Clinical Markers												
Seropositive (%)	63	64	58	9 (0.1)	60	61	60	142 (11)	62	63	59	142 (11)
Baseline DAS (Mean (SD))	6.32 (1.33)	6.32 (1.33)	5.02 (1.26)	13 (0.1)	4.53 (1.58)	4.5 (1.64)	4.5 (1.5)	46 (4)	5.51 (1.7)	5.84 (1.62)	4.66 (1.49)	46 (4)
Baseline ESR (Median (IQR))	37 (44)	38 (44)	36 (40)	7 (0.1)	24 (29)	21 (28)	24 (30)	183 (15)	30 (39)	34 (41)	26 (32)	183 (15)
Baseline HAQ (Mean (SD))	1.15 (0.8)	1.15 (0.8)	1.14 (0.8)	5 (0.1)	1.08 (0.8)	1.03 (0.8)	1.11 (0.76)	37 (3)	1.12 (0.8)	1.12 (0.8)	1.11 (0.77)	37 (3)
Baseline Anaemia (%)	41	42	49	5 (0.1)	28	24	29	32 (3)	35	37	32	32 (3)
Months from symptom onset to First Visit (Median (IQR))	6 (7)	6.5 (7)	6 (8)	0 (0)	6 (8)	6 (8)	6 (9)	91 (7)	6 (8)	6 (8)	6 (8)	91 (7)
Observations	1465	1216	249	-	1236	446	790	-	2701	1662	1039	-

TABLE 6.1: Summary statistics for each cohort, both with and without radiographic data

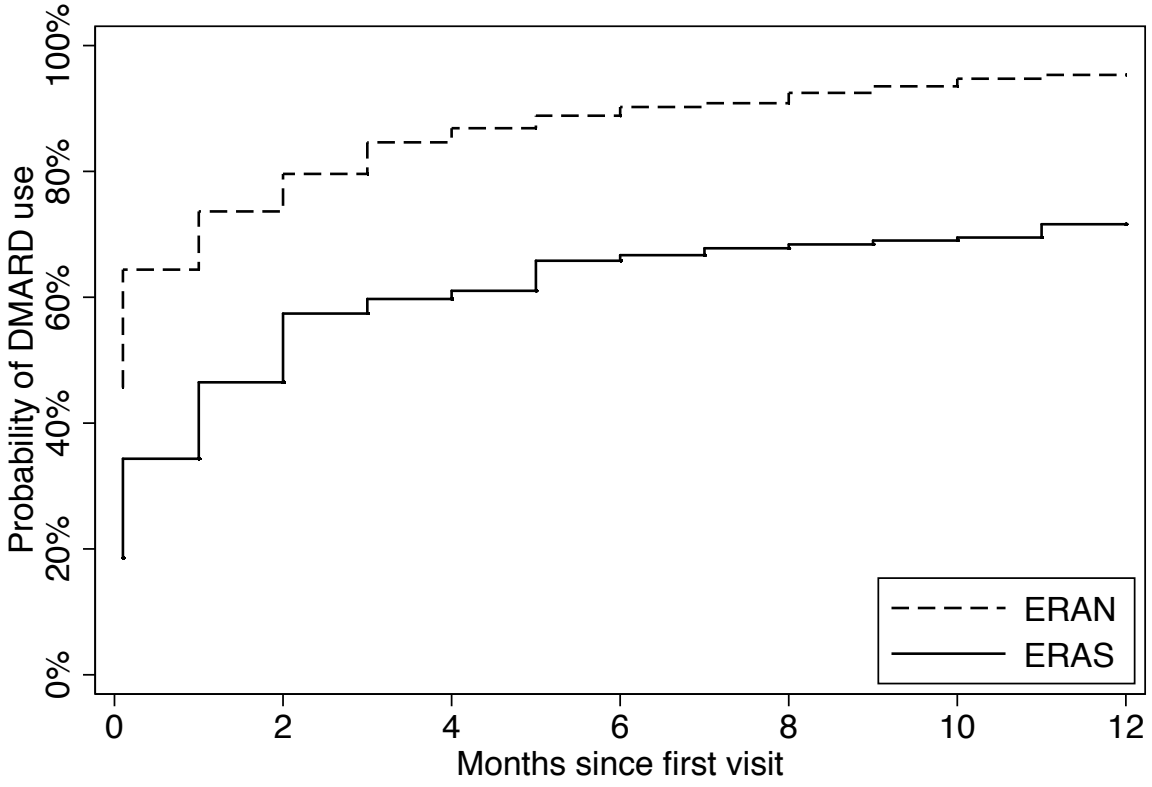


FIGURE 6.1: 12-month cumulative incidence of DMARD use for ERAS and ERAN

Similar results were seen for the JSN score, with ERAN participants displaying lower scores at baseline (Exponentiated coefficient 0.49 [95%CI 0.41-0.58], p<0.001) and a slower annual rate of progression over the first 5-years compared to ERAS (Exponentiated coefficient 0.31 [95%CI 0.21-0.42], p<0.001) (See Figure 6.3A). For the erosion score, the score at baseline was similar for both cohorts (Exponentiated coefficient 0.94 [95%CI 0.73-1.19], p=0.593), however, ERAN exhibited a slower annual rate of progression over the first 5-years compared to ERAS (Exponentiated coefficient 0.43 [95%CI 0.25-0.61], p<0.001) (See Figure 6.3B). See Table 6.2 for absolute and relative changes in both JSN and erosion scores between the two cohorts.

6.3.2 Association of seropositivity with radiographic progression

The absolute and relative difference in total SvdH scores for seropositive and seronegative patients in both cohorts are given in Table 6.3 and displayed graphically in Figure 6.4. For the total SvdH score, seropositive RA was not significantly associated with increased radiographic damage at baseline, compared to seronegative RA in either ERAS or ERAN. Seropositive RA

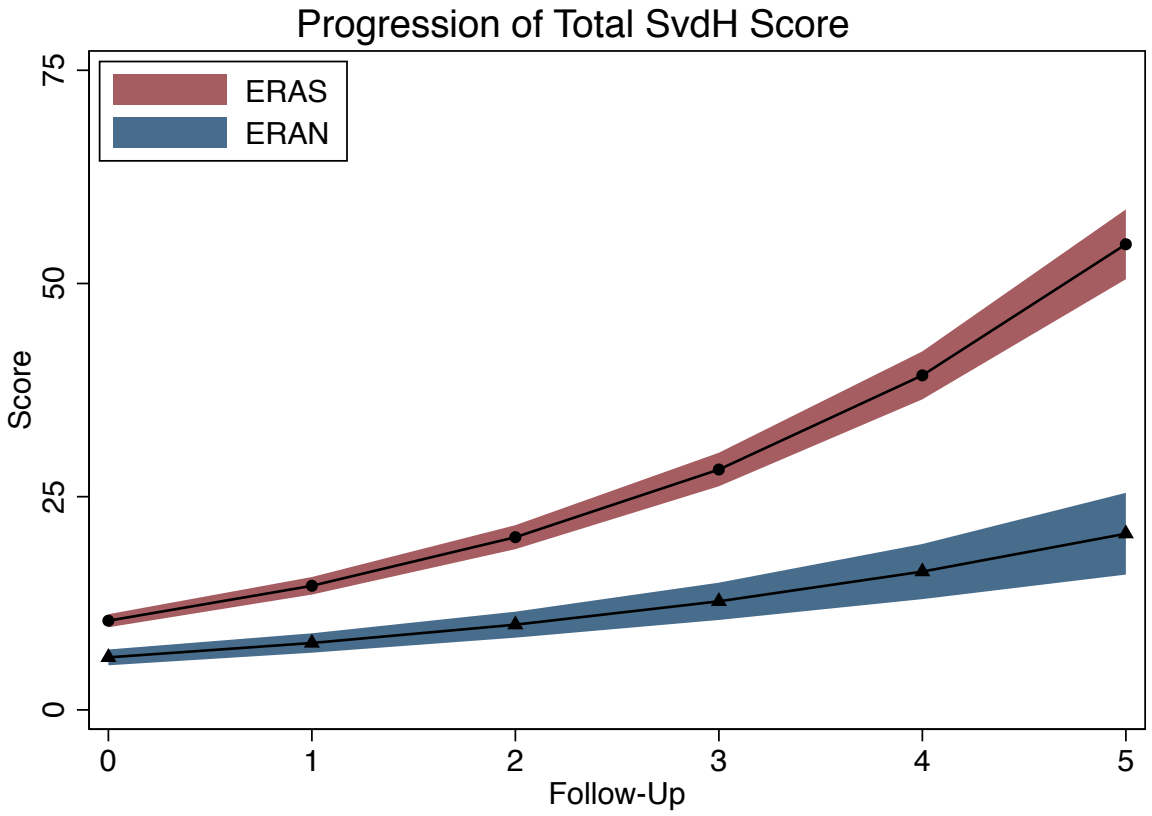


FIGURE 6.2: Progression of Total SvdH score for ERAS and ERAN. Shaded red and blue area denotes the 95% Confidence Intervals for ERAS and ERAN respectively

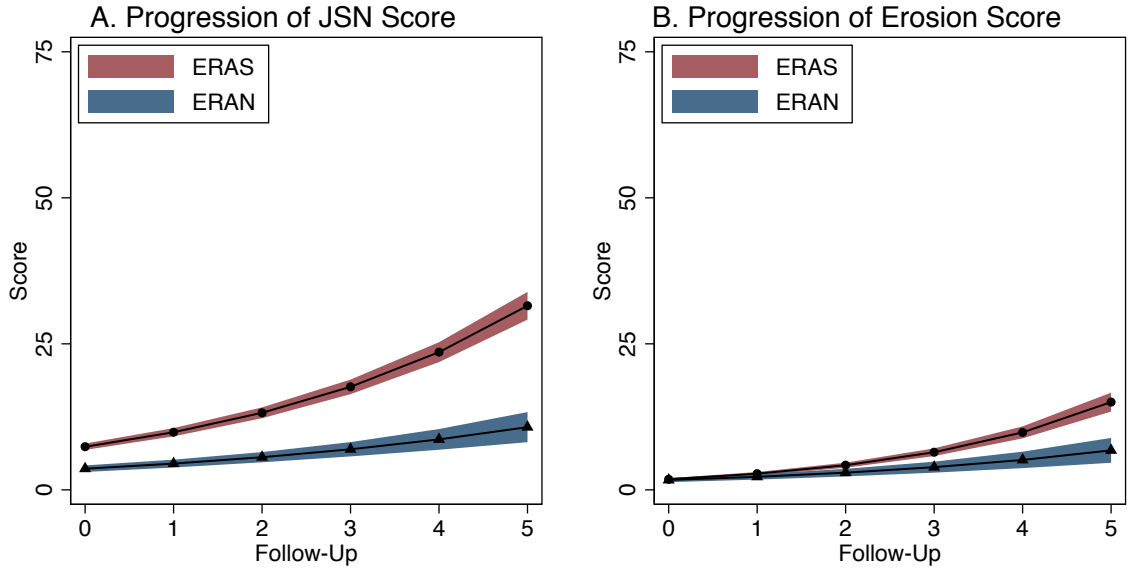


FIGURE 6.3: Progression of JSN and erosion score for ERAS and ERAN. Shaded red and blue area denotes the 95% Confidence Intervals for ERAS and ERAN respectively

was associated with a 70% increased annual rate of progression, compared to seronegative RA, in ERAS, which was statistically significant (Exponentiated coefficient 1.70 [95% CI 1.42-1.97],

Estimated means	ERAS	ERAN	Absolute Difference	Relative Difference (IRR) [95% CI]	P-Value
Total SvdH at baseline	10.5	6.2	4.3	**0.59 [0.50-0.70]**	**<0.001**
Total SvdH annual rate	6.9	2.5	4.5	**0.35 [0.24-0.47]**	**<0.001**
JSN score at baseline	7.4	3.6	3.8	**0.49 [0.41-0.58]**	**<0.001**
JSN score annual rate	3.9	1.2	2.7	**0.31 [0.21-0.42]**	**<0.001**
Erosion score at baseline	1.8	1.7	0.1	0.94 [0.73-1.19]	0.593
Erosion score annual rate	1.9	0.8	1.1	**0.43 [0.25-0.61]**	**<0.001**

TABLE 6.2: Mean and relative difference in baseline level and annual rate of progression for Total SvdH, JSN and erosion scores between ERAS and ERAN

p<0.001). The annual rate of progression for seropositive RA, compared to seronegative RA, in ERAN was increased by 9%, which was not significant (Exponentiated coefficient 1.09 [95% CI 0.51-1.67], p=0.855). This relates to decreases in the relative impact of seropositive RA on the annual rate of progression of 36% for ERAN compared to ERAS, which although considerable was non-significant, as indicated by the three-way interaction in the model (Exponentiated coefficient 0.64 [95%CI 0.29-1.07], p=0.224). The estimated proportion of seropositive patients with an annual progression greater than the MCID was 80% for ERAS, and just 29% for ERAN.

		RF-	RF+	Difference	Relative Difference (IRR) [95% CI]	P-Value
ERAS	Total SvdH at baseline	9.5	11	1.5	1.16 [1.00-1.35]	0.056
	Total SvdH Annual rate	5.1	8.6	3.6	**1.70 [1.42-1.97]**	**<0.001**
ERAN	Total SvdH at baseline	6.0	6.2	0.2	1.04 [0.76-1.42]	0.811
	Total SvdH Annual rate	1.9	2.0	0.2	1.09 [0.51-1.67]	0.855

TABLE 6.3: Mean and relative difference in baseline level and annual rate of progression for Total SvdH score between seropositive and seronegative patients by both ERAS and ERAN

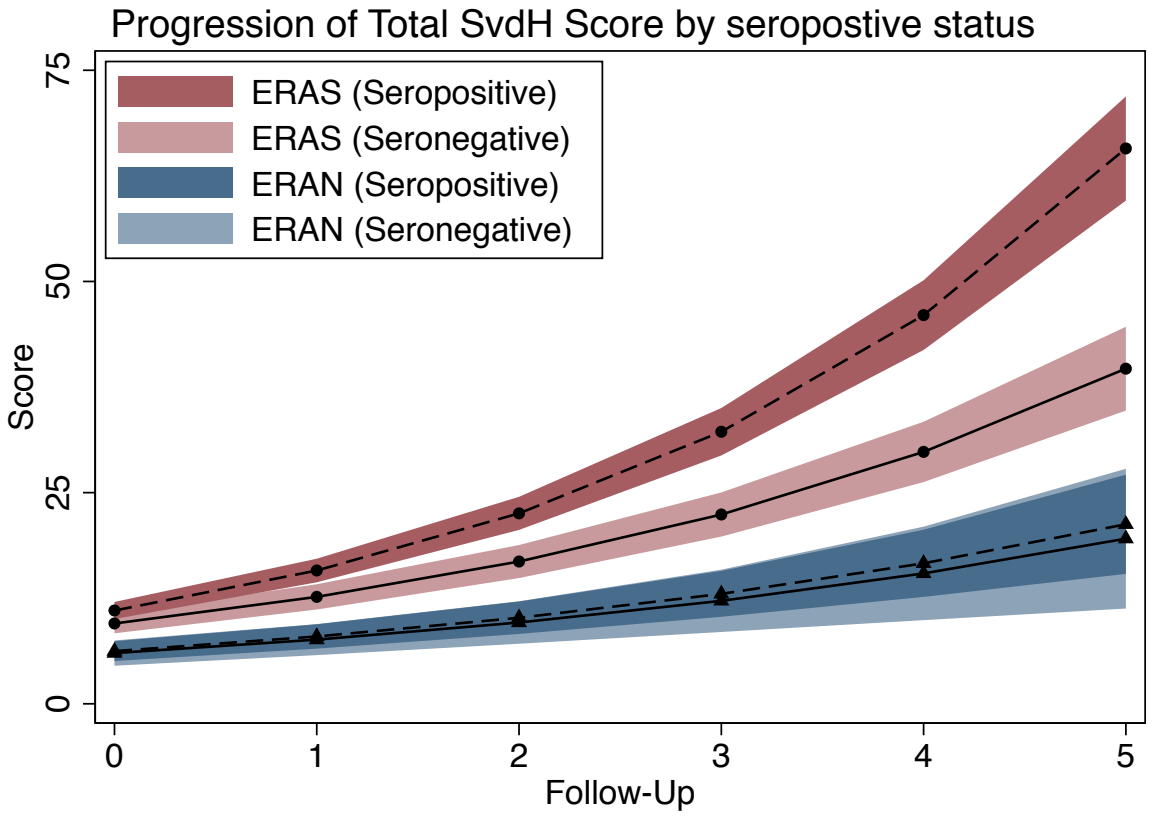


FIGURE 6.4: Progression of Total SvdH score for ERAS and ERAN by seropostive status. Shaded red and blue area denotes the 95% Confidence Intervals for ERAS and ERAN respectively

The influence of seropositive RA on the separate JSN compared to seronegative RA indicated a similar pattern to the total SvdH score (See Table 6.4 and Figure 6.5A). Seropositive RA was not associated with increased JSN scores at baseline for either ERAS (Exponentiated coefficient 1.12 [95% CI 0.96-1.31], p=0.152) or ERAN (Exponentiated coefficient 0.92 [95% CI 0.67-1.27], p=0.619), but was associated with a relative increase in the annual rate of progression. When compared to seronegative RA, seropositive RA indicated a statistically significant increase of 58% (Exponentiated coefficient 1.58 [95% CI 1.32-1.84], p<0.001) and a statistically non-significant increase of 19% (Exponentiated coefficient 1.19 [95% CI 0.35-2.02], p=0.480) for ERAS and ERAN respectively. The relative difference in the increased annual rate of progression of JSN scores for seropositive RA in ERAS compared to ERAN was not significantly different, as indicated by the three-way interaction term in the model (Exponentiated coefficient 0.72 [95% CI 0.26-1.18], p=650).

Differing from the JSN score, seropositive RA was associated with a 40% increased erosion score at baseline, compared to seronegative RA, in ERAS (Exponentiated coefficient 1.40 [95%

CI 1.12-1.75], p=0.003). The relative difference was reduced and non-significant for ERAN at baseline (Exponentiated coefficient 1.17 [95% CI 0.74-1.85], p=0.495). Seropositive RA was associated with relative increases in the annual rate of progression, compared to seronegative RA, with a statistically significant increase of 107% (Exponentiated coefficient 2.07 [95% CI 1.61-2.52], p<0.001) and statistically non-significany increase of 21% (Exponentiated coefficient 1.21 [95% CI 0.28-2.14], p= 0.929) for ERAS and ERAN respectively (See Table 6.4 and Figure 6.5B). The relative impact of seropositive RA was reduced in ERAN, however this was non-significantly different when compared to ERAS, as indicated by the three-way interaction term in the model (Exponentiated coefficient 0.92 [95% CI 0.81-1.05], p=0.234).

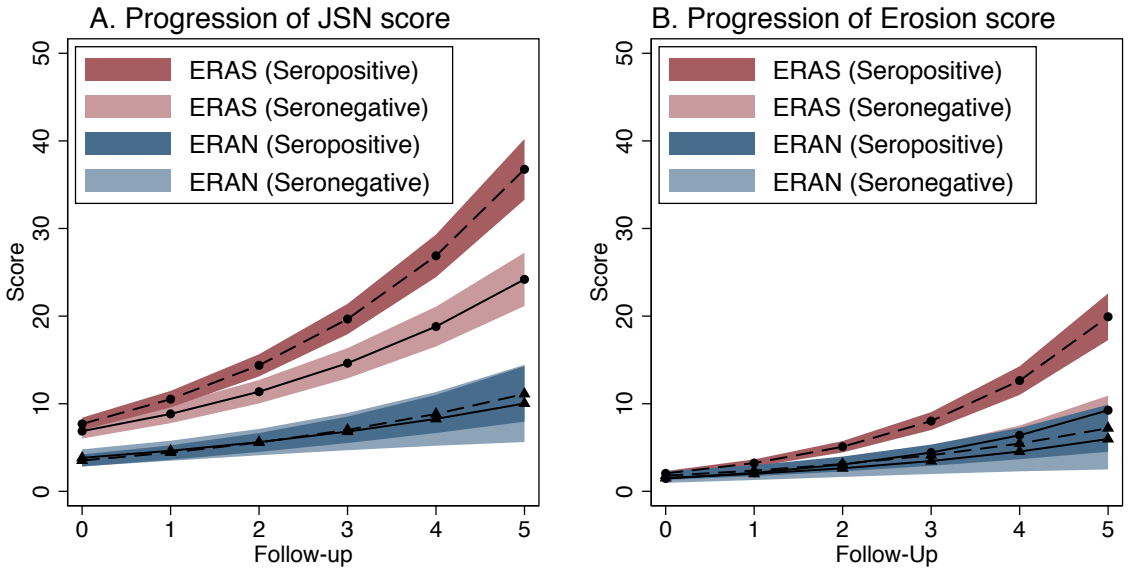


FIGURE 6.5: Progression of the JSN and erosion score for ERAS and ERAN by seropostive status. Shaded red and blue area denotes the 95% Confidence Intervals for ERAS and ERAN respectively

6.4 Discussion

The findings from the analysis indicates that patients with early RA with onset from 2002-2013 (ERAN) had significantly lower baseline and annual rates of radiographic progression compared to those with onset from 1986-2001 (ERAS). Examination of the separate erosion and JSN scores indicate that the reduction in the total SvdH score was largely driven by reductions in JSN. Strikingly, the strong association of seropositivity and increased radiographic progression in ERAS was markedly diminished in ERAN. The reduction in the impact of seropositive RA was such that those with seropositive status in the ERAN cohort had markedly better radiographic outcomes at 5 years than those with seronegative RA in ERAS.

		RF-	RF+	Absolute Difference	Relative Difference (IRR) [95% CI]	P-Value
ERAS	Total JSN at baseline	6.9	7.7	0.8	1.12 [0.96-1.31]	0.152
	Total JSN Annual rate	2.9	4.6	1.7	**1.58 [1.32-1.84]**	**<0.001**
ERAN	Total JSN at baseline	3.8	3.5	0.3	0.92 [0.67-1.27]	0.619
	Total JSN Annual rate	1.1	1.3	0.2	1.19 [0.35-2.02]	0.480
ERAS	Total Erosion at baseline	1.5	2.0	0.6	**1.40 [1.12-1.75]**	**0.003**
	Total Erosion Annual rate	1.2	2.4	1.3	**2.07 [1.61-2.52]**	**<0.001**
ERAN	Total Erosion at baseline	1.5	1.8	0.3	1.17 [0.74-1.85]	0.495
	Total Erosion Annual rate	0.7	0.9	0.2	1.21 [0.28-2.14]	0.929

TABLE 6.4: Mean and relative difference in baseline level and annual rate of progression for Total SvdH score between seropositive and seronegative patients by both ERAS and ERAN

Previous research has indicated that a change of 5 SvdH units indicates a minimal clinically important difference[247, 283], therefore a difference of 5 units per year for ERAN compared to ERAS on total SvdH score observed in this study demonstrates not only a statistically significant change in progression, but also a clinically meaningful reduction. Whereas 74% of patients in ERAS progressed, on average, >5 units per year over the 5 year period of follow-up considered, just 27% of patients in ERAN exhibited similar levels of progression.

The data extends previous findings of reductions in radiographic damage in RA over recent decades[224, 226, 275]. There are two plausible explanations for these findings, both of which are likely to contribute to the reduction in radiographic damage over time. Firstly, RA may have become milder, and secondly, earlier more intensive treatment may have improved disease outcomes. Our models adjusted for disease severity at baseline, but it remains possible that lower rates of progression in the more recent cohort reflect milder disease. This is supported by the observation of lower SvdH scores in ERAN compared to ERAS at baseline, prior to DMARD initiation for the majority of patients. However, the dramatic reductions in radiographic progression, particularly the reduced impact of seropositive RA, is likely to also reflect improvements in the treatment of RA, given the earlier and increased use of methotrexate as

the first line DMARD observed in ERAN. This is in line with other reports[224, 226, 275]. Increasing evidence from RCTs also support the hypothesis that early, intensive treatment has an important effect on reducing radiographic progression[140, 141, 143, 150, 153].

Separate investigation of the erosion and JSN components of radiographic damage scores showed that JSN was the primary driver for the overall reductions seen in the total SvdH score between the two cohorts. This finding reiterates the importance of reporting both the erosion and JSN score separately in clinical trials. Data from ASPIRE show that more patients with early RA have either erosions alone (8.5%) or JSN alone (4.4%), than both (3.7%) at baseline visit[280], and that JSN may be more strongly associated with irreversible disability[279]. Despite this, the separate scores are still rarely reported[278]. If early treatment with methotrexate was the primary cause for the reduction in total SvdH in ERAN, this could indicate that the mechanism by which this is achieved is through the reduction of JSN and preservation of the surrounding cartilage. However, what is not clear is whether the JSN is directly attributable to RA JSN, or OA JSN. A high prevalence of radiographic OA has been documented at baseline in the ERAN cohort in the hands and feet, indicating that high levels of comorbid OA could potentially confound any radiographic assessment of RA[281]. High JSN scores are strongly associated with increased severity of OA osteophytosis and OA JSN[284]. More studies are needed to quantify the exact effect that comorbid OA could be having on RA radiographic scoring.

Seropositive RA has been consistently associated with increased radiographic damage[38, 282]. This analysis also found that seropositive RA was highly associated with increased radiographic progression. While the later cohort indicated non-significant increases in radiographic damage when comparing seropositive and seronegative patients, the difference between seropositivity between the two cohorts was non-significant. This is most likely due to a lack of power, with reduced patient numbers in the later cohort. However, when investigating the absolute change in radiographic score between seropositive and seronegative patients across the two cohorts, seropositive patients in the later cohort no longer represented a patient sub-group with clinically meaningful increases in radiographic progression, at least within the first 5 years of disease. Aletaha et al.[285] analysed the effect of seropositive status on radiographic progression and found seropositive patients displayed higher radiographic progression, compared to seronegative patients. The estimated change in median SvdH score of 0.6 units per year for seropositive over that of seronegative patients provides an estimate similar to this study. It should be noted that seropositive status was primarily based on RF positivity, with only a small number of ERAN patients having data on ACPA positivity. Research is beginning to illuminate the advantage

of ACPA positivity over RF as a more specific antibody in predicting increased radiographic progression[286–289].

6.4.1 Strengths and limitations

Many RCTs are restricted to seropositive patients only, and previous research has not focused on the effect of seropositivity in the context of reduced radiographic progression in more recent years. The two long-term observational cohorts examined in this study provide a 'real-world' account of patients typically seen in secondary care, and the high patient numbers over the full 5 year follow-up also provides a unique opportunity to provide precise estimates using the modelling techniques outlined[126]. The use of the SvdH score also provides a first look at the two principle components of radiographic damage, erosions and JSN, in detail. Further data from observational studies are needed to ascertain whether reductions in radiographic progression have also resulted in the diminished association with seropositive status, particularly in the context of anti-CCP seropositive RA, which could be more predictive of radiographic progression when compared to RF[286–288].

Our research is subject to a number of limitations inherent in cohort studies. Recruiting centres were hosted by enthusiastic clinicians within the UK and, although they might not necessarily reflect people with RA in other contents, or subjected to different treatment regimens, the multi-centre recruitment for these cohorts from district general hospitals is likely to be representative of people with RA in the UK. Radiographs were not available for all participants, and it is possible that those with more severe disease were more likely to have x-rays, increasing the risk of selection bias in our study. However, baseline variables indicated minimal differences between the whole cohorts, and those for whom radiographic data was available. The impact of such a selection bias would overestimate rates of progression, particularly for ERAN, where data was less complete; hence our estimates should be treated as conservative.

6.5 Concluding remarks

The aim of this chapter was to address the third aim of the thesis; to examine the natural progression of radiographic damage. This analysis provides further evidence into the marked reduction in radiographic damage over the last 30 years, while providing accurate, quantified estimates of the extent of that reduction. JSN was the major driver for the overall reductions

seen, and highlights the importance of investigating JSN and erosions separately when investigating radiographic damage. Advances in treatment are likely to be the main cause for the decline, and adequate DMARD treatment might remove the predictive value of seropositivity for radiographic progression in early RA. There is also evidence that RA is getting milder as a disease. Further research should seek other predictors and mediators if residual radiographic progression despite DMARD treatment is to be halted. The significance of these reductions on patients of varying disease severity, and whether these reductions have led to improved long-term functional disability will be crucial in fully realising the impact of these results on clinical care.

Chapter 7

Radiographic Progression, Disease Activity and Functional Disability

7.1 Introduction

The systematic review conducted in Chapter 3 identified several major predictors of radiographic progression. There is overwhelming evidence to suggest that seropositive RA (RF and Anti-CCP) and increased acute phase markers (ESR and CRP) are associated with increased radiographic progression. In Chapter 6 radiographic progression rates were examined in both the ERAS and ERAN cohorts, indicating that overall rates of radiographic progression have significantly decreased between the two cohorts, possibly due to a combination of milder disease and improved treatments. One key finding from this analysis was that seropositive RA was significantly associated with increased radiographic damage. However, large reductions in overall rates of radiographic damage in ERAN meant that the effect of seropositive RA on increased radiologic damage was relatively modest when compared to ERAS.

The aim of this chapter is to address the fourth and final aim of the thesis; to expand the models from Chapter 6 and investigate the long-term relationship with two important clinical outcomes; disease activity and functional disability, with radiographic progression over time. The most common measure of disease activity in RA is the DAS, which incorporates an acute phase marker, along with other clinical and patient reported measures of disease activity. The aim is to quantify disease activity as a simple clinical tool to guide therapeutic management. Disease activity indices (DAI) have become more popular in day-to-day clinical practice, as well

as being used as a primary outcome measure in clinical trials, since they are more reliable as an overall measure of disease activity compared to their individual components[177]. They have also grown particularly fast in the last few years with the advocation of the 'Treat-2-Target' (T2T) initiative in the UK and across Europe[49, 57], where DAS cut-offs are the target.

In the UK, disease states that are defined by thresholds of the DAS28 form the basis of biologic DMARD prescription decisions, with only those patients with a DAS28 >5.1 eligible for biologic DMARDS. Those patients with moderate disease (i.e. those falling just below the 5.1 threshold) are of particular interest, as there has been emerging evidence that these patients exhibit significant levels of radiographic progression, with some reports indicating that they exhibit the same level as those with high disease[176, 290]. Combè et al. found that those patients with sustained moderate disease (DAS28 3.2-5.1) continue to progress radiographically[176] compared to those in low disease states, and other studies have found that those patients with sustained moderate RA on methotrexate continued to exhibit significant radiological progression (SvdH >3 from baseline to years 2 and 3), with one third of patients (n=33) showing a change in SvdH score from baseline to 3 years of >3 units[290]. It is currently unclear whether standard conventional DMARD therapies alone are appropriate in controlling the disease for this patient sub-group, and forms the basis of the TITRATE project; a multi-centre trial funded by a National Institute for Health Research (NIHR) Programme Grant. Along with the collection of primary data from RCTs, this project will also be supplemented using data from observational cohorts, including ERAS and ERAN, to ascertain if there is currently an unmet clinical need in this patient group. Furthermore, recent evidence from the ERAS and ERAN cohorts highlight the consequences of sustained moderate disease on functional disability and orthopaedic surgical rates[74], demonstrating the need to examine the use of biologic DMARDs in moderate disease activity based on these outcomes alone. Studies have already shown the benefit of the early use of biologics in both moderate and high disease activity groups in other countries applying the T2T philosophy[73, 291], although the economic argument of whether this is cost-effective is not clear.

Early theories regarding the relationship between disease activity, radiographic damage and functional disability were suggestive of a monotonic relationship[85]. Increased inflammation caused by the disease would lead to an increase in the destruction of the bones and cartilage. Over time, this would lead to increased disability through painful deformities in the joints, ultimately limiting the patients ability to carry out tasks in everyday living. This monotonic relationship can be illustrated using a box and arrow diagram, as displayed in Figure 7.1.

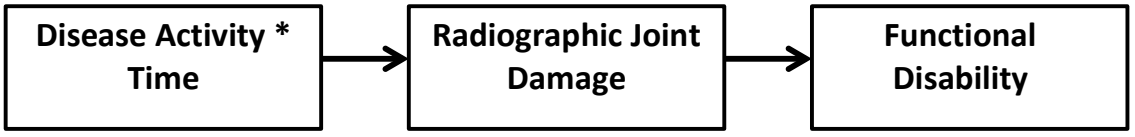


FIGURE 7.1: Box and arrow diagram presenting a single pathway monotonic relationship between disease activity, radiographic damage and functional disability

However, it became apparent from longitudinal analysis that radiographic progression and functional disability had very different patterns of progression over time. Radiographic assessments using the Larsen and SvdH scores were found to increase steadily over time[38, 85], while functional disability measured using the HAQ was found to follow a 'J-shaped' trajectory over time. Unlike radiographic joint damage, HAQ scores improved in the first year as treatments were introduced, then steadily rose over the coming years as the disease progressed[38, 85, 292]. These differences in long-term trajectories, coupled with weak correlations between the two outcomes in early disease, are suggestive of two distinct disease processes, which differ between early onset RA and established RA[293, 294]. This was originally conceptualised by Kirwan in 1992[295], who described the interrelationship between disease processes (e.g. inflammation) and disease outcomes (e.g. disability) and how they interact in the prognosis of RA. Kirwan explained why it is important to describe this interrelationship over different stages of the disease; from early onset, through established, to late disease[295], as the disease progresses and its interaction with disease outcomes evolves over time. As a result, the prevailing theory today is that acute inflammation causing pain and swelling in the joints during early RA leads to increased functional disability at disease onset, which improves as treatment using NSAIDs, steroids and DMARDs are introduced. When not suitably controlled, the increased disease activity then leads to radiographic damage over time, resulting in increased functional disability in later disease[214, 296].

In 2002, Escalante and Ricón[297] presented the theory of two distinct pathways that lead to the development of disability in RA; one caused by pain, and the other by damage. Figure 7.2 indicates how both the pain and damage pathway exist as separate mechanisms of disability within the main pathway. These pathways can be further developed into a biopsychosocial model, where these distinct stages of the disease are also moderated by other external factors, particularly psychosocial factors. In this model, depression, stress and social support can influence all stages of the disease process and ultimately impact on the severity of functional

disability[297].

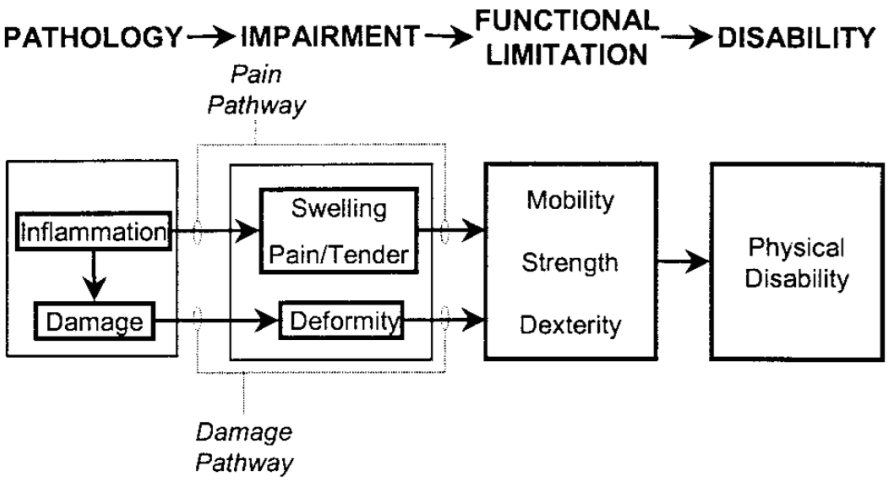


FIGURE 7.2: Box and arrow diagram presenting the revised pathway to disability from Escalante and Ricón

Two systematic reviews examining the relationship between radiographic damage with disease activity[177] and with functional disability[86] have identified a tendency for the associations in both to increase over disease duration. However, both note a paucity of studies using effective multivariate analysis methods, indicating the need for a more uniform approach to allow direct comparisons.

Studies by Welsing et al.[214] and Salaffi et al.[211] depict the two popular methods in which the longitudinal relationship between disease activity and radiographic progression can be analysed. Salaffi et al.[211] investigated the association between radiographic progression and disease activity by using time-averaged measures of disease activity. This allows for large longitudinal data to be condensed into one explanatory variable. This has the benefit of being easier to model, as well as being easier to interpret. However, this reductionist approach means that trends on a more sensitive level are not possible. Yearly variations in disease activity may be relatively low over a 3 year follow-up and using time-averaged approaches over longer time periods may neglect important variation in disease activity over time, which could be important in understanding the temporal relationship[214].

Welsing et al.[214] acknowledged these limitations and looked at the use of time-varying covariates. Using GEE techniques have non-independence of assessments within individuals correlational structure and can be modelled via a residual correlational structure with respect to radiographic outcomes over time. This allows for more sensitive time trends to be examined,

and therefore the association between disease activity and radiographic damage can be examined on a yearly basis. However, Welsing et al.[214] modelled the radiographic data based on a link identity function for their GEE, which does not accurately account for the non-normal distribution, and therefore it is likely that the model estimates are biased by the overestimation of mean scores.

The aim of this chapter is to explore the longitudinal relationship between radiographic joint damage with disease activity and then with functional disability. Using the large sample of radiographic data from the ERAS cohort, the first analysis will investigate the association between 10 year radiographic damage with disease activity and functional disability, expressed as a time-averaged covariate. The second analysis will then investigate these associations using disease activity and functional disability as time-dependent covariates. In contrast to the study by Welsing et al.[214], both modelling techniques will use multi-level negative binomial regression as explained in Chapter 5. Finally, to investigate how the declines in radiographic progression identified in Chapter 6 have impacted on these associations, a sub-analysis restricted to 5 years will incorporate the sub-sample of radiographic data from the ERAN cohort. With regards to the association between radiographic progression and disease activity, those patients with moderate disease activity signify a particularly interesting clinical group and therefore will be the focus of this analysis.

7.2 Methods

7.2.1 Sample

For the ERAS cohort, a total of 1,216 patients (83%) have radiographic data. While radiographic data is available for up to 19 years, and disease activity and functional disability data is available for up to 15 years, there is substantial data attrition beyond year 10. The original aim of the cohort was to collect follow-up data for up to a minimum of 5 years, which was subsequenty altered to 10 years. The latter was achieved by all of the centres involved, but several ceased follow-up after 10 years. This is highlighted in Figure 7.3, which indicates the proportion of missing data, and the proportion of patients with no data due to no follow-up, for both the DAS and HAQ score over the first 15 years. For the DAS score, the number of patients with data at year 10 is 604 (41%), whilst at year 11 this decreases to just 328 (22%). A similar rate of attrition is seen for the HAQ score, where the number of patients with complete data at year

10 is 686 (47%), which then decreases by a similar magnitude to just 380 (26%) by year 11. As such, the decision was made to limit the analyses to the first 10 years.

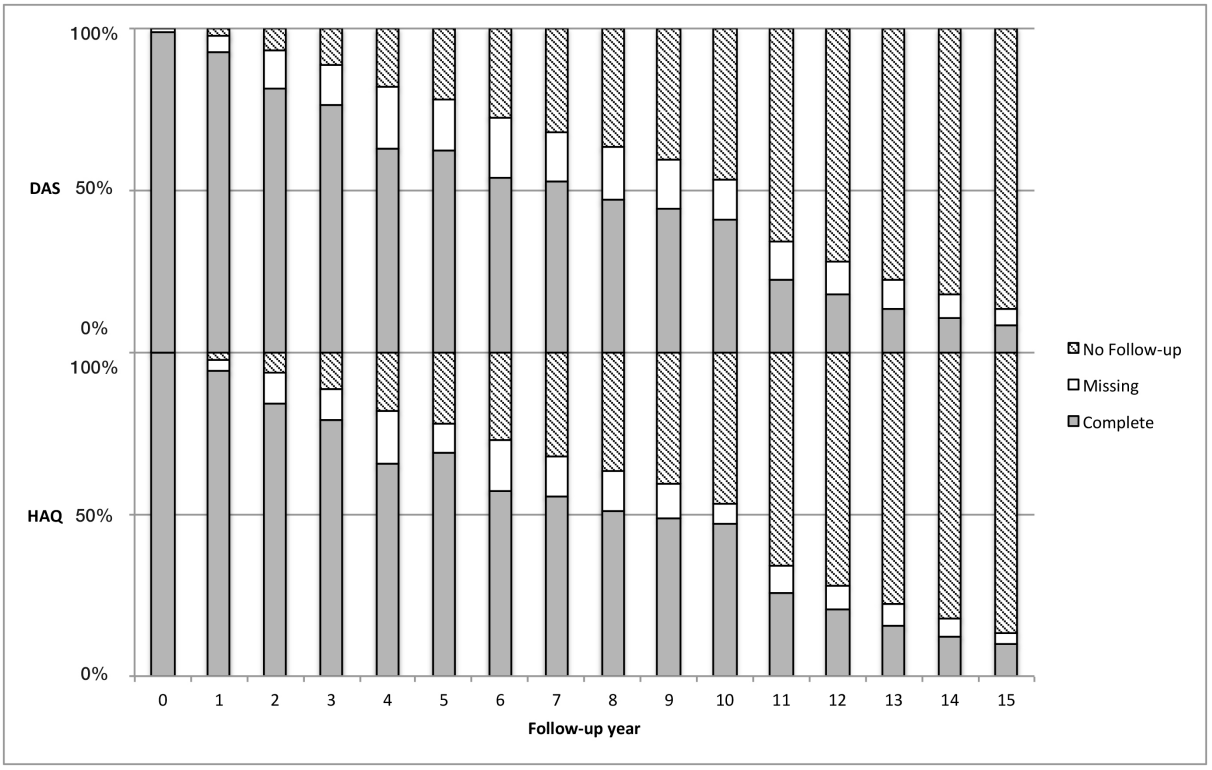


FIGURE 7.3: Complete and missing data for the DAS and HAQ score

The secondary validation analysis extends the models to include a sub-sample of 446 (36%) ERAN patients with radiographic data. While the level of missing data for the DAS and HAQ is similar to the ERAS cohort for all the patients, the relatively small sample with radiographic data restricts the follow-up data to just 5 years.

DAS measure and DAS EULAR Categories

The disease activity score (DAS) was collected for each patient at baseline, 6 months and yearly follow-up. For ERAS, the 3 variable DAS 44-count method was used. This is an overall disease score that is calculated from a tender joint count (TJC), swollen joint count (SJC) and acute phase marker (ESR) using a pre-defined equation[64, 298]. While typically cited as a continuous measure, EULAR proposed pre-defined cut-points for the DAS that could allow for clinicians to categorise patients into low, moderate and high disease groups[70]. For the DAS-44 measure, the groups are defined as <2.4 for low DAS, 2.4-3.7 for moderate DAS and >3.7 for high DAS.

For the ERAN cohort, the four variable DAS28 scoring method was used, which calculates the score using the 28-count version of SJC and TJC, along with an acute phase marker (ESR) and a patient global score of pain[65, 299]. A pain score was also collected in ERAS, but was not used as part of the calculation for the DAS-44. As with the analysis in Chapter 6, to combine the data from both ERAS and ERAN, the DAS-44 score was transformed to the DAS28 score using the formula outlined by van Gestel et al.[67].

Functional disability

Functional disability was measured using the anglicised version Health Assessment Questionnaire (HAQ) devised by Kirwan and Reeback in 1986[233]. The HAQ score measures disability using a 20-item questionnaire over 8 core domains. These questions assess the patient's ability to perform daily activities. Each item is scored from 0 to 3, where 0 indicates 'without difficulty', 1 indicates 'with some difficult', 2 indicates 'with much difficulty' and 3 indicates 'not able to do the activity'. The questionnaire also includes information about aids and devices used, and the score is adjusted to take these into account. As with the DAS score, HAQ was assessed at baseline, 6 months, 12 months and then yearly thereafter.

7.3 Analysis

7.3.1 Radiographic damage and disease activity

The first stage of the analysis was to examine the correlation structure between the SvdH score and the DAS, including the separate components that make up the DAS (ESR, TJC and SJC) over the first 10 years follow-up. Chapter 5 found that the SvdH score was best analysed using the negative-binomial model, which analyses the score based on the log count. As such, the correlations examine the total SvdH based on the natural log transformation (+1) (lnSvdH). Similarly, the distribution of the ESR indicates that the correlation should be based on the log transformed score (lnESR). The pairwise correlation matrices between the lnSvdH score, the DAS score and the lnESR over the first 10 years are shown in Tables 7.1. In order to establish patterns in the correlation matrix, those correlations defined as having a 'moderate' effect (0.3-0.4) where highlighted. Those with a Pearson's r of greater or equal to 0.3 but less than 0.4

are highlighted in light blue, and those with a Pearson's r of greater than or equal to 0.4 are highlighted in dark blue.

Overall, the correlations between DAS and lnESR with lnSvdH ranged from weak (0.1-0.2) to moderate (0.3-0.4). There was a tendency for the correlations to increase over time, with stronger correlations being seen between DAS and SvdH in the latter years. This was observed for lnSvdH scores in year 4 and 9, which saw moderate strength correlations with DAS scores from year 2 and year 5 respectively. The pairwise correlation matrix between the lnESR and lnSvdH score shows a higher number of moderate correlations between the two, while also highlighting the tendency for these stronger correlations to occur in later disease.

The second set of matrices in Table 7.2 indicates the correlations between the lnSvdH score and the separate TJC and SJC. Unlike the DAS and lnESR, the majority of correlations were <0.3, with just one correlation of 0.3 between lnSvdH score at year 9 and the TJC at year 5.

The separate components of the DAS can be split into objective and subjective markers. ESR and SJC are typically seen as objective markers of RA as they relate to quantifiable measures of inflammation in the blood or the clinician's assessment of swollen joints, rather than any subjective perception from the patient[300]. Conversely, the TJC (and in the case of the DAS-28, the Patient Global Assessment (PGA) for pain) is regarded as a subjective marker of inflammation, as it relies on patient reported symptoms relating to inflammation and disease. As such, the objective and subjective components of the DAS are sometimes examined individually when investigating radiographic damage, as theoretically radiographic damage would be more strongly associated with the objective markers, rather than the subjective markers.

The relative strength of the correlations between lnESR and lnSvdH further supports the existence of a relationship between inflammation and radiographic progression in later disease. The relative strength of this relationship in terms of number of moderate correlation coefficients over the 10 year period also provides evidence that, of all the measures used to calculate the DAS, ESR is the strongest for radiological damage. The tendency for these correlations to increase over time is also suggestive that there is a lagged association with radiographic damage, and does not occur serially, but rather a few years after the occurrence of increased inflammation.

When comparing the correlation matrices for the SJC and TJC there was little difference between the correlation structures. While the increased relationship between lnESR and lnSvdH, and the absence of a relationship between the TJC and SvdH may be of little surprise, the SJC

	lnSvdH yr0	lnSvdH yr1	lnSvdH yr2	lnSvdH yr3	lnSvdH yr4	lnSvdH yr5	lnSvdH yr6	lnSvdH yr7	lnSvdH yr8	lnSvdH yr9	lnSvdH yr10
DAS yr0	0.06	0.11	0.09	0.11	0.15	0.14	0.10	0.08	0.10	0.21	0.14
DAS yr1	0.05	0.14	0.17	0.21	0.20	0.21	0.12	0.16	0.18	0.24	0.14
DAS yr2	0.07	0.15	0.21	0.29	0.30	0.30	0.16	0.22	0.24	0.29	0.26
DAS yr3	0.06	0.11	0.14	0.23	0.30	0.24	0.23	0.22	0.15	0.25	0.20
DAS yr4	0.08	0.10	0.17	0.25	0.34	0.30	0.29	0.26	0.17	0.26	0.22
DAS yr5	0.06	0.12	0.18	0.20	0.25	0.22	0.22	0.28	0.22	0.33	0.14
DAS yr6	0.13	0.18	0.18	0.23	0.30	0.24	0.25	0.35	0.18	0.32	0.31
DAS yr7	0.08	0.08	0.15	0.19	0.25	0.23	0.26	0.24	0.17	0.29	0.28
DAS yr8	0.13	0.14	0.18	0.22	0.27	0.22	0.23	0.20	0.19	0.23	0.15
DAS yr9	0.16	0.15	0.20	0.25	0.32	0.27	0.28	0.27	0.28	0.22	0.27
DAS yr10	0.18	0.14	0.17	0.23	0.28	0.24	0.22	0.26	0.21	0.11	0.17
lnESR yr0	0.18	0.27	0.24	0.25	0.23	0.26	0.28	0.18	0.26	0.10	0.16
lnESR yr1	0.17	0.30	0.31	0.33	0.33	0.32	0.35	0.32	0.43	0.31	0.29
lnESR yr2	0.11	0.21	0.25	0.31	0.31	0.29	0.27	0.20	0.32	0.32	0.23
lnESR yr3	0.08	0.15	0.16	0.24	0.35	0.30	0.38	0.32	0.36	0.26	0.23
lnESR yr4	0.12	0.14	0.18	0.26	0.35	0.32	0.46	0.34	0.35	0.36	0.32
lnESR yr5	0.07	0.11	0.15	0.21	0.25	0.23	0.30	0.29	0.35	0.29	0.29
lnESR yr6	0.14	0.21	0.17	0.23	0.26	0.22	0.32	0.35	0.38	0.44	0.32
lnESR yr7	0.11	0.14	0.16	0.20	0.23	0.20	0.30	0.24	0.18	0.24	0.29
lnESR yr8	0.16	0.17	0.18	0.21	0.25	0.18	0.33	0.23	0.33	0.38	0.28
lnESR yr9	0.12	0.23	0.23	0.25	0.27	0.25	0.30	0.30	0.40	0.28	0.30
lnESR yr10	0.11	0.16	0.16	0.21	0.17	0.20	0.20	0.18	0.34	0.23	0.23

TABLE 7.1: Pairwise correlation matrix of total lnSvdH with the DAS and lnESR over the first 10-years follow-up in ERAS

	lnSvdH yr0	lnSvdH yr1	lnSvdH yr2	lnSvdH yr3	lnSvdH yr4	lnSvdH yr5	lnSvdH yr6	lnSvdH yr7	lnSvdH yr8	lnSvdH yr9	lnSvdH yr10
TJC yr0	0.05	0.07	0.05	0.07	0.10	0.09	0.04	0.03	0.03	0.18	0.06
TJC yr1	0.00	0.08	0.08	0.12	0.12	0.12	0.04	0.09	0.07	0.21	0.02
TJC yr2	0.03	0.09	0.14	0.19	0.21	0.22	0.10	0.14	0.15	0.23	0.18
TJC yr3	0.05	0.09	0.10	0.17	0.22	0.17	0.16	0.16	0.10	0.25	0.15
TJC yr4	0.06	0.08	0.13	0.18	0.24	0.21	0.18	0.20	0.08	0.20	0.14
TJC yr5	0.05	0.09	0.13	0.13	0.17	0.15	0.13	0.18	0.18	0.30	0.05
TJC yr6	0.11	0.15	0.14	0.17	0.21	0.17	0.19	0.27	0.09	0.27	0.24
TJC yr7	0.07	0.04	0.08	0.12	0.15	0.15	0.17	0.15	0.09	0.26	0.17
TJC yr8	0.09	0.15	0.17	0.21	0.21	0.19	0.14	0.15	0.10	0.20	0.08
TJC yr9	0.10	0.09	0.09	0.18	0.23	0.21	0.18	0.19	0.21	0.12	0.14
TJC yr10	0.18	0.12	0.14	0.21	0.28	0.21	0.21	0.26	0.16	0.10	0.10
SJC yr0	0.00	0.03	0.03	0.05	0.09	0.07	0.05	0.05	0.02	0.17	0.12
SJC yr1	0.00	0.07	0.09	0.14	0.12	0.14	0.06	0.07	0.01	0.13	0.09
SJC yr2	0.05	0.10	0.16	0.23	0.26	0.26	0.13	0.18	0.16	0.23	0.19
SJC yr3	0.07	0.08	0.11	0.19	0.25	0.19	0.18	0.17	0.05	0.19	0.15
SJC yr4	0.07	0.07	0.13	0.19	0.25	0.24	0.21	0.20	0.10	0.15	0.14
SJC yr5	0.04	0.10	0.14	0.16	0.21	0.18	0.18	0.23	0.12	0.21	0.13
SJC yr6	0.10	0.13	0.14	0.17	0.21	0.20	0.18	0.27	0.07	0.19	0.24
SJC yr7	0.07	0.06	0.11	0.14	0.20	0.19	0.23	0.24	0.16	0.18	0.25
SJC yr8	0.07	0.09	0.12	0.17	0.24	0.23	0.21	0.18	0.15	0.16	0.10
SJC yr9	0.12	0.09	0.10	0.19	0.24	0.22	0.20	0.22	0.21	0.11	0.19
SJC yr10	0.18	0.09	0.14	0.21	0.29	0.24	0.21	0.27	0.15	0.03	0.17

TABLE 7.2: Pairwise correlation matrix of total lnSvdH with the TJC and SJC over the first 10-years follow-up in ERAS

is regarded as more closely related to objective levels of inflammation. Thus, one would hypothesise that it would resemble a correlation structure more similar to the lnESR, rather than the TJC. This differentiation between objective and subjective markers of pain and inflammation will be explored in more detail in the full analysis.

7.3.1.1 Time-averaged analysis

The first analysis will model the effect of disease activity on radiographic progression as a time-averaged covariate. The first 10 years of DAS data was summarised into a mean score for each patient. Based on pre-defined EULAR cut-points, patients were then grouped into low DAS (<2.4), moderate DAS (2.4-3.7) and high DAS (>3.7) categories using the mean DAS over 10 years. Only those patients with at least two DAS scores from years 1 to 10 were included (n=1,335, 91%). The 10 year radiographic progression of patients in each of the three DAS categories was then estimated using a mixed-effects negative binomial regression analysis. The primary analysis looked at the total SvdH as the dependent variable, while secondary models examined the separate JSN and erosions components, and the Larsen score as the dependent variables. The DAS categories, along with year of follow-up entered as a continuous variable, were included as the primary predictors. As well as being entered as main effects, it was also important to include an interaction term between the DAS categories and follow-up year, so that any differences in the rate of change between the groups over time could be accounted for. To control for possible confounding effects, each model controlled for age, sex, RF status, baseline HAQ score, baseline HB level, months from symptom onset to first out-patient visit and use of DMARDs within the first 12months. The Bonferroni adjustment was used to adjust for multiple testing across groups.

The patient population by DAS categories was 466 (35%) in the low group, 456 (34%) in the moderate group and 413 (31%) in the high group respectively. For the sub-sample with SvdH data, there were 402 (34%), 401 (34%) and 372 (32%) patients for the low, moderate and high categories respectively. Figure 7.4 indicates the range of DAS scores at follow-up years 1 to 10 for each of the three DAS categories. The horizontal line denotes the cut-points for the DAS categories (2.4 and 3.7). The figure shows the relative stability of the DAS score over 10 years, with just 211 (16%) patients having >50% of their DAS scores outside their allocated DAS category.

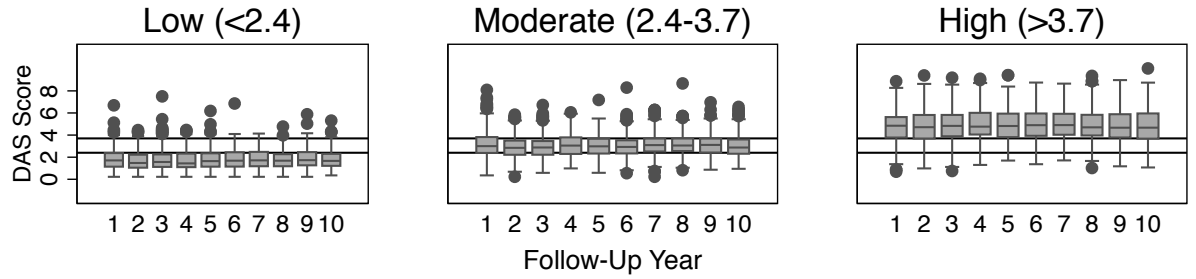


FIGURE 7.4: Box plot of DAS scores over the first 10 years by DAS categories

The basic demographic and clinical characteristics of patients in each DAS category are outlined in Table 7.3. Patients in the low group tended to be slightly younger, male, and RF negative with a lower HAQ and DAS at baseline. They were also more likely to have baseline HB levels within the normal range. In contrast, patients in the high group tended to be older, female, and RF positive with high baseline HAQ and DAS scores. The moderate group were similar to the whole cohort, however they also tended to be slightly older and have low HB at baseline. Months to first visit from referral were similar across all groups.

Table 7.4 provides a full break down of the relative differences between each of the DAS groups from the mixed-effects negative-binomial regression. Alongside these model estimates are the absolute scores calculated from the estimated sample means. Comparing the low to moderate DAS category, the moderate DAS category had higher radiographic damage at year 1 for the total SvdH score (21.9 vs. 13.5, $p<0.001$), JSN score (14.3 vs. 9.5, $p<0.001$), erosions score (5.2 vs. 2.2, $P<0.001$) and Larsen score (3.6 vs. 1.8, $p<0.001$). The relative differences ranged from 1.5 to 1.6 fold increase for the JSN and total SvdH score respectively, while increases of 2 to 2.3 fold were seen for the Larsen and erosion scores. Similar patterns were seen for the annual rate of change, whereby the moderate category indicated increased annual rate of change for the total SvdH (6.4 vs. 3.4, $p<0.001$), JSN score (14.3 vs. 9.5, $p<0.001$), erosion score (2.3 vs. 0.9, $p<0.05$) and Larsen score (1.2 vs. 0.5, $p<0.05$). With all scores, the moderate DAS category indicated around a 2 times increase in annual progression compared to the low category.

Comparisons between the low category and high category indicated a greater distinction, with the high category having higher radiographic damage at year 1 and greater increases in the annual rate of change than the low DAS category.

Interestingly, the model estimated similar levels of radiographic progression between the moderate and high categories. The model indicated that the total SvdH, JSN, erosion and Larsen

	Low		**Moderate**		**High**		**Total**	
	Total	With SvdH	Total	With SvdH	Total	With SvdH	**Total**	**With SvdH**
Demographics								
Age (Mean (SD))	53.7 (14.1)	53.4 (14.1)	55.8 (14.2)	55.3 (14.1)	55.3 (14.3)	55.0 (14.6)	54.9 (14.2)	54.5 (14.3)
Female (%)	54	54	68	68	78	78	67	66
Clinical Markers								
RF+ (%)	59	59	66	66	65	66	63	64
Baseline HAQ (Median (IQR))	0.75 (1.13)	0.75 (1.13)	1.00 (1.00)	1.00 (0.88)	1.38 (1.13)	1.38 (1.13)	1.00 (1.13)	1.00 (1.13)
Baseline DAS (Mean (SD))	3.41 (1.32)	3.40 (1.34)	4.14 (1.44)	4.13 (1.44)	5.16 (1.62)	5.14 (1.60)	4.20 (1.62)	4.20 (1.62)
Low HB (%)	37	37	44	43	44	44	41	41
Months to First Visit (Median (IQR))	6 (7)	6 (6)	7 (8)	7 (8)	7 (6)	6 (6)	6 (7)	6 (7)
Observations	466	402	456	401	413	372	1335	1175

TABLE 7.3: Summary statistics for patients with and without SvdH data stratified by DAS EULAR categories

score at year 1 were similar between the two patient groups. Only the total SvdH indicated an increased annual rate of change for the high DAS category (8.8 vs. 6.6, p<0.05) compared to the moderate category, with the JSN, erosion and Larsen score also indicating similar levels of increased annual rate of change (between a 21%-34% increase). However, none of these differences reached statistical significance (p>0.05).

	Absolute Score			Relative Difference		
	Low	Moderate	High	Moderate vs. Low	High vs. Low	High vs. Moderate
Total SvdH at yr1	13.5	21.9	24	1.59 [1.32-1.92] p<0.001	1.70 [1.39-2.08] p<0.001	1.06 [0.86-1.33] p=0.995
Total SvdH Annual rate	3.4	6.4	8.8	1.90 [1.41-2.38] p<0.001	2.61 [1.93-3.29] p<0.001	1.38 [1.04-1.72] p=0.019
JSN score at yr1	9.5	14.3	15.9	1.50 [1.24-1.81] p<0.001	1.63 [1.33-2.00] p<0.001	1.09 [0.88-1.36] p=0.743
JSN score Annual rate	1.8	3.2	4.3	1.73 [1.28-2.18] p<0.001	2.32 [1.70-2.94] p<0.001	1.34 [1.00-1.68] p=0.089
Erosion score at yr1	2.2	5.2	5.4	2.32 [1.76-3.07] p<0.001	2.33 [1.73-3.13] p<0.001	1.00 [0.73-1.38] p=0.999
Erosion score Annual rate	0.9	2.3	3	2.67 [1.78-3.54] p<0.001	3.48 [2.28-4.68] p<0.001	1.31 [0.88-1.73] p=0.410
Larsen score at yr1	1.8	3.6	3.6	1.95 [1.48-2.56] p<0.001	1.92 [1.44-2.58] p<0.001	0.99 [0.72-1.36] p=0.999
Larsen score Annual rate	0.5	1.2	1.4	2.17 [1.40-2.94] p<0.001	2.63 [1.67-3.59] p<0.001	1.21 [0.79-1.63] p=0.330

TABLE 7.4: Estimated absolute means and relative differences of the Total SvdH score stratified by each DAS category over the first 10 years. Relative differences expressed as Incidence Rate Ratios (IRR). 95% Confidence Intervals in square brackets.

A graphical representation of the estimated sample means over the first 10 years for each DAS category is shown in Figure 7.5 for the total SvdH score, in Figure 7.6 for the JSN and erosion score, and Figure 7.7 for the Larsen score. These graphs highlight how the low category has consistently low radiographic damage over the first 10 years, whereas the moderate and high categories have similar progression over time. While the high category does show marginally higher progression compared to the moderate category, this is not statistical significant, with overlapping 95% CIs in the grey shaded area around the point estimates at each time point.

Additionally, the mean DAS over the first 10 years was also modelled as a continuous variable, rather than categorised into low, moderate and high DAS groups. This provided similar results, indicating that each additional follow-up year resulted in a 14% increase in radiographic damage (IRR 1.14 [95% CI 1.12-1.17], p<0.001), while a 1 unit increase in mean DAS indicated a 16% increase in radiographic damage (IRR 1.16 [95% CI 1.09-1.23], p<0.001). There was also a

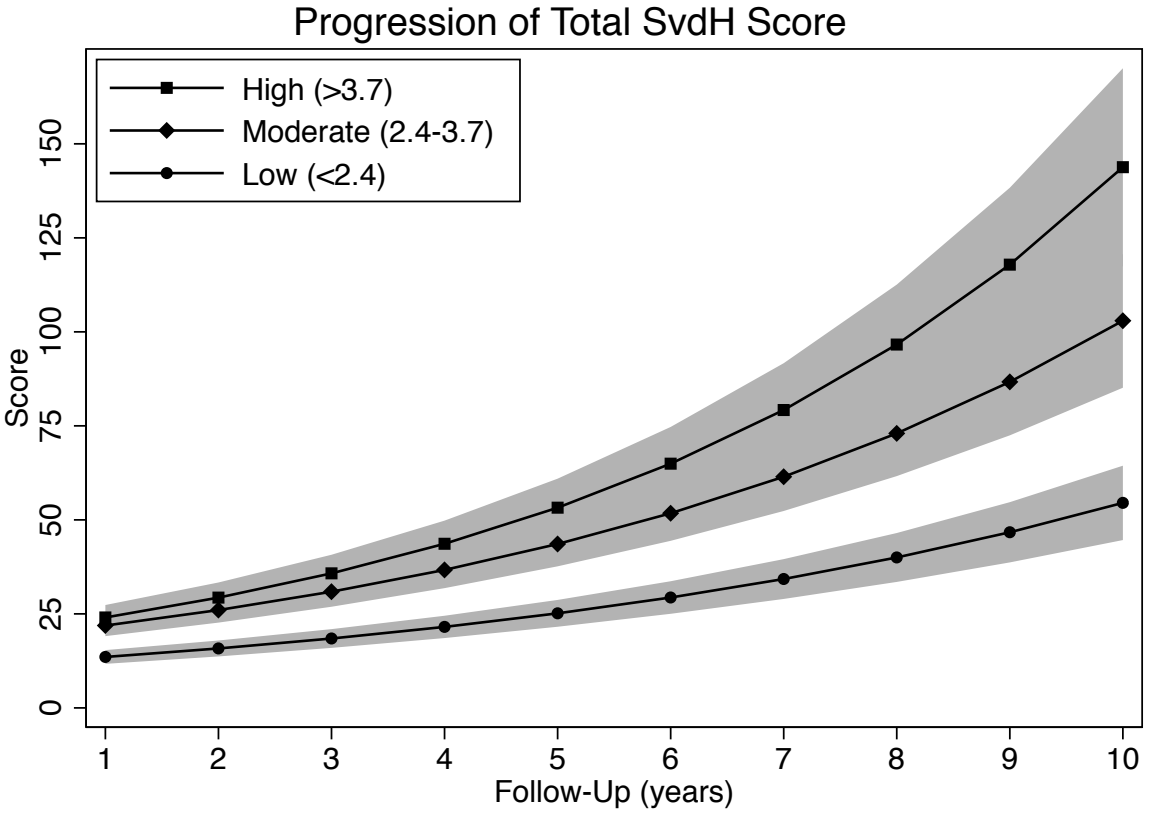


FIGURE 7.5: Estimated absolute means over the first 10 years for the Total SvdH score stratified by DAS categories. Grey area denotes 95% Confidence Intervals

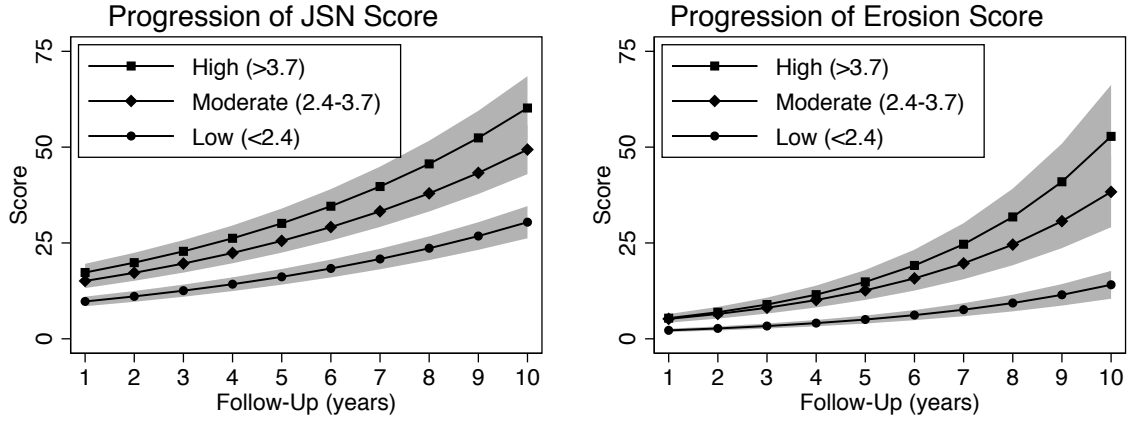


FIGURE 7.6: Estimated absolute means over the first 10 years for the JSN and erosion score stratified by DAS categories. Grey area denotes 95% Confidence Intervals

significant interaction effect between mean DAS and follow-up year, showing that higher than average DAS levels were related to an accelerated rate of progression over time (IRR 1.01 [95% CI 1.01-1.02], p<0.001). A sensitivity analysis including the standard deviation of each patients mean DAS score was also included, however this had no association with the total SvdH score.

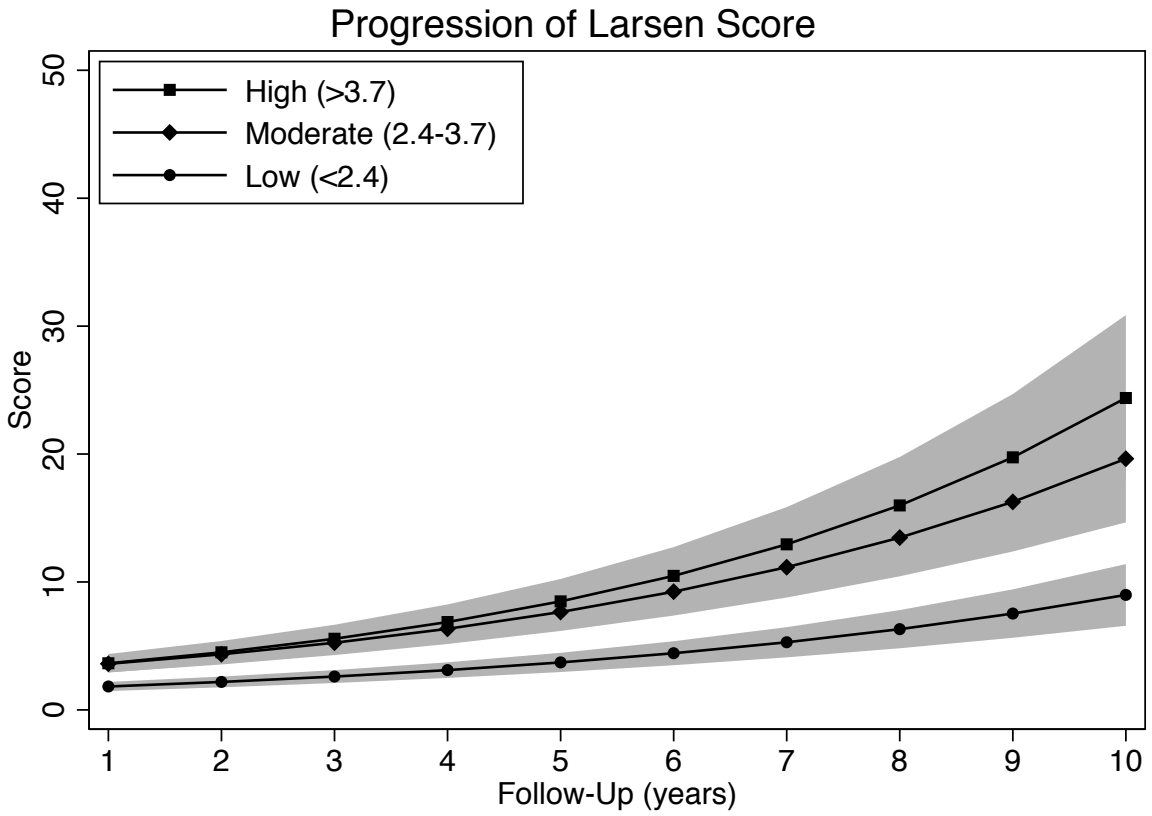


FIGURE 7.7: Estimated absolute means over the first 10 years for the Larsen score stratified by DAS categories. Grey area denotes 95% Confidence Intervals

7.3.1.2 Time-varying analysis

Using DAS as a time-averaged score, the analysis has found that those with sustained low mean DAS over 10 years had reduced radiographic progression compared to those patients with sustained moderate and high mean DAS over 10 years. Furthermore, those patients with sustained moderate DAS indicated similar levels of radiographic progression compared to those patients with sustained high DAS.

The advantage of expressing DAS collected over time as a time-averaged variable is that the results are intuitive, and provide a clear indication of the effect of sustained levels of disease activity on long-term radiographic progression. In this particular case, the mean DAS over the first 10 years provided a suitable proxy marker of disease activity dose over this time period, as demonstrated in the box plots presented in Figure 7.4. It showed how patient's DAS over the 10-year period was relatively stable, and there was little deviation of the yearly DAS score outside the groups pre-defined EULAR DAS categories. However, this method is not without its limitations. While intuitive, summarising the DAS over the 10-year follow-up could be argued

to be too simplistic in its approach. For instance, there are instances where patients' disease activity does fluctuate over time, and it is not clear how this may affect the relationship with radiographic progression. Accounting for this fluctuation has been found to be an important factor in understanding the relationship between disease activity and radiographic damage over time in previous studies[214].

As such, the second part of this analysis looks at modelling the yearly DAS score at each follow-up as a Time-Varying Covariate (TVC) in a mixed-effects NB regression model. The analysis will look at the potential time effect between radiographic damage and DAS by modelling the two measures collected at different time points. The first will examine the association when both are measured at the same time, while the next set of models will investigate the effect of DAS on future radiographic damage scores as a time-lagged effect. This involves modelling the DAS scores that occur earlier in the follow-up period on radiographic scores that are measured later in the follow-up period. The first time-lagged model will look at a 1-year lagged effect, while the second model will look at a 2-year lagged effect. The aim of comparing analyses using time-lagged associations is to investigate whether the association between DAS and radiographic damage increases as the lag time increases. Given the correlation structure seen earlier in the chapter, it is hypothesised that higher DAS will have a stronger association with radiographic damage progression occurring later on in the disease time period, rather than concurrent measures. As before, these models will control for key confounding variables, including age, sex, RF status, baseline functional disability, level of HB, months from referral to first visit and whether DMARDS were prescribed within the first 12-months.

The impact of time-lagged effects

The correlation matrices outlined in Figure 7.1 highlight how the strength of association between radiographic damage and disease activity increases over time. It is hypothesised that the predictive ability of the DAS on total SvdH will increase as the time-lag between the two variables increase. The three TVC models (concurrent, 1 year time lagged and 2 year time lagged) were performed on the DAS, as well as the separate DAS components; ESR, SJC and TJC. The DAS calculation requires the ESR to be log transformed, and the TJC to be square-root transformed. These transformations were therefore included in the models, rather than the raw scores, to ensure direct comparisons with the DAS.

The results of these models are shown in Table 7.5.

	(1) Time-Varying IRR / SE	(2) Time-Varying (Lagged) IRR / SE	(3) Time-Varying (Lagged-2) IRR / SE
DAS	1.01	1.03**	1.05***
	(0.010)	(0.011)	(0.014)
DAS#Follow-up Year	1.00	1.00	1.00
	(0.002)	(0.002)	(0.002)
DAS Components			
ESR (ln)	1.01	1.06***	1.12***
	(0.017)	(0.019)	(0.023)
SJC	1.00	1.00	1.01*
	(0.002)	(0.002)	(0.003)
TJC (Sqrt)	0.99	0.99	0.98
	(0.014)	(0.015)	(0.020)
Follow-up Year#ESR (Ln)	1.00	1.00	0.99
	(0.003)	(0.003)	(0.003)
Follow-up Year#SJC	1.00	1.00	1.00*
	(0.000)	(0.000)	(0.000)
Follow-up Year#TJC (Sqrt)	1.00	1.00	1.01
	(0.003)	(0.003)	(0.003)
Observations	4645	3737	3024
N	1162	1125	1108

* p <0.05, ** p <0.01, *** p <0.001. Standard errors in parentheses.
Note: IRR = Incidence Rate Ratios

TABLE 7.5: Modelling DAS and its separate components as time-varying predictors

Model (1) indicated that there was no significant association between disease activity and radiographic damage when measured at the same follow-up time point. In contrast, models (2) and (3) showed a significant association with DAS by including a time lagged effect. Model (2) with a 1 year lagged-effect, had a 3% increase in total SvdH (IRR 1.03 [95% CI 1.01-1.05], p<0.01), while model (3) for the 2 year time lagged effect had a 5% increase in total SvdH (IRR 1.05 [95% CI 1.03-1.07], p<0.001), for every one unit increase in DAS.

When this was repeated for the separate components of the DAS, the ESR(ln) had a similar trend across the 3 models. There was no significant association with concurrent measures of total SvdH, but an increasing statistically significant association with total SvdH in the 2 year lagged effect model. In this model, a 1 unit increase in lnESR showed a 12% increase in total SvdH (IRR 1.12 [95%CI 1.07-1.16], p<0.001). To provide a clearer interpretation, a proportional increase in ESR can be used to calculate its relative increase on total SvdH. For example, a 10% increase in ESR can be estimated to increase total SvdH by 34% ($1.10^{3.06} = 1.34$, where

3.06=exp(1.12)). Only the 2 year lagged effect model shows a significant association between SJC and total SvdH.

As with the analysis in Chapter 6, an interaction effect between the covariate of interest and follow-up year estimates whether the association between the covariate and the outcome measure changes as follow-up year increases. Starting with DAS model (1), we see that DAS was not associated with total SvdH at baseline when DAS and total SvdH are measured concurrently. The non-significant interaction effect also tells us that this association does not change as follow-up time increases. The inclusion of time-lagged effects in model (2) and (3) does not change the interpretation of the interaction effect, it merely shifts the time point to which these effects are related. In model (2), which examines the association between the previous years DAS and total SvdH, there is a significant association, where increased DAS in the previous year results in an increase of 3% in total SvdH. Like model (1) though, the interaction effect is still non-significant, and therefore this time-lagged effect between DAS and total SvdH also does not change as a product of increased follow-up. The only exception was the SJC in model (3) with a 2 year time lagged effect. However, the model estimate for the interaction effect showed a less than 1% change for each additional follow-up year (IRR 1.00 [95%CI 1.00-1.00], p=0.045).

Modelling DAS using both ERAS and ERAN data over 5 years

As was reported in Chapter 6, the level of radiographic progression has significantly decreased over time between the ERAS and ERAN cohorts, with ERAN reporting much lower levels of radiographic progression over 5 years. While the analysis presented in this chapter demonstrates how this progression is largely determined by the level of the patients disease activity over time, the next natural question is whether increased disease activity has a similar impact on radiographic progression in ERAN, despite the overall lower levels of radiographic damage observed.

It is inappropriate to repeat all the analysis performed in the ERAS cohort with both cohorts combined given the relatively small amount of radiographic data in the ERAN cohort. Categorising the time-averaged DAS into EULAR cut points in ERAN is unsuitable due to the low levels of data over the 5 year period. This is shown in Table 7.6, where the number of patients with SvdH data over the first 5 years are given for each mean DAS category. With many cells within the table at the years 3, 4 and 5 indicating single figure observations, it is clear any

modelling would not be sufficiently powered to provide robust results, such as those seen in the ERAS cohort above.

DAS Group	SvdH yr1	SvdH yr2	SvdH yr3	SvdH yr4	SvdH yr5
Remission	32	22	16	9	9
Low	18	15	7	5	5
Moderate	57	56	33	19	15
High	16	13	9	9	4
Total	123	106	65	42	33

TABLE 7.6: Number of observations with radiographic data stratified by DAS groups over the first 5 years in ERAN

However, using a time-averaged model where the mean DAS score is included as a continuous outcome reduces the parameters needed in the model. While the relative power of the effect estimates for the ERAN cohort is much lower than those for ERAS, it does provide a means of investigating whether the association between radiographic damage and DAS has changed as a result of lower overall radiographic progression in ERAN.

Combining both the ERAS and ERAN data, a total of 1,085 patients (ERAS=978, ERAN=107), contributing 3,587 observations (ERAS=3,397, ERAN=190), were entered into a multi-level NB regression. The lower number of observations was due to missing data, as patients needed to have at least two DAS scores between years 1 and 5, as well as radiographic data over this time period. Year of follow-up and the mean DAS score from years 1-5 were entered as continuous covariates. As with the RF analysis in Chapter 6, a three-way interaction effect between year of follow-up, mean DAS score and cohort membership was also included, along side the main effect. Age at onset, sex, RF positivity, Baseline HAQ and low HB, months from symptom onset to first out-patient visit and use of DMARDs within the first 12months. Additionally, some patients in ERAN received steroids or DMARDs prior to first visit, so these were included as binary variables. The estimated sample rates were calculated for each cohort over a mean DAS score of 2, 3 ,4 and 5 to indicate remission, low, moderate and high DAS groups respectively. The Bonferroni adjustment was used to adjust for multiple testing across groups.

The estimated sample means from the model for both the ERAS and ERAN cohorts representing patients in sustained remission, low, moderate and high mean DAS groups are presented in Figure 7.8.

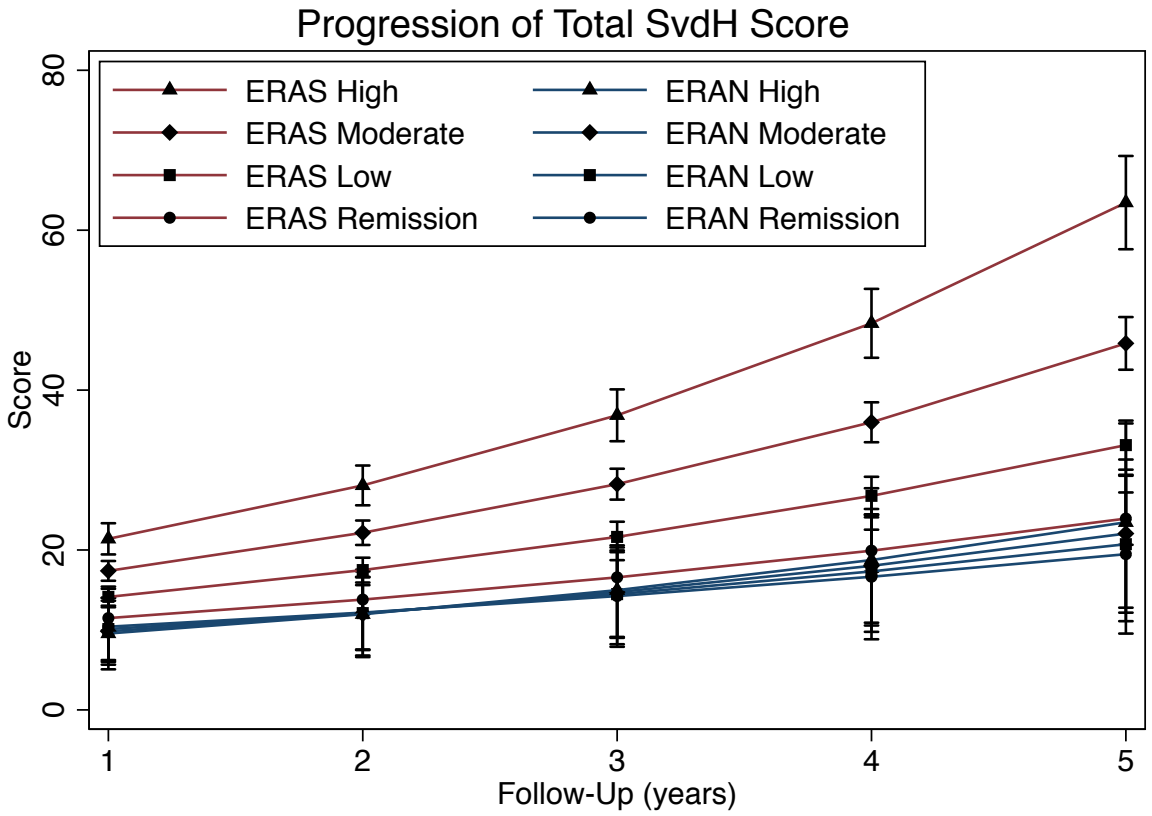


FIGURE 7.8: Estimated sample means for the mean DAS score at 2 (Remission), 3 (Low), 4 (Moderate) and 5 (High) for ERAS and ERAN. Black error bars denote 95% Confidence Intervals

The estimated sample means [95%CI] for the yearly increase of the total SvdH score in ERAS were 2.9 [2.4-3.4], 4.4 [3.9-4.8], 6.5 [6.0-7.2] and 9.7 [8.7-10.7] for patients with sustained remission, low, moderate or high mean DAS respectively over the 5 years. In contrast, the yearly increase in total SvdH score in ERAN was 1.8 [0.5-3.1], 2.2 [1.0-3.3], 2.6 [1.3-3.8] and 3.0 [1.2-4.9] for patients with sustained remission, low, moderate and high mean DAS over the 5 years. The model indicates that a patient in ERAN with a high mean DAS over the first 5 years of disease had similar a rate of radiographic progression compared to an ERAS patient in sustained DAS remission (Δ-0.2 [95%CI -2.1-1.8], p=0.864).

While the estimated sample rates for ERAN indicate no differences in radiographic progression across the different mean DAS values, a non-significant three-way interaction effect in the model suggests there was not sufficient power to conclude that the association of mean DAS between the two cohorts over time was statistically significantly different (IRR 0.99 [95%CI 0.94-1.05], p=0.804). While disease activity may still be associated with increased radiographic progression, the relatively low rates do not signify clinically meaningful increases in radiographic rates.

7.3.2 Radiographic damage and functional disability

The first objective was to examine the association between disease activity and radiographic progression using two methods for longitudinal data; time-averaged approach and the time-varying approach. The second objective of this analysis was to investigate the association between radiographic progression and functional disability using the same methods.

As with the DAS analysis, the first stage of analysing this association is to examine the correlation structure between the two measures. The first pairwise correlation matrix in Table 7.7 shows the pairwise correlation between the HAQ and the total SvdH score over the first 10 years. Those cells with correlations >0.3 are highlighted in light blue, while those cells with a correlation of >0.4 are highlighted in dark blue. As with the DAS, it is clear that the correlations are stronger towards the latter stages of the disease.

The model proposed by Escalante and Ricón[297] in Figure 7.2 denotes that disease activity causes radiographic damage, and that radiographic damage in turn causes functional disability. As such, to model the impact of radiographic damage on functional disability, the outcome measure (dependent variable) needs to be functional disability, while the covariate of interest (independent variable) is the total SvdH score. The data distributions of the HAQ are different to the total SvdH (See Figure 7.9), so a multi-level NB regression model is not required. Instead, the HAQ score can be modelled as a continuous linear outcome, assuming a normal distribution. The coefficient estimates from the model denote a mean change in HAQ.

7.3.2.1 Time-averaged analysis

To investigate the time-averaged effect of radiographic damage on HAQ scores, a summary statistic that encapsulates the average rate over years 1-10 is needed. As has been documented throughout this thesis, summarising this as a mean score (as was done for the DAS in the previous analysis) would be inappropriate. Therefore, to estimate the rate of change for each patient, a separate multi-level NB model is used, where the random-effects slope from years 1-10 for each patient is estimated and used in the final HAQ model. Rates were obtained for 1,192 (81%) ERAS patients, with a median rate of change of 3.7 units per year (IQR 5.3), which ranged from 0.2 to 62.7 units per year. These estimated rates can now be used in the HAQ model to investigate how the rate in change of radiographic progression can predict functional disability over time.

	SvdH yr1	SvdH yr2	SvdH yr3	SvdH yr4	SvdH yr5	SvdH yr6	SvdH yr7	SvdH yr8	SvdH yr9	SvdH yr10
HAQ yr1	0.17	0.17	0.20	0.22	0.19	0.13	0.16	0.16	0.30	0.16
HAQ yr2	0.14	0.19	0.23	0.23	0.23	0.17	0.20	0.23	0.31	0.24
HAQ yr3	0.12	0.15	0.22	0.23	0.23	0.18	0.19	0.27	0.39	0.19
HAQ yr4	0.15	0.18	0.25	0.28	0.25	0.21	0.26	0.30	0.37	0.23
HAQ yr5	0.14	0.15	0.23	0.25	0.24	0.24	0.27	0.26	0.41	0.23
HAQ yr6	0.12	0.17	0.21	0.28	0.23	0.27	0.31	0.30	0.45	0.28
HAQ yr7	0.09	0.13	0.22	0.25	0.24	0.30	0.32	0.29	0.43	0.27
HAQ yr8	0.11	0.16	0.23	0.30	0.26	0.33	0.34	0.36	0.41	0.31
HAQ yr9	0.12	0.17	0.24	0.32	0.27	0.36	0.36	0.42	0.39	0.29
HAQ yr10	0.17	0.21	0.28	0.39	0.33	0.40	0.38	0.40	0.37	0.26

TABLE 7.7: Pairwise correlation matrix of total SvdH and HAQ score over the first 10 years follow-up in ERAS. Light blue cells indicate correlations >0.3 and dark blue cells indicate correlations >0.4.

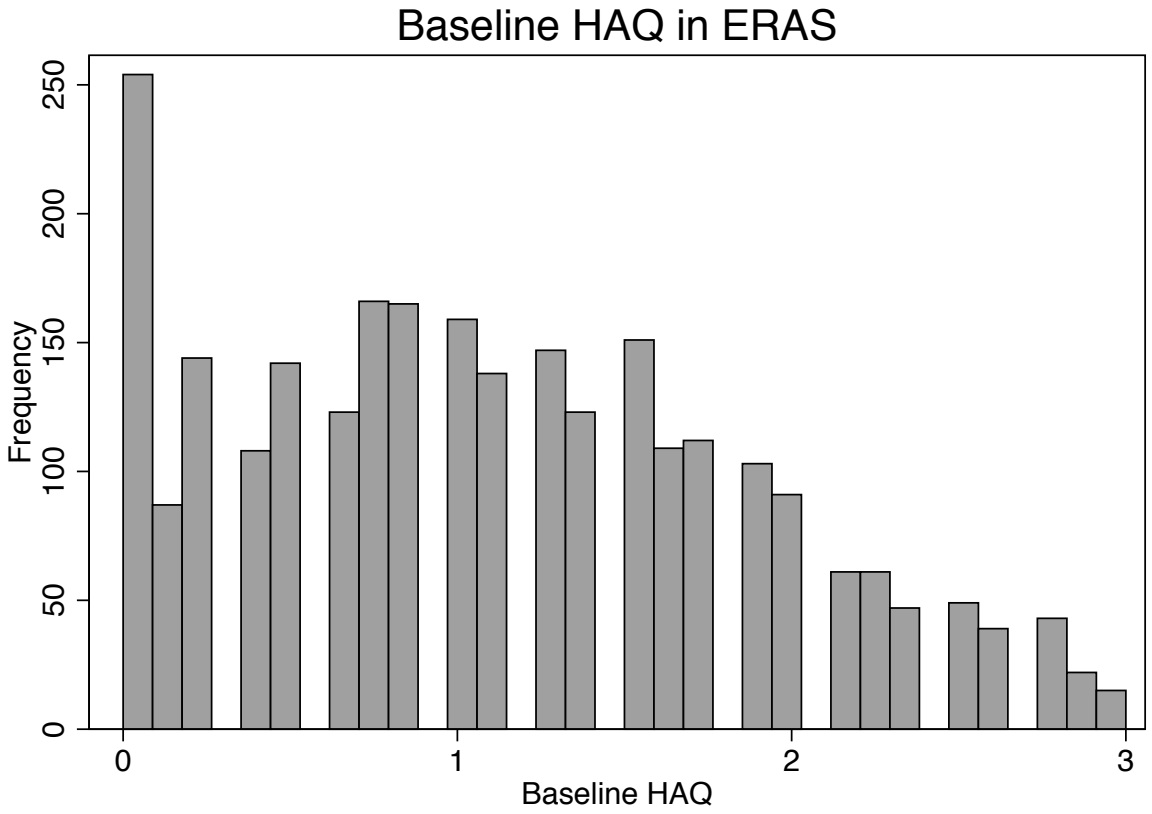


FIGURE 7.9: Histogram of the baseline HAQ values for the ERAS cohort

The baseline SvdH, along with the estimated rates calculated in the separate NB model, were entered as covariates into the model as continuous variables. As before, an interaction effect between the radiographic rate and follow-up year was also included, and the model controlled for age at onset, sex, RF status, baseline HAQ score, baseline HB level, months from symptom onset to first out-patient visit and use of DMARDs within the first 12 months.

A total of 1,016 patients contributed 7,330 observations to the multi-level linear model (mean = 7.2 observations per patient). The model showed that HAQ scores increased by 0.4 units per year (95%CI 0.03-0.04, $p<0.001$) and that each additional unit increase in the rate of radiographic progression was associated with a 0.1 unit increase in HAQ score (95%CI 0.01-0.02, $p<0.01$). A significant interaction between follow-up year and rate of radiographic change shows that the effect of increased radiographic progression was stronger over time ($p<0.001$). Baseline SvdH score was not associated with HAQ scores ($p=0.341$).

To depict the model results graphically, the estimated sample means of the HAQ score over years 1-10 were estimated for patients with a rate of 0, 5 and 10 SvdH units per year. While

no formal cut-points exist for the SvdH score, these denote patients with no damage, complete destruction of 1 joint, and complete destruction of 2 joints respectively.

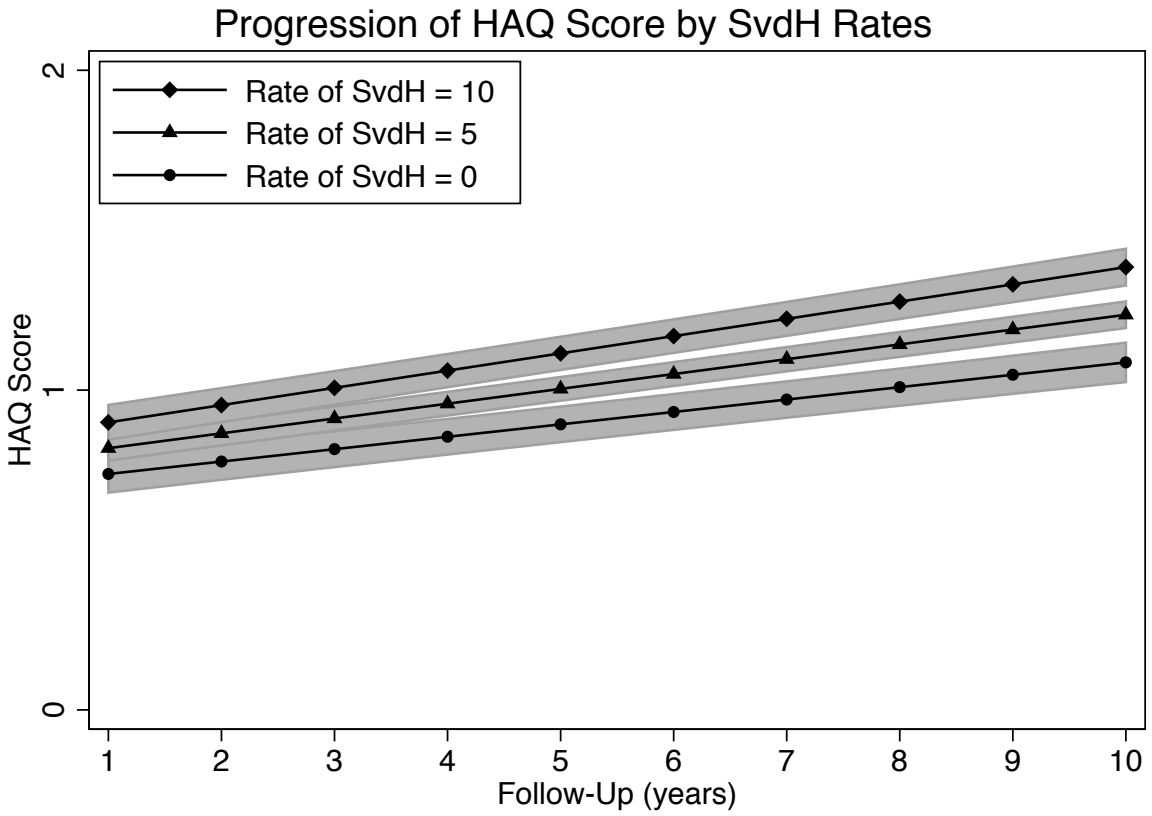


FIGURE 7.10: Estimated sample means over the first 10 years for the HAQ score stratified by annual rates of SvdH. Grey area denotes 95% Confidence Intervals

The same models were then applied based on the rate of JSN, erosions and Larsen score. The models showed that the effect of increased rates of JSN had a marginally higher impact on HAQ over time (0.03 [95%CI 0.02-0.05], p<0.001), whilst increased rates of erosion score showed similar effects to the total SvdH score (0.02 [95%CI 0.00-0.04], p=0.023). The higher effect estimate could be suggestive that JSN is more strongly associated with HAQ than the erosion score. The Larsen model showed similar results to the SvdH score, although direct comparisons are difficult due to the differences in the scoring methods.

7.3.2.2 Time-varying analysis

Following the time-averaged approach detailed above, the next stage of the analysis was to look at HAQ as a TVC. As with the TVC analysis for the DAS, three models were conducted to look at the association of radiographic damage as a TVC with functional disability. The first

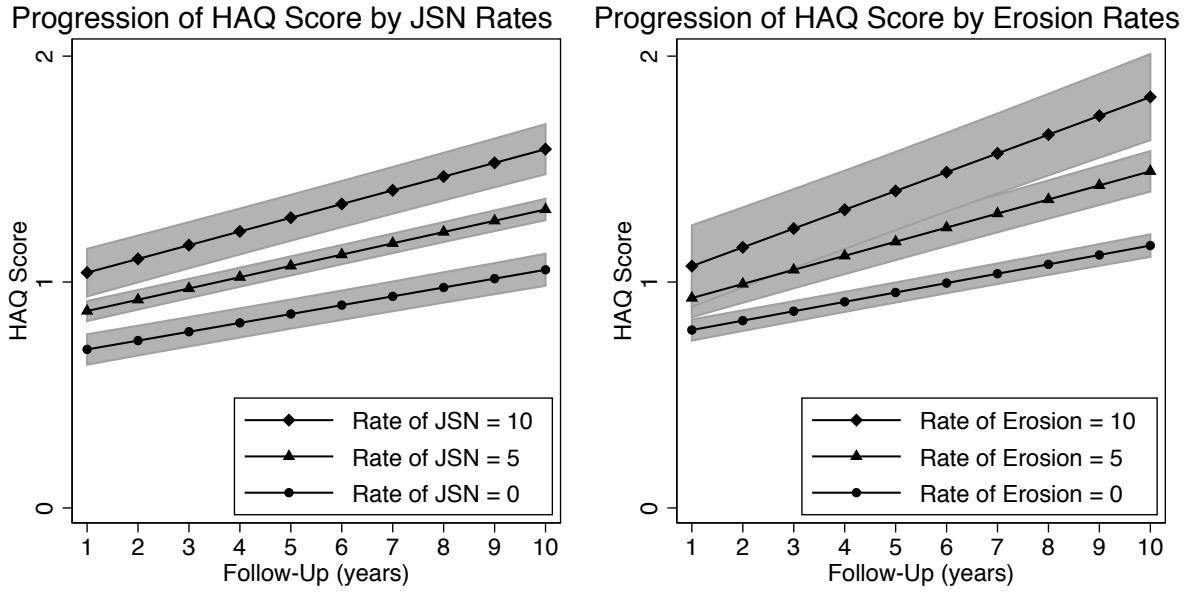


FIGURE 7.11: Estimated sample means over the first 10 years for the HAQ score stratified by annual rates of JSN and Erosions. Grey area denotes 95% Confidence Intervals

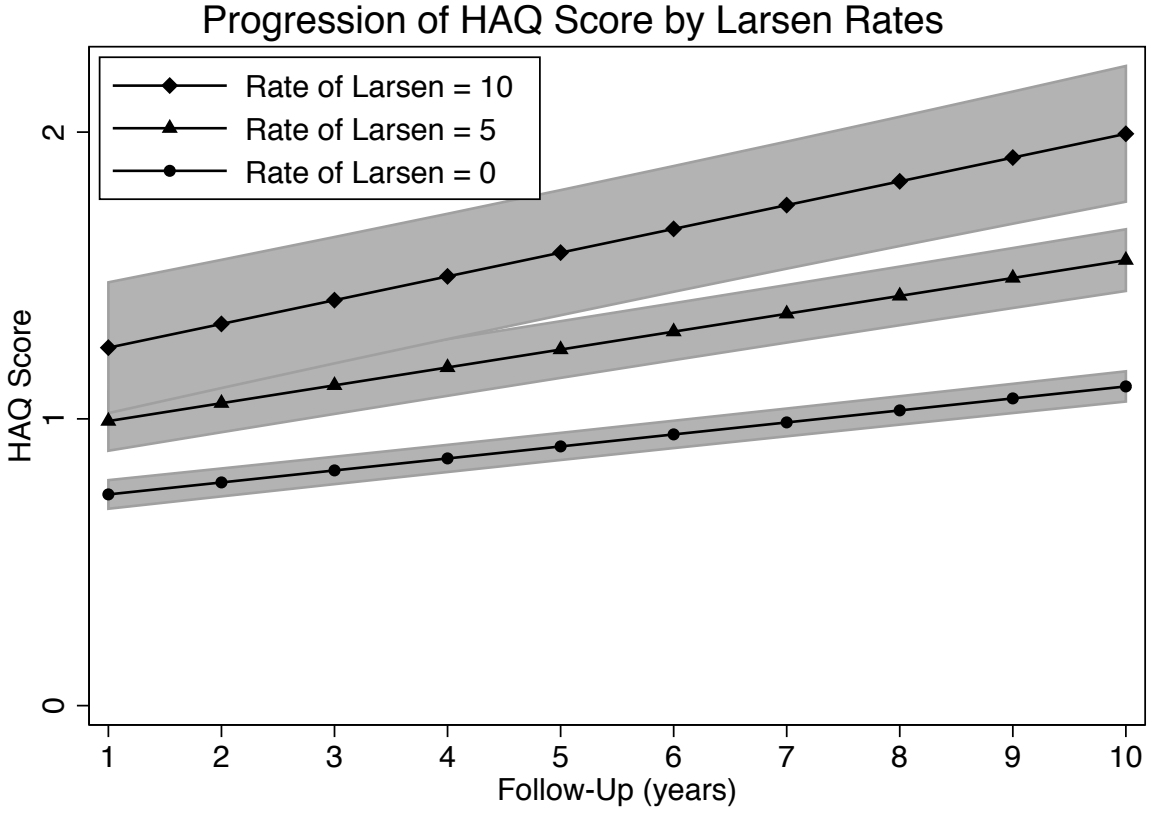


FIGURE 7.12: Estimated sample means over the first 10 years for the HAQ score stratified by annual rates of Larsen. Grey area denotes 95% Confidence Intervals

	(1) Time-Varying β/ SE	(2) Time-Varying (Lagged) β/ SE	(3) Time-Varying (Lagged-2) β/ SE
Total SvdH Score			
Follow-up Year	0.031***	0.047***	0.032***
	(0.005)	(0.006)	(0.006)
Total SvdH	0.002***	0.001	0.001
	(0.000)	(0.000)	(0.001)
Follow-up Year#Total SvdH	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
JSN Score			
Follow-up Year	0.027***	0.045***	0.028***
	(0.006)	(0.006)	(0.007)
JSN	0.002***	0.001	0.000
	(0.001)	(0.001)	(0.001)
Follow-up Year#JSN	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Erosions Score			
Follow-up Year	0.035***	0.050***	0.036***
	(0.005)	(0.005)	(0.005)
Erosion Score	0.003***	0.001	0.001
	(0.001)	(0.001)	(0.001)
Follow-up Year#Erosion Score	−0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Larsen Score			
Follow-up Year	0.037***	0.044***	0.036***
	(0.004)	(0.005)	(0.005)
Larsen Score	0.003***	0.002	0.002
	(0.001)	(0.001)	(0.002)
Follow-up Year#Larsen Score	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Observations	4782	4064	3691
N	1161	1091	1040

* p <0.05, ** p <0.01, *** p <0.001. Standard errors in parentheses.

TABLE 7.8: HAQ models using radiographic outcomes as (1) TVC, (2) TVC lagged 1 year and (3) TVC lagged 2 years

model looked at the association of HAQ with the total SvdH, JSN, erosion and Larsen score as a TVC when measured at the same follow-up time-point. The second two models then focus on the time-lagged effect of 1 year and 2 years. All the models also included follow-up year as a main effect along with an interaction between the radiographic measure and follow-up year. As before, age at onset, sex, RF positivity, DAS at baseline, low HB at baseline, months from first visit to referral and use of DMARDs within first 12 months were controlled for in the models. The results of these models are shown in Table 7.8.

Overall the HAQ score increased over time, indicating an increase of between 0.03 - 0.05 HAQ units per year over the 10 year period. For the first model looking at the association between HAQ and radiographic outcomes measured at the same time-point, all radiographic outcomes indicated a similar level of increase in HAQ. For the total SvdH, 1-unit increase indicated an increase of 0.002 HAQ units ([95%CI 0.001-0.002], p<0.001), while the Larsen score indicated an increase of 0.003 HAQ units ([95%CI 0.001-0.005], p<0.001) for every 1-unit increase. Separate models for the JSN and erosion score reported similar increases in HAQ of 0.002 [95%CI 0.001-0.004] and 0.003 [95%CI 0.001-0.004] respectively. A graphical depiction of the predicted HAQ score over the range of the total SvdH, while held at the mean year of follow-up, is displayed in Figure 7.13. It highlights the linear association between the two constructs, along with increased variance at the higher end of the radiographic score (heteroscedasicity). Additionally, none of the models indicated a significant interaction effect with follow-up year, suggesting any associations were linear over time. In contrast to the disease activity models, both of the time-lagged effect models indicated no significant association with HAQ scores.

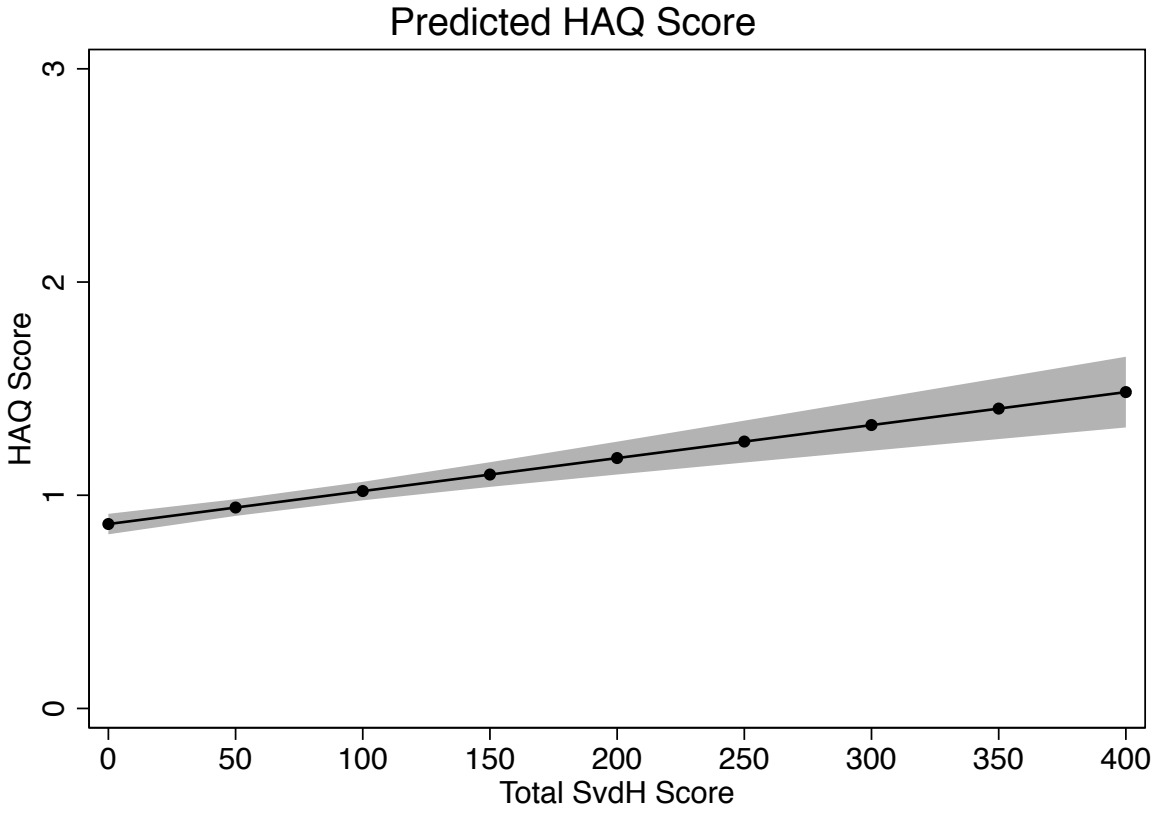


FIGURE 7.13: Estimated sample mean of the HAQ score over the range of the Total SvdH score, held at the mean follow-up year. Grey area denotes 95% Confidence Intervals

The small effect estimates presented represent the large scale of the predictor variable. In the

case of the total SvdH score, a 1 unit increase over a scale of 0-448 does not necessarily indicate a clinically meaningful increase in the total SvdH. As the effect of radiographic damage was proven to be linear, the effect estimate could be multiplied by any magnitude of change in the SvdH to reflect the increase in HAQ score. This was demonstrated in a sensitivity analysis looking at the total SvdH entered as a categorical variable based on the reported quartiles. This highlighted the linear increase over the quartiles, with an increase from the 25th percentile to the 75th percentile estimating an increase of 0.14 HAQ units ([95%CI 0.04-0.24],$p<0.001$).

Modelling HAQ using both ERAS and ERAN data over 5 years

As with the DAS analysis, a validation was conducted on the time-averaged models to investigate how the association between radiographic damage and HAQ has changed in light of reduced radiographic progression in ERAN. Given the relatively small amount of patients with radiographic data in ERAN, the models were restricted to just 5 years.

Combining both the ERAS and ERAN data, a total of 1,140 patients (ERAS=1013, ERAN=127) contributing 4,677 observations(ERAS=4,297, ERAN=380) were entered into a multi-level linear model. Year of follow-up and the mean rate of the total SvdH score from years 1-5 (estimated using the random-effects from a multi-level NB model), along with the baseline SvdH score, were entered as continuous covariates. As before, a three-way interaction effect between year of follow-up, rate of SvdH and cohort membership are also included along side their main effect. Age at onset, sex, RF positivity, Baseline HAQ and low HB, months from symptom onset to first out-patient visit and use of DMARDs within the first 12 months. Additionally, some patients in ERAN received steroids or DMARDs prior to first visit, so these were included as binary variables. The estimated sample rates were calculated for each cohort with a total SvdH rate of 0, 5 and 10 to indicate patients with 0, complete destruction of one joint, and complete destruction of two joints respectively. The Bonferroni adjustment was used to adjust for multiple testing across groups.

The estimated sample means from the model for both the ERAS and ERAN cohorts at SvdH rates of 0, 5 and 10 are presented in Figure 7.14.

The model indicates that there was no difference in the association between rates of SvdH and HAQ scores over the first 5 years between the ERAS and ERAN cohorts ($p>0.05$). There is an

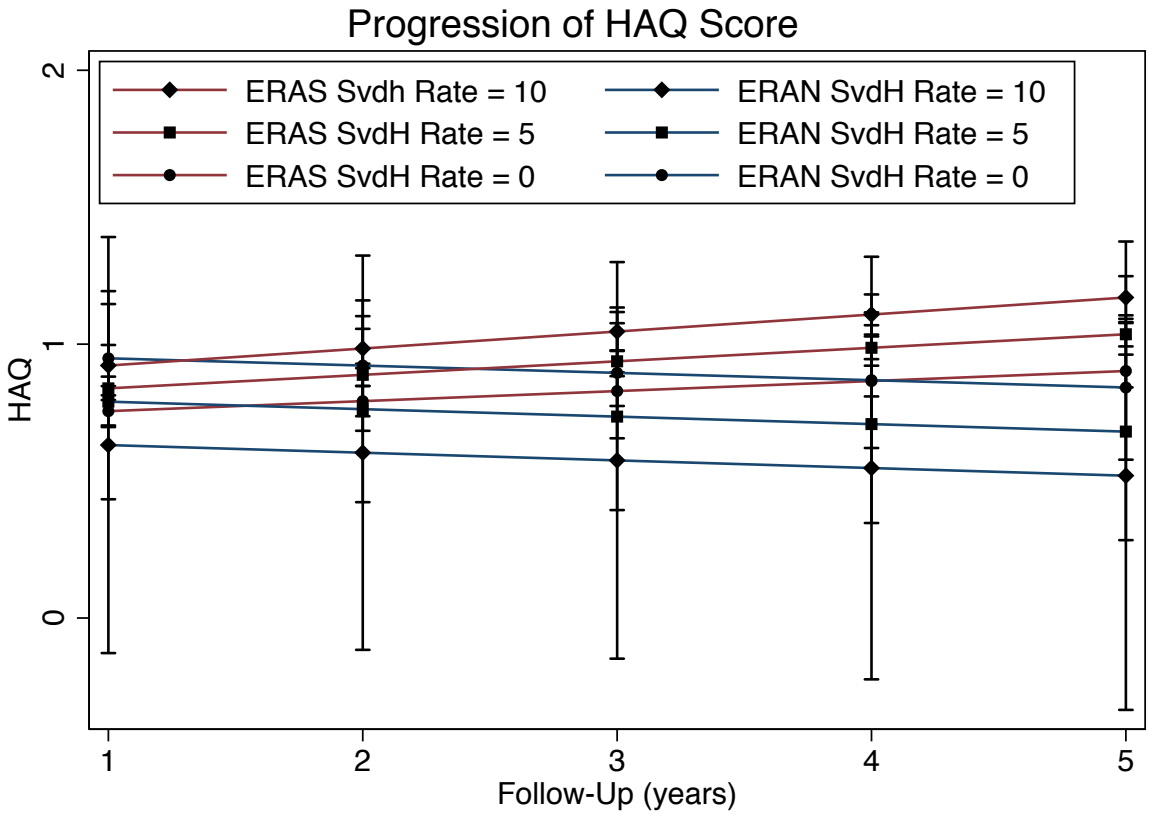


FIGURE 7.14: Estimated sample means for the mean HAQ score stratified by annual rates of SvdH for both ERAS and ERAN. Black error bars denote 95% Confidence Intervals

indication that HAQ scores for patients in ERAN were improving, however the limited data, along with restricted follow-up limits the ability to draw any definitive conclusions on this trend.

7.4 Discussion

This chapter explored the complex, longitudinal association between disease activity, radiographic joint damage, and functional disability. Using two different methods for modelling longitudinal data; time-averaged and time-varying. The aim was to use the modelling techniques described in Chapter 5 to test whether the current model proposed by Escalante and Ricón[297] fits with the ERAS data. Unlike previous studies, this analysis ensures more precise estimation by using the modelling methods described in Chapter 5 and explores these relationships over 10 years. Furthermore, validation analysis using a sub-sample of data from the ERAN dataset was used to investigate whether the secular declines in overall radiographic progression identified in Chapter 6 had any impact on the associations found in the ERAS data. Each

analysis is broken down into sub-sections to detail the findings, and assess how they fit in with the published literature.

DAS modelled as a time-averaged covariate

The first analysis looked at disease activity as a time-averaged covariate, and found that those patients recruited into the ERAS cohort between 1986-2001 with sustained moderate disease activity experienced similar levels of radiographic progression over the first 10 years as those patients with sustained high disease activity. This finding is in agreement with recent studies, one of which examined a similar patient group and found those patients with sustained moderate and high disease activity were also at risk of increased radiographic progression over the first 3 years[301] and 5 years[302]. A further study found that patients in persistent high disease activity states were found to have a 2.5 times increase in rapid radiological progression (>SDD) over the first 3 years, compared to those with persistent low disease or in remission[211]. In the UK, under the current T2T paradigm, prescription of biologics, in combination with conventional DMARDs, is reserved only for those patients with the most severe disease in order to achieve low disease states. The main ERAS analysis found equal levels of radiographic damage in those patients with sustained moderate disease. This gives support to the argument that a large proportion of patients with equally disabling disease are potentially being denied access to more effective therapies.

However, this is a historical cohort, treated with arguably less effective treatment strategies. Chapter 6 demonstrated that radiographic progression has halved between ERAS and ERAN, and as a result, the impact of seropositive RA on clinically meaningful radiographic progression had diminished. The validation analysis in this chapter using the sub-sample of radiographic data from ERAN also demonstrated that while the relative association between disease activity and radiographic progression was still present, the absolute change did not result in clinically meaningful increases in progression rates. That is, patients in ERAN with high disease activity indicated similar levels of radiographic progression over 5 years as those patients in ERAS with low disease activity. These findings have important implications on the argument of prescribing biologics DMARDs in those patients with moderate disease activity[72–74]. Using ERAN data, studies have shown that achieving low disease targets with conventional DMARD therapies in the first 3 years is low[72], and that functional disability is high in those patients with sustained moderate disease[72, 74]. In contrast, this analysis suggests that secular declines in

overall radiographic damage do not support the introduction of biologics in those patients with moderate disease, as radiographically they are already progressing at a clinically low rate.

Comparisons of these results with other observational cohorts is difficult, since variation in patient characteristics and treatment strategies is likely to be high. Most long-term observational studies, looking at radiographic outcomes and disease activity longitudinally, were recruited during the pre-biologic era, where results are similar to those found from the ERAS analysis[211, 214, 301]. In contrast, data from RCTs may provide some insight into the findings present in this analysis. The BeST study reported on radiographic outcomes over 5 years while categorising patients into disease activity levels using a range of DAIs[302], however the analysis looking at radiological progression across these disease groups was controlled to only report on those patients randomised to the monotherapy treatment arm. It is perhaps unsurprising the estimated rates of SvdH are similar to those reported in the ERAS cohort, since the treatment strategy is likely to be similar. If the analysis were to be adjusted to look at those rates in the combination and biologics trial arms, this would go some way in illuminating the findings presented here with the ERAN analysis. Other RCTs have also documented very low rates of radiological progression in methotrexate only arms, in line with rates reported in the ERAN sub-analysis. While the biologic treatment arms are reported as being statistically significantly lower compared to the methotrexate arms, they represent very small effects, much less than the clinically meaningful change[291, 303]. However, these have been restricted to just 1 year, where the association between disease activity and radiographic progression is unlikely to be fully realised due to the delayed onset of radiographic damage[2, 177].

DAS modelled as a time-varying covariate

The second analysis looking at disease activity as a TVC in this chapter found that disease activity was not associated with radiographic damage when measured at the same follow-up time point, but increased disease activity in the preceding 1 and 2 years were associated with increased radiographic damage. This was in largely in agreement with Welsing et al.[214], however this analysis also investigates the separate JSN and erosion scores, and looks at the individual components of the DAS. Two findings from the analysis by Welsing et al.[214] were not replicated in this analysis; that variation around the mean DAS was significantly associated with radiographic progression, and that radiographic progression was non-linear over time. The differences are likely due to differences in the models used in each study. Welsing et al.

applied an identity link function, which assumes that the dependent variable follows a Gaussian distribution (Normal). Chapter 5 demonstrates how this is rarely achieved with radiographic outcome data, and therefore it is likely that the non-linear associations would have seemingly provided improved model fit, by attempting to account for the large preponderance of zero scores. Despite the improved model fit statistics, the lack of model fit is highlighted in Figure 3 of Welsing's et al. study, which graphs the observed versus the predicted SvdH scores. It shows a large number of zero scores being predicted to range from 0 - 50, thereby underestimating the zero count.

The relationship between DAI's and radiological progression has been the subject of a number of studies, 58 of which were summarised by Navarro-Compàn et al.[177] in their systematic review (40 observational studies and 18 RCTs). The DAIs that were studied were predominantly the DAS and DAS-28 measures, but studies also looked at the simplified DAI (SDAI), the clinical DAI (CDAI), the RA DAI (RADAI) and the routine assessment of patient index data (RAPID). As was noted in Chapter 3, large heterogeneity in study populations, statistical methods and classification of 'significant' radiological progression meant that direct comparisons between this analysis, and indeed between the studies reported in the review, is extremely difficult. Furthermore, restriction to cross-sectional analysis, and the lack of multivariate techniques, did not allow for any longitudinal associations to be examined. Nevertheless, there were some trends emerging through pooling of these studies' results. Those studies that looked at time-integrated measures of DAI and their separate components were more likely to indicate significant associations with radiographic damage compared to those concentrating on baseline measures only. Only 4 of the 16 studies looking at the association between DAS at baseline and radiographic progression found a significant association [77, 304–306], whilst all studies looking at DAS as a time-integrated measure found a significant association with increased radiographic progression[75, 104, 211, 214, 285, 291, 301, 302, 307, 308].

The results from this study adds to the growing body of evidence that radiographic progression is a delayed consequence of increased disease activity over time. The results are also highly suggestive of the fact that this is largely driven by increased inflammatory markers over time, in this particular study ESR, rather than the subjective markers such as TJC. This was also highlighted in the systematic review paper and led to the recommendation that DAIs should, at the very least, include a measure of SJC[177], but these results would suggest acute phase markers, particularly ESR, is of higher importance. As discussed above, whether these prove to

be as important in the context of overall reductions in radiographic progression needs further review.

Radiographic progression and its association with HAQ

The next stage of this chapter's analysis was to investigate the association between radiographic progression and functional disability. As with the DAS analyses, the first models looked at the time-averaged effect of functional disability with radiographic progression. The results found that those patients with increased rates of radiographic progression over the 10 years were likely to have increase functional disability over the first 10 years. The effect was also found to increase significantly over time. This, coupled with the increasing correlation over time, indicated that HAQ is indeed associated with radiographic progression, but only in later disease. This was further substantiated in the TVC analysis, which indicates that HAQ scores increase over time as radiographic damage increases. However, unlike the DAS, there is little evidence of a lagged-effect. This result could suggest that while the effect of radiographic damage on patients is only apparent in established disease, once the disease is established, increases in radiographic progression have an immediate impact on increasing functional disability[294].

Interestingly, despite large reductions in radiographic damage in ERAN, the association with functional disability over 5 years remained largely unchanged. There was an indication of slight improvement by year 5 in ERAN compared to ERAS, but the reduced sample made it difficult to draw any definitive conclusions. Earlier research using both the ERAS and ERAN cohorts found a strong association between increased disease activity and increased functional disability[74]. While it is possible that improvements would be realised with more data, and longer follow-up beyond 5 years, the lack of any immediate effect between reduced radiographic progression and functional disability is suggestive of a second pathway involved in the development of functional disability in RA[297]. The second pathway is one that is driven primarily by pain, rather than inflammation. It could be theorised that functional disability is more highly associated with psychosocial factors, including pain perception[309]. Although inflammation has been adequately controlled, as evidenced by reductions in radiographic damage, pain is still largely driving the association between the subjective components (TJC and PGA) and functional disability[300].

Drawing parallels from this study with the previous literature is difficult, since there are large variations in the statistical methods. Furthermore, multivariate statistical methods were rarely

used[86]. The systematic review by Bombardier et al. does however highlight four main methods for which radiographic damage and functional disability were compared; baseline comparisons, correlations at specific follow-up time points, comparisons based on change scores of the radiographic damage over time with either the mean or final HAQ score at follow-up, and finally correlations between changes in radiographic damage and changes in HAQ over the follow-up. The review found inconclusive evidence for the cross-sectional association between radiological damage and functional disability at baseline, but did find evidence that the association increased as disease duration increased. Those studies that assessed either the change in radiographic damage with mean HAQ or HAQ at final follow-up, as well as those looking at changes in radiographic damage with changes in disability over time indicated significant associations, with four studies using multivariate analyses to control for important confounding effects[87, 310–312]. The authors argue these relationships are more clinically relevant, as they are suggestive that changes in radiographic progression through improved treatment can alter the severity of functional disability over time[86]. These findings are in direct agreement with the findings presented in this analysis. Furthermore, the authors suggest that high variability in radiological and inflammatory markers in early disease is likely to confound the association with functional disability. This is true, particularly when, on average, patient's HAQ score decreases in the first year of disease as treatment is introduced[38]. The conscious effort in this analysis to omit baseline measures limits the possible confounding effects they may have had on the model results.

Of the two components that make up the SvdH, JSN and erosions, many studies have reported an increased association between JSN and functional disability, as opposed to erosive damage[313–316]. There was small evidence to suggest that JSN is more strongly associated with long-term functional disability comapred to erosions, but the effect was small. As noted previously, there remains a paucity of studies looking at the time-dependant association between functional disability and the separate JSN and erosions components, with much of the literature focusing on either cross-sectional radiographic damage at baseline, or time-averaged effects[304, 315].

Strengths and Limitations

The findings from this study provide a detailed understanding of the long-term associations between radiographic progression and core RA outcomes in patients treated in a natural clinical setting between 1986-2001. However, its main focus on data from the ERAS cohort restricts the

ability to generalise these findings to RA patients treated using more contemporary treatment strategies. Validation analysis looking at the sub-sample of radiographic data available from the ERAN cohort, is highly suggestive of a reduced association between disease activity and radiographic progression, although caution is needed due to the reduced statistical power, due to the relatively low sample size. It is clear that in the context of T2T and biologic prescription, more up-to-date data is sorely needed in order to establish whether the finding in this study is replicated in larger datasets, with longer follow-up assessments.

Nevertheless, the analysis is one of the first to look at this relationship in such detail using the appropriate statistical models to ensure any bias in the analysis is minimised. The extensive analysis from the ERAS cohort serves as a comprehensive and worthwhile historical account of how treatment during the pre-biologic period of RA management affected radiographic progression, and its relationship with disease activity and functional disability. Any documented changes in more contemporary observational cohorts can therefore be compared and contrasted to these results.

Concluding remarks

In summary, the findings presented here suggest that for those patients treated during the pre-biologic era, radiographic progression is largely determined by increased disease activity from early on in the disease. If left uncontrolled, this increased radiographic progression was associated with increased functional disability in later disease. However, sub-analysis suggests that large reductions in radiographic damage in recent years may have dramatically reduced the clinical impact of increased disease activity on radiographic progression. Under current NICE guidelines, only those patients with high disease warrant use of more effective biologic DMARD therapies, due to increased disease burden. It is not clear whether widespread adoption of methotrexate early on in the disease has led to sufficient control of radiological progression, as has been demonstrated in some RCTs[56]. The association with functional disability was found to be largely unchanged by the reductions in radiographic progression, suggesting that functional disability may be more primarily driven by psychosocial factors, such as pain perception, rather than acute inflammation.

Further research is needed to collect long-term radiographic data in more contemporary cohorts to establish whether substantial treatment changes, particularly the use of very early methotrexate in combination with other conventional DAMRDs, is sufficient at reducing the

increased radiographic progression rates demonstrated here. There is also a need to look more closely at psychosocial factors and their relationship with functional disability, to assess how non-pharmacological interventions may be able to improve patients quality of life over and above disease control[309].

Chapter 8

Discussion

8.1 Introduction

The overarching objective of this thesis was to explore radiographic damage in patients with early RA. Four main aims were established to address this objective: (1) explore current methods for measuring radiographic damage, (2) assess the most suitable statistical techniques to evaluate radiographic damage longitudinally, (3) evaluate the 'natural, true-to-life' progression of radiographic damage using observational data, and finally (4) investigate the longitudinal relationship between radiographic damage and two core RA outcomes; disease severity and functional disability.

There are few observational studies with long-term radiographic data with large sample sizes, transcending both the pre and post biologic era, and the statistical methods typically used were inappropriate. The longitudinal data on radiographic outcomes that were available from both ERAS and ERAN provided a unique opportunity to explore the natural (treated) progression of radiographic damage over time in two cohorts representing distinct eras in the therapeutic management of RA.

This chapter will collate the findings from each of the chapters and look at how they address the specific aims of the thesis set out above, and how they fit with current evidence base. The main strengths and limitations of the cohorts and statistical methods used will be discussed, and the thesis will conclude by exploring what possible implications the findings from this thesis have on the clinical management of RA, and possible directions for future research in this area.

8.2 Aims of the thesis

8.2.1 Measurement of radiographic progression

The use of radiography in clinical practice is valuable to the clinician, as it can provide an objective 'snap-shot' of joint damage at any single point in time, and indeed any progression of this damage over time[119]. Chapter 2 provided a summary of all the scoring methods currently used in studies on RA, however both the Larsen and SvdH were found to be the most commonly used. Chapter 3 also confirmed that in more recent studies, the SvdH was the preferred method of choice. This is likely due to its ability to measure both JSN and erosions as separate scores, rather than being limited to just a single composite score, as with the Larsen method.

The Outcome Measures in RA (OMERACT) group recommends radiographic joint damage as a core outcome measure in RA research, since it achieves all three component criteria set by the OMERACT committee; 1) truth, in that it measures what it intends to measure with good face, content, construct and criterion validity, 2) discrimination, in that it can differentiate between two different situations of interest, be they time or states, and 3) feasibility, in that the plain radiographs can be easily obtained and interpreted in a cost-effective manner[131]. However, despite its advantages, understanding the limitations of radiographic scoring is crucial in ensuring radiographic scoring remains the 'gold standard' and does not become 'fool's gold'[118]. One of the most notable limitations of radiographic scoring is the time and training required to score each radiograph[118, 317]. While the use of expert radiologist to score radiographs has proved to result in little variability between scorers[97, 318], this does restrict the ease in which these measures can be recorded for both observational trials and clinical practice. This is particularly true when compared with other clinical and patient reported RA outcomes, such as the DAS and HAQ.

The Simple Erosion Narrowing Score (SENS) was developed to decrease the time taken to score radiographs in a clinical setting[243], with further attempts made to simplify the score further by omitting joints that do not have a large effect on the measurement performance[319]. Although some studies have found the SENS to have good agreement with the modified SvdH score[103, 243, 320, 321], most of these studies have been restricted to early disease and very small samples. Rau et al.[322] argue that since the SENS can only increase over time as new joints become affected, it lacks any reliability to detect increased radiographic damage in late disease, when no more new joints are involved.

Another limitation is inter-reader variation. Chapter 4 highlighted how variation increased at the higher end of the scale, and that heteroskedasicity needed to be accounted for in any analysis of radiographic outcomes. In the context of longitudinal studies, research has shown that inter-reader variability can be reduced by reading radiographs for each patient in chronological order[102, 323, 324]. However, this is at the expense of potentially over-estimating the progression of radiographic damage, as the reader might inherently be expecting to see increased progression over time[102]. This has important implications on the measurement of erosive healing. Bone erosions, or osteolysis (bone loss) occurs from an imbalance whereby bone resorption (osteoclasts) occurs more than bone formation (osteoblasts)[325]. Although physiologically possible, erosive healing remains a rare occurrence, even in those patients treated with biologics. In an RCT looking at adalimumab treated over 1 year, erosive healing was rare[326], and appears to occur exclusively in those joints with little to no joint swelling[327]. Although, it is noted that there remains a lack of longitudinal studies investigating its prevalence in contemporary cohorts[325].

While this study, and indeed other studies, have reported high inter-reader reliability scores, disagreements still do occur[118]. The impact of other co-morbid diseases, particularly OA on the interpretation of JSN, can easily lead to differences in how radiographic progression is assessed over time[328]. Since osteoporosis is more difficult to quantify, it is generally excluded from radiographic scoring methods[95].

While not without its faults, radiographic scores still represent one of the most important outcome measures in RA, providing an objective indication of the pathological damage caused by increased and prolonged inflammation[322]. While scores like the SENS seek to reduce the time burden associated with the application of radiographic scoring techniques, the reduction in sensitivity at the end end of the scores[104], and in late disease[322], do not outweigh the advantages in time saved. While new methods that account for reading order, co-morbid conditions and erosive healing are welcomed, these methods will still inevitably produce data with non-normal distributions. As such, it could be argued that more focus should be placed on appropriate analysis techniques, rather than development of new scoring methods.

8.2.2 Modelling of radiographic damage longitudinally

The systematic review conducted in Chapter 3 identified large heterogeneity in the statistical methods used to look at the progression and predictors of long-term radiographic progression.

This heterogeneity has been highlighted in previous systematic reviews[86, 313], where Bombardier et al.[86] specifically point out the lack of multivariate techniques in previous research. However, of equal concern is the apparent use of statistical methods, which are not appropriate given the properties of radiographic outcome data. Chapter 4 demonstrated how the distribution of scores are highly positively skewed due to large numbers of patients with early RA with no perceivable radiographic damage. The result is an excess of zero scores that causes over-dispersion.

Round table discussions in 2002[237] led to the development of guidelines on the reporting of radiographic results in RCTs. The lack of consistency in the reporting of mean or median values as the primary endpoint hinders the ability to produce uniform datasets for wider comprehension of the data[237]. The recommendations includes the following; the use of two readers, the reporting of the SDD as quality control, presentation of absolute numbers, focus on group level estimates as the primary analysis and the use of both the mean/SD and box plots for the median/IQR. The use of cumulative probability plots have also been presented as a means of improving interpretability[249]. This plots the cumulative probability of the score against the actual score, in ranked order. This method is advantageous as it is an effective method of summarising the range of scores for either an individual over time, or for specific sub-groups of patients (e.g. trial arms in an RCT). While these recommendations have been pivotal at highlighting the statistical challenges involved with radiographic data in RA, they are only useful as an aid to interpreting radiographic damage, and are not a replacement for statistical models[249]. As such, more is needed to improve the use of statistical models in the RA literature on radiographic progression.

The non-normal properties of radiographic data distributions is likely to explain the large variation in statistical methods used[329]. Transformation of the radiographic score into binary groups has been seen as the most effective way of dealing with the non-normal distribution, however, defining these groups has been far from easy. One of the core initiatives as part of the OMERACT V meeting was to assess the application of a MCID threshold, whereby a certain degree of radiographic progression would determine whether a patient had 'clinically meaningful' radiological damage or not[122]. The initaitve failed to yield any decisive definitions for this cut-point, however it was agreed that the SDD was an appropriate starting point. While the SDD can reflect the point at which a change in score is no longer due to measurement error[124, 283], it might not reflect clinically relevant changes[330]. As such, a change in score that also reflects changes in other important related outcomes, such as HAQ, would ensure clinical relevance

of any defined thresholds[330]. What has been overlooked however, is that current publicised methods for calculating the SDD rely on the standard deviation as a measure of variance. Given the highly positive skew of radiographic data demonstrated in Chapter 4, the asymmetry of the data distribution renders the SD inappropriate, since the variance below the mean is likely to be lower than the variance above the mean. For these reasons, this thesis looked at alternative methods of modelling radiographic damage that did not rely on categorisation or forms of data transformations.

Chapter 5 demonstrated why linear models, even with data transformation methods applied, are not appropriate for radiographic outcomes. Given that radiographic scores are more accurately defined as a weighted count outcome, the chapter explored the suitability of count based regression methods. This would allow the radiographic score to be modelled without any transformation, while still accounting for its non-normal distribution. The two main count distributions investigated were the Poisson and NB. The Poisson distribution failed to adequately account for the high frequency of zero scores, since it assumed the data had equidispersion (that is the mean was equal to the variance). It was clear from the data that radiographic outcomes suffer from overdispersion (variance was much larger than the mean), therefore the NB was needed. The NB method provided very good model fit and was able to accurately predict the large proportion of zero scores.

Zero-inflated Poisson and NB were also explored in Chapter 5, where a two-component or two-part modelling process is undertaken. The first component models the probability of the count being a zero or non-zero score (binary model), while the second component models the frequency count of all non-zero scores (count model). Park et al.[126] conducted a similar study looking at the suitability of count regression methods on radiographic data, using data from 190 RA patients with follow-up for up to 3 years. They also found the NB method to be superior to the Poisson in terms of model fit. However, when including the zero-inflated parameter in both Poisson and NB models, they concluded that the zero-inflated Poisson model provided the best model fit over both the NB and zero-inflated NB.

For the analysis conducted in this thesis as part of Chapter 5, both the Poisson and zero-inflated Poisson indicated the worst model fit. There were only marginal gains in predicted zero counts and model fit statistics by adopting the NB zero-inflated model over the standard NB model. This was evidenced by the predictive probability plots in Figure 5.8, which demonstrated that both the Poisson and zero-inflated Poisson did not accurately predict the lower proportion of

the scores as well as either the NB or the zero-inflated NB. The differences could be explained by differences in the mean-variance relationship in the ERAS data versus the data used in the study by Park et al.[126].

Further support for the suitability of the models is provided by Xie et al.[331]. They note how for instances where the mean count is low due to rare events, zero-adjusted models are not always necessary, and the adoption of zero-inflated methods can lead to over-fitting and less parsimonious models. Additionally, while zero-inflated methods are an effective means of dealing with excess zero scores, the application of both of these modelling techniques make theoretical assumptions about the nature of the zero-scores. In the context of radiographic progression, a zero-inflated model assumes that non-erosive forms of RA exist. That is, there is a probability that some patients have a form of RA where radiographic damage will never occur. There is little evidence of non-erosive forms of RA[263], and it is believed that all forms of RA have the capacity to incur radiographic damage if not suitably controlled. Conceptually, trying to tease apart those patients that have non-erosive forms of RA, compared to those that have milder forms of RA but are suitably controlled by DMARD therapies, becomes very difficult in the observational study setting. This thesis therefore argues that the use of zero-inflated methods should be decided based on conceptual grounds, rather than purely statistical grounds[331], and as yet, there is little theoretical evidence for the use of zero-inflated in the application of radiographic outcomes.

Finally, the importance of using either GEE or multi-level models in longitudinal data was outlined. While both are acceptable methods for modelling data over time, the advantages of looking at within-group estimates, and the use of FIML estimators to handle missing data, led to the conclusion that multi-level NB regression is the preferred modelling technique of choice. It was therefore subsequently used in the primary analysis of this thesis.

8.2.3 Progression of radiographic joint damage

The third aim of the thesis was 'to understand the natural (but treated) progression of radiographic damage in RA using data from two UK inception cohorts'. Chapter 6 outlined the first primary analysis of radiographic data from the ERAS and ERAN cohorts, using the statistical models outlined in Chapter 5. The results found that patients in ERAN (treated from 2002-2012) have approximately half the radiographic damage at baseline, and approximately half the annual rate of radiographic progression over the first 5 years, compared to patients in ERAS

(treated from 1986-2001). This was largely driven by reductions in JSN, with reductions in erosions contributing in later disease only.

Despite reflecting slightly different periods, the meta-analysis in Chapter 3 indicated similar results, with patients treated post-1990 estimated to have half the yearly rate of radiographic damage compared to those patients treated pre-1990. This trend towards decreasing radiographic progression has been documented in other studies[224, 226, 275], with decreased radiographic progression in more recent years. In contrast, the difference in baseline radiographic rates in Chapter 6 was only demonstrated in the study by Fiehn et al.[275]. The systematic review in Chapter 3 and the studies by Finckh et al.[226] and Sokka et al.[224] demonstrated similar rates of radiographic damage at baseline between the periods analysed. It is likely that these differences between baseline rates reflects a secular trend. The systematic review in Chapter 3 and the studies by Finckh et al and Sokka et al. all compared patients from the 1980s to those patients from the 1990s. The analysis in Chapter 6 and the study by Fiehn et al. primarily compare patients from the 1990s to those recruited from the 2000s. Overall, the evidence suggests that baseline rates of radiographic damage was similar for those patients treated in the 80s and 90s, but patients treated in the 2000s and onwards have significantly lower radiographic damage at presentation.

While the non-randomised nature of the study means it is not possible to directly test the causal nature of this secular decline, it is hypothesised that the early use of methotrexate in the ERAN cohort, during the so-called 'window-of-opportunity'[54], is likely to explain these improvements. However, the reduction of radiographic damage at baseline observed between ERAS and ERAN could be evidence of less severe disease. While direct comparisons between the DAS is complicated by the use of the DAS-44 in ERAS and the DAS28 in ERAN, the median ESR at baseline, the proportion of patients with low HB and the proportion of seropositive; RA in ERAN was substantially lower. Fiehn et al.[275] also demonstrated reduced radiographic damage in their later cohort, along with significantly reduced ESR at baseline.

While the model controlled for the use of DMARD and steroid treatment in the ERAN cohort, capturing the full extent of actual use and dosage of these drugs is complex. It is also possible that the reduction in radiographic damage seen at baseline in the ERAN patients is a reflection of early initiation of steroids in primary care in a small proportion of patients, for which duration and dosage was not fully captured, and which occurred much less in the ERAS cohort. It is also possible that the effect of milder disease at presentation may have also contributed

to the decreased radiographic progression rates over the following 5 years. The systematic review in Chapter 3 found that acute phase markers were consistently associated with increased radiographic progression over the long-term, insinuating that the decreased progression is due to a combination of both a milder form of disease at presentation and improved treatment early on.

The second major finding in Chapter 6 was the changing contribution of seropositive RA on the progression of radiographic damage between the two cohorts. While seropositivity was associated with increased radiographic progression in both cohorts, the absolute change in score was much lower in ERAN, compared to ERAS. This resulted in seropositive patients in ERAN demonstrating significantly lower radiographic progression over the first 5 years compared to seronegative patients in ERAS. It should be noted that seropositivity in both cohorts was restricted to RF positivity, with only a very small proportion of patients being tested for ACPA positivity. In a recent study by Hecht et al.[332], the impact of RF and ACPA on erosive damage was evaluated using high-resolution peripheral quantitative CT (HR-pQCT). They found an interdependence of both RF and ACPA anti-bodies on bone erosions, where the presence of both led to an additive effect on the cumulative prevalence and size of bone erosions. However, when van Steenbergen et al.[333] investigated this hypothesis using radiographic data scored using the SvdH method from two early RA cohort studies in the Netherlands, they found that the rate of radiographic progression for those patients who were ACPA positive had similar levels of radiographic progression regardless of RF status. This does not support the additive effect, as suggested by Hecht et al.[332]. These studies do indicate increased radiographic damage in those patients that are ACPA positive compared to just RF positivity, a finding that has been replicated in other studies[288, 289]. Both studies utilise linear longitudinal methods, and were restricted to just the erosion score. Future research needs to investigate the added association of ACPA positivity over RF positivity on total radiographic damage, while using appropriate methods described in this thesis. The impact of secular changes on ACPA positive RA is also currently unknown.

8.2.4 Radiographic damage, disease activity and functional disability

The final chapter, Chapter 7, addresses the fourth aim of the thesis; to investigate the longitudinal relationship between radiographic damage and two core RA outcomes; disease severity and functional disability. Disease activity measures are crucial in the therapeutic management

of RA, as they provide the targets which clinicians aim for under the current T2T paradigm[49, 57]. It is important to understand the progression of radiographic damage, and its relationship with disease severity, as previous studies have shown it to be associated with important patient outcomes, namely functional disability[86]. The analysis presented in Chapter 7 indicated that increased disease activity is associated with increased long-term radiographic progression and functional disability. Looking at the longitudinal relationship, there was evidence of a time-lagged effect between markers of inflammation, namely ESR and SJC, and radiographic progression. This finding is in agreement with previous systematic reviews investigating the association between radiographic progression and DAIs[177].

In the earlier ERAS cohort, moderate disease is associated with similar rates of radiographic progression as those patients with high disease. However, sub-analysis looking at patients from the ERAN cohort indicated that dramatic reductions in overall radiographic rates means that the progression of radiographic damage was relatively low across all disease groups. This has very important implications on the debate about whether biologic therapies need to be introduced to those patients in moderate disease. RCTs, looking at the addition of combination biologic DMARDs in patients with moderate disease, have concluded that the inclusion of biologics lead to significant improvements in radiological damage[291, 303]. However, investigation of the reported annual rate of radiographic damage over 1 year was just 0.8 units for the methotrexate only group[303], and just 0.6 units per year over 7 years[291]. As was shown in this analysis, while the association with disease activity is statistically significant, it indicates only very minor increases in radiological progression, much lower than reported rates deemed clinically meaningful.

While the introduction of biologic therapies will undoubtedly lead to improved radiographic outcomes, the economic argument for introducing biologics in moderate patients needs to be explored. From purely a radiographic standpoint, early combination therapy with methotrexate is evidenced to reduce radiographic progression to a similar level to that of combination biologic therapies[56, 334]. The need to use biologic therapies in relation to other RA outcomes is more striking. Studies using the data from both ERAS and ERAN have demonstrated that those patients with persistent moderate disease over the first 5 years were found to have increased functional disability and increased incidence of orthopaedic surgical rates[74], and that ERAN patients with moderate disease at baseline, were unlikely to achieve clinical remission by year 3 on conventional DMARD therapies[72]. Data from the BSRBR has also demonstrated that the

use of biologic DMARD's can significantly reduce the level of functional disability over 1 year in patients with moderate RA[73].

The reduction in radiographic damage in ERAN also correlates with improvements in other outcomes related to radiography. Nikiphorou et al. found that the incidence of intermediate orthopaedic surgical procedures (involving the small joints of the hands and feet) have significantly declined in more recent years[274], which may be explained in part by reductions in radiological progression. However, although the analysis shows a significant association between radiographic damage and functional disability, one which does not occur until later in the disease, there was little, if any, evidence that the association improved in the ERAN cohort. These findings are supportive of the model proposed by Escalante and Ricón in 2002, where functional disability can be a result of either inflammation and radiographic damage, or through pain, mediated by psychosocial factors[297]. It could be hypothesised that the increased relationship between disease activity and functional disability seen in this patient group[74], is being largely driven by the subjective markers of DAS (TJC and PGA), and that other aspects of the biopsychosocial model proposed by Escalante and Ricón should be targeted to reduce functional disability[297].

8.3 Strengths and Limitations

The data comes from two of the largest early RA cohorts in the UK, which recruited a combined total of 2,701 patients followed-up for up to 25 years. All patients were treated in rheumatology outpatient clinics across the UK, based on published clinical guidelines at the time. The patients were recruited from centres all over the UK, ensuring high generalisability to a large number of patient groups. The analyses presented within concentrates on radiographic data. A total of 7,100 plain x-rays collected from 1,678 patients were scored using the SvdH method over the first 10 year period, while 5,763 radiographs from 1,662 patients were scored using the SvdH method over the first 5 year period . This large data sample allowed for multi-level, multivariate modelling techniques to provide precise estimates on the natural progression of radiographic progression transcending over 3 decades.

Although the use of evidence hierarchies in evidence based medicine has helped clarify the quality and usefulness of a specific study design in answering a research question, its simplified approach has also resulted in a lot of misconceptions[335]. Clinical trials are widely regarded as the most

effective means of determining the effect of an intervention[134]. However, different research questions require different research designs in order to provide the most valid results[133]. In the case of this thesis, observational cohort data is the most appropriate means of examining the natural progression of the disease in specific patient sub-groups over a long time period[336].

However, understanding the potential causal links surrounding the results found is difficult without the randomised nature of clinical trials. Potentially important confounding effects, both known and unknown, can be minimised through random allocation of patients, and blinded assessment of outcomes[133]. Although, that is not to say that causal effects are impossible to investigate using observational data. Statistical techniques, such as propensity scoring, can be an effective means of looking at treatment effects in observational data[337]. However, as with all methods, it is imperative that the methods are correctly applied[337]. The approach taken by this thesis was not to apply methods such as propensity scoring, but to ensure the analysis made every effort to control for potential confounding effects through the use of multivariate regression techniques. The analysis avoided automated covariate selection techniques, such as stepwise selection, as those have been shown to introduce bias[338]. Instead, the selection of covariates was selected based on prior research and theory about which factors are important in the progression of radiographic damage. Potential causal associations, namely treatment effects and milder disease, were highlighted based on the findings, and the weight of this evidence was supplemented by findings from experimental studies. The result is that the conclusions regarding why these patterns are emerging are subject to further scrutiny, and can only be confirmed or indeed denied by further research.

Another key limitation of observational data is missing data. The analyses in Chapter 7 largely utilise the radiographic data from the ERAS cohort, for which the proportion of patients with missing data is small. Missing data beyond 10 years was high due to non-participation of centres beyond the 10 year mark, leading to the decision to restrict analyses to 10 years only. Missing radiographic data within these 10 years was accounted for by the use of multi-level modelling, which utilises FIML estimation, allowing for radiographic rates over the full 10 years to be estimated for each patients with at least 1 available data point[339]. Furthermore, the characteristics of those patients with and without radiographic outcomes was found to be very similar, leading to the conclusion that the bias from missing data was likely to be small, and it's likely the MAR assumption is satisfied. Nevertheless, the possibility that those patients with greater number of radiographs represent a sub-group of patients with more aggressive forms of RA (hence the need for more x-rays), is still present. In contrast, there was a significantly

reduced amount of radiographic data for the ERAN cohort. While radiographic data was available for all centres, scoring of those radiographs using the SvdH method was only conducted for 6 of the 23 centres. Although, it was found that patients from these 6 specific centres were similar to the whole cohort with regards to demographic and baseline characteristics, suggesting that the data was MAR. As with the ERAS cohort, those patients with radiographic data were similar in demographic and clinical characteristics to those without radiographic data. The reduced data sample did however restrict the ability to apply longitudinal models over 5 years, whilst maintaining high levels of precision around the model estimates. Further scoring of the digitised radiographs from the ERAN cohort would allow for more accurate estimates from the ERAN cohort, as well as extending the analysis to longer follow-up periods of up to 10 years. Missing data from the covariates included in the model were fairly low, and driven more by missing radiographic data. While the incorporation of multiple imputation methods into multi-level models is possible, it is complex[340]. Given the relatively low missing data on DAS and HAQ measurements compared to the radiographic data, it is unlikely that the incorporation of these complex methods would have yielded remarkably different findings.

8.4 Clinical Implications

The clinical utility of radiographic assessment has been challenged on several grounds[119]. Many clinicians question their use in routine clinical practice due to the lagged effect between pathological changes and the manifestation of erosions and JSN that can be detected on plain radiographs[119]. However, elsewhere it is regarded as an indispensable and a readily available tool in the clinical setting to guide therapeutic management[119]. EULAR recommendations in 2013[84] on the use of imaging in the clinical management of RA indicated strong support of periodic radiographs to monitor disease progression. In light of the secular declines displayed in this thesis, showing relatively low rates of radiographic progression, the continued need for routine x-rays is debatable. While the thesis has focused primarily on the use of plain x-rays, EULAR does highlight the use of other imaging modalities, such as MRI and ultrasound, in detecting soft tissue inflammation. However, the availability, practicality and cost effectiveness of these modalities in routine care is still yet to be determined[84].

At the other end of the spectrum, sub-clinical inflammation has been shown to progress, even in those patients in clinical remission[341, 342]. Clinical remission defined using DAS thresholds is not uncommon, even in historical cohorts, such as ERAS[343–345]. While not the focus of

this thesis, radiographic progression was evident in patients with sustained remission in both ERAS and ERAN, albeit at a much lower rate compared to other disease activity groups. This is supported by additional research that looked at the Larsen score in the same data[346]. As a result, many argue that there is a need to incorporate radiographic measures within the 'clinical remission' definition set by disease activity scores[347, 348].

The analysis looking at the association between disease severity and radiographic progression has also provided unique insight into the debate surrounding the use of biologic DMARDS in the UK. Current guidelines permit the use of biologics in patients with sustained high disease activity, and there have been recent calls to reduce this threshold to include moderate disease in order to achieve the target of remission[72–74]. Data from the ERAS study provided more evidence that those patients with moderate RA follow a similar level of radiographic progression compared to patients with high RA. However, the large secular decline seen in the ERAN cohort again suggests that, from purely a radiographic perspective, patients overall are not progressing at a rate seen in historical cohorts. While the introduction of biologic DMARDs in more patients with less severe disease would undoubtedly lead to improvement in radiographic outcomes, the magnitude of this effect needs to be established. This has been shown in RCTs[56, 334], where tight management through the use of early combination DAMRDs with methotrexate, according to T2T principles, has been proven to be as effective as combination methotrexate with biologic therapies.

The possibility of a two pathway model driving functional disability highlights the pressing need to look at other psychosocial interventions in the management of RA[297, 349]. Patients with RA report higher levels of depression and anxiety when compared to healthy individuals[350]. Impaired cognition brought about by depression and anxiety can reduce the ability for an RA patient to manage pain, bringing about a cyclical pattern whereby depression causes increased pain, and increased pain causes further depression[309]. Pharmacological treatments are aimed at modifying the disease, which lead to pain through inflammation. However, patients typically require the additional use of analgesics to further manage their symptoms[300]. The assessment of quality of life measures in the ERAN cohort has shown improvements in the first 12 months of up to 20%. Although, rates remained the same over the next 4 years, at lower levels than the rest of the UK[300]. A systematic review has shown that non-pharmacological interventions can lead to improvement in RA outcomes, including functional disability[351], although the methodological quality of these studies was found to be relatively poor. Further emphasis is needed in understanding how interventions can improve patients' quality of life[309, 352, 353],

and how psychosocial factors can influence disease outcomes, such as functional disability, so that more holistic treatment strategies can be implemented in routine care to address all aspects of the disease; from inflammation to quality of life.

8.5 Future Research

The natural next step in continuing this research is to apply the models developed within this thesis to larger, more contemporary, cohort data. In particular, increased radiographic data from patients treated in the biologic era would enable the findings from this thesis to be extended.

RA encompasses a wide spectrum of disease severity, ranging from mild and remitting, to severe. This thesis has highlighted the high variation in radiographic progression between patients over time, and the subsequent need to apply modelling techniques to account for this. While these methods are appropriate for estimation of single group trajectories, there is perhaps an argument to define sub-group trajectories of patients that progress in different ways over time. Research by Park et al.[208] examined the possibility of different patterns of radiographic progression by using statistical clustering methods. Using this method, they identified 3 distinct patterns of radiographic progression using data from 190 early RA patients examined over 3 years; increasing, increasing then decreasing and flat. These patients represent a relatively small sub-sample of RA patients, treated during the pre-biologic era only. Future research could extend these findings to see whether these sub-groups of patients exist in both a larger cohort in the pre-biologic era (such as ERAS), and whether differences in disease severity and treatment in the post-biologic era, has led to differences in the patterns of radiographic progression.

Further still, the increased patient numbers and follow-up could also allow for more novel statistical methods to be utilised to identify these sub-group trajectories. A recent paper by Norton et al.[292] looked at the use of Growth Mixture Modelling (GMM), a form of latent class modelling, to investigate distinct trajectories of functional disability (HAQ) over 10 years using the ERAS data. The traditional methods of multi-level modelling used in this thesis assume that a single trajectory can adequately approximate the entire sample[292]. Latent based modelling relaxes this assumption and allows for distinct sub-group trajectories to be examined. As a result, a total of 4 distinct HAQ trajectories were identified; low stable, moderate stable, moderate increasing and high stable. Understanding the potential factors that could predict patient membership into these sub-groups could prove to be an invaluable tool in shaping treatment

and management of RA. Norton et al.[292] highlighted how the Larsen score was significantly increased at baseline, 3 and 5 years in the moderate increasing and high stable HAQ groups. Increased radiographic data through scoring of more radiographs in the ERAN cohort would also allow these more complex models to be conducted in both cohorts, allowing for a more detailed look at radiographic progression between the two.

The long-term association with other important factors also need to be investigated. Smoking, for example, has also been shown to be related to radiographic damage[34, 354, 355]. However, there was insufficient data to investigate this relationship within this thesis. Collection of smoking data in ERAS was not started until the 1990s resulting in firstly, missing data, and secondly bias towards those still alive.

8.6 Final conclusions

RA is a disabling and chronic disease, which can lead to joint destruction and disability. Adequate control is key, and recent adoption of T2T principles have shown remarkable success in reducing the disease burden for patients now living with the disease. There is emerging evidence that this has had a large impact on reducing radiographic damage. However, functional disability has shown little improvement. While more evidence is needed to justify the increased use of biologic DMARDs in those patients with less severe disease from a radiographic standpoint, it could be pivotal in controlling long-term functional disability. However, it is possible that a secondary mechanism causing disability exists, which could be entirely separate from inflammation, and therefore more focus should be applied to looking at non-pharmacological interventions. A more holistic treatment approach, incorporating psychosocial interventions, with continued improvements on pathophysiological aspects, will likely lead to better disease control, as well as improvements to patients' overall quality of life.

Bibliography

[1] Choy, E. H. and Panayi, G. S. "Cytokine pathways and joint inflammation in rheumatoid arthritis." *The New England journal of medicine* 344.12 (2001), pp. 907–16.

[2] Scott, D. L., Wolfe, F, and Huizinga, T. W. J. "Rheumatoid arthritis". *The Lancet* 376.9746 (2010), pp. 1094–1108.

[3] Scott, D. L. et al. "Long-term outcome of treating rheumatoid arthritis: results after 20 years". *Lancet* 1.8542 (1987), pp. 1108–1111.

[4] National Rheumatoid Arthritis Society. *The Economic Burden of Rheumatoid Arthritis.* Tech. rep. March. 2010, pp. 1–20.

[5] Symmons, D. et al. "The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century". *Rheumatology* 41.7 (2002), pp. 793–800.

[6] Ropes, R. M. W. "Diagnostic Criteria for Rheumatoid Arthritis". *Annals of the rheumatic diseases* 18.1 (1959), pp. 49–53.

[7] Arnett, F. C. et al. "The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis." *Arthritis and rheumatism* 31.3 (1988), pp. 315–324.

[8] Harrison, B. J. et al. "The performance of the 1987 ARA classification criteria for rheumatoid arthritis in a population based cohort of patients with early inflammatory polyarthritis. American Rheumatism Association." *Journal of Rheumatology* 25.12 (1998), pp. 2324–2330.

[9] Alamanos, Y. and Drosos, A. A. "Epidemiology of adult rheumatoid arthritis". *Autoimmunity Reviews* 4.3 (2005), pp. 130–6.

[10] Carmona, L et al. "The prevalence of rheumatoid arthritis in the general population of Spain." *Rheumatology (Oxford, England)* 41.1 (2002), pp. 88–95.

[11] Cimmino, M. A. et al. "Prevalence of rheumatoid arthritis in Italy: the Chiavari Study." *Annals of the rheumatic diseases* 57.5 (1998), pp. 315–318.

[12] Stojanović, R et al. "Prevalence of rheumatoid arthritis in Belgrade, Yugoslavia." *British journal of rheumatology* 37.7 (1998), pp. 729–732.

[13] Spindler, A. et al. "Prevalence of rheumatoid arthritis in Tucumán, Argentina." *The Journal of rheumatology* 29.6 (2002), pp. 1166–70.

[14] Pountain, G. "The prevalence of rheumatoid arthritis in the Sultanate of Oman." *British journal of rheumatology* 30.1 (1991), pp. 24–8.

[15] Akar, S et al. "The prevalence of rheumatoid arthritis in an urban population of Izmir-Turkey." *Clinical and experimental rheumatology* 22.4 (2004), pp. 416–20.

[16] Abdel-Nasser, A. M., Rasker, J. J., and Valkenburg, H. A. "Epidemiological and clinical aspects relating to the variability of rheumatoid arthritis." *Seminars in arthritis and rheumatism* 27.2 (1997), pp. 123–40.

[17] Silman, A. J. et al. "Absence of rheumatoid arthritis in a rural Nigerian population." *The Journal of rheumatology* 20.4 (1993), pp. 618–22.

[18] Rothschild, B. M. "Rheumatoid arthritis at a time of passage." *Journal of Rheumatology* 28.2 (2001), pp. 45–50.

[19] Symmons, D. P. et al. "The incidence of rheumatoid arthritis in the United Kingdom: results from the Norfolk Arthritis Register." *British journal of rheumatology* 33.8 (1994), pp. 735–739.

[20] Kaipiainen-Seppänen, O et al. "Incidence of rheumatoid arthritis in Finland during 1980-1990." *Annals of the rheumatic diseases* 55.9 (1996), pp. 608–611.

[21] Boots, A. M. H. et al. "The influence of ageing on the development and management of rheumatoid arthritis." *Nature reviews. Rheumatology* 9.10 (2013), pp. 604–13.

[22] Perruccio, A. V., Power, J. D., and Badley, E. M. "Revisiting arthritis prevalence projections–it's more than just the aging of the population." *The Journal of rheumatology* 33.9 (2006), pp. 1856–62.

[23] Sokka, T and Pincus, T. "Ascendancy of weekly low-dose methotrexate in usual care of rheumatoid arthritis from 1980 to 2004 at two sites in Finland and the United States". *Rheumatology (Oxford)* 47.10 (2008), pp. 1543–1547.

[24] Humphreys, J. H. et al. "Mortality trends in patients with early rheumatoid arthritis over 20 years: results from the Norfolk Arthritis Register." *Arthritis care & research* 66.9 (2014), pp. 1296–301.

[25] Dadoun, S. et al. "Mortality in rheumatoid arthritis over the last fifty years: systematic review and meta-analysis." *Joint, bone, spine : revue du rhumatisme* 80.1 (2013), pp. 29–33.

[26] Choi, H. K. et al. "Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study." *Lancet (London, England)* 359.9313 (2002), pp. 1173–7.

[27] Choy, E. "Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis". *Rheumatology.(Oxford)* 51 Suppl 5.1462-0332 (Electronic) (2012), pp. v3–11.

[28] Klareskog, L., Catrina, A. I., and Paget, S. "Rheumatoid arthritis". *The Lancet* 373.9664 (2009), pp. 659–672.

[29] Klareskog, L. et al. "A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination." *Arthritis and rheumatism* 54.1 (2006), pp. 38–46.

[30] Holoshitz, J. "The Rheumatoid Arthritis HLA-DRB1 Shared Epitope". *Current opinion in rheumatology* 22.3 (2010), pp. 293–298.

[31] Weyand, C. M. et al. "The Influence of HLA-DRB1 Genes on Disease Severity in Rheumatoid Arthritis". *Annals of Internal Medicine* 117.10 (1992), pp. 801–806.

[32] MacGregor, A. J. et al. "Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins". *Arthritis & Rheumatism* 43.1 (2000), pp. 30–37.

[33] Alamanos, Y., Voulgari, P. V., and Drosos, A. A. "Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review." *Seminars in arthritis and rheumatism* 36.3 (2006), pp. 182–8.

[34] Silman, A. J., Newman, J., and MacGregor, A. J. "Cigarette smoking increases the risk of rheumatoid arthritis: Results from a nationwide study of disease-discordant twins". *Arthritis and rheumatism* 39.5 (1996), pp. 732–735.

[35] Baka, Z., Buzás, E., and Nagy, G. "Rheumatoid arthritis and smoking: putting the pieces together." *Arthritis research & therapy* 11.4 (2009), p. 238.

[36] Nishimura, K. "Meta-analysis: Diagnostic Accuracy of Anti–Cyclic Citrullinated Peptide Antibody and Rheumatoid Factor for Rheumatoid Arthritis". *Annals of Internal Medicine* 146.11 (2007), p. 797.

[37] Jónsson, T, Thorsteinsson, J, and Valdimarsson, H. "Elevation of only one rheumatoid factor isotype is not associated with increased prevalence of rheumatoid arthritis–a population based study." *Scandinavian journal of rheumatology* 29.3 (2000), pp. 190–1.

[38] Scott, D. L., Smith, C, and Kingsley, G. "Joint damage and disability in rheumatoid arthritis: an updated systematic review." *Clinical and experimental rheumatology* 21.5 Suppl 31 (2003), S20–7.

[39] Rantapaa-Dahlqvist, S. et al. "Antibodies Against Cyclic Citrullinated Peptide and IgA Rheumatoid Factor Predict the Development of Rheumatoid Arthritis". *Arthritis and Rheumatism* 48.10 (2003), pp. 2741–2749.

[40] Szodoray, P. et al. "Anti-citrullinated protein/peptide autoantibodies in association with genetic and environmental factors as indicators of disease outcome in rheumatoid arthritis." *Autoimmunity reviews* 9.3 (2010), pp. 140–3.

[41] Van Heemst, J. et al. "HLA and rheumatoid arthritis: How do they connect?" *Annals of Medicine* 46.5 (2014), pp. 304–310.

[42] Van Der Woude, D. and Huizinga, T. W. J. "Genetics: Every shared epitope allele for itself?" *Nature Reviews Rheumatology* 5.9 (2009), pp. 477–478.

[43] Symmons, D and Harrison, B. "Early inflammatory polyarthritis: results from the norfolk arthritis register with a review of the literature. I. Risk factors for the development of inflammatory polyarthritis and rheumatoid arthritis." *Rheumatology (Oxford, England)* 39.8 (2000), pp. 835–43.

[44] Symmons, D. P. M. "Epidemiology of rheumatoid arthritis: Determinants of onset, persistence and outcome". *Best Practice and Research: Clinical Rheumatology* 16.5 (2002), pp. 707–722.

[45] Brennan, P et al. "Oral contraceptives and rheumatoid arthritis: results from a primary care-based incident case-control study." *Seminars in arthritis and rheumatism* 26.6 (1997), pp. 817–23.

[46] Silman, A, Kay, A, and Brennan, P. "Timing of pregnancy in relation to the onset of rheumatoid arthritis." *Arthritis and rheumatism* 35.2 (1992), pp. 152–5.

[47] Smolen, J. S. et al. "A simplified disease activity index for rheumatoid arthritis for use in clinical practice". *Rheumatology (Oxford)* 42.2 (2003), pp. 244–257.

[48] Panayi, G. S. "B cells: a fundamental role in the pathogenesis of rheumatoid arthritis?" *Rheumatology (Oxford, England)* 44 Suppl 2.July 2004 (2005), pp. ii3–ii7.

[49] Smolen, J. S., Aletaha, D, and J. S. A. Smolen, D. "The assessment of disease activity in rheumatoid arthritis". *Clin Exp Rheumatol* 28.3 SUPPL. 59 (2010), S18–S27.

[50] Sokka, T. M. et al. "Conventional monotherapy compared to a "sawtooth" treatment strategy in the radiographic progression of rheumatoid arthritis over the first eight years". *Clin Exp Rheumatol* 17.5 (1999), pp. 527–532.

[51] Egsmose, C et al. "Patients with rheumatoid arthritis benefit from early 2nd line therapy: 5 year followup of a prospective double blind placebo controlled study". *J Rheumatol* 22.12 (1995), pp. 2208–2213.

[52] Van Der Heijde, A et al. "The effectiveness of early treatment with "second-line" antirheumatic drugs. A randomized, controlled trial". *Annals of internal medicine* 124.8 (1996), pp. 699–707.

[53] Choy, E. H. et al. "Treating rheumatoid arthritis early with disease modifying drugs reduces joint damage: a randomised double blind trial of sulphasalazine vs diclofenac sodium". *Clin Exp Rheumatol* 20.3 (2002), pp. 351–358.

[54] O'Dell, J. R. "Treating rheumatoid arthritis early: a window of opportunity?" *Arthritis and rheumatism* 46.2 (2002), pp. 283–5.

[55] Emery, P et al. "Clinical identification and treatment of a rapidly progressing disease state in patients with rheumatoid arthritis". *Rheumatology* 47.4 (2008), pp. 392–398.

[56] Goekoop-Ruiterman, Y. P. et al. "Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial". *Arthritis Rheum* 52.11 (2005), pp. 3381–3390.

[57] Smolen, J. S. et al. "Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force." *Annals of the rheumatic diseases* 75.1 (2016), pp. 3–15.

[58] Smolen, J. S. and Aletaha, D. "What should be our treatment goal in rheumatoid arthritis today?" *Clin Exp Rheumatol* 24.6 SUPPL. 43 (2006), S7–S13.

[59] Pinals, R. S., Masi, A. T., and Larsen, R. A. "Preliminary criteria for clinical remission in rheumatoid arthritis". *Arthritis & Rheumatism* 24.10 (1981), pp. 1308–1315.

[60] Felson, D. T. et al. "American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials". *Arthritis Rheum* 63.3 (2011), pp. 573–586.

[61] Boers, M et al. "World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials." *The Journal of rheumatology. Supplement* 41 (1994), pp. 86–9.

[62] Felson, D. T. et al. "The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials". *Arthritis Rheum* 36.6 (1993), pp. 729–740.

[63] Felson, D. T. et al. "American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis." *Arthritis and rheumatism* 38.6 (1995), pp. 727–35.

[64] Van Der Heijde, D et al. "Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score." *Annals of the Rheumatic Diseases* 49.11 (1990), pp. 916–920.

[65] Prevoo, M. L. L. et al. "Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis". *Arthritis & Rheumatism* 38.1 (1995), pp. 44–48.

[66] Van Der Heijde, D and Jacobs, J. W. "The original "DAS" and the "DAS28" are not interchangeable: comment on the articles by Prevoo et al." *Arthritis and rheumatism* 41.5 (1998), pp. 942–5.

[67] Van Gestel, A. M., Haagsma, C. J., and Riel, P. L. van. "Validation of rheumatoid arthritis improvement criteria that include simplified joint counts". *Arthritis Rheum* 41.10 (1998), pp. 1845–1850.

[68] Aletaha, D. et al. "Remission and active disease in rheumatoid arthritis: Defining criteria for disease activity states". *Arthritis and Rheumatism* 52.9 (2005), pp. 2625–2636.

[69] Anderson, J. et al. "Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice." *Arthritis care & research* 64.5 (2012), pp. 640–7.

[70] Fransen, J. and Van Riel, P. "The Disease Activity Score and the EULAR response criteria." *Clinical and experimental rheumatology* 23.Supp. 39 (2005), S93–S99.

[71] National Institute for Clinical Excellence. *Rheumatoid arthritis*. Tech. rep. 2990.

[72] Kiely, P. et al. "Outcome in rheumatoid arthritis patients with continued conventional therapy for moderate disease activity–the early RA network (ERAN)". *Rheumatology (Oxford, England)* 50.5 (2011), pp. 926–931.

[73] Hyrich, K. L. et al. "Benefit of anti-TNF therapy in rheumatoid arthritis patients with moderate disease activity." *Rheumatology (Oxford, England)* 48.10 (2009), pp. 1323–7.

[74] Nikiphorou, E. et al. "Association between rheumatoid arthritis disease activity, progression of functional limitation and long-term risk of orthopaedic surgery: combined analysis of two prospective cohorts supports EULAR treat to target DAS thresholds". *Annals of the Rheumatic Diseases* (2016), annrheumdis–2015–208669.

[75] Machold, K. P. et al. "Very recent onset rheumatoid arthritis: clinical and serological patient characteristics associated with radiographic progression over the first years of disease". *Rheumatology (Oxford)* 46.2 (2007), pp. 342–349.

[76] Harrison, B and Symmons, D. "Early inflammatory polyarthritis: results from the Norfolk Arthritis Register with a review of the literature. II. Outcome at three years." *Rheumatology (Oxford, England)* 39.9 (2000), pp. 939–949.

[77] Dixey, J, Solymossy, C, and Young, A. "Is it possible to predict radiological damage in early rheumatoid arthritis (RA)? A report on the occurrence, progression, and prognostic factors of radiological erosions over the first 3 years in 866 patients from the Early RA Study (ERAS)". *J Rheumatol Suppl* 69 (2004), pp. 48–54.

[78] Bathon, J. M. et al. "A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis". *N Engl J Med* 343.22 (2000), pp. 1586–1593.

[79] Keystone, E and Haraoui, B. "Adalimumab therapy in rheumatoid arthritis". *Rheum Dis Clin North Am* 30.2 (2004), pp. 349–64, vii.

[80] Klareskog, L. et al. "Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial". *Lancet* 363.9410 (2004), pp. 675–681.

[81] Möttönen, T. T., Hannonen, P. J., and Boers, M. "Combination DMARD therapy including corticosteroids in early rheumatoid arthritis". *Clin Exp Rheumatol* 17.6 SUPPL. 18 (1999), S59–S65.

[82] Van Der Heijde, D et al. "Inhibition of radiographic progression with combination etanercept and methotrexate in patients with moderately active rheumatoid arthritis previously treated with monotherapy." *Annals of the rheumatic diseases* 68.7 (2009), pp. 1113–8.

[83] Van Vollenhoven, R. F. et al. "Conventional combination treatment versus biological treatment in methotrexate-refractory early rheumatoid arthritis: 2 year follow-up of the randomised, non-blinded, parallel-group Swefot trial". *Lancet* 379.9827 (2012), pp. 1712–1720.

[84] Colebatch, A. N. et al. "EULAR recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis". *Ann Rheum Dis* 72.6 (2013), pp. 804–814.

[85] Scott, D. L. et al. "The links between joint damage and disability in rheumatoid arthritis". *Rheumatology* 39.2 (2000), pp. 122–132.

[86] Bombardier, C. et al. "The relationship between joint damage and functional disability in rheumatoid arthritis: a systematic review." *Annals of the rheumatic diseases* 71.6 (2012), pp. 836–44.

[87] Van Der Heijde, D et al. "Level of radiographic damage and radiographic progression are determinants of physical function: a longitudinal analysis of the TEMPO trial". *Ann Rheum Dis* 67.9 (2008), pp. 1267–1270.

[88] Kapetanovic, M. C. et al. "Orthopaedic surgery in patients with rheumatoid arthritis over 20 years: prevalence and predictive factors of large joint replacement." *Annals of the rheumatic diseases* 67.10 (2008), pp. 1412–1416.

[89] Van Der Heijde, D. "Impact of imaging in established rheumatoid arthritis". *Best Practice and Research: Clinical Rheumatology* 17.5 (2003), pp. 783–790.

[90] Van Der Heijde, D. "Erosions versus joint space narrowing in rheumatoid arthritis: what do we know?" *Annals of the rheumatic diseases* 70 Suppl 1.Suppl 1 (2011), pp. i116–8.

[91] Aletaha, D and Smolen, J. S. "Joint damage in rheumatoid arthritis progresses in remission according to the Disease Activity Score in 28 joints and is driven by residual swollen joints". *Arthritis Rheum* 63.12 (2011), pp. 3702–3711.

[92] Smolen, J. S. and Aletaha, D. "Developments in the clinical understanding of rheumatoid arthritis". *Arthritis Research and Therapy* 11.1 (2009).

[93] Redlich, K. et al. "Osteoclasts are essential for TNF-alpha-mediated joint destruction." *The Journal of clinical investigation* 110.10 (2002), pp. 1419–27.

[94] Goldring, S. R. "Pathogenesis of bone and cartilage destruction in rheumatoid arthritis." *Rheumatology (Oxford, England)* 42 Suppl 2.suppl_2 (2003), pp. ii11–6.

[95] Boini, S and Guillemin, F. "Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages." *Annals of the rheumatic diseases* 60.9 (2001), pp. 817–827.

[96] Sharp, J. T. et al. "Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities". *Arthritis & Rheumatism* 14.6 (1971), pp. 706–720.

[97] Sharp, J. T. et al. "How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis?" *Arthritis & Rheumatism* 28.12 (1985), pp. 1326–1335.

[98] Fries, J. F. et al. "Assessment of radiologic progression in rheumatoid arthritis. A randomized, controlled trial." *Arthritis and rheumatism* 29.1 (1986), pp. 1–9.

[99] Van Der Heijde, D et al. "Prognostic factors for radiographic damage and physical disability in early rheumatoid arthritis. A prospective follow-up study of 147 patients". *Br J Rheumatol* 31.8 (1992), pp. 519–525.

[100] Plant, M. J. et al. "Prediction of radiographic damage in hands and feet in rheumatoid arthritis by clinical evaluation". *Clin Rheumatol* 13.3 (1994), pp. 487–491.

[101] Van Der Heijde, D. "How to read radiographs according to the Sharp/van der Heijde method". *The Journal of Rheumatology* 26.3 (1999), pp. 743–745.

[102] Van Der Heijde, D, Boers, M, and Lassere, M. "Methodological issues in radiographic scoring methods in rheumatoid arthritis." *The Journal of rheumatology* 26.3 (1999), pp. 726–30.

[103] Barnabe, C et al. "Comparison of radiographic scoring methods in a cohort of RA patients treated with anti-TNF therapy". *Rheumatology (Oxford)* 51.5 (2012), pp. 878–881.

[104] Klarenbeek, N. B. et al. "A comparison between the simplified erosion and narrowing score and the Sharp-van der Heijde score: post hoc analysis from the Best study". *Ann Rheum Dis* 70.4 (2011), pp. 714–716.

[105] Larsen, A. "A radiological method for grading the severity of rheumatoid arthritis". PhD thesis. Helsinki, 1974.

[106] Larsen, A, Dale, K, and Eek, M. "Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films". *Acta Radiologica: Diagnosis* 18.4 (1977), pp. 481–491.

[107] Larsen, A. "How to apply Larsen score in evaluating radiographs of rheumatoid arthritis in long-term studies". *J Rheumatol* 22.10 (1995), pp. 1974–1975.

[108] Scott, D. L., Houssien, D. A., and Laasonen, L. "Proposed modification to Larsen's scoring methods for hand and wrist radiographs." *British journal of rheumatology* 34.1 (1995), p. 56.

[109] Rau, R and Herborn, G. "A modified version of Larsen's scoring method to assess radiologic changes in rheumatoid arthritis". *J Rheumatol* 22.10 (1995), pp. 1976–1982.

[110] Genant, H. K. "Methods of assessing radiographic change in rheumatoid arthritis". *Am J Med* 75.6a (1983), pp. 35–47.

[111] Genant, H. K. et al. "Assessment of rheumatoid arthritis using a modified scoring method on digitized and original radiographs." *Arthritis and rheumatism* 41.9 (1998), pp. 1583–90.

[112] Nance, E. P. et al. "Observer variation in quantitative assessment of rheumatoid arthritis. Part I. Scoring erosions and joint space narrowing." *Investigative radiology* 21.12 (1986), pp. 922–7.

[113] Rau, R et al. "A new method of scoring radiographic change in rheumatoid arthritis". *J Rheumatol* 25.11 (1998), pp. 2094–2107.

[114] Wolfe, F, Van Der Heijde, D, and Larsen, A. "Assessing radiographic status of rheumatoid arthritis: introduction of a short erosion scale". *J Rheumatol* 27.9 (2000), pp. 2090–2099.

[115] Hetland, M. L. et al. "MRI bone oedema is the strongest predictor of subsequent radiographic progression in early rheumatoid arthritis. Results from a 2-year randomised controlled trial (CIMESTRA)". *Ann Rheum Dis* 68.3 (2009), pp. 384–390.

[116] Dohn, U. M. et al. "Rheumatoid arthritis bone erosion volumes on CT and MRI: reliability and correlations with erosion scores on CT, MRI and radiography". *Ann Rheum Dis* 66.10 (2007), pp. 1388–1392.

[117] Boyesen, P et al. "MRI in early rheumatoid arthritis: synovitis and bone marrow oedema are independent predictors of subsequent radiographic progression". *Ann Rheum Dis* 70.3 (2011), pp. 428–433.

[118] Brower, A. C. "Use of the radiograph to measure the course of rheumatoid arthritis". *Arthritis & Rheumatism* 33.3 (1990), pp. 316–324.

[119] Van Der Heijde, D. "Radiographic imaging: The 'gold standard' for assessment of disease progression in rheumatoid arthritis". *Rheumatology* 39.SUPPL. 1 (2000), pp. 9–16.

[120] Mewa, A. A. et al. "Observer differences in detecting erosions in radiographs of rheumatoid arthritis. A comparison of posteroanterior, Nørgaard and Brewerton views." *The Journal of rheumatology* 10.2 (1983), pp. 216–21.

[121] Van Der Heijde, D et al. "Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials." *Rheumatology (Oxford, England)* 38.12 (1999), pp. 1213–1220.

[122] Van Der Heijde, D et al. "Minimal clinically important difference in plain films in RA: group discussions, conclusions, and recommendations. OMERACT Imaging Task Force". *J Rheumatol* 28.4 (2001), pp. 914–917.

[123] Bruynesteyn, K. et al. "Progression of rheumatoid arthritis on plain radiographs judged differently by expert radiologists and rheumatologists." *The Journal of rheumatology* 31.6 (2004), pp. 1088–94.

[124] Lassere, M. et al. "Smallest detectable difference in radiological progression". *J Rheumatol* 26.3 (1999), pp. 731–739.

[125] Navarro-Compán, V et al. "Measurement error in the assessment of radiographic progression in rheumatoid arthritis (RA) clinical trials: the smallest detectable change (SDC) revisited." *Annals of the rheumatic diseases* 73.6 (2014), pp. 1067–70.

[126] Park, G. S. et al. "Examining radiographic outcomes over time". *Rheumatology International* 34.2 (2014), pp. 271–279.

[127] Van Der Heijde, D. et al. "Expert agreement confirms that negative changes in hand and foot radiographs are a surrogate for repair in patients with rheumatoid arthritis". *Arthritis Research and Therapy* 9.4 (2007), R62.

[128] Rau, R et al. "Identification of radiologic healing phenomena in patients with rheumatoid arthritis". *J Rheumatol* 28.12 (2001), pp. 2608–2615.

[129] Sokka, T and Hannonen, P. "Healing of erosions in rheumatoid arthritis." *Annals of the rheumatic diseases* 59.8 (2000), pp. 647–9.

[130] Landewé, R and Van Der Heijde, D. "Presentation and analysis of radiographic data in clinical trials and observational studies". *Ann Rheum Dis* 64 Suppl 4 (2005), pp. iv48–51.

[131] Tugwell, P. et al. "OMERACT: an international initiative to improve outcome measurement in rheumatology". *Trials* 8 (2007), p. 38.

[132] Swinkels, H. L. et al. "Modified sharp method: factors influencing reproducibility and variability". *Semin Arthritis Rheum* 31.3 (2001), pp. 176–190.

[133] Sackett, D. L. and Wennberg, J. E. "Choosing the best research design for each question." *BMJ (Clinical research ed.)* 315.7123 (1997), p. 1636.

[134] Campbell, D. T., Stanley, J. C., and Gage, N. L. *Experimental and quasi-experimental designs for research.* Boston, MA, US: Houghton, Mifflin and Company, 1963.

[135] Strand, V, Landewé, R, and Van Der Heijde, D. "Using estimated yearly progression rates to compare radiographic data across recent randomised controlled trials in rheumatoid arthritis". *Ann Rheum Dis* 61 Suppl 2 (2002), pp. ii64–6.

[136] Sigler, J. W. et al. "Gold salts in the treatment of rheumatoid arthritis. A double-blind study". *Annals of internal medicine* 80.1 (1974), pp. 21–26.

[137] Van Der Heijde, D et al. "Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis". *Lancet* 1.8646 (1989), pp. 1036–1038.

[138] Weinblatt, M. E. et al. "The effects of drug therapy on radiographic progression of rheumatoid arthritis. Results of a 36-week randomized trial comparing methotrexate and auranofin". *Arthritis Rheum* 36.5 (1993), pp. 613–619.

[139] Jeurissen, M. E. et al. "Influence of methotrexate and azathioprine on radiologic progression in rheumatoid arthritis. A randomized, double-blind study". *Annals of internal medicine* 114.12 (1991), pp. 999–1004.

[140] Boers, M et al. "Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis". *Lancet* 350.9074 (1997), pp. 309–318.

[141] Landewé, R. B. et al. "COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention". *Arthritis Rheum* 46.2 (2002), pp. 347–356.

[142] Dougados, M et al. "Combination therapy in early rheumatoid arthritis: a randomised, controlled, double blind 52 week clinical trial of sulphasalazine and methotrexate compared with the single components". *Ann Rheum Dis* 58.4 (1999), pp. 220–225.

[143] Maillefert, J. F. et al. "Long term structural effects of combination therapy in patients with early rheumatoid arthritis: five year follow up of a prospective double blind controlled study". *Ann Rheum Dis* 62.8 (2003), pp. 764–766.

[144] Rantalaiho, V et al. "Early combination disease-modifying antirheumatic drug therapy and tight disease control improve long-term radiologic outcome in patients with early rheumatoid arthritis: the 11-year results of the Finnish Rheumatoid Arthritis Combination Therapy trial". *Arthritis Res Ther* 12.3 (2010), R122.

[145] Kirwan, J. R. "The effect of glucocorticoids on joint destruction in rheumatoid arthritis. The Arthritis and Rheumatism Council Low-Dose Glucocorticoid Study Group". *N Engl J Med* 333.3 (1995), pp. 142–146.

[146] Svensson, B. et al. "Low-dose prednisolone in addition to the initial disease-modifying antirheumatic drug in patients with early active rheumatoid arthritis reduces joint destruction and increases the remission rate: a two-year randomized trial". *Arthritis Rheum* 52.11 (2005), pp. 3360–3370.

[147] Kirwan, J. R. et al. "Effects of glucocorticoids on radiological progression in rheumatoid arthritis". *Cochrane Database Syst Rev* 1 (2007), p. Cd006356.

[148] Verstappen, S. M. M. et al. "Intensive treatment with methotrexate in early rheumatoid arthritis: aiming for remission. Computer Assisted Management in Early Rheumatoid Arthritis (CAMERA, an open-label strategy trial)." *Annals of the rheumatic diseases* 66.11 (2007), pp. 1443–9.

[149] Emery, P et al. "Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial". *Lancet* 372.9636 (2008), pp. 375–382.

[150] Emery, P et al. "Combination etanercept and methotrexate provides better disease control in very early (<=4 months) versus early rheumatoid arthritis (>4 months and <2 years): post hoc analyses from the COMET study". *Ann Rheum Dis* 71.6 (2012), pp. 989–992.

[151] Breedveld, F. C. et al. "The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previo". *Arthritis Rheum* 54.1 (2006), pp. 26–37.

[152] St Clair, E. W. et al. "Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial". *Arthritis Rheum* 50.11 (2004), pp. 3432–3443.

[153] Smolen, J. S. et al. "Predictors of joint damage in patients with early rheumatoid arthritis treated with high-dose methotrexate with or without concomitant infliximab: results from the ASPIRE trial". *Arthritis Rheum* 54.3 (2006), pp. 702–710.

[154] Tak, P. P. et al. "Inhibition of joint damage and improved clinical outcomes with rituximab plus methotrexate in early active rheumatoid arthritis: the IMAGE trial". *Ann Rheum Dis* 70.1 (2011), pp. 39–46.

[155] Nishimoto, N et al. "Study of active controlled monotherapy used for rheumatoid arthritis, an IL-6 inhibitor (SAMURAI): evidence of clinical and radiographic benefit from an x ray reader-blinded randomised controlled trial of tocilizumab". *Ann Rheum Dis* 66.9 (2007), pp. 1162–1167.

[156] Van Vollenhoven, R. F. et al. "Addition of infliximab compared with addition of sulfasalazine and hydroxychloroquine to methotrexate in patients with early rheumatoid arthritis (Swefot trial): 1-year results of a randomised trial". *The Lancet* 374.9688 (2009), pp. 459–466.

[157] Van Den Broek, M et al. "Rapid radiological progression in the first year of early rheumatoid arthritis is predictive of disability and joint damage progression during 8 years of follow-up". *Ann Rheum Dis* 71.9 (2012), pp. 1530–1533.

[158] Kellgren, J and Lawrence, J. "Rheumatoid arthritis in a population sample." *Annals of the rheumatic diseases* 15.1 (1956), pp. 1–11.

[159] Ropes, M. W. et al. "1958 Revision of diagnostic criteria for rheumatoid arthritis." *Bulletin on the rheumatic diseases* 9.4 (1958), pp. 175–6.

[160] Brook, A and Corbett, M. "Radiographic changes in early rheumatoid disease". *Ann Rheum Dis* 36.1 (1977), pp. 71–73.

[161] Young, A. "Short-term outcomes in recent-onset rheumatoid arthritis." *British journal of rheumatology* 34 Suppl 2 (1995), pp. 79–86.

[162] Young, A. et al. "How does functional disability in early rheumatoid arthritis (RA) affect patients and their lives? Results of 5 yr follow up in 781 patients from the Early RA Study (ERAS)". *Rheumatology (Oxford)* 39 (2000), pp. 603–11.

[163] Kiely, P. D. et al. "Contemporary treatment principles for early rheumatoid arthritis: a consensus statement". *Rheumatology* 48.7 (2009), p. 765.

[164] Symmons, D. P. and Silman, A. J. "The Norfolk Arthritis Register (NOAR)". *Clinical and experimental rheumatology* 21.5 Suppl 31 (2003), S94–99.

[165] Wiles, N. J. et al. "One year followup variables predict disability 5 years after presentation with inflammatory polyarthritis with greater accuracy than at baseline." *The Journal of rheumatology* 27.10 (2000), pp. 2360–6.

[166] Cooper, N. J. et al. "Total costs and predictors of costs in individuals with early inflammatory polyarthritis: a community-based prospective study." *Rheumatology (Oxford, England)* 41.7 (2002), pp. 767–74.

[167] Goodson, N. J. et al. "Mortality in early inflammatory polyarthritis: cardiovascular mortality is increased in seropositive patients." *Arthritis and rheumatism* 46.8 (2002), pp. 2010–9.

[168] Silman, A. et al. "British Society for Rheumatology Biologics Register". *Annals of the Rheumatic Diseases* 62.suppl 2 (2003), pp. ii28–ii29.

[169] Dixon, W. G. et al. "Serious infection following anti-tumor necrosis factor alpha therapy in patients with rheumatoid arthritis: lessons from interpreting data from observational studies." *Arthritis and rheumatism* 56.9 (2007), pp. 2896–904.

[170] Van Aken, J et al. "The Leiden Early Arthritis Clinic". *Clinical and experimental rheumatology* 21.5 Suppl 31 (2003), S100–105.

[171] Welsing, P. and Van Riel, P. "The Nijmegen inception cohort of early rheumatoid arthritis." *The Journal of rheumatology. Supplement* 69 (2004), pp. 14–21.

[172] Welsing, P., Fransen, J., and Van Riel, P. "Is the disease course of rheumatoid arthritis becoming milder? Time trends since 1985 in an inception cohort of early rheumatoid arthritis". *Arthritis and Rheumatism* 52.9 (2005), pp. 2616–2624.

[173] Hetland, M. L., Jensen, D. V., and Krogh, N. S. "Monitoring patients with rheumatoid arthritis in routine care: experiences from a treat-to-target strategy using the DANBIO registry." *Clinical and experimental rheumatology* 32.5 Suppl 85 (2014), S–141–6.

[174] Lindqvist, E et al. "Ten year outcome in a cohort of patients with early rheumatoid arthritis: health status, disease process, and damage." *Annals of the rheumatic diseases* 61.12 (2002), pp. 1055–1059.

[175] Sokka, T. "Early rheumatoid arthritis in Finland." *Clinical and experimental rheumatology* 21.5 Suppl 31 (2005), S133–7.

[176] Combe, B et al. "Comparison of the long-term outcome for patients with rheumatoid arthritis with persistent moderate disease activity or disease remission during the first year after diagnosis: data from the ESPOIR cohort." *Annals of the rheumatic diseases* 74.4 (2015), pp. 724–9.

[177] Navarro-Compán, V. et al. "Relationship between disease activity indices and their individual components and radiographic progression in RA: a systematic literature review." *Rheumatology (Oxford, England)* (2014), pp. 1–14.

[178] Combe, B. "Progression in early rheumatoid arthritis". *Best Practice and Research: Clinical Rheumatology* 23.1 (2009), pp. 59–69.

[179] Young, A et al. "Can we predict aggressive disease?" *Baillieres Clin Rheumatol* 11.1 (1997), pp. 27–48.

[180] Downs, S. H. and Black, N. "The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions". *Journal of Epidemiology & Community Health* 52.6 (1998), pp. 377–384.

[181] Hozo, S. P., Djulbegovic, B., and Hozo, I. "Estimating the mean and variance from the median, range, and the size of a sample". *BMC Medical Research Methodology* 5.1 (2005), p. 13.

[182] Finckh, A et al. "Performance of an automated computer-based scoring method to assess joint space narrowing in rheumatoid arthritis: a longitudinal study". *Arthritis Rheum* 54.5 (2006), pp. 1444–1450.

[183] Lassere, M. "Pooled metaanalysis of radiographic progression: Comparison of Sharp and Larsen methods". *Journal of Rheumatology* 27.1 (2000), pp. 269–276.

[184] Bridges, S. L. et al. "Radiographic severity of rheumatoid arthritis in African Americans: results from a multicenter observational study". *Arthritis Care Res (Hoboken)* 62.5 (2010), pp. 624–631.

[185] Courvoisier, N. et al. "Prognostic factors of 10-year radiographic outcome in early rheumatoid arthritis: a prospective study". *Arthritis Res Ther* 10.5 (2008), R106.

[186] James, D et al. "Orthopaedic intervention in early rheumatoid arthritis. Occurrence and predictive factors in an inception cohort of 1064 patients followed for 5 years". *Rheumatology (Oxford)* 43.3 (2004), pp. 369–376.

[187] Kaarela, K and Kautiainen, H. "Continuous progression of radiological destruction in seropositive rheumatoid arthritis". *J Rheumatol* 24.7 (1997), pp. 1285–1287.

[188] Kapetanovic, M. C. et al. "Early changes in bone mineral density measured by digital X-ray radiogrammetry predict up to 20 years radiological outcome in rheumatoid arthritis". *Arthritis Res Ther* 13.1 (2011), R31.

[189] Knevel, R et al. "Evaluating joint destruction in rheumatoid arthritis: is it necessary to radiograph both hands and feet?" *Ann Rheum Dis* 72.3 (2013), pp. 345–349.

[190] Kuper, H. H. et al. "Radiographic damage in large joints in early rheumatoid arthritis: relationship with radiographic damage in hands and feet, disease activity, and physical disability". *Br J Rheumatol* 36.8 (1997), pp. 855–860.

[191] Viatte, S et al. "Investigation of rheumatoid arthritis genetic susceptibility markers in the early rheumatoid arthritis study further replicates the TRAF1 association with radiological damage". *J Rheumatol* 40.2 (2013), pp. 144–156.

[192] Tanaka, N et al. "Synovial membrane enhancement and bone erosion by magnetic resonance imaging for prediction of radiologic progression in patients with early rheumatoid arthritis". *Rheumatol Int* 25.2 (2005), pp. 103–107.

[193] Van Aken, J et al. "Radiological outcome after four years of early versus delayed treatment strategy in patients with recent onset rheumatoid arthritis". *Annals of the Rheumatic Diseases* 63.3 (2004), pp. 274–279.

[194] Bukhari, M et al. "Rheumatoid factor is the major predictor of increasing severity of radiographic erosions in rheumatoid arthritis: results from the Norfolk Arthritis Register Study, a large inception cohort". *Arthritis Rheum* 46.4 (2002), pp. 906–912.

[195] Combe, B et al. "Prognostic factors for radiographic damage in early rheumatoid arthritis: a multiparameter prospective study". *Arthritis Rheum* 44.8 (2001), pp. 1736–1743.

[196] Constantin, A et al. "Stromelysin 1 (matrix metalloproteinase 3) and HLA-DRB1 gene polymorphisms: Association with severity and progression of rheumatoid arthritis in a prospective study". *Arthritis Rheum* 46.7 (2002), pp. 1754–1762.

[197] Constantin, A et al. "Collagenase-1 (MMP-1) and HLA-DRB1 gene polymorphisms in rheumatoid arthritis: a prospective longitudinal study". *J Rheumatol* 29.1 (2002), pp. 15–20.

[198] De Vries, N et al. "A T cell receptor beta chain variable region polymorphism associated with radiographic progression in rheumatoid arthritis". *Ann Rheum Dis* 52.5 (1993), pp. 327–331.

[199] Van Gaalen, F. A. et al. "Association between HLA class II genes and autoantibodies to cyclic citrullinated peptides (CCPs) influences the severity of rheumatoid arthritis". *Arthritis Rheum* 50.7 (2004), pp. 2113–2121.

[200] Gourraud, P. A. et al. "A new classification of HLA-DRB1 alleles differentiates predisposing and protective alleles for rheumatoid arthritis structural severity". *Arthritis Rheum* 54.2 (2006), pp. 593–599.

[201] Houseman, M et al. "Baseline serum MMP-3 levels in patients with Rheumatoid Arthritis are still independently predictive of radiographic progression in a longitudinal observational cohort at 8 years follow up". *Arthritis Res Ther* 14.1 (2012), R30.

[202] Kroot, E. J. et al. "The prognostic value of anti-cyclic citrullinated peptide antibody in patients with recent-onset rheumatoid arthritis." *Arthritis and rheumatism* 43.8 (2000), pp. 1831–5.

[203] Kuiper, S et al. "Influence of sex, age, and menopausal state on the course of early rheumatoid arthritis". *J Rheumatol* 28.8 (2001), pp. 1809–1816.

[204] McQueen, F. M. et al. "Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis". *Arthritis Rheum* 48.7 (2003), pp. 1814–1827.

[205] Meyer, O et al. "Anticitrullinated protein/peptide antibody assays in early rheumatoid arthritis for predicting five year radiographic damage". *Ann Rheum Dis* 62.2 (2003), pp. 120–126.

[206] Meyer, O et al. "Serial determination of cyclic citrullinated peptide autoantibodies predicted five-year radiological outcomes in a prospective cohort of patients with early rheumatoid arthritis". *Arthritis Res Ther* 8.2 (2006), R40.

[207] Nyhall-Wahlin, B. M. et al. "The presence of rheumatoid nodules at early rheumatoid arthritis diagnosis is a sign of extra-articular disease and predicts radiographic progression of joint destruction over 5 years". *Scand J Rheumatol* 40.2 (2011), pp. 81–87.

[208] Park, G. S. et al. "Patterns of radiographic outcomes in early, seropositive rheumatoid arthritis: a baseline analysis". *Contemp Clin Trials* 32.2 (2011), pp. 160–168.

[209] Posthumus, M. D. et al. "Serum matrix metalloproteinase 3 in early rheumatoid arthritis is correlated with disease activity and radiological progression". *J Rheumatol* 27.12 (2000), pp. 2761–2768.

[210] De Rooy, D. P. C. et al. "Predicting arthritis outcomes-what can be learned from the Leiden Early Arthritis Clinic?" *Rheumatology* 50.1 (2011), pp. 93–100.

[211] Salaffi, F. et al. "Relationship between time-integrated disease activity estimated by DAS28-CRP and radiographic progression of anatomical damage in patients with early rheumatoid arthritis". *BMC Musculoskelet Disord* 12 (2011), p. 120.

[212] Van Der Helm-Van Mil, A. H. et al. "A high body mass index has a protective effect on the amount of joint destruction in small joints in early rheumatoid arthritis". *Ann Rheum Dis* 67.6 (2008), pp. 769–774.

[213] Van Der Helm-Van Mil, A. H. M. et al. "An independent role of protective HLA class II alleles in rheumatoid arthritis severity and susceptibility". *Arthritis Rheum* 52.9 (2005), pp. 2637–2644.

[214] Welsing, P. M. J. et al. "The relationship between disease activity and radiologic progression in patients with rheumatoid arthritis: a longitudinal analysis". *Arthritis Rheum* 50.7 (2004), pp. 2082–2093.

[215] Wolfe, F and Sharp, J. T. "Radiographic outcome of recent-onset rheumatoid arthritis: a 19-year study of radiographic progression". *Arthritis Rheum* 41.9 (1998), pp. 1571–1582.

[216] Fex, E et al. "Development of radiographic damage during the first 5-6 yr of rheumatoid arthritis. A prospective follow-up study of a Swedish cohort". *Br J Rheumatol* 35.11 (1996), pp. 1106–1115.

[217] Fex, E, Eberhardt, K, and Saxne, T. "Tissue-derived macromolecules and markers of inflammation in serum in early rheumatoid arthritis: relationship to development of joint destruction in hands and feet". *Br J Rheumatol* 36.11 (1997), pp. 1161–1165.

[218] Kaltenhauser, S et al. "Antibodies against cyclic citrullinated peptide are associated with the DRB1 shared epitope and predict joint erosion in rheumatoid arthritis". *Rheumatology (Oxford)* 46.1 (2007), pp. 100–104.

[219] Kaltenhauser, S et al. "Immunogenetic markers and seropositivity predict radiological progression in early rheumatoid arthritis independent of disease activity". *J Rheumatol* 28.4 (2001), pp. 735–744.

[220] Kraan, M. C. et al. "T cells, fibroblast-like synoviocytes, and granzyme B+ cytotoxic cells are associated with joint damage in patients with recent onset rheumatoid arthritis". *Ann Rheum Dis* 63.5 (2004), pp. 483–488.

[221] Lindqvist, E et al. "Prognostic laboratory markers of joint damage in rheumatoid arthritis". *Ann Rheum Dis* 64.2 (2005), pp. 196–201.

[222] Mustila, A et al. "Antineutrophil cytoplasmic antibodies in patients with early rheumatoid arthritis: an early marker of progressive erosive disease". *Arthritis Rheum* 43.6 (2000), pp. 1371–1377.

[223] Roux-Lombard, P et al. "Cytokines, metalloproteinases, their inhibitors and cartilage oligomeric matrix protein: relationship to radiological progression and inflammation in early rheumatoid arthritis. A prospective 5-year study". *Rheumatology (Oxford)* 40.5 (2001), pp. 544–551.

[224] Sokka, T et al. "Radiographic progression is getting milder in patients with early rheumatoid arthritis. Results of 3 cohorts over 5 years". *J Rheumatol* 31.6 (2004), pp. 1073–1082.

[225] Wagner, U et al. "Prospective analysis of the impact of HLA-DR and -DQ on joint destruction in recent-onset rheumatoid arthritis". *Rheumatology (Oxford)* 42.4 (2003), pp. 553–562.

[226] Finckh, A., Choi, H. K., and Wolfe, F. "Progression of radiographic joint damage in different eras: trends towards milder disease in rheumatoid arthritis are attributable to improved treatment." *Annals of the rheumatic diseases* 65.9 (2006), pp. 1192–1197.

[227] Finckh, A, Dehler, S, and Gabay, C. "The effectiveness of leflunomide as a co-therapy of tumour necrosis factor inhibitors in rheumatoid arthritis: a population-based study". *Ann Rheum Dis* 68.1 (2009), pp. 33–39.

[228] Rahman, M. U. et al. "Changes in patient characteristics in anti-tumour necrosis factor clinical trials for rheumatoid arthritis: results of an analysis of the literature over the past 16 years". *Ann Rheum Dis* 70.9 (2011), pp. 1631–1640.

[229] Pincus, T, Sokka, T, and Kautiainen, H. "Patients seen for standard rheumatoid arthritis care have significantly better articular, radiographic, laboratory, and functional status in 2000 than in 1985". *Arthritis Rheum* 52.4 (2005), pp. 1009–1019.

[230] Pincus, T and Sokka, T. "Complexities in the quantitative assessment of patients with rheumatic diseases in clinical trials and clinical care". *Clin Exp Rheumatol* 23.5 SUPPL. 39 (2005), S1–S9.

[231] Viatte S et al. "Association of HLA-DRB1 haplotypes with rheumatoid arthritis severity, mortality, and treatment response." *JAMA* 313.16 (2015), pp. 1645–1656.

[232] Van Der Heijde, D et al. "Influence of prognostic features on the final outcome in rheumatoid arthritis: a review of the literature". *Seminars in Arthritis and Rheumatism* 17.4 (1988), pp. 284–292.

[233] Kirwan, J. R. and Reeback, J. S. "Stanford Health Assessment Questionnaire modified to assess disability in British patients with rheumatoid arthritis." *British journal of rheumatology* 25.2 (1986), pp. 206–9.

[234] Zigmond, A. S. and Snaith, R. P. "The hospital anxiety and depression scale." *Acta psychiatrica Scandinavica* 67.6 (1983), pp. 361–70.

[235] Ware, J. E. and Sherbourne, C. D. "The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection." *Medical care* 30.6 (1992), pp. 473–83.

[236] Pincus, T et al. "Methotrexate as the "anchor drug" for the treatment of early rheumatoid arthritis". *Clinical and experimental rheumatology* 21.5 Suppl 31 (2003), S179–185.

[237] Van Der Heijde, D. et al. "How to report radiographic data in randomized clinical trials in rheumatoid arthritis: guidelines from a roundtable discussion". *Arthritis Rheum* 47.2 (2002), pp. 215–218.

[238] Martin Bland, J. and Altman, D. "Statistical methods for assessing agreement between two methods of clinical measurement". *The Lancet* 327.8476 (1986), pp. 307–310.

[239] Giavarina, D. "Understanding Bland Altman analysis." *Biochemia medica* 25.2 (2015), pp. 141–51.

[240] Müller, R. and Büttner, P. "A critical discussion of intraclass correlation coefficients". *Statistics in Medicine* 13.23-24 (1994), pp. 2465–2476.

[241] Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Academic Press, 1969.

[242] Descalzo, M. A. Á. et al. "Tackling missing radiographic progression data: multiple imputation technique compared with inverse probability weights and complete case analysis". *Rheumatology (Oxford)* 52.2 (2013), pp. 331–336.

[243] Van Der Heijde, D et al. "Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis". *Rheumatology (Oxford)* 38.10 (1999), pp. 941–947.

[244] Landewe, R and Van Der Heijde, D. "Is radiographic progression a realistic outcome measure in clinical trials with early inflammatory arthritis?" *Clin Exp Rheumatol* 21.5 Suppl 31 (2003), S37–41.

[245] Ødegård, S. et al. "Association of early radiographic damage with impaired physical function in rheumatoid arthritis: a ten-year, longitudinal observational study in 238 patients". *Arthritis Rheum* 54.1 (2006), pp. 68–75.

[246] Brealey, S. "Measuring patient health status in rheumatoid arthritis – what is a minimal clinically important difference?" *The Journal of rheumatology* 31.6 (2004), pp. 1026–1028.

[247] Bruynesteyn, K et al. "Detecting radiological changes in rheumatoid arthritis that are considered important by clinical experts: influence of reading with or without known sequence". *J Rheumatol* 29.11 (2002), pp. 2306–2312.

[248] Lassere, M. N. et al. "Robustness and generalizability of smallest detectable difference in radiological progression". *J Rheumatol* 28.4 (2001), pp. 911–913.

[249] Landewé, R. and Van Der Heijde, D. "Radiographic progression depicted by probability plots: presenting data with optimal use of individual values." *Arthritis and rheumatism* 50.3 (2004), pp. 699–706.

[250] Rau, R. "Healing of radiologic damage in rheumatoid arthritis". *Current Rheumatology Reviews* 3.1 (2007), pp. 57–65.

[251] Van Der Heijde, D. "Radiographic progression in rheumatoid arthritis: does it reflect outcome? Does it reflect treatment?" *Ann Rheum Dis* 60 Suppl 3 (2001), pp. iii47–50.

[252] Lassere, M et al. "Reference curves of radiographic damage in patients with rheumatoid arthritis: application of quantile regression and fractional polynomials". *J Rheumatol* 24.7 (1997), pp. 1288–1294.

[253] Salaffi, F et al. "Progression of erosion and joint space narrowing scores in rheumatoid arthritis assessed by nonlinear models". *J Rheumatol* 21.9 (1994), pp. 1626–1630.

[254] Salaffi, F et al. "Methotrexate-induced pneumonitis in patients with rheumatoid arthritis and psoriatic arthritis: report of five cases and review of the literature". *Clin Rheumatol* 16.3 (1997), pp. 296–304.

[255] Duncan, C., Jones, K., and Moon, G. "Context, composition and heterogeneity: Using multilevel models in health research". *Social Science & Medicine* 46.1 (1998), pp. 97–117.

[256] Wang, M. and Wang, M. "Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments". *Advances in Statistics* 2014 (2014), pp. 1–11.

[257] Rubin, D. B. "Inference and missing data". *Biometrika* 63.3 (1976), pp. 581–592.

[258] Baraldi, A. N. and Enders, C. K. "An introduction to modern missing data analyses." *Journal of school psychology* 48.1 (2010), pp. 5–37.

[259] Van Buuren, S. "Multiple imputation of multilevel data". *Handbook of advanced multilevel analysis*. 2011, pp. 173–196.

[260] Walter, S. and Tiemeier, H. "Variable selection: current practice in epidemiological studies." *European journal of epidemiology* 24.12 (2009), pp. 733–6.

[261] Greenland, S. "Methods for epidemiologic analyses of multiple exposures: A review and comparative study of maximum-likelihood, preliminary-testing, and empirical-bayes regression". *Statistics in Medicine* 12.8 (1993), pp. 717–736.

[262] Steyerberg, E. W., Eijkemans, M. J. C., and Habbema, J. D. F. "Stepwise Selection in Small Data Sets: A Simulation Study of Bias in Logistic Regression Analysis". *Journal of Clinical Epidemiology* 52.10 (1999), pp. 935–942.

[263] Amaya-Amaya, J. et al. "Does non-erosive rheumatoid arthritis exist? A cross-sectional analysis and a systematic literature review". *Seminars in Arthritis and Rheumatism* 44.5 (2015), pp. 489–498.

[264] Harrell, F. E. *Regression Modeling Strategies*. Springer Series in Statistics. New York, NY: Springer New York, 2001.

[265] Doran, M. F. et al. "Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a forty-year period". *Arthritis and Rheumatism* 46.3 (2002), pp. 625–631.

[266] Odegård, S., Kvien, T. K., and Uhlig, T. "Incidence of clinically important 10-year health status and disease activity levels in population-based cohorts with rheumatoid arthritis." *The Journal of Rheumatology* 35.1 (2008), pp. 54–60.

[267] Silman, A. J. "Are there secular trends in the occurrence and severity of rheumatoid arthritis?" *Scandinavian Journal of Rheumatology. Supplement* 79 (1989), pp. 25–30.

[268] Hochberg, M. C. "Changes in the incidence and prevalence of rheumatoid arthritis in England and Wales, 1970-1982." *Seminars in arthritis and rheumatism* 19.5 (1990), pp. 294–302.

[269] Spector, T. D., Hart, D. J., and Powell, R. J. "Prevalence of rheumatoid arthritis and rheumatoid factor in women: evidence for a secular decline." *Annals of the Rheumatic Diseases* 52.4 (1993), pp. 254–257.

[270] Dugowson, C. E. et al. "Rheumatoid arthritis in women. Incidence rates in group health cooperative, Seattle, Washington, 1987-1989." *Arthritis and rheumatism* 34.12 (1991), pp. 1502–7.

[271] Bergström, U. et al. "Lower disease activity and disability in Swedish patients with rheumatoid arthritis in 1995 compared with 1978". *Scandinavian Journal of Rheumatology* 28.3 (1999), pp. 160–165.

[272] Aho, K. et al. "Is seropositive rheumatoid arthritis becoming less severe?" *Clinical and Experimental Rheumatology* 7.3 (1989), pp. 287–290.

[273] Porter, D. R. et al. "Is Rheumatoid Arthritis Becoming a Milder Disease? Or Are We Starting Second-Line Therapy in Patients with Milder Disease?" *Rheumatology* 35.12 (1996), pp. 1305–1308.

[274] Nikiphorou, E. et al. "Hand and foot surgery rates in rheumatoid arthritis have declined from 1986 to 2011, but large-joint replacement rates remain unchanged: Results from two UK inception cohorts". *Arthritis and Rheumatology* 66.5 (2014), pp. 1081–1089.

[275] Fiehn, C et al. "Improved radiological outcome of rheumatoid arthritis: The importance of early treatment with methotrexate in the era of biological drugs". *Clin Rheumatol* 32.12 (2013), pp. 1735–1742.

[276] Emery, P et al. "A comparison of the efficacy and safety of leflunomide and methotrexate for the treatment of rheumatoid arthritis". *Rheumatology (Oxford, England)* 39.6 (2000), pp. 655–665.

[277] Möttönen, T. et al. "Comparison of combination therapy with single-drug therapy in early rheumatoid arthritis: a randomised trial". *The Lancet* 353.9164 (1999), pp. 1568–1573.

[278] Lillegraven, S et al. "What is the clinical relevance of erosions and joint space narrowing in RA?" *Nature Reviews Rheumatology* 8.2 (2012), pp. 117–120.

[279] Aletaha, D., Funovits, J., and Smolen, J. S. "Physical disability in rheumatoid arthritis is associated with cartilage damage rather than bone destruction". *Annals of the rheumatic diseases* 70.5 (2011), pp. 733–739.

[280] Smolen, J. S. et al. "Progression of radiographic joint damage in rheumatoid arthritis: independence of erosions and joint space narrowing". *Ann Rheum Dis* 68.10 (2009), pp. 1535–1540.

[281] McWilliams, D. et al. "Erosive and osteoarthritic structural progression in early rheumatoid arthritis". *Rheumatology (Oxford)* In Press (2016).

[282] Scott, D. L. "Prognostic factors in early rheumatoid arthritis". *Rheumatology (Oxford)* 39 Suppl 1 (2000), pp. 24–29.

[283] Bruynesteyn, K. et al. "Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference." *Arthritis and rheumatism* 46.4 (2002), pp. 913–920.

[284] Khanna, D et al. "Increased radiographic damage scores at the onset of seropositive rheumatoid arthritis in older patients are associated with osteoarthritis of the hands, but not with more rapid progression of damage". *Arthritis Rheum* 52.8 (2005), pp. 2284–2292.

[285] Aletaha, D., Alasti, F., and Smolen, J. S. "Rheumatoid factor determines structural progression of rheumatoid arthritis dependent and independent of disease activity". *Ann Rheum Dis* 72.6 (2013), pp. 875–880.

[286] Forslind, K et al. "Prediction of radiological outcome in early rheumatoid arthritis in clinical practice: Role of antibodies to citrullinated peptides (anti-CCP)". *Ann Rheum Dis* 63.9 (2004), pp. 1090–1095.

[287] Vallbracht, J et al. "Diagnostic and clinical value of anti-cyclic citrullinated peptide antibodies compared with rheumatoid factor isotypes in rheumatoid arthritis". *Ann Rheum Dis* 63.9 (2004), pp. 1079–1084.

[288] Van Der Woude, D. and Huizinga, T. W. "The battle between anti-cyclic citrullinated peptide and rheumatoid factor tests—a winner at last?" *Nature Clinical Practice Rheumatology* 3.12 (2007), pp. 696–697.

[289] Van Der Woude, D. et al. "The ACPA isotype profile reflects long-term radiographic progression in rheumatoid arthritis". *Ann Rheum Dis* 69.6 (2010), pp. 1110–1116.

[290] Fautrel, B et al. "Identifying patients with rheumatoid arthritis with moderate disease activity at risk of significant radiographic progression despite methotrexate treatment". *RMD Open* 1.1 (2015), e000018.

[291] Smolen, J. S. et al. "Maintenance, reduction, or withdrawal of etanercept after treatment with etanercept and methotrexate in patients with moderate rheumatoid arthritis (PRESERVE): A randomised controlled trial". *The Lancet* 381.9870 (2013), pp. 918–929.

[292] Norton, S et al. "A study of baseline prevalence and cumulative incidence of comorbidity and extra-articular manifestations in ra and their impact on outcome". *Rheumatology (United Kingdom)* 52.1 (2013), pp. 99–110.

[293] Guillemin, F, Briançon, S, and Pourel, J. "Functional disability in rheumatoid arthritis: two different models in early and established disease." *The Journal of rheumatology* 19.3 (1992), pp. 366–9.

[294] Welsing, P. M. et al. "The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis". *Arthritis Rheum* 44.9 (2001), pp. 2009–2017.

[295] Kirwan, J. R. "A theoretical framework for process, outcome and prognosis in rheumatoid arthritis". *J Rheumatol* 19.3 (1992), pp. 333–336.

[296] Dawes, P. T. et al. "Prediction of progressive joint damage in patients with rheumatoid arthritis receiving gold or D-penicillamine therapy". *Ann Rheum Dis* 45.11 (1986), pp. 945–949.

[297] Escalante, A. and Del Rincón, I. "The disablement process in rheumatoid arthritis". *Arthritis Care & Research* 47.3 (2002), pp. 333–342.

[298] Prevoo, M. L. et al. "Validity and reliability of joint indices. A longitudinal study in patients with recent onset rheumatoid arthritis." *British journal of rheumatology* 32.7 (1993), pp. 589–94.

[299] Smolen, J. S. et al. "Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity." *Arthritis and rheumatism* 38.1 (1995), pp. 38–43.

[300] McWilliams, D. F. et al. "Predictors of change in bodily pain in early rheumatoid arthritis: an inception cohort study." *Arthritis care & research* 64.10 (2012), pp. 1505–13.

[301] De Punder, Y. M. R. et al. "Should we redefine treatment targets in rheumatoid arthritis? Low disease activity is sufficiently strict for patients who are anticitrullinated protein antibody-negative". *Journal of Rheumatology* 40.8 (2013), pp. 1268–1274.

[302] Klarenbeek, N. B. et al. "Association with joint damage and physical functioning of nine composite indices and the 2011 ACR/EULAR remission criteria in rheumatoid arthritis". *Ann Rheum Dis* 70.10 (2011), pp. 1815–1821.

[303] Aletaha, D., Alasti, F., and Smolen, J. S. "Rituximab dissociates the tight link between disease activity and joint damage in rheumatoid arthritis patients". *Annals of the Rheumatic Diseases* (2012), pp. 7–12.

[304] Drossaers-Bakker, K. W. et al. "Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time". *Arthritis Rheum* 42.9 (1999), pp. 1854–1860.

[305] Lindqvist, E et al. "Course of radiographic damage over 10 years in a cohort with early rheumatoid arthritis". *Ann Rheum Dis* 62.7 (2003), pp. 611–616.

[306] Contreras-Yáñez, I. et al. "Radiographic outcome in Hispanic early rheumatoid arthritis patients treated with conventional disease modifying anti-rheumatic drugs". *Eur J Radiol* 79.2 (2011), e52–7.

[307] Nishina, N. et al. "Reduction of plasma IL-6 but not TNF-α by methotrexate in patients with early rheumatoid arthritis: a potential biomarker for radiographic progression." *Clinical rheumatology* 32.11 (2013), pp. 1661–6.

[308] Bakker, M. F. et al. "Early clinical response to treatment predicts 5-year outcome in RA patients: follow-up results from the CAMERA study". *Ann Rheum Dis* 70.6 (2011), pp. 1099–1103.

[309] Walsh, D. A. and McWilliams, D. F. "Mechanisms, impact and management of pain in rheumatoid arthritis". *Nature Reviews Rheumatology* 10.10 (2014), pp. 581–592.

[310] Clarke, A. E. et al. "Radiographic damage in rheumatoid arthritis correlates with functional disability but not direct medical costs". *J Rheumatol* 28.11 (2001), pp. 2416–2424.

[311] Rupp, I et al. "Disability and health-related quality of life among patients with rheumatoid arthritis: association with radiographic joint damage, disease activity, pain, and depressive symptoms". *Scand J Rheumatol* 35.3 (2006), pp. 175–181.

[312] Gonzalez-Alvaro, I et al. "Trends towards an improved disease state in rheumatoid arthritis over time: influence of new therapies and changes in management approach: analysis of the EMECAR cohort." *Arthritis research & therapy* 10.6 (2008), R138.

[313] Navarro-Compán, V. et al. "Relationship between types of radiographic damage and disability in patients with rheumatoid arthritis in the EURIDISS cohort: a longitudinal study." *Rheumatology (Oxford, England)* 54.1 (2015), pp. 83–90.

[314] Smolen, J. S. et al. "Association of joint space narrowing with impairment of physical function and work ability in patients with early rheumatoid arthritis: protection beyond disease control by adalimumab plus methotrexate". *Ann Rheum Dis* 72.7 (2013), pp. 1156–1162.

[315] Maillefert, J. F. et al. "The 5-yr HAQ-disability is related to the first year's changes in the narrowing, rather than erosion score in patients with recent-onset rheumatoid arthritis". *Rheumatology (Oxford)* 43.1 (2004), pp. 79–84.

[316] Gherghe, A. M. et al. "Association of radiographic damage with physical function in patients with rheumatoid arthritis-results from the Rapid1 trial". *Annals of the rheumatic diseases* 73 (2014), pp. 1–8.

[317] Landewé, R. and Van Der Heijde, D. "Radiographic progression in rheumatoid arthritis". *Clin Exp Rheumatol* 23.5 Suppl 39 (2005), S63–8.

[318] Weisman, M. H. "Use of radiographs to measure outcome in rheumatoid arthritis". *Am J Med* 83.4b (1987), pp. 96–100.

[319] Oude Voshaar, M. A. H. et al. "Further Simplification of the Simple Erosion Narrowing Score With Item Response Theory Methodology". *Arthritis Care & Research* 68.8 (2016), pp. 1206–1210.

[320] Dias, E. M. et al. "Reliability and sensitivity to change of the Simple Erosion Narrowing Score compared with the Sharp-van der Heijde method for scoring radiographs in rheumatoid arthritis". *Ann Rheum Dis* 67.3 (2008), pp. 375–379.

[321] Guillemin, F et al. "Reproducibility and sensitivity to change of 5 methods for scoring hand radiographic damage in patients with rheumatoid arthritis". *J Rheumatol* 32.5 (2005), pp. 778–786.

[322] Rau, R. and Wassenberg, S. "Reliability of scoring methods to measure radiographic change in patients with rheumatoid arthritis." *The Journal of rheumatology* 32.5 (2005), pp. 766–8.

[323] Ferrara, R. et al. "Clinical trials in rheumatoid arthritis: Methodological suggestions for assessing radiographs arising from the GRISAR* study". *Ann Rheum Dis* 56.10 (1997), pp. 608–612.

[324] Salaffi, F et al. "Quantitation of radiologic progression in rheumatoid arthritis: Controversies and perspectives". *La Radiologia Medica* 93.3 (1997), pp. 174–184.

[325] Schett, G. and Gravallese, E. "Bone erosion in rheumatoid arthritis: mechanisms, diagnosis and treatment." *Nature reviews. Rheumatology* 8.11 (2012), pp. 656–64.

[326] Moller Dohn, U et al. "Erosive progression is minimal, but erosion healing rare, in patients with rheumatoid arthritis treated with adalimumab. A 1 year investigator-initiated follow-up study using high-resolution computed tomography as the primary outcome measure". *Ann Rheum Dis* 68.10 (2009), pp. 1585–1590.

[327] Lukas, C et al. "Repair of erosions occurs almost exclusively in damaged joints without swelling". *Ann Rheum Dis* 69.5 (2010), pp. 851–855.

[328] Weinblatt, M. E. et al. "Long-term prospective trial of low-dose methotrexate in rheumatoid arthritis". *Arthritis & Rheumatism* 31.2 (1988), pp. 167–175.

[329] Landewé, R. B. M. et al. "How to interpret radiological progression in randomized clinical trials?" *Rheumatology* 42.1 (2003), pp. 2–5.

[330] Welsing, P. M., Borm, G. F., and Van Riel, P. "Minimal clinically important difference in radiological progression of joint damage. A definition based on patient perspective". *J Rheumatol* 33.3 (2006), pp. 501–507.

[331] Xie, H. et al. "Comparing statistical methods for analyzing skewed longitudinal count data with many zeros: An example of smoking cessation". *Journal of Substance Abuse Treatment* 45.1 (2013), pp. 99–108.

[332] Hecht, C. et al. "Additive effect of anti-citrullinated protein antibodies and rheumatoid factor on bone erosions in patients with RA." *Annals of the rheumatic diseases* 74.12 (2015), pp. 2151–6.

[333] Van Steenbergen, H. W. et al. "Association of Valine and Leucine at HLA-DRB1 Position 11 With Radiographic Progression in Rheumatoid Arthritis, Independent of the Shared Epitope Alleles but Not Independent of Anti-Citrullinated Protein Antibodies". *Arthritis & Rheumatology* 67.4 (2015), pp. 877–886.

[334] Visser, K et al. "A matrix risk model for the prediction of rapid radiographic progression in patients with rheumatoid arthritis receiving different dynamic treatment strategies: post hoc analyses from the BeSt study". *Ann Rheum Dis* 69.7 (2010), pp. 1333–1337.

[335] Glasziou, P. et al. "Assessing the quality of research." *BMJ (Clinical research ed.)* 328.7430 (2004), pp. 39–41.

[336] Von Elm, E. et al. "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies." *BMJ (Clinical research ed.)* 335.7624 (2007), pp. 806–8.

[337] Lunt, M. et al. "Different Methods of Balancing Covariates Leading to Different Effect Estimates in the Presence of Effect Modification". *American Journal of Epidemiology* 169.7 (2009), pp. 909–917.

[338] Harrell Jr, F. E, Lee, K. L, and Mark, D. B. "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors". *Statistics in medicine* 15.4 (1996), pp. 361–387.

[339] Sterne, J. et al. "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls." *BMJ (Clinical research ed.)* 338 (2009), b2393.

[340] Hox, J. and Robots, K. J. *Handbook of advanced multilevel analysis*. New Yprk: Taylor & Francis Group, 2011.

[341] Symmons, D. et al. "Patients with stable long-standing rheumatoid arthritis continue to deteriorate despite intensified treatment with traditional disease modifying anti-rheumatic drugs - Results of the British Rheumatoid Outcome Study Group randomized controlled clinical tr". *Rheumatology* 45.5 (2006), pp. 558–565.

[342] Saleem, B et al. "Should imaging be a component of rheumatoid arthritis remission criteria? A comparison between traditional and modified composite remission scores and imaging assessments". *Ann Rheum Dis* 70.5 (2011), pp. 792–798.

[343] Jayakumar, K. et al. "Sustained clinical remission in rheumatoid arthritis: prevalence and prognostic factors in an inception cohort of patients treated with conventional DMARDS". *Rheumatology* 51.1 (2012), pp. 169–175.

[344] Van Der Woude, D. et al. "Prevalence of and predictive factors for sustained disease-modifying antirheumatic drug-free remission in rheumatoid arthritis: results from two large early arthritis cohorts". *Arthritis Rheum* 60.8 (2009), pp. 2262–2271.

[345] Van Der Woude, D. et al. "Sustained drug-free remission in rheumatoid arthritis after DAS-driven or non-DAS-driven therapy: a comparison of two cohort studies". *Rheumatology* 51.6 (2012), pp. 1120–1128.

[346] Jayakumar, K. "Clinical Disease Activity and Radiological Damage in Early Rheumatoid Arthritis". PhD thesis. University of Hertfordshire, 2011.

[347] Govoni, M and Caporali, R. "Predicting and managing the progression of structural damage in rheumatoid arthritis: Where do we stand?" *Clin Exp Rheumatol* 30.4 (2012), pp. 459–463.

[348] Van Der Heijde, D. "Remission by imaging in rheumatoid arthritis: should this be the ultimate goal?" *Annals of the rheumatic diseases* 71 Suppl 2.Supp II (2012), pp. i89–92.

[349] Escalante, A. and Del Rincón, I. "How much disability in rheumatoid arthritis is explained by rheumatoid arthritis?" *Arthritis and Rheumatism* 42.8 (1999), pp. 1712–1721.

[350] Edwards, R. R. et al. "Pain, catastrophizing, and depression in the rheumatic diseases." *Nature reviews. Rheumatology* 7.4 (2011), pp. 216–24.

[351] Astin, J. A. et al. "Psychological interventions for rheumatoid arthritis: a meta-analysis of randomized controlled trials." *Arthritis and rheumatism* 47.3 (2002), pp. 291–302.

[352] Norton, S. et al. "Distinct psychological distress trajectories in rheumatoid arthritis: Findings from an inception cohort". *Journal of Psychosomatic Research* 71.5 (2011), pp. 290–295.

[353] Graves, H. et al. "Illness beliefs predict disability in rheumatoid arthritis." *Journal of psychosomatic research* 67.5 (2009), pp. 417–23.

[354] Hutchinson, D et al. "Heavy cigarette smoking is strongly associated with rheumatoid arthritis (RA), particularly in patients without a family history of RA". *Annals of the Rheumatic Diseases* 60.3 (2001), pp. 223–7.

[355] Westhoff, G, Rau, R, and Zink, A. "Rheumatoid arthritis patients who smoke have a higher need for DMARDs and feel worse, but they do not have more joint damage than non-smokers of the same serological group". *Rheumatology (Oxford)* 47.6 (2008), pp. 849–854.

Appendix A

ERAS Case Report From

RCT 0

FIRST VISIT FORM

CARD 1 | HOSP | ID NUMBER | SEX | DOB (YEAR) | ETHNIC

1 = Caucasoid
2 = Asian
3 = Negroid
4 = Mongoloid
5 = Semitic
6 = Other

INITIAL DATA FORM

STICKY LABEL or

PATIENT's NAME

ONSET OF DISEASE MONTH | YEAR

HAND (R = 1, L = 2, Both = 3)

HEIGHT (CM)

SPEED
1 = Acute
2 = Insidious
3 = Epidsodic
4 = Other

PATTERN OF JOINT SYMPTOMS

	Small	Large	Both
Asymmetrical	1	2	3
Symmetrical	4	5	6
Neck	7		
Palindromic	8		
Soft Tissue	9		
Monoarticular	0		

PRODROME

Flu like	= 1
Weight loss	= 2
Malaise/Fatigue	= 3
Depression	= 4
Soft tissue	= 5
Rash	= 6
Myalgia	= 7
Other	= 0

TRIGGER

Illness	= 1
Infection	= 2
Trauma	= 3
Surgery	= 4
Drugs	= 5
Hormonal	= 6
Bereavement	= 7
Emotional	= 8
Pregnancy	= 9
Other	= 0

FAMILY HIST

RA	= 1
Other CTD	= 2
AI Thyroid	= 3
IDDM	= 4
Other AID	= 5

PAST MED HIST

Seroneg spond	= 6
Psoriasis	= 7
Inflam. Bowel	= 8
Ocular	= 9
Other	= 0

Drg 1 | Drg 2 | Crap

DATE OF VISIT MONTH | YEAR

TREATMENT CURRENT 1 2 | PREVIOUS 1 2

0= None
1= Analgesia
2= NSAIDS
3= 1+2
4= Short course Steroids
5= IA Cortisone
9= Other

EFFICACY 1 2

SIDE EFFECTS 1 2

(See follow up form for codes)

ARA CRITERIA

EMS
0 = <1/5 hour
1 = <1 hour
2 = <2 hour
3 = <3 hour etc

JOINT SCORE 00-59

JOINTS INVOLVED
1 = Hand/Wrists
2 = Feet
3 = Both
4 = Large Jts Only
5 = Small + Large Jts
6 = Mono Articular
7 = Soft Tissue Only
8 = Neck Only
9 = Other

SYMMETRY
1 = Symmetry
2 = Asymm

CLINICAL MEASUREMENTS

Nocturnal Waking — No. times night

FUNCTIONAL GRADE (I-IV = 1 - 4)

GRIP (R) --0/30 | GRIP (L) --0/30 | RITCHIE 00-78 | PAIN SCORE 00-99mm | HAQ 00-24 | WEIGHT (KG)

CLINICAL SUBSETS

TYPE OF SYNOVITIS
0 = None
1 = Mild Proliferative
2 = Marked Proliferative
3 = Dry/Atrophic
4 = Soft Tissue
5 = Myalgic
6 = Tenderness Only
9 = Other

RA / Other

COURSE
0 = Too Early
1 = Episodic/Palin
2 = Nonrecurrent
3 = Relapse/Remit
4 = Chronic Persistent
5 = Transient Synovitis
6 = Other specific diagnosis
7 = Diagnosis not known
9 = Other

PAIN THRESHOLD
1 = Low
2 = Normal
3 = High

EXTRA ARTICULAR DISEASE 1 2 3
1 = Nodules
2 = Sjögren's
3 = Raynaud's
4 = Lung
5 = CNS/Muscle
6 = Non Specific
7 = Feltys
8 = Cut/Ocular Vasculitis
9 = Systemic Vasculitis
0 = Other

X-RAYS HANDS | FEET | C/S
0=Normal 1=OA only 9=Not done
2 = Joint Space Loss
3 = Juxta OP
4 = Erosions
5 = Ankylosis
2 = AAS
3 = Subaxial Subluxation
4 = Both
5 = Ankylosis

LABORATORY

WCC /Cumm^2 | HB Gm/l | PLATELETS x10/l^9 | ESR mm/hr

LX | SCAT | ANA

0 = Negative
1 = Equivocal
2 = Low Positive
3 = High Positive
(or titre eg 1=1/10 2=1/20 3=1/40 etc)

LABORATORY STORAGE

0 = None
1 = Serum
2 = Cells
3 = Urine
4 = 1 + 2
5 = 1 + 3
6 = 2 + 3
7 = 1, 2, 3

FOLLOW UP FORM

CARD

HOSP ID NUMBER

DATE OR VISIT

DNA

IF DNA
1 = Moved
2=Can't
3=Won't
4=Remission
5=since died
6=Discharged
9=Don't know

CURRENT MEDS 1 2

0= NSAIDs only
1= Chloroquin
2= SZP
3=IM Gold
4=Oral G
5=DPM
6=AZA
7=CYCLO
8=MTX
9=Other

CURRENT STEROIDS

Dose Mgs
0= None
1=2.5
2=5
3=7.5
4=10
5=12-5
6=15
7=20
8=>20
9=Other

PREVIOUS STEROIDS

Dose Mgs

PREVIOUS MEDS (Since last visit)

1 DRUG 2

EFFICACY

TOXICITY

CODES FOR EFFICACY & SEVERITY OF TOXICITY

Effective — Ineffective

1……No toxicity…….5
2……Tolerable …….6
3……Intolerable ……..7
4……Severe ……..8
9=fatal
0=don't know

CODES FOR TYPE OF TOXICITY

1= Skin rash
2=Mucocutaneous
3=Renal
4=Haematology
5=Dyspepsia
6=Peptic Ulcer
7=Hepatic
8=Ocular
9=CNS
10=Other

LIST NSAIDS
1
2
3

LIST TYPE
1
2
3

DETAILS
1
2
3

ARA CRITERIA

EMS
0=<1/5hour
1=<1hour
2=<2hour
3-3hour etc

JOINT SCORE
00 - 59

JOINTS INVOLVED
1= Hand/Wrist
2=Feet
3=Both
4=Large Jts only
5=Small&Large
6=Mono Articular
7=Soft Tissue only
8=Neck only
9=Other

SYMMETRY
1=Symmetry
2=Asymm

CLINICAL MEASUREMENTS

Nocturnal Waking No. times night

Functional Grade (I-IV=1.4)

Grip(R) –0/30

Grip (R) –0/30

Ritchie 00 – 78

Pain Score 00 – 99mm

HAQ 00-24

Weight (KG)

CLINICAL SUBSETS

TYPE OF SYNOVITIS

0=None
1=Mild Proliferative
2=Marked Proliferative RA
3=Dry/Atrophic
4=Soft Tissue
5=Myalgic Other
6=Tenderness only
9=Other

COURSE

0=To early
1=Episodic/Palin
2=Nonrecurrent
3=Relapse/Remit
4=Chronic Persistent
5=Transient Synovitis
6=Other specific diagnosis
7=Diagnosis not known
9=Other

PAIN THRESHOLD

1=Low
2=Normal
3=High

EXTRA ARTICULAR DISEASE 1 2 3

1=Nodules
2=Sjögern's
3=Raynaud's
4=Lung
5=CNS/Muscle
6=Non Specific
7=Feltys
8=Cut/Ocular Vasculitis
9=Other

X-RAY

HANDS FEET C/S

0=Normal 1=OA only 9=Not done
2=Jt . space loss 2=AAS
3=Juxta OP 3=Subaxial Subluxation
4=Erosions 4=Both
5=Ankylosis 5=Ankylosis

LABORATORY

WCC /Cumm²

HB Gm/l

PLATELETS x10/1 /9

ESR mm/hr

LX **SCAT** **ANA**

LABORATORY STORAGE

0= None
1=Serum
2=Cells
3=Urine
4=1+2
5=1+3
6-2+3
7=1,2,3

0=Negative
1=Equicoxal
2=Low Positive
3=High Positive
(or titre eg 1=1/10, 2=1/20, 3=1/40)

HAD SCALE
A D

ERAS OUTCOME 3 / 5 OR 10YEAR FOLLOW UP

Eras ID number: ☐ **Date:** ☐

1. SOCIAL DETAILS – at time of entry to ERAS and yearly outcome visits
2. Visit date:

	Fup year	Fup year	Fup year
Fup year *Tick if no change:*	☐	☐	☐
1a. Occupation details reasons for change spouses occupation	☐	☐	☐
1b. Av. Hours per day	☐☐	☐☐	☐☐
1c. Months off (RA)	☐	☐	☐
1d. Marital status	☐	☐	☐
1e. Support FAMILY	☐	☐	☐
1f. Support LOCAL	☐	☐	☐
Accomodation 1g. Type	☐	☐	☐
1h. Numbers bedrooms	☐	☐	☐
Details & reason if different:-			
1i. Education	☐	☐	☐
1j. Social Class (1-5)	☐	☐	☐
1k Allowances:-	☐	☐	☐
1w. Other:-	☐	☐	☐

1x. Comorbid Conditions
1.
2.
3.
4.
5.
6.

Cause of Death:-

Current Medication:-

1.	2	3
4	5	6
7	8	9
10		

CODES

Occupation
1= manual
2= semi manual
3= semi sedentary
4= sedentary
5= housewife
6= employed
7= retired
8= student
9= other
0= none

If out of work at present put in code for normal job and in hours/day:98=redundant, 99=unemployed

Marital Status
1= married 2=cohabit
3=single 4= divorced
5=separated 6=widowed
7=single parent 8= other
9=don't know

Family & local support NONE = 0

NON PERSONAL	PERSONAL
1=< once a week	4=<once a week
2=>once a week	5=>once a week
3= daily	6 = daily

Acomodation

Owned	Rented
1 = Room	=5
2 = Flat	=6
3 = Bungalow	=7
4 = House	=8
Part III	=9
Institute	=10

Education
0=None 1=No CSE
2= CSE/O's 3=CSE/A's
4= Training 5= Nat.Diploma
6=University Degree 7= Other
8=Don't know

Social Class
1=Professional – Doctor etc
2=Professional – Teacher etc
3=Skilled
4=Semi-skilled
5=Unskilled

Enter details of comorbid conditions & medication in text, abbreviations or acronyms.

Smoking Data
Current smoker

Amount per day	Start date	Todays date

Ex-smoker

Amount per day	Start date	Stopped date

2. IN PATIENT EPISODES

Fup year _____

(please enter)	DATE	REASON	Number of days	Codes op'n	Codes jt/site
2l:					
2m:					
2n:					
2o:					
2p:					
2q:					
2r:					
2s:					
on W/List for:					

Codes for med/orth IP

(Columns 1 & 2)
0 1 = RA medical
0 2 = iatrogenic
0 4 = non RA
0 5 = rehab
0 6 = other

OPERATION CODE
(Column 1)
1 = joint replacement
2 = revision replacement
3 = excision
4 = CT decompression
5 = surgical synovector
6 = soft tissue operation
7 = arthodesis
8 = medical synovector
9 = other

JOINT/SITE CODES
(Column 2)
See jt/site in sect. 4 ROM

3. OUT PATIENT EPISODES

Fup year:	____ year	____ year
3t. Endoscopies		
3u. OT/PT/Hand Rx		
3v. Chiropody		
3w. Appliances		
3x. shoe fitter		
3y. Wheel chair		
Home adapt		

Codes for OP episodes

Write in number of episodes/ sessions (eg 0-8,9=9 ore more)

APPLIANCES
0 = none
1 = wrist splint/collars
2 = kitchen/home aids
3 = walking aids
4 = calipers etc
5 = hoists
6 = other

WHEEL CHAIR
0 = never 1 = > 1/year
2 = > 1/month 3 = >1/week
4 = daily 5 = constant

4. RANGE OF JOINT MOVEMENT (ROM)

Fup year:	____ year		____ year	
Joint site	R	L	R	L
1. Shoulder				
2. Elbow				
3. Wrist				
4. MCP/PIP				
5. Hip				
6. Knee				
7. Ankle				
8. Hindfoot				
9. MTP				
0. Cervical spine				

Codes for ROM

0 = Normal ROM

1 = up to 25% loss

2 = up to 50% loss

3= up to 75% loss

4 = up to 95% loss

5 = complete ankylosis

Appendix B

ERAS Disease Activity Score Form

Formulae to calculate DAS with 4 or 3 variables and with ESR or CRP

DAS = $0.54 \times \sqrt{(\text{Ritchie})} + 0.065 \times \text{SJC44} + 0.33 \times \log_{nat}(\text{ESR}) + 0.0072 \times \text{GH}$

DAS-CRP = $0.54 \times \sqrt{(\text{Ritchie})} + 0.065 \times \text{SJC44} + 0.17 \times \log_{nat}(\text{CRP}+1) + 0.0072 \times \text{GH} + 0.45$

DAS-3 = $0.54 \times \sqrt{(\text{Ritchie})} + 0.065 \times \text{SJC44} + 0.33 \times \log_{nat}(\text{ESR}) + 0.224$

DAS-3 CRP = $0.54 \times \sqrt{(\text{Ritchie})} + 0.065 \times \text{SJC44} + 0.17 \times \log_{nat}(\text{CRP}+1) + 0.65$

Formulae to calculate DAS28 with 4 or 3 variables and with ESR or CRP

DAS28 = $0.56 \times \sqrt{(\text{TJC28})} + 0.28 \times \sqrt{(\text{SJC28})} + 0.70 \times \log_{nat}(\text{ESR}) + 0.014 \times \text{GH}$

DAS28-CRP = $0.56 \times \sqrt{(\text{TJC28})} + 0.28 \times \sqrt{(\text{SJC28})} + 0.36 \times \log_{nat}(\text{CRP}+1) + 0.014 \times \text{GH} + 0.96$

DAS28 -3 = $[0.56 \times \sqrt{(\text{TJC28})} + 0.28 \times \sqrt{(\text{SJC28})} + 0.70 \times \log_{nat}(\text{ESR})] \times 1.08 + 0.16$

DAS28 -3 CRP = $[0.56 \times \sqrt{(\text{TJC28})} + 0.28 \times \sqrt{(\text{SJC28})} + 0.36 \times \log_{nat}(\text{CRP}+1) \times 1.10 + 1.15$

EULAR criteria based on DAS

DAS < 1.60	-	remission
DAS ≥1.60 and ≤ 2.40	-	low disease activity
DAS >2.40 and ≤ 3.70	-	moderate disease activity
DAS> 3.70	-	high disease activity

EULAR criteria based on DAS 28

DAS 28 < 2.6	-	remission
DAS 28 ≥ 2.6 and ≤ 3.2	-	low disease activity
DAS 28 >3.2 and ≤ 5.1	-	moderate disease activity
DAS 28 > 5.1	-	high disease activity

Appendix C

ERAS Health Assessment Questionnaire

HEALTH ASSESSMENT QUESTIONNAIRE (HAQ)

Name: …… Date: ……

We are interested in learning how your illness affects your ability to function in daily life.
Please feel free to add any comments at the end of this form.

PLEASE TICK THE ONE RESPONSE WHICH BEST DESCRIBES YOUR USUAL ABILITIES OVER THE PAST WEEK

	Without **ANY** difficulty (0)	With **SOME** difficulty (1)	With **MUCH** difficulty (2)	**Unable** to do (3)
1. DRESSING & GROOMING Are you able to:				
Dress yourself, including tying shoelaces	☐	☐	☐	☐
and doing buttons?	☐	☐	☐	☐
Shampoo your hair?				
2. RISING Are you able to:				
Stand up from an armless straight chair?	☐	☐	☐	☐
Get in and out of bed?	☐	☐	☐	☐
3. EATING Are you able to:				
Cut your meat?	☐	☐	☐	☐
Lift a full cup or glass to your mouth?	☐	☐	☐	☐
Open a new carton of milk (or soap powder)	☐	☐	☐	☐
4. WALKING Are you able to:				
Walk outdoors on flat ground?	☐	☐	☐	☐
Climb up five steps?	☐	☐	☐	☐

PLEASE TICK ANY AIDS OR DEVICES THAT YOU USUALLY USE FOR ANY OF THESE ACTIVITIES :

☐ Cane ☐ Walking frame ☐ Crutches ☐ Wheelchair

☐ Built up or special utensils ☐ Special or built-up chair
☐ Other (specify) ……..

PLEASE TICK ANY CATEGORIES FOR WHICH YOU USUALLY NEED HELP FROM ANOTHER PERSON :

☐ Dressing and Grooming ☐ Eating ☐ Rising ☐ Walking

PLEASE TICK THE ONE RESPONSE WHICH BEST DESCRIBES YOUR USUAL ABILITIES OVER THE PAST WEEK

	Without **ANY** difficulty (0)	With **SOME** difficulty (1)	With **MUCH** difficulty (2)	**Unable** to do (3)
5. HYGIENE Are you able to:				
Wash and dry entire body?	☐	☐	☐	☐
Take a bath?	☐	☐	☐	☐
Get on and off the toilet?	☐	☐	☐	☐
6. REACH Are you able to:				
Reach and get down a 5 lb object (e.g. a bag of potatoes) from just above your head?	☐	☐	☐	☐
Bend down to pick up clothing from the floor?	☐	☐	☐	☐
7. GRIP Are you able to:				
Open a car door?	☐	☐	☐	☐
Open jars which have been previously opened?	☐	☐	☐	☐
Turn taps on and off?	☐	☐	☐	☐
8. ACTIVITIES Are you able to:				
Run errands and shop?	☐	☐	☐	☐
Get in and out of the car?	☐	☐	☐	☐
Do chores such as vacuuming, housework or light gardening?	☐	☐	☐	☐

PLEASE TICK ANY AIDS OR DEVICES THAT YOU USUALLY USE FOR ANY OF THESE ACTIVITIES :

☐ Raised toilet seat ☐ Bath rail ☐ Bath seat ☐ Jar opener (for jars previously opened)
☐ Long handled appliances for reach ☐ Other (specify) ……

PLEASE TICK ANY CATEGORIES FOR WHICH YOU USUALLY NEED HELP FROM ANOTHER PERSON :

☐ Hygiene ☐ Gripping and opening things ☐ Reach ☐ Errands and housework

We are also interested in learning whether or not you are affected by pain because of your illness.

HOW MUCH PAIN HAVE YOU HAD BECAUSE OF YOUR ILLNESS IN THE PAST WEEK ?

Place a mark on the line to indicate the severity of the pain

No pain _____ Very severe pain

Appendix D

Radiographic Scoring Sheets

Following joints are assessed in the SvdH method for erosions:

a. 10 MCP joints
b. 8 PIP joints
c. 2 IP joints of the thumbs
d. right and left 1st metacarpal bone
e. right and left radius and ulnar bones
f. right and left trapezium and trapezoid as one unit
g. right and left navicular bones
h. 10 MTP joints
i. 2 IP joints of the big toes

Erosions are scored 1 if they are discrete and 2 or 3 depending on the surface area of the joint involved. In the carpal bones it is sometimes very difficult to score erosions as the bone collapses completely and in this case the collapsed area is given a score according to the surface area involved and a complete collapse is scored as 5.

In each hand including the wrists, 16 joint areas are scored for erosions and a maximum erosion score for each joint is 5, whereas, in the feet 6 joint areas are scored for erosions in each foot with a maximum erosion score of 10 for each joint area, to increase weight of the feet joints in the total erosion score. Therefore, erosion score ranges from 0 to 160 in the hands and 0 to 120 in the feet with a total erosion score ranging from 0 to 280.

Following joints are assessed in the SvdH method for joint space narrowing:

a. 10 MCP joints
b. 8 PIP joints
c. right and left 3rd, 4th and 5th carpometacarpal joints
d. right and left multangular-navicular joints
e. right and left capitate-navicular-lunate joints
f. right and left radio carpal joints
g. 10 MTP joints
h. 2 IP joints of the big toes

Joint space narrowing is combined with score for (sub)luxation and is scored as:

0 = normal,

1 = focal or doubtful

2 = generalised but less than 50% of the original joint space,

3 = generalised and more than 50% of the original joint space or subluxation

4 = bony ankylosis or complete luxation

JSN is assessed in 15 joint areas in each hand including the wrists and in the feet 6 joint areas in each foot are scored. Therefore, JSN score in the hands ranges from 0 to 120 and in the feet it ranges from 0 to 48 with a total JSN score ranging between 0 and 168.

Erosion score and JSN score are added together to give a total Sharp score, which ranges from 0 to 448 in the SvdH method. SvdH method has been used widely in several studies and is currently the most common method used in clinical trials.

Following joints are assessed in this modified Larsen method:

Proximal interphalangeal(PIP) joints of both hands	-	8
Interphalangeal(IP) joints of both thumbs	-	2
Metacarpophalangeal (MCP) joints of both hands	-	10
Both wrists (score multiplied by 5)	-	2
Metatarsophalangeal joints (MTP) of 2^{nd} -5^{th} toes on both sides	-	8
Interphalangeal (IP) joint of big toes on both sides	-	2

Grading of radiographic abnormalities in this modified Larsen method:

Grade 0: Normal finding

Grade 1: Soft tissue swelling, juxta-articular osteoporosis, possibly with slight narrowing of the joint space

Grade 2: Early but definite abnormality consisting of bone erosion and distinct narrowing of the joint space.

Grade 3: Medium destructive abnormality with marked narrowing of the joint space

Grade 4: Severe destructive abnormality. Only minor parts of the articular surfaces remain

Grade 5: Mutilating lesions

In this modified Larsen method, 20 joint areas in the hands and 10 joint areas in the feet are assessed with a maximum score of 5 for each joint area. The wrist is assessed as one unit and then multiplied by 5, which gives a maximum score of 25 for each wrist. Therefore, the total score in this method ranges from 0 to 200.

Appendix E

Published Paper - Clinical Practice

Review

Importance of registries in informing clinical practice for arthritis

Lewis Carpenter[1], Elena Nikiphorou[2] & Adam Young*[1,3]

Practice Points

- Cohort studies and registries provide information on clinical guidelines and quality of care, drug safety and benefit–risk data, and health-related outcomes, including quality of life and clinical effectiveness.
- The prevalence of rheumatoid arthritis (RA) from registries in the UK in the 1990s was 1.16% in women and 0.44% in men, which had decreased in women and increased in men from the 1950s.
- RA has a very variable course, but is generally set in the first few years.
- Reduced functional ability is frequently an early feature.
- Rheumatoid factor remains the most useful prognostic factor.
- The long-term adverse effects of biologics include a slight increase in opportunistic infections.
- The identification of 'early RA' is critical and drives forward improvements in management.
- The continuous use of contemporary data is important for national policy making organizations (the National Audit Office and NICE).
- Current databases underpin future and more ambitious national initiatives.

SUMMARY: The gold standard in research for evidence that underlies clinical practice is the randomized controlled trial. In recent years it has been accepted that observational studies, which include disease and drug registries and cohort studies, are very important sources of data not available from randomized clinical trials, and the two different approaches complement one another. In rheumatology, the development of clinical guidelines, standards of care and health policies, and appraisal of new drugs by NICE, all rely on clinical outcomes, prognostic factors and responses to drug therapies provided by both sources. Observational

[1]The Center for Lifespan & Chronic Illness Research, University of Hertfordshire, Hatfield, UK
[2]Department of Rheumatology, Addenbrookes Hospital, Cambridge, UK
[3]Department of Rheumatology, City Hospital, St Albans, AL3 5PN, UK
*Author for correspondence: adam.young@nhs.net

Future Medicine part of fsg

studies and registries in arthritis have promoted greater collaborations between academics and clinicians, and with patient support groups and public health. The main strengths of observational studies are that, first, they reflect 'real-world' practice and, second, they can achieve prolonged follow-up. As the management of chronic conditions such as arthritis becomes more complex and health economic issues more important in the 21st century, it is probable that more reliance will be placed on these types of studies.

Disease or patient registries are collections of secondary data related to patients with a specific condition or intervention. In previous decades, a disease registry in its simplest form would consist of a diagnostic index of patients collected at one point in time on paper cards kept by an individual physician. Now registries vary in sophistication from simple computer spreadsheets, with confined access, to very complex databases, available online across multiple institutions. Although they may just provide a snapshot of a condition or drug, many have extended to include regular yearly follow-ups and collection of outcomes, either directly or through database linkage via unique personal identification codes, resulting in similar aims and designs to observational and longitudinal cohort studies. Cancer registries have been in use in the UK and internationally for many years, whereas in rheumatology most registries have been developed only recently to monitor the new biologic therapies.

The term 'cohort' is used to describe a group of people who have something in common when they are first assembled. A cohort is usually established based on a specific diagnosis, and individuals may be recruited either at the time of diagnosis ('inception') or any time in the course of their disease ('prospective'). They are often referred to as long-term observational studies and used to describe and record the course and long-term complications of disease and its therapy. Ideally, they should be inception cohorts and provide assessments at specific and regular time-points for prolonged periods with as complete follow-up as possible.

Patient registries and cohort studies are very different from clinical trials in terms of design, logistics, approvals and site expectations. Success demands different expertise and core competencies. The main evidence for the efficacy and safety of medical products and therapies are provided by well-conducted randomized controlled trials (RCTs), considered the 'gold standard' research method. They generally study carefully selected groups of patients under controlled conditions, possibly over periods of months but not years. In a world of limited resources and patients with diverse risk factors and health conditions, clinicians, patients and commissioners need to know which products and services are safer, more effective and adhered to in a variety of 'true-to-life' settings that reflect the populations of interest, in both the short- and the long-term. Longer follow-up is possible through open-label extensions of RCTs, but these again include positive selections of patients originating from the trials.

Registries and cohort studies imply different processes to each other, and although they usually have different aims and designs, many ultimately come to resemble one another closely since most examine chronic conditions requiring prolonged follow-up. Most are related to conditions responsible for the bulk of follow-up in clinical rheumatology, namely adult inflammatory arthritis (AIA) and its subset rheumatoid arthritis (RA). Many have been designed for multiple purposes and not confined necessarily to answer just one research question. This article will examine the rationale, content and results of both registries and cohort studies for RA in the UK, although the principles and conclusions could apply to any medical field. Space restricts detailed inclusion of the many European and USA registries and cohort studies, which will be the subject of a future article.

Rationale for registries & cohort studies

Whether disease-based or product-focused, registries and cohort studies have been designed to capture data and evidence for both scientific and clinical governance issues. They provide the health community with invaluable data about the natural history of a disease or intervention under standard care practices, over periods that cover the development of most disease patterns and/or drug effects. Most aspire to regional if

not national coverage and many are voluntary. Clinical registries or cohort studies designed to capture operational clinical data as part of routine clinical care have the potential to promote better quality of treatment in general and in the individual patient, and specifically provide:

- Prevalence and incidence figures, geographical variations and secular change;
- Drug safety and benefit–risk data;
- Clinical guidelines and quality of care: development, improvement, monitoring and adherence;
- Health-related outcomes, including quality of life;
- Clinical and cost–effectiveness;
- Immediate access to well-presented longitudinal patient records generated from little additional work for the clinician;
- Data for research questions and, with time, a powerful research database.

The medical care costs of chronic diseases account for most of the National Health Service (NHS) budget, and some of the more common ones covered by national registries, for example cancer and diabetes. Registries are being increasingly developed for novel interventions, for example drug products, such as the biologic agents.

Currently, the most common registries in rheumatology are the Biologic Registries, developed in European countries since the introduction in 2001 of the new, more effective and expensive TNF-blocking agents for a number of autoimmune inflammatory conditions. At this time, the main clinical issues were not efficacy or short-term toxicity, which were not disputed, but were safety in the long-term and the choice of when the optimal stage of RA to introduce these novel agents was. The main reason for their initiation was because data on the long-term intended and unintended effects of biologics were relatively scarce. There were theoretical reasons to suspect that patients on biologics have increased risk of both malignancies (especially lympho-proliferative) and opportunistic infections over time. Another reason in some countries was the requirement to register patients prior to securing funding for these agents. Initially the clinical criteria for eligibility for biologics varied between countries, although agreement has now been achieved in Europe [1]. Registries have the potential for improving the understanding efficacy of therapies in the long-term, and provide information on whether a specific drug is clinically effective in real-world situations.

Registries can be associated with pay-for-performance quality-based contracts for individual, groups of or all doctors in a country. For example, the UK now rewards physicians according to 146 quality measures related to ten chronic diseases that are tracked electronically, and linked to the best-practice tariff.

In the USA, many registries are for surgical procedures or devices to monitor both long-term efficacy and healthcare expenses. The UK, Norway, Sweden and Australia have national patient registries that track patients with artificial joints in order to assess performance over time. Regulators can use such information to force manufacturers to justify why poorly performing hip or knee prostheses should remain available, and products have been withdrawn as a result.

Ethical issues vary according to individual national laws. Generally, no ethical approval is needed for the publication of the results of clinical audits that are based on routine collection of data. Research and other projects involving linkage to other registries or biobanks require ethical approval. Registries generally clear their methodology with data protection agencies, which are based on the Act of Processing of Personal Data that ensures that data security and protection of individual rights, among others, are dealt with correctly.

Historical account of registries & prospective cohorts of RA in the UK

The early population-based studies in the UK and USA in the 1950s–1960s provided data concerning the prevalence of RA and rheumatoid factor (RF) [2–4]. The findings suggested that self-limited polyarthritis was more common than progressive RA in general populations. RA inception cohorts in clinical settings followed, initially in the UK in the 1960s [5,6]. These cohorts provided, for the first time, valuable information on the course of hospital-based RA as they included strategies to follow-up the majority of patients using the same standard observations for at least 5 years. From these it became recognized that RA in the clinical setting differed from population studies. A far higher proportion of these patients had progressive

disease, supporting clinical experience. These differences were not widely recognized until the 1970s–1980s when these studies had sufficient follow-up to report the wide spectrum of RA. Unfavorable outcomes were reported early on in the disease course and, in significant proportions, were irreversible, resulting in questioning of management strategies.

Therapies at this time were limited to steroids and NSAIDs, and a small number of slow-acting disease-modifying antirheumatic drugs (DMARDs). The more effective agents, such as intramuscular gold and D-penicillamine, had significant and sometimes severe toxicity, so the need for reliable predictors of severe RA became increasingly more important.

At this time, no single marker or set of markers could be used to predict with certainty which patients were most likely to fare worse. The initial cohorts were disadvantaged by being single site and tertiary referral centers with limited numbers at follow-up, and not all assessments had been standardized. Scientific data from observational studies were not well regarded compared with randomized studies at this time, and it took several more years before it was recognized that if well designed and performed to a high standard, inception cohorts can provide clinical effectiveness and prognostic data to complement the results of RCTs [7].

Possible sources of bias in inception cohorts include small sample sizes at follow-up, left censoring (milder RA not being referred), right censoring (severe RA not surviving long enough for follow-up), and treatment effects. Assessment of therapies is limited in observational studies with nonrandom assignment of drug therapy. Study of drug efficacy is more reliably achieved with RCTs. However, newer agents can only be described as disease-modifying if demonstrated to alter objective measures in the long-term, namely x-ray damage. However, inception cohorts may permit comparison of the broader issue of clinical effectiveness of conventional and newer drugs in well-described historical cohorts.

A small number of single and multicenter hospital- and community-based inception observational cohorts in RA were designed in the UK and northern Europe in the 1980s in order to address these issues. With greater numbers at follow-up, some cohorts included the less common but important outcomes of clinical remission, work disability and orthopedic intervention. By the late 1990s most countries in northern Europe had established inception cohorts.

Most cohorts continue in follow-up, providing valuable comparative data on variations in therapeutic practice as well as other outcomes. Accumulated evidence from all RA inception cohorts has suggested that the course of the disease is highly variable, but is established early and that the most important phase for therapy is in the first 2 years. Most, but not all, have reported increased mortality compared with normal populations, mainly from cardiovascular disease. These factors were the main drivers for a more focused approach to the management of early RA.

Since the initiation of these cohorts, several major epidemiological advances in RA have been achieved in the last 15 years and included the revised classification criteria for RA [8], core sets of disease activity (DAS) measures [9], response criteria for the assessment of drug efficacy [10], and agreement on a core set of measures for longitudinal observational studies [11]. Another important development has been the formation of patient support groups in rheumatology. In the UK, both the National Rheumatoid Arthritis Society and the Arthritis and Musculoskeletal Alliance have become active not only politically, but have also contributed to the formation and running of studies and interpretation of results [101,102].

An important advance of the 1990s was evidence to support early intervention with disease-modifying therapies in RA [12]. This, and the success of inception cohorts, just described above, led to the development of early arthritis clinics, which are now part of standard services in many rheumatology departments. One challenge in establishing early arthritis clinics is to collect data continuously during routine care when patients are seen for the first time, or to have a clinical research facility attached to the clinical unit with the capacity to perform immediate on-demand data collection [13].

The more effective and expensive biological agents became available at the start of the millennium and this major therapeutic development resulted in the formation of product registries in the UK and many European countries including The Netherlands, Sweden, Norway, Denmark, France, Germany and Spain in order to monitor long-term adverse effects. The British Society of Rheumatology (BSR) established

the first Biologics Register (BSR-BR) in 2001 providing invaluable data on these agents in several disease areas, and the model for similar registries in other specialties, for example dermatology [14].

Details of the main RA cohorts & biologic registries

Table 1 summarizes the basic details of RA cohorts and registries initiated in the UK from the late 1950s that have stood the test of time and have reported on important outcomes with adequate follow-up. These are also described in detail below, followed by brief descriptions of the more recent initiatives not in the table.

Bath cohort

The first hospital-based early RA cohort recruited patients at the Royal National Hospital for Rheumatic Diseases in Bath, UK, between 1957 and 1963, and included 100 patients who met the American Rhematology Association criteria for definite or classical RA [15] and were first seen within 1 year of their initial arthritis symptoms prior to disease-modifying therapy [5]. Follow-up of these patients continued for up to 40 years and, although limited by progressively smaller numbers for analysis, it was the first to report a significant decline in functional capacity in as many as a third of the patients over the first 3 years, as well as high disability rates in the longer term.

The Middlesex Hospital cohort

The RAPS study was established in 1966 at Middlesex Hospital (London, UK) and enrolled consecutive patients with the same entry criteria as the Bath study, except that wider American Rheumatology Association criteria for RA were accepted to include less severe RA. The aim was to gather detailed information on the characteristics of disease onset in 100 patients in order to develop prognostic factors [6,16]. A novel finding was that serial x-rays of hands and feet demonstrated early changes, and nearly a third had structural damage by 1 year, rising to 71% by 5 years [17]. A subgroup of erosive patients was identified in whom no new erosions developed or progressed after approximately 3 years. This study was the first to show the importance of foot involvement in early RA, both clinically and radiographically

Only a few standardized and validated assessments were available to these first two cohorts and sample sizes were small, but they did achieve 15–25 years follow-up, and provided insights into early RA, which in this era was treated relatively late, mainly with intramuscular gold therapy as the first disease-modifying drug. Both demonstrated considerable fluctuation in the course of early RA and introduced the importance of serial follow-up of functional measures and x-rays of hands and feet.

The ERAS study

In the UK the proposal for a new inception cohort of RA arose from the recognition by a group of clinical rheumatologists in the 1980s that the optimal management for RA was a major challenge and any improvements were unlikely to result from RCTs alone. Important advances in the care of RA from the 1980–1990s included new drug therapies and standardized disease assessments. Large joint replacement surgery had become more routine and available. In order to develop prognostic factors and to capture the wide variations in clinical outcomes and health status, and in therapies offered in clinical practice, larger numbers of patients were required than hitherto possible.

The ERAS study was designed to recruit RA patients from NHS hospital rheumatology outpatients in nine different regions of England from 1986: a modernized version of RAPS. The aims were to establish a database of long-term clinical data on 1000 patients in order to monitor and compare management and outcomes between centers, and develop prognostic factors. Standard clinical assessments by research nurses included DAS, function (Health Assessment Questionnaire) and x-rays of hands and feet at baseline, 6 months and yearly for up to 20 years. Outcomes included validated assessments of functional and radiological progression, and mortality and comorbidity, with greater numbers and an improved management era compared with earlier cohorts [18]. The larger sample size allowed examination of both standard and less well-documented outcomes, which included clinical remission, work disability and orthopedic interventions (joint replacement and reconstruction) in RA patients treated with conventional DMARDs of the era [19–21]. Sulphasalazine was the first-choice DMARD of clinicians in this study, followed by methotrexate and intramuscular gold, which reflected common UK practice of this era. Severe toxicity from these conventional

Table 1. UK Cohort/registries with years of follow-up and publications.

Cohort	Location	Type of cohort	Recruitment	Maximum follow-up	Sample	Inclusion criteria	DAS	HAQ	QoL	Radiographic	Genetic	Comorbidity	Surgery	Ref.
Bath cohort	Bath	Inception	1957–1963	40 years	100	ARA, symptoms <1 year	No	No	No	Yes	No	No	No	[5]
RAPS	London	Inception	1966–1971	20 years	102	ARA, symptoms <2 years	No	No	No	Yes	No	No	No	[6]
ERAS	Multicenter	Inception	1986–2001	25 years	1465	Physician, symptoms <2 years	Yes	Yes	Yes	Yes	Yes	Yes	Yes	[18]
NOAR	Norfolk	Inception	1989–present	Ongoing	1657	Physician	Yes	Yes	Yes	Yes	Yes	Yes	No	[24]
ERAN	Multicenter	Inception	2001–2012	11 years	1236	Physician, symptoms <3 years	Yes	Yes	Yes	Yes	Yes	Yes	Yes	[30]
GORA	Sheffield	Cross-sectional	1999–2006	NA	873	Physician, symptoms <3 years	Yes	Yes	No	Yes	Yes	No	No	[110]
BSR-BR	Manchester	Registry	2001–present	Ongoing	~20,000	Physician, starting biologics	Yes	Yes	Yes	No	No	Yes	No	[43]
YEAR	Yorkshire	Inception	Cohort B: 1998–2003; Cohort C: 2000–2009	15 years	~1600	Physician, symptoms <1 years	Yes	Yes	Yes	Yes	No	Yes	No	[49]
BRAGGS	Multicenter	Prospective	2009–present	Ongoing	2520	Physician	No	No	No	No	Yes	No	No	[109]
CARA	Scotland	Prospective	2005–2006	1 year	465	Physician	Yes	Yes	Yes	No	No	Yes	No	[111]
RAMS	Multicenter	Inception	2008–2013	Ongoing	1100	Physician	Yes	Yes	Yes	Yes	No	Yes	No	[112]

ARA: American Rheumatology Association; DAS: Disease activity; HAQ: Health Assessment Questionnaire; NA: Not applicable; QoL: Quality of life.

DMARDs was uncommon, and very rarely related to mortality, which was more probable from nonsteroidal and steroidal drugs [22].

This study demonstrated that it was possible to collect standardized assessments in ordinary clinical settings over time without huge expense. However, despite a larger sample size, predictive markers powerful enough to be used routinely in clinical settings remained elusive. The power and relevance of predictive factors depended considerably on the outcome measure of interest, although RF remained consistent for most outcomes except function [23].

The Norfolk Arthritis Registry

NOAR [24] is a community-based study in one region of the UK designed to establish the incidence of inflammatory polyarthritis and RA in the 1990s [25], following on from the population studies conducted mainly in the industrial north of the UK in the 1950s. Patients recruited between 1989 and 1994 and between 2000 and 2008 in Norfolk were clinically assessed at baseline and at 1, 2, 3, 5, 7, 10 and 15 years by community research nurses. Patients recruited between 1994 and 2000 were followed for 2 years. NOAR reported an overall minimum prevalence of 1.16% in women and 0.44% in men [26] and, in comparison with the first UK studies in the 1960s [2,3], it was evident that prevalence in women was decreasing in all age groups (except in the 75+ group) whereas the prevalence in men has increased.

This study reported that even patients with AIA have unfavorable outcomes, significant morbidity and functional loss [27]. Patients who developed RA were followed-up in secondary care as a cohort study, and in these patients NOAR has reported on similar outcomes as ERAS. It is one of only a few UK studies that collected information on direct, indirect and intangible costs of the disease, based on specially designed questionnaires to capture costs of the disease from a personal, NHS and societal perspective [28].

Both NOAR and ERAS confirmed the finding of previous cohorts of irreversible disease in significant proportions of patients within the first few years. Despite improved therapies, reduced function early on was associated with later disability. Both studies took advantage of linkage to national datasets and reported increased standard mortality rates in RA [22,29]. The causes of

death in patients with RA were similar to that of the normal population, although infectious, cardiovascular and respiratory conditions were more common. These two cohorts represent the full spectrum of RA, and some differences in the results could be explained by generally milder RA from community sources compared with hospital-based patients.

The Early Rheumatoid Arthritis Network

The 21st century brought in a number of relevant developments: international and national guidelines on the management of RA had been published; patient support groups had become well organized and proactive; equitable access to appropriate care had become a greater issue in rheumatology. It was recognized that the introduction of clinical governance within the NHS had created a need for the collation of data on activity, contemporary treatment patterns and outcomes at national, as well as at local levels. Such data were required to facilitate planning and provision of healthcare for RA patients, and to inform the development of appropriate and realistic standards against which future activity could be audited.

ERAN had similar aims, design, and clinical assessments and outcome measures as ERAS, which ceased recruiting in 1999, but in a wider geographical area that included more centers. ERAN also had the intent to contribute to the development of good clinical practice, guidelines and clinical governance issues, and the added facility to conduct nested studies. ERAN recruited patients between 2002 and 2012 from 23 UK centers and was the first to report the variations in clinical management in the UK in the biologic era [30]. Follow-up data from this cohort provided data to support the BSR view that eligibility criteria for the initiation of biologics may have been set too high by NICE guidance [31], and that even patients with 'moderate' disease do badly within the first few years [32].

ERAS and ERAN are the only RA cohorts that reflect clinical practice in different regions of the UK prior to and during the biologic era, with data on 5–20-year outcomes. Strengths include a rapid reporting system of feedback loops with participating centers, allowing review of individual performance and comparison with national guidelines, and in some centers there was evidence of change in practice [33]. Recruitment fluctuated at certain times because some ERAN centers were not always able to recruit sequential patients, and some centers opted to stop recruiting new patients once a critical mass was achieved for that center in order to concentrate resources on follow-up. This is not an uncommon event in observational studies. Occurance of missing data was generally low and acceptable, but highest in drug start and stop dates (10%).

Combining the two cohorts has allowed examination of secular change in management of RA from 1986 to the present. This analysis has demonstrated several important trends: the earlier use of DMARDs once referred into secondary care; changes to the recommended practice of more intensive approaches in drug therapies in early RA in the UK has been slower than generally perceived and expected [33]; with improving therapies there has been the expected decline in orthopedic interventions, but only in reconstructive surgery of hands and feet, and not in large joint replacement arthroplasties (mainly hip and knees) [34].

Summary of results from UK inception cohorts

The early hospital-based inception cohort studies of the 1980s in the UK, Sweden and The Netherlands broadened the spectrum of RA following the pioneering population-based research in the UK of Kellgren and Lawrence in the 1950s and explained the discrepancy between the two approaches. These cohorts have provided valuable information on the natural (but treated) history of early RA and insights into the etiology, pathogenesis and outcomes of RA. They identified the significant proportion of patients who exhibited serious complications of RA at early stages of disease, not necessarily in those with conventional clinical features of moderate-to-severe disease. Measures of function (Health Assessment Questionnaire and work disability), structural damage (serial x-rays and orthopedic surgery) and morbidity and mortality have now become standard outcomes in RCTs, registries and cohort studies. Some of the subsequent north European cohorts have been multicenter with larger sample sizes, allowing subgroup analysis and, most importantly, had more complete and longer follow-up. They confirmed, refined and widened the earlier findings, highlighting the importance of both standardized and validated assessments, and the inclusion of all important outcome measures [35].

Several studies have provided important information on clinical effectiveness of disease-modifying therapies over time in the real world, which RCTs cannot do, and also on treatment variations, both regionally and between countries. Identifying the optimal management of RA still remains the most important challenge for clinical rheumatology. The BSR has used these data to inform published guidelines on management of RA in the 21st century [36,37].

At present optimal management depends on the use of best clinical practice/guidelines currently available, translating results of research studies into routine clinical practice, and identifying patients with poor prognostic factors and inadequate responses to initial therapies. The main UK cohorts identified the degree of delay from, first, onset of symptoms to hospital referral and, second, to start of disease-modifying therapies, and the possibility of improving on this. In the 1960–1970s patients were managed with NSAIDs for up to 2 years, often by primary care physicians, and slow-acting disease-modifying drugs only started in secondary care once erosions had developed. During this interval it was now postulated that the optimal window of opportunity to treat RA inflammation may be lost. By the 1980s, rheumatologists from the USA were advocating earlier intervention, based on clinical experience and small prospective studies [38]. Referral times into secondary care in the UK have improved only minimally over the last 20 years, the main delay being patient self-referral to primary care. Time to initiation of disease-modifying therapies once in secondary care has improved (from a median of 2 months to less than a month), although use of intensive therapies at outset was lower than expected [33].

The National Audit Office (NAO; London, UK) used extensive data from ERAS, ERAN and NOAR in their report on RA, commissioned by parliament in 2008 that highlighted these variations in current clinical practice and the disappointing outcomes, a key resource used in the HM Government Public Accounts Committee (tenth report published in 2010) [39]. Extensive use was made of the NAO report in subsequent NICE guidelines [103]. The profile of RA had certainly been raised by the NAO report [104], along with the availability and health economic issues surrounding the expensive biologics, ultimately to the benefit of patients. NICE guidelines for eligibility for funding of biologics include specific clinical criteria that are based on health economic analysis as well as clinical evidence [105] and are more stringent compared with the rest of Europe [31]. These issues will need further exploration in UK cohorts and registries, as long as they continue and provide contemporary data.

The current therapeutic 'treat-to-target' strategies for RA have demonstrated improvements in radiological change [40], but need to show consistent improvements in two other important outcomes: function and mortality. Age, sex and functional ability are the most consistent risk factors reported for mortality in these RA studies, followed by RF and acute phase response [22,29], supporting other evidence that the inflammatory process itself may play an important role in the development of ischemic heart disease. This is an important area as it raises the possibility of specific interventions to reduce mortality in RA. The beneficial effects of the more intensive therapies on mortality in RA are not yet proven, but these studies highlight the need for rheumatologists to treat RA patients with active disease early and effectively, identify those at risk from coexisting conditions and treat them actively or with preventative measures accordingly [41]. Pulmonary fibrosis is a well-recognized extra-articular feature of RA and poses an uncommon but severe risk, since this condition was responsible for 6% of deaths in one cohort [22]. This report prompted the formation of a register of RA-associated interstitial lung disease by rheumatologists in the UK to explore the course and predictive factors of this condition and possible therapeutic approaches to improve the poor prognosis [42].

UK product registries

The BSR-BR was set up to register all patients with RA newly starting biologic therapy from January 2002 [14]. The project includes a comparison cohort of RA patients treated with standard DMARDs. The registry records basic demographic characteristics including disease duration, function, DAS and associated European League for Arthritis and Rheumatism response, adverse events and quality-of-life scores, at baseline and at 6-monthly intervals for 3 years [43]. It now has data on several thousands of patients, estimated at 80% of patients starting DMARDs. The project was powered to detect a twofold increased risk in lymphoma, its primary aim, as RA patients already carry an increased

risk of lymphomas, thought to be linked to the abnormal immune system in RA. Biologics have a profound effect on immune mechanisms, so it has been reassuring that BSR-BR found no increased risk, although nonmelanoma skin cancers and opportunistic infections such as tuberculosis were increased [44]. The registry has been expanded to include patients with psoriatic arthritis and ankylosing spondylitis on biologics.

The results of BSR-BR and other national biologic registers in Europe have provided important evidence for evaluations of not only efficacy and toxicity, but also regional variations in the access and use of biologics. The BSR-BR has been able to answer its primary aim concerning the risk of lymphoproliferative conditions because this study had a control group of biologic-naive RA patients treated with conventional DMARDs. After more than 10 years of widespread use, it is still debated whether treatment with biologic agents is associated with an increase in solid tumor cancer incidence, and longer follow-up linked to independent cancer registers is needed. A meta-analysis reported no overall increase compared with nonbiologic treated RA [45]. The main drawback of biologic registries is expense due to labor-intensive data entry methodology and processes to minimize nonmissing data sets. One solution is to improve operational data capture methods by convincing clinicians and/or health professionals to engage at this level by recording data at the point of clinical contact. Another limitation of biologic registries is in the interpretation of drug efficacy, owing to nonrandomization of therapies. Misleading results may arise from channeling bias and confounding by indication, as well as variations in data quality of influences such as comorbidity.

The National Joint Registry (NJR) of England and Wales was established in 2002 to monitor, define, improve and maintain the quality of care of individuals receiving hip, knee and ankle joint replacement surgery across the NHS and the independent healthcare sector [106]. In 2008, the NJR was incorporated into the National Clinical Audit and Patients' Outcomes Program, both managed by the Healthcare Quality Improvement Partnership. The NJR is funded through a levy raised on the sale of hip, knee and ankle replacement implants available and used in the NHS and independent healthcare sectors across England and Wales.

■ Data linkage

The association between biologic-treated patients and lymphoma was made possible following data linkage with national databases, including the Cancer Registries, and mortality data from the Medical Research Information Service [107]. A number of other national databases in the UK provide invaluable data for cohort studies linked by NHS numbers. These include NHS hospital-based interventions from hospital episode statistics and the NJR [108], the General Practice Research Database, a primary care research databank started in 1987 covering 7% of the UK population (a proportion that is increasing) [46], and the NJR [106].

Several other important findings have resulted from linkage between separate and often quite different databases. For example, linkage between two unique registries, the NOAR and EPIC databases, revealed possible links between diet and the development of inflammatory arthritis in Norfolk [47]. Another example was the high incidence of pancreatic cancer in patients with RA exposed to leflunomide, which was observed in the German biologics register [48]. A concerted analysis with the national biologics registers in the UK and Sweden was performed, and the results of the replication analyses did not support the original finding [48]. Orthopedic surgery in RA is considered a surrogate marker for structural damage, not normally measured in large joints in cohort studies, and linkage of two consecutive RA inception cohorts with Hospital Episode Statistics and NJR has allowed an analysis of the frequency of and prognostic factors for this outcome. ERAS and ERAN cover the management of RA in the UK from 1986 to 2012 and demonstrated that over 20 years only certain types of orthopedic surgical rates have declined over this time [34]. The reasons for this are speculative, but highlight the problem associated with the interpretation of other long-term outcomes that are subject to variable and both modifiable and nonmodifiable influences, for example, mortality and work disability.

Other UK RA cohorts & registries recently initiated and/or that have limited publications

■ BRAGGS

Prospective cohort of RA patients from 49 clinical and academic centers in England designed

to collect clinical information and biologic and genetic samples from patients being treated with biologic drugs, in order to investigate treatment response predictors [109].

GORA

Genetic study in Sheffield (UK) that started in 1938, currently comparing genetic samples of established RA patients with controls [110].

Yorkshire Early Arthritis Register

An inception cohort for AIA (American College of Rheumatology criteria) followed yearly was set up in Leeds (UK) to cover NHS hospitals in Yorkshire, from 1998 to 2003 (Year B; 14 centers) and 2000–2009 (Year C; eight centers) [49].

Clinical Audit of RA

Initiated in Scotland in 2005 in eight centers to monitor management of early RA mainly as an audit exercise, from 2005–2008 [111].

RA methotrexate study

A recent initiative to collect and monitor data on physician diagnosis of RA or early undifferentiated polyarthritis, about to start methotrexate as monotherapy or in combination with other DMARDs for the first time. Included 32 centers between 2008 and 2013 [112].

RA-MAP

A recent initiative to identify predictors of remission in RA using patient-level data from patients who were either in the placebo arm of recently published RCTs examining nonbiologic DMARDs or biological agents, or in a longitudinal observational cohort in the UK designed to generate a model to predict remission. Sponsored by the Association of the British Pharmaceutical Industry (London, UK) and Arthritis Research UK (Derbyshire, UK). Includes a literature review; inventory of all RCTs and long-term observational studies in patients with RA that have remission among their outcome measures; and a survey to map cohort characteristics (size, entry criteria, baseline data, duration of follow-up, clinical and other end points, comparator and control/placebo treatments) [113].

MATURA consortium

A culmination of several separate initiatives with the primary aim to identify biomarkers (to include genetic and genomic tissue responses) in order to stratify medicines for RA to enable patients to be treated with the drug they are most likely to respond to earlier in their disease course. It builds on and complements other Medical Research Council-funded cohorts including the Pathobiology of Early Arthritis Cohort [114], BRAGGS [109] and RA-MAP Consortia [113], and represents an ideal opportunity for rheumatologists to collaborate on several strategic national and international initiatives in 2013. It underscores the cooperative philosophy, aims and values in developing a national arthritis network, similar to the cancer network that enables high-quality research to be translated into patient benefit [113].

ARUK INBANK

The ARUK INBANK initiative will provide the musculoskeletal research community with web-based software for standardized clinical data collection (with linked biosamples) to facilitate rapid and efficient acquisition and sharing of high-quality data for research across a network of UK NHS collection centers from 2014. It is proposed to support collection of data once and data reuse for multiple purposes, including research, clinical management, clinical audit and so on. Several musculoskeletal disease areas are proposed: the first and exemplar initiative is the AIA Hub (to include RA), and subsequent Hub developments will include other subspecialties in rheumatology [115].

Non-UK longitudinal cohorts & registries

Several north European countries and the USA initiated inception cohorts in the late 1980s and 1990s with similar designs and aims as those in the UK. This article cannot do justice to the importance of these numerous studies in detail owing to limited space. However, these cohorts reported on the same outcomes and prognostic factors in the same timeframe as already discussed, expanding at national levels on similar findings. Some outcomes are less easy to compare owing to socioeconomic variations, for example those studies that collected data on work disability, while others have exposed significant differences concerning management issues. It became clear that several European countries were treating early RA with more intensive combination drug strategies and biologics than the UK. Most other European

countries now also have biologic registries, mainly in northern Europe, with many based on the BSR-BR model. The accumulated evidence from European registries confirms the relative safety of biologics in the long term, and that the initial fears of malignancy have not been justified, although risk of infection is higher than expected.

Conclusion & future perspective

The disease and product registries and cohort studies in RA in the UK have generally been successful and have largely achieved their initial declared aims, outlined in the rationale list shown earlier in the opening section. The extent to which cohort studies have benefited the rheumatology community is detailed elsewhere [33,50], most importantly by identifying 'early RA' as a critical phase of the disease, which has driven the agenda for improved management, and by defining important outcomes.

The impressive results from registries and cohort studies have readdressed the previous imbalance perceived between observational studies and RCTs [7]. RCTs are limited by inclusion and exclusion criteria and, owing to this selection bias, do not reflect the 'real' world, one of strengths of registries and cohort studies. Biologic registries may suffer from 'channeling bias' because inclusion is often based on the more severe end of the disease spectrum. Well-designed inception cohorts that recruit all consecutive patients based on specific diagnostic criteria minimize bias, but do depend on low attrition rates, which should be accounted for, at least by notifications of mortality. The latter is possible in the UK due to linkage to the National Death Register. Missing data is a common issue that should be reported and to some extent can be addressed with modern statistical methods.

An encouraging development is the much greater extent of meaningful collaborations than previously seen between clinicians involved in the current registries and inception cohorts in both the UK [51] and Europe [52] in order to validate and strengthen problem areas of research findings, clearly the future for rheumatology research. Despite the large numbers in the inception cohorts described, the development of powerful prognostic factors has been disappointing, lacking the robustness needed for routine standard practice. In fact, it is uncertain how the predictive factors rheumatologist have at present are being used. Genetic analysis in particular needs large numbers, and is only available by combining cohorts. Great faith is being put into the development of biomarkers to replicate the models used in cancer.

After several years of planning, a number of national initiatives involving registries/cohorts in the management of AIA and RA are at last coming to fruition, underpinned by collaborations between key consortia with sound track records, including genomics [109], pathobiology [114], clinical trials for remission with pooled clinical trial datasets [113], a stratified medicines project [116]. ARUK has invested in the development of the national platform of clinical data in rheumatology, INBANK, linked to a central archive of stored biological samples and to NHS data from both primary and secondary care, for the acquisition and sharing of high-quality data for research across the NHS [115]. The pilot is planned to commence in 2014 and will bring together all the national cohorts, registries and current initiatives in RA described above.

The initiatives described here in rheumatology could promote the long-held and laudable view of the National Institute of Health Research to offer all patients the opportunity to be involved in research if appropriate, and patients inputting their own data is not far away.

Rheumatologists have been in the forefront of observational research, and operational data capture methodology has been developed as bespoke systems in some rheumatology departments, but as yet not on a national basis. This will be needed in order to accommodate the many new and established initiatives described here. Convincing clinicians to become involved at this level is a major challenge. The future brings multipurpose, user-friendly and cost-effective databases that will accommodate the needs of busy clinicians, clinical audit requirements, research, quality standards and commissioners.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:

- ■ of interest
- ■■ of considerable interest

1 Combe B, Landewe R, Lucas C *et al.* EULAR recommendations for the management of early arthritis: report of the Standing Committee of ESCISIT. *Ann. Rheum. Dis.* 66(1), 34–45 (2007).

2 Kellgren JH, Lawrence JS. Rheumatoid arthritis in a population sample. *Ann. Rheum. Dis.* 15, 1–11 (1956).

3 Lawrence JS, Bennett PH. Benign polyarthritis. *Ann. Rheum. Dis.* 19, 20–30 (1960).

4 Mikkelsen WM, Dodge H. A four year follow-up of suspected rheumatoid arthritis: the Tecumseh, Michigan, community health study. *Arthritis Rheum.* 12, 87–91 (1969).

5 Jacoby RK, Jayson MIV, Cosh JA. Onset, early stages, and prognosis of rheumatoid arthritis: a clinical study of 100 patients with 11-year follow-up. *Br. Med. J.* 2, 96–100 (1973).

6 Fleming A, Crown JM, Corbett M. Early rheumatoid disease: I. Onset. *Ann. Rheum. Dis.* 35, 357–360 (1976).

7 Barton S. Which clinical studies provide best evidence? *Br. Med. J.* 321, 255–256 (2000).

8 Arnett FC, Edworthy SM, Bloch DA *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* 31, 315–312 (1988).

9 Felson DT, Anderson JJ, Boers M *et al.* American College of Rheumatology core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum.* 36, 729–740 (1993).

10 Felson DT, Aderson JJ, Boers M *et al.* American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum.* 38, 727–735 (1995).

11 Wolfe F, Lassere M, van der Heijde D *et al.* Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology. *J. Rheumatol.* 26, 484–489 (1999).

12 van der Heijde D, Jacobs JW, Bijlsma JW *et al.* The effectiveness of early treatment with 'second-line' anti-rheumatic drugs. A randomized, controlled trial. *Ann. Intern. Med.* 124, 699–707 (1996).

■ **One of the first well-designed randomized controlled trials to examine and demonstrate the value of early intervention in recent-onset rheumatoid arthritis (RA).**

13 Gough A, Young A, Bacon P. Setting up an early inflammatory arthritis clinic. In: *Management of Early Inflammatory Joint Disease*. Emery P (Ed.). Elsevier, Amsterdam, The Netherlands (1992).

14 Silman A, Symmons D, Scott DG, Griffiths I. British Society for Rheumatology Biologics Register. *Ann. Rheum. Dis.* 62(Suppl. 2), ii28–ii29 (2003).

15 Ropes MW, Bennett GA, Cobb S, Jacox RF, Jessar RA. 1958 revision of diagnostic criteria for rheumatoid arthritis. *Bull. Rheum. Dis.* 9, 175–176 (1958).

16 Young A, Corbett M, Winfield J *et al.* A prognostic index for erosive changes in the hands, feet, and cervical spines in early rheumatoid arthritis. *Br. J. Rheumatol.* 27, 94–101 (1988).

17 Brook A, Corbett M. Radiographic changes in early rheumatoid disease. *Ann. Rheum. Dis.* 36, 71–73 (1977).

18 Young A. Short term outcomes in recent onset rheumatoid arthritis. *Br. J. Rheumatol.* 34(Suppl. 2), 79–86 (1995).

19 Young A, Dixey J, Cox N *et al.* How does functional disability in early rheumatoid arthritis affect patients and their lives? Results of 5 year follow up in 781 patients from the Early RA Study (ERAS). *Rheumatology (Oxford)* 39, 603–611 (2000).

20 Young A, Dixey J, Kulininska E *et al.* Which patients with early RA stop working? Results from a 5 year inception cohort of 732 patients. *Ann. Rheum. Dis.* 61, 335–340 (2002).

21 James D, Young A, Kulinskaya E *et al.* Orthopaedic intervention in early rheumatoid arthritis. Occurrence and predictive factors in an inception cohort of 1029 patients followed for 5 years. *Rheumatology (Oxford)* 43, 369–376 (2004).

22 Young A, Koduri G, Batley M, Kulinskaya K, Gough A, Dixey J. Mortality in rheumatoid arthritis. Increased in the early course of disease, in ischaemic heart disease and pulmonary fibrosis. *Rheumatology (Oxford)* 46, 350–357 (2006).

■ **One of the few hospital-based inception cohorts examining mortality rates and cause of death in RA in nine different regions of England, which included rural, urban and inner city socioeconomic variations. Standard mortality rates were increased in the first 7 years of RA.**

23 Young A, van der Heijde DF. Can we predict aggressive disease? In: *Early Rheumatoid Arthritis*. Woolf A, van Reil P (Eds). Elsevier, Amsterdam, The Netherlands, 11, 27–48 (1997).

24 Symmons DP, Silman AJ. The Norfolk Arthritis Register (NOAR). *Clin. Exp. Rheumatol.* 21(5 Suppl. 31), S94–S99 (2003).

25 Symmons DP, Barrett EM, Bankhead CR, Scott DG, Silman AJ. The incidence of rheumatoid arthritis in the United Kingdom: results from the Norfolk Arthritis Register. *Br. J. Rheumatol.* 33, 735–739 (1994).

26 Symmons D, Turner G, Webb R *et al.* The prevalence of rheumatoid arthritis: new estimates for a new century. *Rheumatology (Oxford)* 41, 793–800 (2002).

27 Wiles NJ, Dunn G, Barrett EM, Harrison BJ, Silman AJ, Symmons DP. One year follow up variables predict disability 5 years after presentation with inflammatory polyarthritis with greater accuracy than at baseline. *J. Rheumatol.* 27, 2360–2366 (2000).

28 Cooper NJ, Mugford M, Symmons DP, Barrett EM, Scott DG. Total costs and predictors of costs in individuals with early inflammatory polyarthritis: a community-based prospective study. *Rheumatology (Oxford)* 41(7), 767–774 (2002).

■ **One of few cost-effective studies from RA inception cohorts, and highlighted the importance of community costs.**

29 Goodson NJ, Wiles NJ, Lunt M, Barrett EM, Silman A, Symmons DP. Mortality in early inflammatory polyarthritis. *Arthritis Rheum.* 2010–2019 (2002).

30 Kiely P, Williams R, Walsh D, Young A; for the Early Rheumatoid Arthritis (ERAN). Contemporary patterns of care and disease activity outcome in early rheumatoid arthritis: the ERAN cohort. *Rheumatology* 48, 57–60 (2009).

■ **Only contemporary hospital-based inception cohort in the UK examining current patterns of care in RA in over 20 rheumatology units. The results have informed national bodies in deliberations concerning management of RA in the future.**

31 Deighton C, Hyrich K. Why do the French get much greater access to anti-TNF than the British? Vive la différence? Pas nécessairement. *Rheumatology* 47, 1600–1602 (2008).

■ **Highlighted the difference in eligibility criteria for biologics for RA between the UK and the rest of Europe.**

32 Kiely P, Walsh D, Williams R, Young A. Outcome in RA patients with continued conventional therapy for moderate disease

activity. The Early Rheumatoid Arthritis Network (ERAN). *Rheumatology (Oxford)* 50(5), 926–931 (2011).

■ **Demonstrated that even patients with 'moderate' RA who are therefore ineligible for biologics could have poor outcomes.**

33 Young A, Dixie J, Williams P *et al.* An evaluation of the strengths and weaknesses of a register of newly diagnosed rheumatoid arthritis, 1986–2010. *Rheumatology (Oxford)* 50, 176–183 (2011).

■■ **Summary of two early RA registries in the UK that demonstrated that most therapeutic recommendations for good clinical practice in RA have been followed, but for some, the expected changes over time have been considerably slower than anticipated.**

34 Nikiphorou E, Carpenter L, Kiely P, Walsh D, Dixey J, Young A. Has orthopaedic intervention for rheumatoid arthritis changed in line with combination and anti-TNF therapies? An evaluation of joint surgery rates and prognostic factors in two UK inception cohorts (1986–2011). *Rheumatology* 51(Suppl. 3), iii27 (2012).

35 Curtis JR, Jain A, Askling J *et al.* A comparison of patient characteristics and outcomes in selected european and US rheumatoid arthritis registries. *Semin. Arthritis Rheum.* 40(1), 2–14 (2010).

36 Luqmani S, Hennell C, Estrach F *et al.*; on behalf of the British Society for Rheumatology and British Health Professionals in Rheumatology Standards, Guidelines and Audit Working Group British Society for Rheumatology and British Health Professionals in Rheumatology. Guideline for the management of rheumatoid arthritis (the first 2 years). *Rheumatology* 45, 1167–1169 (2006).

37 Luqmani R, Hennell S, Estrach C *et al.*; on behalf of the British Society for Rheumatology and British Health Professionals in Rheumatology Standards, Guidelines and Audit Working Group. British Society for Rheumatology and British Health Professionals in Rheumatology guideline for the management of rheumatoid arthritis (after the first 2 years). *Rheumatology* 48, 375–384 (2009).

38 Pincus T, Callahan L. Taking mortality in rheumatoid arthritis seriously – predictive markers, socio-economic status and comorbidity. *J. Rheumatol.* 13, 841–845 (1986).

39 National Audit Office. *Services for People with Rheumatoid Arthritis.* The Stationery Office, London, UK (2009).

40 Schoels M, Knevel R, Aletaha D *et al.* Evidence for treating rheumatoid arthritis to target: results of a systematic literature search. *Ann. Rheum. Dis.* 69, 638–643 (2010).

■■ **Provided the evidence to date of the value of targeted treatment schedules in achieving better outcomes in RA.**

41 Pincus T, Gibofsky A, Weinblatt ME. Urgent care and tighter control of rheumatoid arthritis as in diabetes and hypertension: better treatments but a shortage of rheumatologists. *Arthritis Rheum.* 46, 851–854 (2002).

42 Chan E, Saravanan V, Dawson J *et al.* Rheumatoid arthritis related interstitial lung disease – which factors predict its development? *Rheumatology (Oxford)* 52(Suppl. 1), ii75 (2013).

43 Watson K, Symmons D, Griffiths I, Silman A. The British Society for Rheumatology biologics register. *Ann. Rheum. Dis.* 64(Suppl. 4), iv42–iv43 (2005).

44 Dixon WG, Symmons DP, Lunt M, Watson KD, Hyrich KL, Silman AJ. Serious infection following anti-tumor necrosis factor α therapy in patients with rheumatoid arthritis: lessons from interpreting data from observational studies. *Arthritis Rheum.* 56, 2896–2904 (2007).

■ **Careful analysis of the British Society for Rheumatology Biologic Register showing that the occurrence of opportunistic infections was slightly increased in patients treated with biologics compared with controls, the most important being tuberculosis.**

45 Mariette X, Metucci-Cerinic M, Pavelka K *et al.* Malignancies associated with tumour necrosis factor inhibitors in registries and prospective observational studies: a systematic review and meta-analysis. *Ann. Rheum. Dis.* 70, 1895–1904 (2011).

46 Watson DJ, Rhodes T, Guess HA. All-cause mortality and vascular events among patients with rheumatoid arthritis, osteoarthritis, or no arthritis in the UK General Practice Research Database. *J. Rheumatol.* 30, 1196–1202 (2003).

47 Pattison DJ, Symmons DP, Lunt M *et al.* Vitamin C and the risk of developing inflammatory polyarthritis: prospective nested case–control study. *Ann. Rheum. Dis.* 63, 843–847 (2005).

48 Strangfeld A, Hyrich K, Askling J *et al.* Detection and evaluation of a drug safety signal concerning pancreatic cancer: lessons from a joint approach of three European biologics registers. *Rheumatology* 50, 146–115 (2011).

49 Morgan AW, Robinson JI, Conaghan PG *et al.*; UKRAG Consortium; YEAR Consortium. Evaluation of the rheumatoid arthritis susceptibility loci HLA-DRB1, PTPN22, OLIG3/TNFAIP3, STAT4 and TRAF1/C5 in an inception cohort. *Arthritis Res. Ther.* 12(2), R57 (2010).

50 Young A. What have we learnt from early RA cohorts? *Best Pract. Res. Clin. Rheumatol.* 23(1), 3–12 (2009).

■■ **Review and summary of the value of inception cohorts in RA.**

51 Norton S, Fu B, Verstappen SM *et al.* Common trajectories of HAQ disability progression over 15-years in the Early Rheumatoid Arthritis Study and the Norfolk Arthritis Register. *Ann. Rheum. Dis.* 71(Suppl. 3), 508 (2012).

52 van der Woude D, Young A, Jayakumar K *et al.* Prevalence of and predictive factors for sustained DMARD-free remission in RA; results from two large early arthritis cohorts. *Arthritis Rheum.* 60, 2262–2227 (2009).

■ Websites

101 National Rheumatoid Arthritis Society. Early Rheumatoid Arthritis Study. www.nras.org.uk/about_rheumatoid_arthritis/established_disease/new_developments_clinical_research/archive_clinical_research/early_rheumatoid_arthritis_study_eras.aspx

102 Arthritis and Musculoskeletal Alliance. http://arma.uk.net/wp-content/uploads/pdfs/ia06.pdf

103 NICE. www.nice.org.uk/usingguidance/commissioningguides/biologicaltherapies/home.jsp

104 National Audit Office. www.nao.org.uk

105 NICE. Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis. www.nice.org.uk/nicemedia/pdf/ta130guidance.pdf

106 National Joint Registry of England, Wales and Northern Ireland. www.njrcentre.org.uk

107 Medical Research Information Service. http://webarchive.nationalarchives.gov.uk/20080814090357/ic.nhs.uk/our-services/improving-patient-care/medical-research-information-service

108 Hospital Episode Statistics online, 2011. www.hscic.gov.uk/hes

109 University of Manchester. Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate (BRAGGS). www.medicine.manchester.ac.uk/musculoskeletal/research/arc/genetics/pharmacogenetics/braggss

110 The University of Sheffield. www.shef.ac.uk/polopoly_fs/1.123251!/file/GORA-Newsletter.pdf

111 Healthcare improvement Scotland. Clinical Audit of Care in Rheumatoid Arthritis (CARA). www.healthcareimprovementscotland.org/previous_resources/audit_report/cara_rheumatoid_arthritis.aspx

112 RAMS. www.medicine.manchester.ac.uk/musculoskeletal/research/arc/clinicalepidemiology/pharmacoepidemiology

113 TACERA. www.gsttbrc.com/tacera/..%5CTACERA%5CPatInfoDocs/TACERA%20Protocol%20-%20Version%202.0%20-%2029.05.2012.pdf

114 Medical Research Council. Pathobiology of Early Arthritis Cohort (PEAC). www.peac-mrc.mds.qmul.ac.uk

115 Arthritis Research UK. INBANK. www.arthritisresearchuk.org/research/inbank.aspx

116 National Institute for Health Research. Biomedical Research Centres. www.nihr.ac.uk/infrastructure/Pages/infrastructure_biomedical_research_centres.aspx

Appendix F

Published Paper - Rheumatology

Rheumatology Advance Access published March 8, 2016
RHEUMATOLOGY

Original article

doi:10.1093/rheumatology/kew004

Have radiographic progression rates in early rheumatoid arthritis changed? A systematic review and meta-analysis of long-term cohorts

Lewis Carpenter[1], Elena Nikiphorou[2], Rachel Sharpe[1], Sam Norton[3], Kirsten Rennie[1], Frances Bunn[4], David L. Scott[5], Josh Dixey[6] and Adam Young[2,7]

Abstract

Objective. To evaluate, firstly, all published data on baseline and annual progression rates of radiographic damage from all longitudinal observational cohorts, and secondly, the association of standard clinical and laboratory parameters with long-term radiographic joint damage.

Methods. A comprehensive search of the literature from 1975 to 2014, using PubMed, SCOPUS and Cochrane databases, identified a total of 28 studies that investigated long-term radiographic progression, and 41 studies investigating predictors of long-term radiographic progression. This was submitted and approved by PROSPERO in February 2014 (Registration Number: CRD42014007589).

Results. Meta-analysis indicated an overall baseline rate of 2.02%, and a yearly increase of 1.08% of maximum damage. Stratified analysis found that baseline radiographic scores did not differ significantly between cohorts recruiting patients pre- and post-1990 (2.01% *vs* 2.03%; $P > 0.01$); however, the annual rate of progression was significantly reduced in the post-1990 cohorts (0.68% *vs* 1.50%; $P < 0.05$). High levels of acute phase markers, baseline radiographic damage, anti-CCP and RF positivity remain consistently predictive of long-term radiographic joint damage.

Conclusion. Critical changes in treatment practices over the last three decades are likely to explain the reduction in the long-term progression of structural joint damage. Acute phase markers and presence of RF/anti-CCP are strongly associated with increased radiographic progression.

Key words: systematic review, meta-analysis, rheumatoid arthritis, radiographic progression, predictive models

Rheumatology key messages

- Progression of radiographic damage in 1990–2011 is significantly lower compared with 1965-1989 in early RA.
- Acute phase markers and RF/anti-CCP positive RA remain important predictors of erosive disease in RA.
- Longitudinal-studies are needed on whether anti-CCP is superior to RF in predicting radiographic damage in RA.

[1]Centre for Lifespan and Chronic Illness Research, [2]Postgraduate Medicine, University of Hertfordshire, Hatfield, [3]Psychology Department, Institute of Psychiatry, Psychology & Neuroscience, Kings College London, London, [4]Centre for Research in Primary and Community Care, University of Hertfordshire, Hatfield, [5]Department of Rheumatology, Kings College London, London, [6]Department of Rheumatology, New Cross Hospital, Wolverhampton and [7]Rheumatology Department, St Albans City Hospital, St Albans, UK

Submitted 22 May 2015; revised version accepted 11 January 2016

Correspondence to: Adam Young, University of Hertfordshire, 2F420 Health Research Building, College Lane, Hatfield AL10 9AB, UK.
E-mail: adam.young@nhs.net

Introduction

Radiographic damage is an important outcome in observational studies and clinical trials in RA. Chronic synovitis in RA results in irreversible bone and cartilage destruction [1]. Erosions are indicators of failure to control the disease [2, 3] that are associated with increased pain and functional disability [4, 5].

Previous systematic reviews have shown [5, 6] 39–73% of early RA patients to develop one or more erosions in the first 5 years, with radiographic damage progressing at a

CLINICAL SCIENCE

constant rate for the first 20 years of the disease [5]. Subsequent systematic reviews [4, 7] concentrated on specific predictors (functional assessment and disease activity indices) and their relationship with radiological damage. However, to date no review has used quantitative analysis techniques, including meta-analysis, to investigate radiographic progression rates.

As structural damage is irreversible [5, 8], it would be advantageous to identify patients at higher risk of severe damage so their treatment could be tailored earlier on. Predictive modelling is a relevant statistical method to identify factors associated with primary RA outcomes [8, 9]. Previous studies have highlighted relationships between radiographic progression and functional disability [4] and disease activity [7]. Other factors like anti-CCP antibodies and genetic factors have yet to be fully reviewed.

In this systematic review we have evaluated published data on baseline and annual progression rates of radiographic damage from longitudinal observational cohorts, and defined their association with standard clinical and laboratory variables. To date, this is the first review to use appropriate meta-analysis techniques to evaluate both the baseline and annual progression rates of radiographic joint damage scores, as well as the predictive markers identified, for all long-term observation cohort studies.

Methods

A systematic review protocol was developed to ensure the objectives and aims were outlined from the outset. This was approved by PROSPERO in February 2014 (CRD42014007589) (supplementary data, PubMed search section, available at *Rheumatology* Online).

Identifying publications

Publications were identified by computerized searches of PubMed, Cochrane Library (including CENTRAL, CDSR, DARE, HTA) and Scopus. Additional lateral search techniques included checking reference lists, performing key word searches in Google Scholar and using the cited by option in PubMed. Databases were searched from 1 January 1975 to 31 February 2014. The search strategy used key words and MeSH terms on the title/abstract and full text as appropriate.

Inclusion/exclusion criteria

Inclusion criteria to select publications comprised the following: investigated the progression or predictive/prognostic markers of radiographic joint damage; patients had a diagnosis of RA, using validated classification criteria like the EULAR and/or the ACR criteria; baseline assessments occurred no later than 3 years from symptom onset; prospective cohort study design; radiographic follow-up data available for at least 5 years for progression rates, and 3 years for predictive markers; used Larsen or Sharp–van der Heijde (SvdH) method to score radiographic damage; and only publications in English.

Publication screening

One reviewer (L.C.) screened titles/abstracts identified in searches, using the selection criteria to identify potentially relevant papers. A second reviewer (E.N.) independently screened the full text of 10% of all publications identified against agreed inclusion criteria. Agreement was achieved in 97% of cases with disagreements resolved through discussion. Supplementary Fig. S1, available at *Rheumatology* online, shows publications identified, screened and included in this review.

Data extraction

Two reviewers (L.C. and R.S.) extracted data using a predesigned form, piloted to ensure all data necessary were captured. It included cohort name, country of study population, scoring method used, number of patients included, years of recruitment, length of follow-up, sex, mean age, baseline DAS and HAQ scores, proportion of patients on DMARDs, proportion RF positive, number, mean/median and standard deviation/interquartile range of radiographic scores at each follow-up visit, analysis method used, significant and non-significant predictors identified and the effect estimate and 95% CIs. In cases where the raw data were not given in the published paper, the author was contacted to provide these (n = 21).

Quality assessment

Studies were rated using the Downs and Blacks instrument for non-randomized studies of health care interventions [10]. Since the studies did not examine clinical effectiveness, checklist items related to comparative groups (e.g. randomization and blinding procedures) were omitted. One reviewer (L.C.) scored all studies using the amended checklist and another reviewer (R.S.) independently scored 10% of studies drawn at random. Discrepancies between reviewers were discussed and consensus achieved.

Analysis

Means and standard deviations of the Larsen or Sharp score were recorded at each follow-up time for each study. In cases where only a median score was obtained, the median and range was converted into a mean score and standard deviation [11]. To estimate annual rates of change, with standard errors, a linear regression model was conducted with follow-up year as the independent variable. Baseline scores and annual progression rates, with respective standard errors, were transformed into percentage maximum damage for each scoring method [12, 13]. Transformed scores were entered into random effects meta-analysis to calculate pooled effect estimates for both baseline radiographic scores and annual rate of change.

To assess the strength of predictive markers, the regression coefficients and odds ratios (OR), with 95% CIs, were collated. Unadjusted effect estimates were sought. Where these were not reported the adjusted estimates were used. Random effects meta-analysis was used for all models due to the likely high level of

heterogeneity between studies. Analysis used Stata (version 13); significance was assumed at $P < 0.05$.

Heterogeneity

The study entry criteria aimed to include studies as homogeneous as possible to allow appropriate meta-analysis. Heterogeneity between studies was predicted *a priori*, mainly due to differences in when cohorts started and differences in scoring methods. The i^2 statistic for each model was found to be consistently above 80%, and therefore random effects models were used throughout. To investigate possible sources of heterogeneity, scoring method and recruitment year were entered into meta-regression models and were the basis of two separate stratified analyses. Given the low level of studies included in the analysis, the 10 studies were stratified into two recruitment period groups, 1965–89 and 1990–2000. This provided equal groupings for stratified analysis. In addition, this marked a change in the clinical management of RA, were from 1990 the focus moved toward treat-to-target, with more intensive treatment within the first 3 months of disease.

Narrative synthesis of predictive factors

Identified markers were recorded and counted to ascertain common associations with a separate count of significant predictors. Where possible, meta-analysis was used to assess the strength of predictive markers. However, for several predictive markers meta-analysis was not possible as too few studies reported results that could be pooled. When meta-analysis was inappropriate a narrative synthesis of the data was conducted.

Results

Meta-analysis of long-term radiographic progression

Of the 28 studies identified, 10 provided the necessary data for meta-analysis [14–22] (Table 1). Patients were recruited from 1965 to 2000 and follow-up ranged from 5 to 20 years. The number of patients included with baseline radiographic data ranged from 73 to 1121. Four studies used Larsen; six used the SvdH scores. Five recruited patients from 1965 to 1989 and five from 1990 to 2000.

Baseline radiographic score

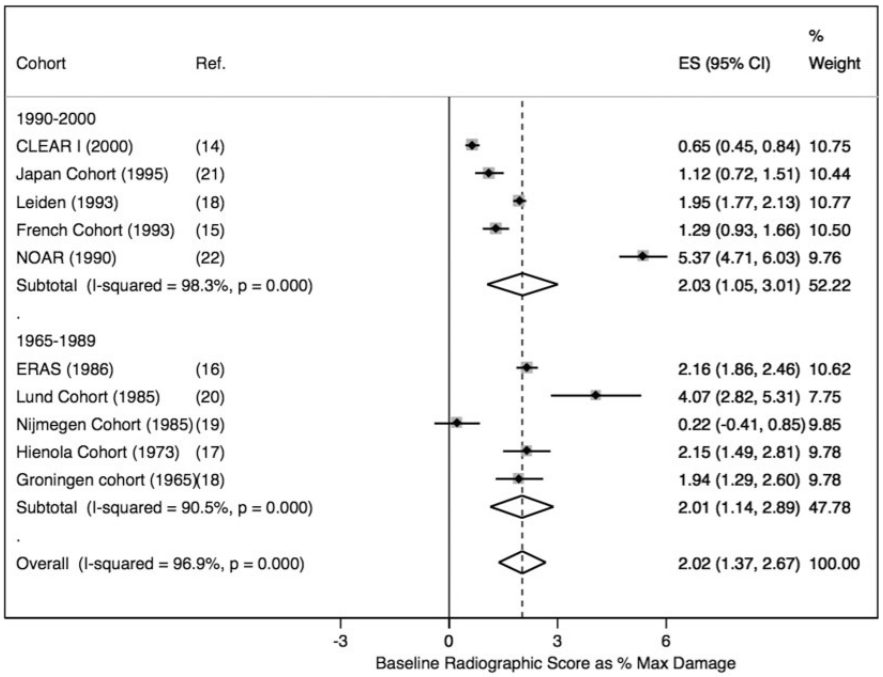
The first analysis examined baseline radiographic score across all studies. The overall rate of damage at baseline was estimated at 2.02% (95% CI: 1.37, 2.67) of maximum damage. The sub-group pooled estimate for Larsen score was 3.41% (95% CI: 1.80, 5.01) of maximum damage (6.82 U); the sub-group pooled estimate for the SvdH score was 1.20% (95% CI: 0.60, 1.80) of maximum damage (5.38 U). Studies recruiting patients between 1965 and 1989 had a sub-group pooled estimate of 2.01% (95% CI: 1.14, 2.89) of maximum damage; studies recruiting between 1990 and 2000

Table 1 Summary of cohorts stratified by recruitment year

Author	Cohort	Country	Scoring method (max)	Sample size	Recruitment year	Years follow-up	% Female	Mean age	% RF+	Radiographic damage	
										Mean baseline (S.D.)	Annual rate (S.E.)
Post-1990											
Bridges *et al.* [14]	CLEAR I	USA	SvdH (448)	357	2000	5	82.4	50	80.1	2.89 (7.65)	1.87 (0.70)
Tanaka *et al.* [21]	Japan Cohort	Japan	SvdH (448)	130	1995	10	69	54	54	5 (10.33)	3 (0.23)
Courvoisier *et al.* [15]	French Cohort	France	SvdH (448)	117	1993	10	80.3	50.4	78.6	5.8 (9)	3.08 (0.42)
Knevel *et al.* [18]	Leiden	Nether-lands	SvdH (448)	678	1993	7	67.4	56.6	57.9	8.74 (10.74)	4.34 (0.11)
Viatte *et al.* [22]	NOAR	UK	Larsen[a] (200)	1446	1990	5	68	56	44	10.74 (13.89)	0.83 (0.61)
Pre-1990											
James *et al.* [16]	ERAS	UK	Larsen[a] (200)	1465	1986	9	66.4	55.3	62.7	4.32 (10.13)	2.44 (0.70)
Kuper *et al.* [19]	Nijmegen Cohort	Nether-lands	SvdH (448)	126	1985	6	64	50	83	1 (16.17)	8 (0.10)
Kapetanovic *et al.* [20]	Lund Cohort	Sweden	Larsen[a] (200)	135	1985	20	62.8	52.1	83	8.13 (1.47)	3.4 (0.31)
Kaarela *et al.* [17]	Hienola Cohort	Finland	Larsen[a] (200)	103	1973	20	68	45	100	4.3 (6.8)	4.12 (0.45)
Knevel *et al.* [18]	Groningen cohort	Nether-lands	SvdH (448)	261	1965	25	67.8	45.1	93.2	3 (56.5)	3.67 (0.50)

[a]Larsen 1977 scoring method. SvdH: Sharp–van der Heijde scoring method.

Fig. 1 Baseline radiographic score pre- and post-1990



Forest plot of baseline radiographic scores stratified by recruitment periods.

reported a sub-group pooled estimate of 2.03% (95% CI: 1.05, 3.01) of maximum damage (Fig. 1).

Annual rate of change

In the second analysis overall annual rate of change was estimated at 1.08% (95% CI: 0.72, 1.44) of maximum damage. The sub-group pooled estimate for Larsen score was 1.38% (95% CI: 1.80, 5.01) of maximum damage (2.76 U/year); the SvdH score was 1.20% (95% CI: 0.88, 1.88) of maximum damage (4.03 U/year). In studies recruiting patients between 1965 and 1989, patients had a sub-group pooled estimate of 1.50% (95% CI: 1.08, 1.92) of maximum damage; for 1990–2000 it was 0.68% (95% CI: 0.47, 0.90) of maximum damage (Fig. 2).

Meta-regression

The small sample size (10 studies) limited the power to conduct meta-regression models with an appropriate number of covariates; however, it was important to investigate possible factors influencing the overall effect estimate given the high levels of heterogeneity between studies (I^2 score ranging from 90.5 to 98.3%).

The meta-regression indicated that there was a statistically non-significant difference for baseline progression rates between recruitment periods ($P > 0.1$), but a statistically significant difference for annual progression rates between recruitment periods ($P < 0.05$), while controlling for scoring method. The models indicated that differences between Larsen and SvdH scoring methods were not statistically significantly different for annual progression rates ($P > 0.1$), suggesting relative increases in either scoring method were comparable. Scoring method was a statistically significant factor for baseline progression rates ($P < 0.05$).

Review of predictive markers of long-term radiographic damage

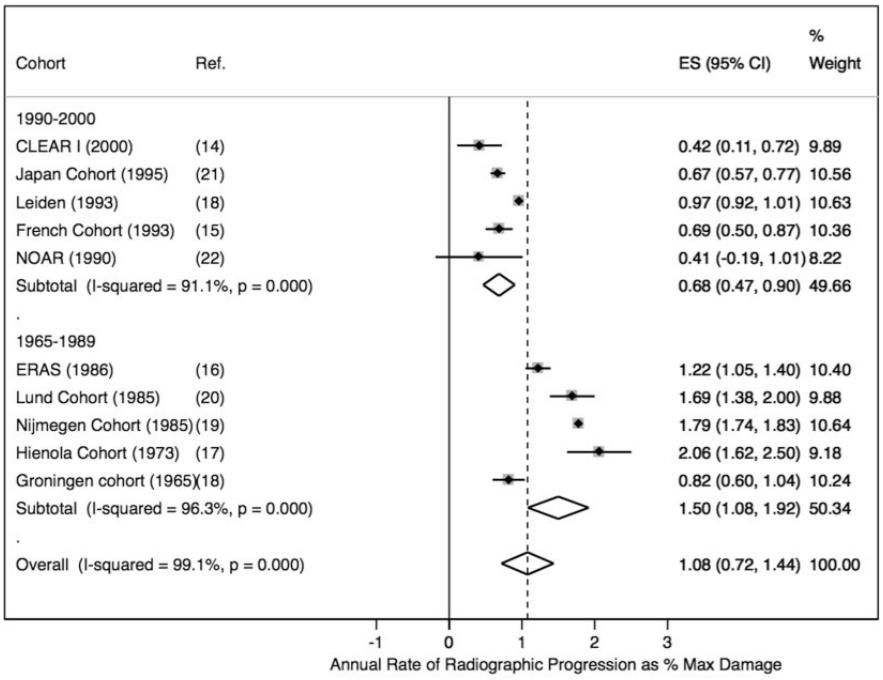
Forty-one papers were identified that examined predictive markers of radiographic joint damage, representing 21 cohort studies. Although several papers were based on the same cohort data (Table 2), the analysis techniques used were sufficiently different from each other to allow their inclusion in the analysis.

Twenty-eight studies used the SvdH method [1, 15, 19, 21, 23–46]; 13 used the Larsen scoring method [20, 47–58]. Twenty-four of 41 studies examined radiographic damage at a single time point, while 17 investigated radiographic damage expressed as a change in score over two time points. Thirteen studies transformed radiographic scores into binary variables and 27 treated the radiographic score as a continuous score. One study treated the radiographic score as an event in a time-to-event analysis [53]. Overall 12 different analysis methods were used (Table 2).

Acute phase markers

Acute phase markers (ESR or CRP) were one of the most frequently reported covariates (Fig. 3). Fifteen studies included the ESR and 13 found it was a statistically significant predictor. Eleven studies included CRP and 10

Fig. 2 Annual rate of radiographic progression pre- and post-1990



Forest plot of annual rates of change stratified by recruitment periods.

found it was a statistically significant predictor. Although there was sufficient data to conduct a meta-analysis, large intra-study differences on how acute phase markers were evaluated made formal meta-analysis inappropriate. While some studies assessed acute phase markers as continuous predictors, others used them as categorical predictors, either using pre-defined cut-points or using quartiles. This made direct comparison between the effect estimates unfeasible.

Courvoisier *et al.* [15] reported that an increased ESR indicated over a 3-fold increased risk of a radiological damage score above the median at 10 years. Similar effect estimates were seen in other studies using similar analysis techniques. An OR of 2.7 (CIs not given) was reported by Fex *et al.* [48] and an OR of 2.9 (95% CI: 1.01, 5.88) was reported by Tanaka *et al.* [21]. Similarly Bukhari *et al.* [23] reported an incidence rate ratio of 2.0 (95% CI: 1.4, 3.0). Using linear regression techniques, Lindqvist *et al.* [51] reported an average increase of 0.42 (95% CI: 0.62, 1.04) units of the Larsen Score for every 1 unit increase in CRP. Mustila *et al.* [52] reported only ESR was significantly associated with radiographic joint damage at 12, 36, 60 and 84 months in univariate analysis, whereas RF was only statistically significant at 36 months, and pANCA, Antikeratin antibodies, antiperinuclear factor and age were not associated at any time.

Anti-cyclic protein antibodies and RF

Anti-cyclic protein antibodies, largely anti-CCP, were evaluated in 16 studies and 14 of these reported statistically significant associations. Using linear regression, Lindqvist *et al.* [51] reported patients positive for anti-CCP had on average an increase of 37 U on the Larsen score compared with anti-CCP negative patients over 10 years. Nyhäll-Wåhlin *et al.* [30] reported an increase of 14.74 over 5 years. Anti-CCP positive patients were also reported to have between a 2.3- and 9.3-fold increase in risk of rapid radiological progression [24, 25].

The predictive role of RF was evaluated in 21 studies and 12 reported statistical significance. Four studies investigating radiographic progression based on low or high radiographic damage groups showed RF positive patients were 1.8–2.8 times more likely to have high rates of long-term radiographic joint damage [21, 23, 24, 55].

To assess the relative strength of anti-CCP and RF, studies reporting OR and 95% CIs were entered into a random effects meta-analysis. Five out of the 13 studies reporting anti-CCP and 10/21 studies reporting RF were included in the meta-analysis. Reasons for exclusion comprised insufficient data, lack of data on measures of variation and no calculated ORs. The overall pooled effect estimate for anti-CCP was 2.49 (95% CI: 1.96, 3.15) and for RF was 2.07 (95% CI: 1.61, 2.65) (Fig. 4). These findings suggest a moderate difference between the two markers, with anti-CCP more strongly associated; but overlapping 95% CIs suggest this difference is statistically non-significant. All five studies included in the meta-analysis for anti-CCP showed an increased risk. Only one reported a statistically non-significant result, which was also the only adjusted effect estimate included [49]. All but two

Table 2 Table of studies investigating predictors of radiographic progression

Author	Year	Sample	Cohort	Scoring method (max)	Max follow-up (years)	Data used f or analysis	Data type	Analysis used	Multivariate
5+ years follow-up									
Bukhari *et al.* [23]	2002	439	NOAR	Larsen[a] (200)	5	Single time point	Continuous	Negative binomial regression	Yes
Courvoisier *et al.* [15]	2008	117	French Cohort	SvdH (448)	10	Single time point	Binary	Logistic regression	Yes
de Rooy *et al.* [24]	2011	676	Leiden Cohort	SvdH (448)	5	Change in score	Continuous	Linear regression	Yes
Fex *et al.* [47]	1997	113	Lund Cohort	Larsen (200)	5	Single time point	Continuous	Linear regression	Yes
Fex *et al.* [48]	1996	113	Lund Cohort	Larsen (200)	5	Change in score	Binary	Logistic regression	Yes
Houseman *et al.* [26]	2012	58	Portsmouth Cohort	SvdH (448)	8	Change in score	Binary	Logistic regression	Yes
Kaltenhauser *et al.* [49]	2007	93	Leipzig Cohort	Larsen (200)	6	Single time points	Binary	Logistic regression	Yes
Kapetanovic *et al.* [20]	2011	183	Lund Cohort	Larsen[a] (200)	10	Single time point	Continuous	Linear regression	Yes
Kraan *et al.* [50]	2004	36	Netherland/Australia Study	Larsen[b] (200)	6	Change in score	Binary	Logistic regression	Yes
Kroot *et al.* [26]	2000	273	Nijmegen Cohort	SvdH (448)	6	Single time point	Continuous	Linear regression	Yes
Kuper *et al.* [19]	1997	157	Nijmegen Cohort	SvdH (448)	6	Single time point	Continuous	Linear regression	Yes
Lindqvist *et al.* [51]	2005	157	Lund Cohort	Larsen[a] (200)	10	Single time points	Continuous	Linear regression	Yes
McQueen *et al.* [27]	2003	31	Auckland Cohort	SvdH (448)	6	Single time point	Continuous	Linear regression	Yes
Meyer *et al.* [28]	2003	156	French Cohort	SvdH (448)	5	Single time points	Binary	Logistic regression	No
Meyer *et al.* [29]	2006	99	French Cohort	SvdH (448)	5	Single time point	Binary	Logistic regression	No
Mustila *et al.* [52]	2000	82	Helsinki Cohort	Larsen[a] (210)	7	Single time points	Binary	Logistic regression	Yes
Nyhall-Wahlin *et al.* [30]	2011	191	BARFOT	SvdH (448)	5	Single time points	Continuous	Linear regression	Yes
Roux-Lombard *et al.* [53]	2001	24	Lund Cohort	Larsen[a] (200)	5	Change in score	Survival	Cox regression model	Yes
Sokka *et al.* [54]	2004	197	Jyväskylä Cohort	Larsen[b] (100)	5	Change in score	Continuous	Quantile regression	Yes
Tanaka *et al.* [21]	2005	114	Japan Cohort	SvdH (448)	10	Change in score	Binary	Logistic regression	Yes
Welsing *et al.* [31]	2004	185	Nijmegen Cohort	SvdH (448)	9	Score at each time point	Continuous	Auto-regressive GEE	Yes
Wolfe *et al.* [32]	1998	256	Wichita Cohort II	SvdH (448)	19	Change in score	Continuous	Linear regression	Yes
3–5 years follow-up									

(continued)

Table 2 Continued

Author	Year	Sample	Cohort	Scoring method (max)	Max follow-up (years)	Data used f or analysis	Data type	Analysis used	Multivariate
Boyesen *et al.* [33]	2010	55	Diakonhjemmet Cohort	SvdH (280)	3	Change in score	Continuous - Square Root	Linear regression	Yes
Combe *et al.* [34]	2001	172	French Cohort	SvdH (448)	3	Single time point	Binary	Logistic regression	Yes
Constantin *et al.* [35]	2002a	96	Rangueil Cohort	SvdH (448)	4	Change in score	Continuous	ANOVA	No
Constantin *et al.* [36]	2002b	96	Rangueil Cohort	SvdH (448)	4	Single time point	Continuous	ANOVA	No
de Vries *et al.* [37]	1993	111	Nijmegen Cohort	SvdH (448)	3	Change in score	Continuous - Square Root	ANOVA	Yes
Dixey *et al.* [55]	2004	866	ERAS Cohort	Larsen[a] (200)	3	Single time point	Binary	Logistic regression	Yes
Gourraud *et al.* [38]	2006	144	Rangueil Midi-Pyrénées Cohort	SvdH (448)	4	Single time points	Continuous	Rank sum tests	No
Kaltenhauser *et al.* [56]	2001	48	Leipzig Cohort	Larsen[b] (200)	4	Change in score at each time point	Continuous	Mixed effects linear regression	Yes
Kuiper *et al.* [39]	2001	332	Nijmegen Cohort	SvdH (448)	3	Single time point	Continuous - Square Root	Linear regression	Yes
Machold *et al.* [57]	2007	55	Austrian Early Arthritis Cohort	Larsen[c] (168)	3	Change in score	Binary	Logistic regression	Yes
Park *et al.* [40]	2010	184	CPR Cohort	SvdH (406)	3	Change in score at each time point	Continuous	Cluster analysis	Yes
Posthumus *et al.* [41]	2000	33	Groningen Cohort	SvdH (448)	3	Change in score	Continuous	Correlation analysis	No
Salaffi *et al.* [42]	2011	48	Italian Cohort	SvdH (448)	3	Change in score	Continuous	Linear regression	Yes
van Aken *et al.* [43]	2004	153	Leiden Cohort	SvdH (448)	4	Change in score	Continuous	Rank sum tests	No
van der Helm-van Mil *et al.* [44]	2005	324	Leiden Cohort	SvdH (448)	4	Single time points	Continuous	Chi-Square	No
van der Helm-van Mil *et al.* [45]	2008	488	Leiden Cohort	SvdH (448)	3	Single time point	Continuous	Linear regression	Yes
van Gaalen *et al.* [46]	2004	268	Leiden Cohort	SvdH (448)	4	Single time point	Continuous	Linear regression	Yes
van Leeuwen *et al.* [1]	1993	110	Groningen/Nijmegen Cohort	SvdH (448)	3	Single time points	Continuous	ANOVA	Yes
Wagner *et al.* [58]	2003	77	Leipzig Cohort	Larsen[b] (200)	4	Single time point	Binary	Logistic regression	Yes

[a]Larsen 1977 scoring method. [b]Larsen 1995 Scoring Method. [c]Larsen–Scott scoring method. SvdH: Sharp–van der Heijde scoring method.

FIG. 3 Number of significant and non-significant predictive factors

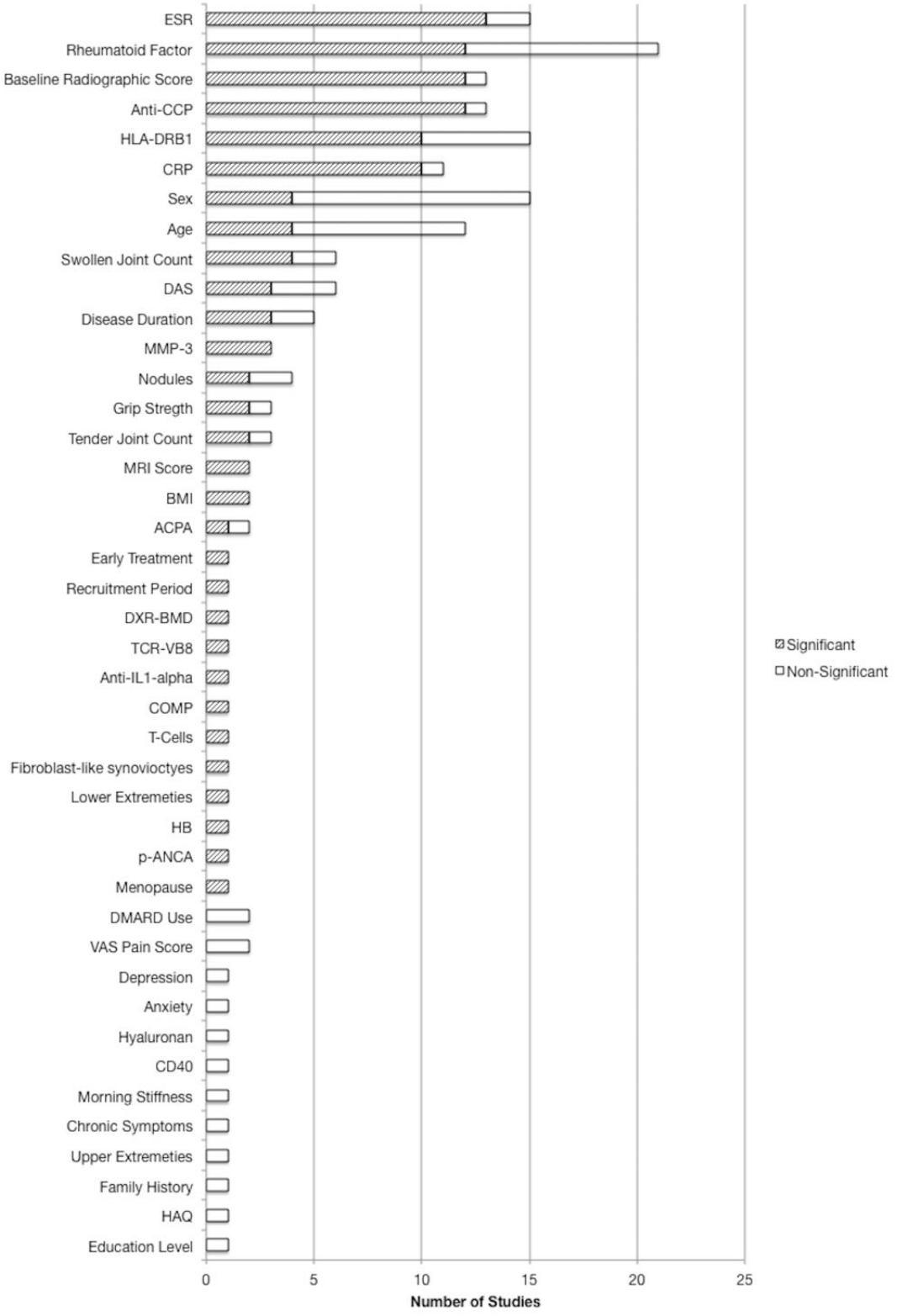
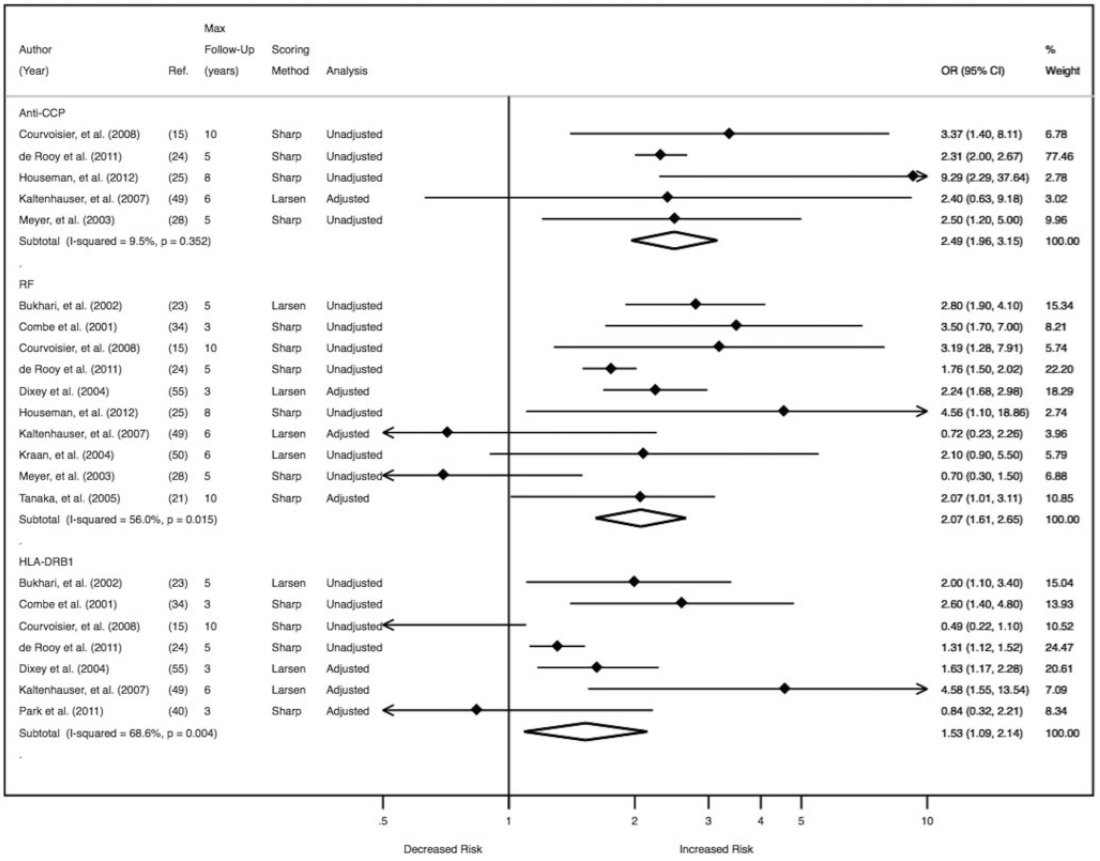


Fig. 4 Forest plot of anti-CCP, RF and HLA-DRB1



studies included in the RF analysis reported an increased risk [28, 49].

Genetic factors

Sixteen studies investigated the influence of genetic factors on radiographic progression and 12 reported statistically significant associations. Four studies used follow-up data of $\geqslant$ 5years; 12 were restricted to 3–4 years follow-up. ORs for the presence of HLA-DRB1 shared epitope (SE) ranged between 1.31 and 2.6 [23, 24, 34]. Two studies by Constantin *et al*. showed HLA-DRB1 was associated with increased radiographic progression over 4 years [35, 36].

Seven of the 16 studies provided sufficient data for meta-analysis. A random effects model showed an overall pooled estimate of 1.53 (95% CI: 1.09, 2.14) (Fig. 4). Two of the seven studies reported a decreased risk [15, 40].

Other factors

There was limited evidence that age and female sex predicted radiographic joint damage. Only 4/12 and 4/15 studies, respectively, reported statistically significant findings. The reported effect sizes of both age and sex were low: age gave 1.14 [24] to 1.2 [23] times increase in risk, while female sex reduced risk by 25% [24]. Few studies evaluated joint counts, DAS, pANCA, MMP-3 and functional disability making it impractical to draw conclusions about their impact on radiographic damage or to undertake meta-analyses.

Quality assessment

All studies were assessed for quality using the Downs and Blacks Quality Assessment Checklist [10] (supplementary Fig. S2 and supplementary Table S1, available at *Rheumatology* Online). Most studies were of good quality. All studies reported clear aims, objectives and outcome measures and recruited representative patients. Only three studies (6%) reported on missing data and only seven (15%) reported on losses to follow-up. The use of appropriate statistical methods was also lacking, particularly in the 3–5 year follow-up predictive studies, where only 13 studies (27%) used appropriate statistical methods.

Discussion

This review is the first to use meta-analysis techniques to provide accurate estimates of overall radiographic

damage at presentation and over a 20-year period in early RA patients. Data from 10 studies shows the overall radiographic damage rate at presentation was 2.02% of maximum damage, and the overall annual progression rate was 1.08% of maximum damage.

Previous reports [5] estimated total annual radiographic progression rates were 1.9% of maximum damage; the Larsen score progressed 3.8 U/year (2.5% maximum damage) and SvdH score progressed 4.3 U/year (1.3% maximum damage) over the first 15 years. The present study found similar rates with an overall progression rate of 1.08% (95% CI: 0.72, 1.44) of maximum damage. Split by scoring method, the Larsen score progressed 2.76 U/year (1.38% maximum damage), and the SvdH score progressed 4.03 U/year (1.20% maximum damage) over the first 20 years of disease. The differences in rates between our findings and previous reports [5] are likely to be multifaceted. Firstly, meta-analytical techniques to calculate pooled effect estimates give different rates from relying on averages. Meta-analysis is a more robust method as larger studies are given a higher weighting, reducing the influence of less precise estimates from smaller studies; it also estimates precision (95% CIs). Secondly, our inclusion criteria focused on observational cohorts of early RA patients. This ensured a more homogeneous study sample as patients in randomised control trials (RCTs) are highly selected with higher levels of disease activity and higher rates of radiographic progression [13, 40]. This review studied patients from true-to-life clinical settings.

Stratifying studies by recruitment year showed annual progression rates in studies recruiting between 1990 and 2000 were more than half the rate reported in studies recruiting between 1965 and 1989. However, baseline radiographic damage was similar across both recruitment periods. The reduction in radiographic progression from 1965 to 2000 is concordant with data from Finckh *et al*. [59], who found decreased progression rates from 1970 to 1990, and Sokka *et al*. [54], who found decreased 5-year radiographic progression rates across three cohorts (1983–85, 88–89 and 1995–96). Finckh *et al.* [59] suggested this was a consequence of more intensive therapies as the temporal effect diminished after controlling for DMARD use. More recent data from RCTs show combinations of synthetic DMARDs and biologics are highly effective in slowing radiographic progression [60], particularly during the window of opportunity [12]. Reduced rates of radiographic progression were also seen in a systematic review of RCTs, where more recent RCTs of patients on MTX had less radiographic progression compared with RCTs conducted earlier [61].

Differences between the two recruitment periods in our review also coincide with changes in clinical management, particularly more intensive treatment in the 1990s with MTX the anchor DMARD [62]. Pincus *et al.* [62] reported that improvements in radiographic outcomes from 1985 to 2000 were associated with better joint scores, functional capacity and mortality outcomes. How much of these changes should be attributed to better treatment strategies, however, remains uncertain due to the non-randomized study designs [54, 62].

Interestingly there is an apparent dearth of new large observational cohort studies of new unselected RA patients. One factor could be the development of national registers of patients treated with biologics, which diverted expertise away from other observational cohorts. Other factors include continuing recruitment to observational studies and less emphasis on collecting radiographic assessments.

The predictive factors we identified are in agreement with previous findings [5] including the importance of acute phase markers and RF positivity. This review also found evidence for the association between anti-CCP positivity and long-term radiological damage. Navarro-Compán *et al*. [7] assessed the relationship between radiographic joint damage and disease activity indices (DAIs) like the DAS. It would appear that while DAIs are clinically useful, the individual components of the DAIs, particularly Swollen Joint Count and acute phase markers, were better predictors.

Our review is the first to summarize associations of anti-CCP and genetic factors with radiographic progression in long-term cohort studies. De Rooy *et al*. [24] found HLA-DRB1 SE increases the risk of radiographic joint damage at 5 years, but they did not include anti-CCP in their models. Recent studies [63, 64] highlight the importance on the dependence of RA-related genetic markers on anti-CCP for associations with radiographic progression. Kaltenhauser *et al.* [49] reported that anti-CCP and DRB1*04 SE, used as a compound marker, was statistically significantly associated with increased radiographic damage at 4 years. However, Kroot *et al.* [26] found anti-CCP but not HLA-DRB4 was statistically significantly associated in multivariate analysis. This evidence suggests an association between SE-positive alleles and anti-CCP antibodies, though the pathogenetic mechanisms remain unclear [49]. Further study of specific HLA-DRB1 haplotypes may show a prognostic role [63]. Currently, genetic markers do not provide much additional prognostic information that can be applied clinically.

Several studies included in our review [28, 46, 49] found RF was not a significant predictor in the presence of anti-CCP, suggesting anti-CCP is the superior marker of long-term radiographic damage. Our meta-analysis suggests that anti-CCP could be more highly associated with increased radiographic damage. However, differences in specific RF antibodies and titre levels may explain variations between studies.

The heterogeneity of the methods and analysis techniques used meant it was impossible to conduct a formal meta-analysis on all predictive markers to allow a direct aggregation of these results. One challenge in comparing studies related to differences in study design [65]. When investigating novel markers in the absence of multivariate methods, the importance of well-established factors like seropositivity and acute phase reactants may not be appropriately considered. Consequently the effect of novel markers may be masked, or over-exaggerated when

already established factors are not considered [9]. Novel markers like MMP-3 [25, 53] have potentially strong associations with radiographic joint damage, but more evidence is needed with large patient samples using appropriate multivariate modelling techniques.

Another limitation is that it was not possible to stratify patients using disease markers like seropositivity when modelling radiographic progression rates, since it would require more detailed and complex data from each cohort, which would be unfeasible to obtain. Consequently, although the review highlighted the potential differences in radiographic progression in patients with anti-CCP positivity, we could not produce separate rates of radiographic progression for seropositive and seronegative RA patients. Furthermore, the direct impact of treatment could not be fully assessed. Evaluating recruitment years provides a surrogate marker of changes in treatment practices, but we could not directly model the effect of treatment. Nevertheless, it is likely patients received standard contemporary care based on published guidelines about treatment regimens from the time they were being studied.

We conclude that the progression of radiographic damage has halved since 1990, with improved treatment providing the most likely cause. RF/anti-CCP, along with increased markers of acute phase reactants, remains strongly associated with radiographic damage, but the value of other novel antibodies needs further study. Finally, while the investigation of different haplotypes is proving hopeful, currently the genetic data are of limited additional prognostic value independent of anti-CCP positivity.

Funding: The paper presents independent research funded by the National Institute for Health Research (NIHR) as one of its Programme Grants For Applied Research (Grant Reference Number: RP-PG-0610-10066). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Disclosure statement: The authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

1 Van Leeuwen MA, van Rijswijk MH, van der Heijde DM, *et al*. The acute-phase response in relation to radiographic progression in early rheumatoid arthritis: a prospective study during the first three years of the disease. Br J Rheumatol 1993;32(Suppl 3):9–13.

2 Landewe R, van de Heijde D. Is radiographic progression a realistic outcome measure in clinical trials with early inflammatory arthritis? Clin Exp Rheumatol 2003;21(Suppl 31):S37–41.

3 Van der Heijde D. Radiographic progression in rheumatoid arthritis: does it reflect outcome? Does it reflect treatment? Ann Rheum Dis 2001;60(Suppl 3):iii47–50.

4 Bombardier C, Barbieri M, Parthan A *et al*. The relationship between joint damage and functional disability in rheumatoid arthritis: a systematic review. Ann Rheum Dis 2012;71:836–44.

5 Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. Clin Exp Rheumatol 2003;21(5 Suppl 31):S20–7.

6 Scott DL, Pugner K, Kaarela K *et al*. The links between joint damage and disability in rheumatoid arthritis. Rheumatology 2000;39:122–32.

7 Navarro-Compán V, Gherghe AM, Smolen JS, Aletaha D, Landewé R, van der Heijde D. Relationship between disease activity indices and their individual components and radiographic progression in RA: a systematic literature review. Rheumatology 2014;54:994–1007.

8 Combe B. Progression in early rheumatoid arthritis. Best Pract Res Clin Rheumatol 2009;23:59–69.

9 Young A, van der Heijde DMFM. Can we predict aggressive disease? Baillieres Clin Rheumatol 1997;11:27–48

10 Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998;52:377–84.

11 Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005;5:13.

12 Finckh A, Liang MH, van Herckenrode CM, de Pablo P. Long-term impact of early treatment on radiographic progression in rheumatoid arthritis: a meta-analysis. Arthritis Care Res 2006;55:864–72.

13 Lassere M. Pooled meta-analysis of radiographic progression: comparison of Sharp and Larsen methods. J Rheumatol 2000;27:269–76.

14 Bridges SL, Causey ZL, Burgos PI *et al*. Radiographic severity of rheumatoid arthritis in African Americans: results from a multicenter observational study. Arthritis Care Res 2010;62:624–31.

15 Courvoisier N, Dougados M, Cantagrel A *et al*. Prognostic factors of 10-year radiographic outcome in early rheumatoid arthritis: a prospective study. Arthritis Res Ther 2008;10:R106.

16 James D, Young A, Kulinskaya E *et al*. Orthopaedic intervention in early rheumatoid arthritis. Occurrence and predictive factors in an inception cohort of 1064 patients followed for 5 years. Rheumatology (Oxford) 2004;43:369–76.

17 Kaarela K, Kautiainen H. Continuous progression of radiological destruction in seropositive rheumatoid arthritis. J Rheumatol 1997;24:1285–7.

18 Knevel R, Kwok KY, de Rooy DP *et al*. Evaluating joint destruction in rheumatoid arthritis: is it necessary to radiograph both hands and feet? Ann Rheum Dis 2013;72:345–9.

19 Kuper HH, van Leeuwen MA, van Riel PL *et al*. Radiographic damage in large joints in early rheumatoid arthritis: relationship with radiographic damage in hands

and feet, disease activity, and physical disability. Br J Rheumatol 1997;36:855–60.

20 Kapetanovic MC, Lindqvist E, Algulin J *et al*. Early changes in bone mineral density measured by digital X-ray radiogrammetry predict up to 20 years radiological outcome in rheumatoid arthritis. Arthritis Res Ther 2011;13:R31.

21 Tanaka N, Sakahashi H, Ishii S, Sato E, Hirose K, Ishima T. Synovial membrane enhancement and bone erosion by magnetic resonance imaging for prediction of radiologic progression in patients with early rheumatoid arthritis. Rheumatol Int 2005;25:103–7.

22 Viatte S, Plant D, Lunt M *et al*. Investigation of rheumatoid arthritis genetic susceptibility markers in the early rheumatoid arthritis study further replicates the TRAF1 association with radiological damage. J Rheumatol 2013;40:144–56.

23 Bukhari M, Lunt M, Harrison BJ, Scott DGI, Symmons DPM, Silman AJ. Rheumatoid factor is the major predictor of increasing severity of radiographic erosions in rheumatoid arthritis: results from the Norfolk Arthritis Register Study, a large inception cohort. Arthritis Rheum 2002;46:906–12.

24 de Rooy DPC, van der Linden MPM, Knevel R, Huizinga TWJ, van der Helm-van Mil AHM. Predicting arthritis outcomes—what can be learned from the Leiden Early Arthritis Clinic? Rheumatology 2011;50:93–100.

25 Houseman M, Potter C, Marshall N *et al*. Baseline serum MMP-3 levels in patients with Rheumatoid Arthritis are still independently predictive of radiographic progression in a longitudinal observational cohort at 8 years follow up. Arthritis Res Ther 2012;14:R30.

26 Kroot E-JJ, De Jong BA, Van Leeuwen MA *et al*. The prognostic value of anti-cyclic citrullinated peptide antibody in patients with recent-onset rheumatoid arthritis. Arthritis Rheum 2000;43:1831–5.

27 McQueen F, Lassere M, Edmonds J *et al*. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Summary of OMERACT 6 MR imaging module. J Rheumatol 2003;30:1387–92.

28 Meyer O, Labarre C, Dougados M *et al*. Anticitrullinated protein/peptide antibody assays in early rheumatoid arthritis for predicting five year radiographic damage. Ann Rheum Dis 2003;62:120–6.

29 Meyer O, Nicaise-Roland P, Santos MD *et al*. Serial determination of cyclic citrullinated peptide autoantibodies predicted five-year radiological outcomes in a prospective cohort of patients with early rheumatoid arthritis. Arthritis Res Ther 2006;8:R40.

30 Nyhäll-Wahlin BM, Turesson C, Jacobsson LTH *et al*. The presence of rheumatoid nodules at early rheumatoid arthritis diagnosis is a sign of extra-articular disease and predicts radiographic progression of joint destruction over 5 years. Scand J Rheumatol 2011;40:1–7.

31 Welsing PMJ, Landewé RBM, van Riel PLCM *et al*. The relationship between disease activity and radiologic progression in patients with rheumatoid arthritis: a longitudinal analysis. Arthritis Rheum 2004;50:2082–93.

32 Wolfe F, Sharp JT. Radiographic outcome of recent-onset rheumatoid arthritis: a 19-year study of radiographic progression. Arthritis Rheum 1998;41:1571–82.

33 Bøyesen P, Haavardsholm EA, Ostergaard M, van der Heijde D, Sesseng S, Kvien TK. MRI in early rheumatoid arthritis: synovitis and bone marrow oedema are independent predictors of subsequent radiographic progression. Ann Rheum Dis 2011;70:428–33.

34 Combe B, Dougados M, Goupille P *et al*. Prognostic factors for radiographic damage in early rheumatoid arthritis: a multiparameter prospective study. Arthritis Rheum 2001;44:1736–43.

35 Constantin A, Lauwers-Cances V, Navaux F *et al*. Collagenase-1 (MMP-1) and HLA-DRB1 gene polymorphisms in rheumatoid arthritis: a prospective longitudinal study. J Rheumatol 2002;29:15–20.

36 Constantin A, Lauwers-Cances V, Navaux F *et al*. Stromelysin 1 (matrix metalloproteinase 3) and HLA-DRB1 gene polymorphisms: Association with severity and progression of rheumatoid arthritis in a prospective study. Arthritis Rheum 2002;46:1754–62.

37 De Vries N, Prinsen CF, Mensink EB, van Riel PL, van't Hof MA, van de Putte LB. A T cell receptor beta chain variable region polymorphism associated with radiographic progression in rheumatoid arthritis. Ann Rheum Dis 1993;52:327–31.

38 Gourraud PA, Boyer JF, Barnetche T *et al*. A new classification of HLA-DRB1 alleles differentiates predisposing and protective alleles for rheumatoid arthritis structural severity. Arthritis Rheum 2006;54:593–9.

39 Kuiper S, van Gestel AM, Swinkels HL, de Boo TM, da Silva JA, van Riel PL. Influence of sex, age, and menopausal state on the course of early rheumatoid arthritis. J Rheumatol 2001;28:1809–16.

40 Park GS, Wong WK, Elashoff DA, Khanna D, Gold RH, Paulus HE. Patterns of radiographic outcomes in early, seropositive rheumatoid arthritis: a baseline analysis. Contemp Clin Trials 2011;32:160–8.

41 Posthumus MD, Limburg PC, Westra J, van Leeuwen MA, van Rijswijk MH. Serum matrix metalloproteinase 3 in early rheumatoid arthritis is correlated with disease activity and radiological progression. J Rheumatol 2000;27:2761–8.

42 Salaffi F, Carotti M, Ciapetti A, Gasparini S, Filippucci E, Grassi W. Relationship between time-integrated disease activity estimated by DAS28-CRP and radiographic progression of anatomical damage in patients with early rheumatoid arthritis. BMC Musculoskelet Disord 2011;12:120.

43 van Aken J, Lard LR, le Cessie S, Hazes JMW, Breedveld FC, Huizinga TWJ. Radiological outcome after four years of early versus delayed treatment strategy in patients with recent onset rheumatoid arthritis. Ann Rheum Dis 2004;63:274–9.

44 Van der Helm-van Mil AHM, Huizinga TWJ, Schreuder GMT, Breedveld FC, de Vries RRP, Toes REM. An independent role of protective HLA class II alleles in rheumatoid arthritis severity and susceptibility. Arthritis Rheum 2005;52:2637–44.

45 Van der Helm-van Mil AHM, van der Kooij SM, Allaart CF, Toes REM, Huizinga TWJ. A high body mass index has a protective effect on the amount of joint destruction in small joints in early rheumatoid arthritis. Ann Rheum Dis 2008;67:769–74.

46 Van Gaalen FA, van Aken J, Huizinga TWJ *et al*. Association between HLA class II genes and autoantibodies to cyclic citrullinated peptides (CCPs) influences the severity of rheumatoid arthritis. Arthritis Rheum 2004;50:2113–21.

47 Fex E, Eberhardt K, Saxne T. Tissue-derived macromolecules and markers of inflammation in serum in early rheumatoid arthritis: relationship to development of joint destruction in hands and feet. Br J Rheumatol 1997;36:1161–5.

48 Fex E, Jonsson K, Johnson U, Eberhardt K. Development of radiographic damage during the first 5-6 yr of rheumatoid arthritis. A prospective follow-up study of a Swedish cohort. Br J Rheumatol 1996;35:1106–15.

49 Kaltenhauser S, Pierer M, Arnold S *et al*. Antibodies against cyclic citrullinated peptide are associated with the DRB1 shared epitope and predict joint erosion in rheumatoid arthritis. Rheumatology (Oxford) 2007;46:100–4.

50 Kraan MC, Haringman JJ, Weedon H *et al*. T cells, fibroblast-like synoviocytes, and granzyme B+ cytotoxic cells are associated with joint damage in patients with recent onset rheumatoid arthritis. Ann Rheum Dis 2004;63:483–8.

51 Lindqvist E, Eberhardt K, Bendtzen K, Heinegård D, Saxne T. Prognostic laboratory markers of joint damage in rheumatoid arthritis. Ann Rheum Dis 2005;64:196–201.

52 Mustila A, Paimela L, Leirisalo-Repo M, Huhtala H, Miettinen A. Antineutrophil cytoplasmic antibodies in patients with early rheumatoid arthritis: an early marker of progressive erosive disease. Arthritis Rheum 2000;43:1371–7.

53 Roux-Lombard P, Eberhardt K, Saxne T, Dayer JM, Wollheim FA. Cytokines, metalloproteinases, their inhibitors and cartilage oligomeric matrix protein: relationship to radiological progression and inflammation in early rheumatoid arthritis. A prospective 5-year study. Rheumatology (Oxford) 2001;40:544–51.

54 Sokka T, Kautiainen H, Häkkinen A, Hannonen P. Radiographic progression is getting milder in patients with early rheumatoid arthritis. Results of 3 cohorts over 5 years. J Rheumatol 2004;31:1073–82.

55 Dixey J, Solymossy C, Young A. Is it possible to predict radiological damage in early rheumatoid arthritis (RA)? A report on the occurrence, progression, and prognostic factors of radiological erosions over the first 3 years in 866 patients from the Early RA Study (ERAS). J Rheumatol Suppl 2004;69:48–54.

56 Kaltenhauser S, Wagner U, Schuster E *et al*. Immunogenetic markers and seropositivity predict radiological progression in early rheumatoid arthritis independent of disease activity. J Rheumatol 2001;28:735–44.

57 Machold KP, Stamm TA, Nell VPK *et al*. Very recent onset rheumatoid arthritis: clinical and serological patient characteristics associated with radiographic progression over the first years of disease. Rheumatology 2007;46:342–9.

58 Wagner U, Kaltenhäuser S, Pierer M *et al*. Prospective analysis of the impact of HLA−DR and −DQ on joint destruction in recent−onset rheumatoid arthritis. Rheumatology 2003;42:553–62.

59 Finckh A, Choi HK, Wolfe F. Progression of radiographic joint damage in different eras: trends towards milder disease in rheumatoid arthritis are attributable to improved treatment. Ann Rheum Dis 2006;65:1192–7.

60 Finckh A, Simard JF, Duryea J *et al*. The effectiveness of anti-tumor necrosis factor therapy in preventing progressive radiographic joint damage in rheumatoid arthritis: a population-based study. Arthritis Rheum 2006;54:54–9.

61 Rahman MU, Buchanan J, Doyle MK *et al*. Changes in patient characteristics in anti-tumour necrosis factor clinical trials for rheumatoid arthritis: results of an analysis of the literature over the past 16 years. Ann Rheum Dis 2011;70:1631–40.

62 Pincus T, Sokka T, Kautiainen H. Patients seen for standard rheumatoid arthritis care have significantly better articular, radiographic, laboratory, and functional status in 2000 than in 1985. Arthritis Rheum 2005;52:1009–19.

63 Viatte S, Plant D, Han B *et al*. Association of HLA-DRB1 haplotypes with rheumatoid arthritis severity, mortality, and treatment response. JAMA 2015;313:1645–56.

64 Van Steenbergen HW, Raychaudhuri S, Rodríguez-Rodríguez L *et al*. Association of valine and leucine at HLA-DRB1 position 11 with radiographic progression in rheumatoid arthritis, independent of the shared epitope alleles but not independent of anti-citrullinated protein antibodies. Arthritis Rheumatol 2015;67:877–86.

65 Van der Heijde DM, van Riel PL, van Rijswijk MH, van de Putte LB. Influence of prognostic features on the final outcome in rheumatoid arthritis: a review of the literature. Semin Arthritis Rheum 1988;17:284–92.

Appendix G

Under Review - Arthritis and Rheumatism

American College of Rheumatology | Arthritis Care & Research

Reductions in radiographic progression in early RA over 25-years: changing contribution from seropositivity in 2 multi-centre UK inception cohorts

Journal:	*Arthritis Care and Research*
Manuscript ID	ACR-16-0745
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	11-Jul-2016
Complete List of Authors:	Carpenter, Lewis; University of Hertfordshire, Centre for Clinical and Health Service Research Norton, Sam; Kings College London, Institute of Psychiatry Nikiphorou, Elena; Whittington Hospital, Department of Rheumatology Jayakumar, Keeran; West Hertfordshire Hospitals NHS Trust, Department of Rheumatology McWilliams, Daniel; University of Nottingham, Academic Rheumatology Rennie, Kirsten; University of Hertfordshire, Centre for Clinical and Health Service Research Dixey, Josh; New Cross Hospital, Rheumatology Kiely, Patrick; St. George's Healthcare NHS Trust, Department of Rheumatology Walsh, David; University of Nottingham, Arthritis Research UK Pain Centre Young, Adam; West Hertfordshire Hospitals NHS Trust, Rheumatology
Key Words: Please choose key words for your manuscript. Choosing the best match of the journal's keywords to your work will likely result in better match to reviewers with expertise in your interest area. It may also help to speed the review process.:	Rheumatoid Arthritis, Radiography, Rheumatoid Factor, Cohort Study, Longitudinal Studies
Optional Terms: You have the option to add additional key words that are not on the journal's list to more specifically categorize your submission.:	

Reductions in radiographic progression in early RA over 25-years: changing contribution from seropositivity in 2 multi-centre UK inception cohorts

Lewis Carpenter[1], Sam Norton[2, 3], Elena Nikiphorou[4], Keeran Jayakumar[5], Daniel F McWilliams[6], Kirsten Rennie[1], Josh Dixey[7], Patrick Kiely[8], David Andrew Walsh[6], Adam Young[5] on behalf of ERAS & ERAN

[1]Centre for Clinical and Health Service Research, University of Hertfordshire, Hatfield, UK

[2]Psychology Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[3]Division of Health & Social Care Research, Faculty of Life and Medical Science, King's College London, London, UK

[4]Department of Rheumatology, Whittington Hospital NHS Trust, London, UK

[5]Early Rheumatoid Arthritis Study, Department of Rheumatology, St Albans City Hospital, St Albans, UK

[6]Arthritis UK Pain Centre, University of Nottingham, Nottingham, UK

[7]Department of Rheumatology, New Cross Hospital, Wolverhampton, UK

[8]Department of Rheumatology, St Georges University Hospitals NHS Foundation Trust, London, UK

Corresponding author:

Mr Lewis Carpenter

2F420, Health Research Building,

College Lane,

University of Hertfordshire,

Hatfield,

AL10 9AB

Email: l.carpenter5@herts.ac.uk

Tel: +447908657562

Word Count: 3137

Objectives: To assess 5-year progression of erosions and Joint Space Narrowing (JSN), and their associations with seropositive status in two large, multi-centre early-RA cohorts spanning 25-years.

Methods: Radiographic joint damage was recorded using the Sharp/van der Heijde (SvdH) method in the Early RA Study (ERAS) 1986-2001, and the Early RA Network (ERAN) 2002-2013. Mixed-effects negative-binomial regression estimated changes in radiographic damage over 5-years, including erosions and JSN separately. Seropositive status, along with age, sex and baseline markers of disease activity were included in the analysis.

Results: A total of 1,216 patients from ERAS and 446 from ERAN had radiographic data. Compared to ERAS, ERAN patients had a lower mean total SvdH score at baseline (ERAN=6.2 vs. ERAS=10.5, p<0.001), and mean annual rate of change (ERAN=2.5 vs. ERAS=6.9 per year, p<0.001). The proportion with progression ≥5 units was 74% for ERAS and 27% for ERAN. Reductions at baseline were largely driven by changes in JSN (ERAS=3.9 vs. ERAN=1.2, p<0.001), rather than erosions (ERAS=1.9 vs. ERAN=0.8, p<0.001). Seropositive status was associated with greater progression in each cohort, but the absolute difference in mean annual rate of change for seropositivity was substantial for ERAS (seropositive=8.6 vs. seronegative=5.1, p<0.001), relative to ERAN (seropositive=2.0 vs. seronegative=1.9, p=0.855).

Conclusion: Radiographic progression has significantly reduced between the two cohorts, with reductions associated with lower baseline damage and other factors, including changes in early DMARD use. The impact of seropositive status as a prognostic marker of clinically meaningful change in radiographic progression has markedly diminished in the context of more modern treatment.

Significance and Innovations

- Radiographic progression at baseline and over the first 5-years has dramatically reduced over the last 25-years.
- Joint space narrowing is the main driver for radiographic reductions early on, with reductions in erosions contributing later in the disease course.
- Seropsotive RA remains an important predictor of increased radiographic damage, however in the context of overall reductions, it is no longer associated with clinically-meaningful changes in radiographic damage.

Published literature has suggested that the incidence of Rheumatoid Arthritis (RA) has declined over the last three decades[1–9]. This corresponds with reports of declines in disease activity[10,11] , functional disability[12,13], orthopaedic surgery[14] and radiographic progression[12,15,16].

While the causal nature of this decline is not entirely clear, it is hypothesised that these declines in disease severity are related to widespread changes in treatment strategies during the 1990s[17]. Data from randomised controlled trials (RCTs) have demonstrated that early initiation of conventional synthetic Disease Modifying Anti-Rheumatic Drugs (DMARDs) can significantly improve patient outcomes, particularly the increased use of methotrexate in combination with other DMARDs[18–21], and indeed biologic DMARDs[22–25].

Radiographic joint damage is often used in RCTs as a primary outcome, and has been shown to be strongly related to levels of functional disability[26] and disease activity[27]. Although commonly expressed as a global score[28], radiographic joint damage comprises of two main components, erosions and joint space narrowing (JSN). While related, they are thought to be the result of two distinct pathophysiological mechanisms[29,30]. Possible causes of erosive joint destruction is the product of invading synovium into the boney structures of the joints, and increased osteoclast activity[31]. Likewise, JSN has been hypothesised to reflect cartilage damage as a result of metalloproteinases, which are upregulated by pro-inflammatory cytokines[32]. JSN is common to a range of pathologies, including osteoarthritis (OA), and is a common comorbid condition in people with RA[33]. Despite this, much of the focus of longitudinal data concerning radiographic damage has reported the combination of these two processes as one composite score[29,30], for example using the radiographic

scoring methods of Ratingen or Larsen, that lack the ability to distinguish progression of erosions and JSN as separate domains[12,15,16].

Further still, seropositive status has been strongly associated with worse radiographic progression[34–36], however, to date no study has looked at whether the relative strength of this association has changed given the wider demographic changes seen in many other aspects of RA, including disease severity. It might be hypothesised that radiographic measures of RA will show significant changes given declines in disease activity, but whether previously demonstrated risk factors for progression continue to be influential remains unclear.

This study therefore aims to investigate long-term radiographic progression by comparing data from two UK, multi-centre inception cohorts, the Early RA Study (ERAS), which collected patient data from 1986-2011, and the Early RA Network (ERAN), which collected data from 2002-2011. Specifically, this study 1) compares the total SvdH, erosion and JSN scores at onset and the rate of progression over the first 5-years between the two cohorts, and 2) estimates the association between seropositive status and radiographic damage at onset and progression over the first 5-years in ERAS and ERAN.

Patients and Methods

The data used for this study were collected from two longitudinal inception cohorts, ERAS and ERAN. ERAS recruited 1,465 patients from 9 centres across the UK between 1986-2001, while ERAN recruited 1,236 patients from 25 centres between 2002 and 2013. Two centres recruited to both cohorts. All patients had a confirmed diagnosis of RA and were recruited within 3 years of symptom onset, typically prior

to conventional DMARD initiation. Maximum follow-up for ERAS was 25 years (median 10 years) and for ERAN was 11 years (median 3 years). All patients were treated based on standard clinical practice of the time.

Standard clinical, laboratory and radiographic data were collected at baseline, 3 to 6 months, 12 months, and then yearly thereafter. These included the original three variable 44-joint Disease Activity Score (DAS) for ERAS and the DAS28 for ERAN, the Health Assessment Questionnaire (HAQ), Rheumatoid Factor Positivity (seropositive RA) and haemoglobin level. To enable comparison of disease activity across the two cohorts, the original DAS in ERAS was converted to DAS28 using the formula DAS28 = (1.072*DAS) + 0.938[37].

Radiographic scoring

Structural joint damage was assessed from plain radiographs using the SvdH scoring method[38]. All 32 centres collected yearly plain x-rays of hands and feet. Radiographs from all 9 centres recruiting from ERAS, and 7/25 (28%) centres from ERAN scored films using the SvdH method.

The SvdH rates radiographic damage based on the prevalence and severity of the erosions in 32 joints in the hands and 12 joints in the feet, and the prevalence and severity of JSN in 30 joints in the hands and 12 joints in the feet. Each joint was rated from 0-5 (or 0-10 for erosions in the joints of the feet) giving a maximum score of 280 for the erosion score and 168 for the JSN score. These scores were combined to give a total SvdH score ranging from 0 to 448.

One person (KJ) scored the radiographs for ERAS, while another person (DMcW) scored the radiographs for ERAN. Each scorer rated the radiographs in chronological

order. To assess agreement between the two, both scored a random sub-sample of thirty-nine radiographs from twenty patients from the ERAS cohort at two time-points (baseline and 5 years). An Intra-class Correlation Coefficient (ICC) of 0.95 (95% confidence intervals 0.90-0.97) was calculated for the erosion score, and 0.98 (95% CI 0.95-0.99) calculated for both the JSN score and total SvdH score. The ICC is an estimate of the proportion of the total variability in ratings for the sample that are due to variability between x-rays, rather variability within x-rays between readers. The high values in our assessment of agreement confirm the risk of systematic bias due to two readers is low, and as such the level of agreement acceptable for the comparison of trends over time.

Statistical analysis

To assess differences in the use of first-line conventional DMARDs between the two cohorts, the cumulative incidence of time to first DMARD within the first 12months from first outpatient appointment was estimated. This was estimated for any DMARD use, as well as separate estimates for the two most commonly used first-line DMARDS, methotrexate and sulphazalasine.

The skewed distributions of radiographic scores derived by the SvdH method renders linear regression inappropriate[39]. Generalised linear regression with a negative-binomial distribution, henceforth negative binomial regression (NBR), was found to achieve best fit to the data, compared with linear and Poisson distributions. Mixed-effects NBR (MENBR) models allowed for the longitudinal structure of the data to be modelled appropriately, whereby random intercept and time slope parameters were estimated. Cohort membership (either ERAS or ERAN) was the

main covariate of interest. Baseline scores, along with yearly measures of SvdH were used in the models to estimate rates at presentation and over the 5-year follow-up. Missing data is inherent in longitudinal studies. To probe potential selection bias based on the availability of radiographs, baseline characteristics of those with and without radiographic data were compared. Furthermore, protecting against confounding due to missing longitudinal data, mixed-effects models use full information maximum likelihood making use of all available data under the *missing at random* assumption, so that all patients with data are included.

Time was defined as years from enrolment and was included as a continuous variable with a random slope to allow for the estimation of the annual rate of progression for each patient. Seropositive RA was the secondary covariate of interest and entered as a main effect, along with a three-way interaction term with cohort and time to allow for progression rates to be estimated separately by seropositivity status for each cohort. Sex, age, DAS28, HAQ, low Hb (<12/13), months from symptom onset to first rheumatology visit, steroid use prior to first assessment and DMARD use within first 12-months were all entered into the model to control for any potential confounding effects.

Exponentiated regression coefficients of an NBR model are incidence rate ratios (IRR), which are interpreted as the relative increase in the log-count of the dependent variable (i.e. the SvdH score) given a one-unit increase in the respective covariate (e.g. age). To aid interpretation, the results from the models were also expressed as an absolute change in the SvdH score using the estimated mean SvdH, along with 95% Confidence Intervals [95% CI]. This allowed for a more direct interpretation of the effect that each factor had in terms of absolute difference in

SvdH units, the percentage of maximum possible damage, and annual progression greater than the minimum clinically important difference of 5 units[40].

These models were estimated separately for the total SvdH score, JSN and erosion score. All analyses were conducted using Stata (version 14; StataCorp LP, USA).

Results

Of the 2,701 total patients recruited, 1,662 had SvdH data: 1,216 from ERAS and 446 from ERAN. The demographic and baseline clinical characteristics of both ERAS and ERAN patients, including only those with radiographic data, are shown in Table 1. Reasons for missing radiographic data included loss of records, unreadable radiographs and loss to follow-up. Patients from ERAS were marginally younger at presentation and had higher DAS28, ESR, HAQ and more likely to be anaemic at baseline. Patient's characteristics with recorded radiographic data were similar to the total number of patients in their respective cohort.

Table 1 - Summary Statistics for each Cohort

Differences in treatment strategies between the two Cohorts

For all DMARDs, ERAS reported a 12-month cumulative incidence of 71.6% [95%CI 69.2-73.8] and for ERAN 95.3% [95%CI 93.9-96.4] (See Figure 1). The 12-month cumulative incidence of sulfasalazine (SSZ) use was higher in ERAS (55% [95%CI 52.4-57.5]) than ERAN (33.1% [95%CI 30.4-35.8]), while methotrexate (MTX) use was substantially lower in ERAS (1.4% [95%CI 0.9-2.1]) compared to ERAN (52.1% [95%CI 49.2-55.0]).

Figure 1. 12-month Cumulative Incidence of DMARD use for ERAS and ERAN

Radiographic progression rates of ERAS and ERAN

For the MENBR analysis a total of 1,508 patients contributing 5,430 observations (mean observations per patient = 3.6) were included. Overall, the ERAN cohort exhibited a 41% lower total SvdH score at baseline compared to ERAS (IRR 0.59 [95%CI 0.50-0.70], p<0.001), along with a 65% slower annual rate of progression over the first 5-years (IRR 0.35 [95%CI 0.24-0.47], p<0.001) (See Figure 2A). The differences in absolute and relative scores for both cohorts are shown in Table 2. When expressed as a proportion of maximum possible damage, the estimated values indicated an increase of 1.5% [95%CI 1.4-1.7] per year for ERAS and 0.6% [95%CI 0.4-0.7] per year for ERAN. The total proportion of patients who had annual progression estimated to be greater than the MCID (≥5 SvdH units) was 74% for ERAS and 27% for ERAN.

Table 2. Mean and relative difference in baseline level and annual rate of progression for Total SvdH, JSN and erosion scores between ERAS and ERAN. Estimates based on fixing the values of the covariates to the sample means. Controlling covariates = age, sex, baseline DAS28, baseline HAQ, low Hb (<12/13) at baseline, months from symptom onset to first rheumatology visit, steroid use prior to first assessment and DMARD use

Similar results were seen for the JSN score, with ERAN participants displaying lower scores at baseline (IRR 0.49 [95%CI 0.41-0.58], p<0.001) and a slower annual rate of progression over the first 5-years compared to ERAS (IRR 0.31 [95%CI 0.21-0.42], p<0.001) (See Figure 2B).

For the erosion score, the score at baseline was similar for both cohorts (IRR 0.94 [95%CI 0.73-1.19], p=0.593), however, ERAN exhibited a slower annual rate of progression over the first 5-years compared to ERAS (IRR 0.43 [95%CI 0.25-0.61],

p<0.001) (See Figure 2C). See Table 2 for absolute and relative changes in both JSN and erosion scores between the two cohorts.

Figure 2 - Progression of A) Total SvdH, B) JSN and C) Erosion score for ERAS and ERAN

Association of seropositivity with radiographic progression in ERAS and ERAN

The absolute and relative difference in total SvdH scores for seropositive and seronegative patients in both cohorts are given in Table 3 and displayed graphically in Figure 3. For the total SvdH score, seropositive RA was not significantly associated with increased radiographic damage at baseline, compared to seronegative RA, in either ERAS or ERAN. Seropositive RA was associated with a 70% increased annual rate of progression, compared to seronegative RA, in ERAS, which was statistically significant. The annual rate of progression for seropositive RA, compared to seronegative RA, in ERAN was increased by 9%, which was not significant. This relates to decreases in the relative impact of seropositive RA on the annual rate of progression of 36% for ERAN compared to ERAS, which although considerable was non-significant (IRR 0.64 [95%CI 0.29-1.07], p=0.224). This related to the proportion of seropositive patients with an annual progression greater than the MCID of 80% for ERAS and just 29% for ERAN.

Investigation of the association between seropositive RA in both the cohorts for the separate JSN and erosion score indicated similar results to the total SvdH (See Supplementary Material 1).

Figure 3 - Progression of Total SvdH score for ERAS and ERAN stratified by seropositivity

Table 3. Mean and relative difference in baseline level and annual rate of progression for Total SvdH based on seropostive (RF+) status between ERAS and ERAN. Estimates based on fixing the values of the covariates to the sample means

Discussion

The findings from the present study indicate that patients with early RA with onset from 2002-2013 (ERAN) had significantly lower baseline and annual rates of radiographic progression compared to those with onset from 1986-2001 (ERAS). Examination of the separate erosion and JSN scores indicate that the reduction in the total SvdH score was largely driven by reductions in JSN. Strikingly, the strong association of seropositivity and increased radiographic progression in the earlier time period (ERAS) was markedly diminished in the later time period (ERAN). The reduction in the impact of seropositive RA was such that those with seropositive status in the ERAN cohort had markedly better radiographic outcomes at 5 years than those with seronegative RA in ERAN.

Previous research has indicated that a change of 5 SvdH units indicates a minimal clinically important difference[40,41], therefore a difference of 5 units per year for ERAN compared to ERAS on total SvdH score observed in this study demonstrates not only a statistically significant change in progression, but also a clinically meaningful reduction. Whereas 74% of patients in the earlier cohort progressed on average ≥5 units per year over the 5-year period of follow-up considered, just 27% of patients in the later cohort did.

Our data extend previous findings of reductions in radiographic damage in RA over recent decades [12,15,16]. There are two plausible explanations for these findings, both of which are likely to contribute to the reduction in radiographic damage over time.

Firstly, RA may have become milder, and secondly, earlier more intensive treatment may have improved disease outcomes. Our models adjusted for disease severity at baseline, but it remains possible that lower rates of progression in the more recent cohort reflect milder disease. This is supported by the observation of lower SvdH scores in ERAN compared to ERAS at baseline, prior to DMARD initiation for the majority of patients. However, the dramatic reductions in radiographic progression, particularly the reduced impact of seropositive RA, is likely to also reflect improvements in the treatment of RA, given the earlier and increased use of methotrexate as the first line DMARD observed in ERAN in this study, which is in line with other reports [12,15,16,36]. Increasing evidence from RCTs also support the hypothesis that early, intensive treatment has an important effect on reducing radiographic progression[42–46].

Separate investigation of the erosion and JSN components of radiographic damage scores showed that JSN was the primary driver for the overall reductions seen in the total SvdH score between the two cohorts. This finding reiterates the importance of reporting both the erosion and JSN score separately in clinical trials. Data from ASPIRE show that more patients with early RA have either erosions alone (8.5%) or JSN alone (4.4%), than both (3.7%) at baseline visit[30], and that JSN may be more strongly associated with irreversible disability[29]. Despite this, the separate scores are still rarely reported[28]. If early treatment with MTX was the primary cause for the reduction in total SvdH in ERAN, this could indicate that the mechanism by which this is achieved is through the reduction of JSN and preservation of the surrounding cartilage. However, what is not clear is whether the JSN is directly attributable to RA JSN, or OA JSN. A high prevalence of radiographic OA has been documented at

baseline in the ERAN cohort in the hands and feet, indicating that high levels of comorbid OA could potentially confound any radiographic assessment of RA[33]. High JSN scores are strongly associated with increased severity of OA osteophytosis and OA JSN[47]. More studies are needed to quantify the exact effect that co-morbid OA could be having on RA radiographic scoring.

Seropositive RA has been consistently associated with increased radiographic damage[35,36]. This study also found that seropositive RA was highly associated with increased radiographic progression. However, when investigating the absolute change in radiographic score between seropositive and seronegative patients across the two cohorts, seropositive patients in the later cohort no longer represented a patient sub-group with clinically meaningful increases in radiographic progression, at least within the first 5-years of disease. Aletaha et al.[48] analysed the effect of seropositive status on radiographic progression and found seropositive patients displayed higher radiographic progression, compared to seronegative patients[39,49]. The estimated change in median SvdH score of 0.6 units per year for seropositive over that of seronegative patients provides an estimate similar to this study.

Many RCTs are restricted to seropositive patients only, and previous research has not focused on the effect of seropositivity in the context of reduced radiographic progression in more recent years. The two long-term observational cohorts examined in this study provide a 'real-world' account of patients typically seen in secondary care, and the high patient numbers over the full 5-year follow-up also provides a unique opportunity to provide precise estimates using the modelling techniques outlined[39]. The use of the SvdH score also provides a first look at the two principle components of radiographic damage, erosions and JSN, in detail. Further

data from observational studies are needed to ascertain whether reductions in radiographic progression have also resulted in the diminished association with seropositive status, particularly in the context of anti-CCP seropositive RA, which could be more predictive of radiographic progression when compared to RF[36,50,51].

Our research is subject to a number of limitations inherent in cohort studies. Recruiting centres were hosted by enthusiastic clinicians within the UK and, although they might not necessarily reflect people with RA in other contents, or subjected to different treatment regimens, the multicentre recruitment for these cohorts from district general hospitals is likely to be representative of people with RA in the UK. Radiographs were not available for all participants, and it is possible that those with more severe disease were more likely to have x-rays, increasing the risk of selection bias in our study. However, baseline variables indicated minimal differences between the whole cohorts, and those for whom radiographic data were available. The impact of such a selection bias would overestimate rates of progression, particularly for ERAN, where data were less complete; hence our estimates should be treated as conservative.

This study provides further evidence into the marked reduction in radiographic damage over the last 30-years, while providing accurate, quantified estimates of the extent of that reduction. JSN was the major driver for the overall reductions seen, and highlights the importance of investigating JSN and erosions separately when investigating radiographic damage. Advances in treatment are likely to be the main cause for the decline, and adequate DMARD treatment might remove the predictive value of seropositivity for radiographic progression in early RA. Further research should seek other predictors and mediators if residual radiographic progression

despite DMARD treatment is to be halted. The impact of these reductions on patients of varying disease severity, and whether these reductions have an impact on improved long-term functional disability will be crucial in fully realising the impact of these results on clinical care.

Declarations

The authors declare no conflicts of interest.

References

1. Doran MF, Pond GR, Crowson CS, O'Fallon WM, Gabriel SE. Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a forty-year period. *Arthritis Rheum*. 2002;46(3):625-631.
2. Odegård S, Kvien TK, Uhlig T. Incidence of clinically important 10-year health status and disease activity levels in population-based cohorts with rheumatoid arthritis. *J Rheumatol*. 2008;35(1):54-60.
3. Silman AJ. Are there secular trends in the occurrence and severity of rheumatoid arthritis? *Scand J Rheumatol Suppl*. 1989;79:25-30.
4. Hochberg MC. Changes in the incidence and prevalence of rheumatoid arthritis in England and Wales, 1970-1982. *Semin Arthritis Rheum*. 1990;19(5):294-302.
5. Spector TD, Hart DJ, Powell RJ. Prevalence of rheumatoid arthritis and rheumatoid factor in women: evidence for a secular decline. *Ann Rheum Dis*. 1993;52(4):254-257.
6. Symmons DPM. Epidemiology of rheumatoid arthritis: Determinants of onset,

persistence and outcome. *Best Pract Res Clin Rheumatol*. 2002;16(5):707-722.

7. Kaipiainen-Seppänen O, Aho K, Isomäki H, Laakso M. Incidence of rheumatoid arthritis in Finland during 1980-1990. *Ann Rheum Dis*. 1996;55(9):608-611.

8. Symmons DP, Barrett EM, Bankhead CR, Scott DG, Silman a J. The incidence of rheumatoid arthritis in the United Kingdom: results from the Norfolk Arthritis Register. *Br J Rheumatol*. 1994;33(8):735-739.

9. Dugowson CE, Koepsell TD, Voigt LF, Bley L, Nelson JL, Daling JR. Rheumatoid arthritis in women. Incidence rates in group health cooperative, Seattle, Washington, 1987-1989. *Arthritis Rheum*. 1991;34(12):1502-1507.

10. Bergström U, Book C, Lindroth Y, Marsal L, Saxne T, Jacobsson L. Lower disease activity and disability in Swedish patients with rheumatoid arthritis in 1995 compared with 1978. *Scand J Rheumatol*. 1999;28(3):160-165.

11. Aho K, Tuomi T, Palosuo T, Kaarela K, Von Essen R, Isomäki H. Is seropositive rheumatoid arthritis becoming less severe? *Clin Exp Rheumatol*. 1989;7(3):287-290.

12. Finckh a, Choi HK, Wolfe F. Progression of radiographic joint damage in different eras: trends towards milder disease in rheumatoid arthritis are attributable to improved treatment. *Ann Rheum Dis*. 2006;65(9):1192-1197.

13. Porter DR, Capell HA, Mcinnes I, et al. Is Rheumatoid Arthritis Becoming a Milder Disease? Or Are We Starting Second-Line Therapy in Patients with Milder Disease? *Rheumatology*. 1996;35(12):1305-1308.

14. Nikiphorou E, Carpenter L, Morris S, et al. Hand & foot surgery rates in RA have declined from 1986-2011, but large joint replacements remain unchanged. Results from two UK inception cohorts. *Arthritis Rheumatol*.

2014:n/a - n/a.

15. Fiehn C, Belke-Voss E, Krause D, Wassenberg S, Rau R. Improved radiological outcome of rheumatoid arthritis: The importance of early treatment with methotrexate in the era of biological drugs. *Clin Rheumatol*. 2013;32(12):1735-1742.

16. Sokka T, Kautiainen H, Hakkinen A, Hannonen P. Radiographic progression is getting milder in patients with early rheumatoid arthritis. Results of 3 cohorts over 5 years. *J Rheumatol*. 2004;31(6):1073-1082.

17. Pincus T, Sokka T, Kautiainen H. Patients seen for standard rheumatoid arthritis care have significantly better articular, radiographic, laboratory, and functional status in 2000 than in 1985. *Arthritis Rheum*. 2005;52(4):1009-1019.

18. Dougados M, Combe B, Cantagrel A, et al. Combination therapy in early rheumatoid arthritis: a randomised, controlled, double blind 52 week clinical trial of sulphasalazine and methotrexate compared with the single components. *Ann Rheum Dis*. 1999;58(4):220-225.

19. Emery P, Breedveld FC, Lemmel EM, et al. A comparison of the efficacy and safety of leflunomide and methotrexate for the treatment of rheumatoid arthritis. *Rheumatology (Oxford)*. 2000;39(6):655-665.

20. Möttönen T, Hannonen P, Leirisalo-Repo M, et al. Comparison of combination therapy with single-drug therapy in early rheumatoid arthritis: a randomised trial. *Lancet*. 1999;353(9164):1568-1573.

21. Choy EH, Scott DL, Kingsley GH, et al. Treating rheumatoid arthritis early with disease modifying drugs reduces joint damage: a randomised double blind

trial of sulphasalazine vs diclofenac sodium. *Clin Exp Rheumatol*. 2002;20(3):351-358.

22. Emery P, Breedveld FC, Hall S, et al. Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. *Lancet*. 2008;372(9636):375-382.
23. Breedveld FC, Weisman MH, Kavanaugh AF, et al. The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previo. *Arthritis Rheum*. 2006;54(1):26-37.
24. St Clair EW, van der Heijde DM, Smolen JS, et al. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. *Arthritis Rheum*. 2004;50(11):3432-3443.
25. Goekoop-Ruiterman YP, Vries-Bouwstra JK, Allaart CF, et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. *Arthritis Rheum*. 2005;52(11):3381-3390.
26. Bombardier C, Barbieri M, Parthan A, et al. The relationship between joint damage and functional disability in rheumatoid arthritis: a systematic review. *Ann Rheum Dis*. 2012;71(6):836-844.
27. Navarro-Compán V, Gherghe AM, Smolen JS, Aletaha D, Landewé R, van der Heijde D. Relationship between disease activity indices and their individual components and radiographic progression in RA: a systematic literature

review. *Rheumatology (Oxford)*. 2014:1-14.

28. Lillegraven S, Van Der Heijde D, Uhlig T, Kvien TK, Haavardsholm EA. What is the clinical relevance of erosions and joint space narrowing in RA? *Nat Rev Rheumatol*. 2012;8(2):117-120.

29. Aletaha D, Funovits J, Smolen JS. Physical disability in rheumatoid arthritis is associated with cartilage damage rather than bone destruction. *Ann Rheum Dis*. 2011;70(5):733-739.

30. Smolen JS, van der Heijde DM, Aletaha D, et al. Progression of radiographic joint damage in rheumatoid arthritis: independence of erosions and joint space narrowing. *Ann Rheum Dis*. 2009;68(10):1535-1540.

31. Redlich K, Hayer S, Ricci R, et al. Osteoclasts are essential for TNF-alpha-mediated joint destruction. *J Clin Invest*. 2002;110(10):1419-1427.

32. Goldring SR. Pathogenesis of bone and cartilage destruction in rheumatoid arthritis. *Rheumatology (Oxford)*. 2003;42 Suppl 2(suppl_2):ii11-ii16.

33. McWilliams D, Michelle M, Jayakumar K, et al. Erosive and osteoarthritic structural progression in early rheumatoid arthritis. *Rheumatol*. 2016;In Press.

34. Scott DL. Prognostic factors in early rheumatoid arthritis. *Rheumatol*. 2000;39 Suppl 1:24-29.

35. Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. *Clin Exp Rheumatol*. 2003;21(5 Suppl 31):S20-S27.

36. Carpenter L, Nikiphorou E, Sharpe R, et al. Have radiographic progression rates in early rheumatoid arthritis changed? A systematic review and meta-analysis of long-term cohorts. *Rheumatology (Oxford)*. 2016:kew004

37. Anderson JK, Zimmerman L, Caplan L, Michaud K. Measures of rheumatoid arthritis disease activity: Patient (PtGA) and Provider (PrGA) Global Assessment of Disease Activity, Disease Activity Score (DAS) and Disease Activity Score With 28-Joint Counts (DAS28), Simplified Disease Activity Index (SDAI), Cl. *Arthritis Care Res (Hoboken)*. 2011;63(S11):S14-S36.

38. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. *J Rheumatol*. 1999;26(3):743-745. http://www.ncbi.nlm.nih.gov/pubmed/10090194. Accessed January 21, 2011.

39. Park GS, Wong WK, Khanna D, Gold RH, Paulus HE. Examining radiographic outcomes over time. *Rheumatol Int*. 2014;34(2):271-279.

40. Bruynesteyn K, van der Heijde DD, Boers M, et al. Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference. *Arthritis Rheum*. 2002;46(4):913-920.

41. Bruynesteyn K, Van Der Heijde D, Boers M, et al. Detecting radiological changes in rheumatoid arthritis that are considered important by clinical experts: influence of reading with or without known sequence. *J Rheumatol*. 2002;29(11):2306-2312.

42. Boers M, Verhoeven AC, Markusse HM, et al. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. *Lancet*. 1997;350(9074):309-318.

43. Landewé RB, Boers M, Verhoeven AC, et al. COBRA combination therapy in

patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. *Arthritis Rheum*. 2002;46(2):347-356.

44. Maillefert JF, Combe B, Goupille P, Cantagrel A, Dougados M. Long term structural effects of combination therapy in patients with early rheumatoid arthritis: five year follow up of a prospective double blind controlled study. *Ann Rheum Dis*. 2003;62(8):764-766.

45. Emery P, Kvien TK, Combe B, et al. Combination etanercept and methotrexate provides better disease control in very early (<=4 months) versus early rheumatoid arthritis (>4 months and <2 years): post hoc analyses from the COMET study. *Ann Rheum Dis*. 2012;71(6):989-992.

46. Smolen JS, Van Der Heijde DMFM, St Clair EW, et al. Predictors of joint damage in patients with early rheumatoid arthritis treated with high-dose methotrexate with or without concomitant infliximab: results from the ASPIRE trial. *Arthritis Rheum*. 2006;54(3):702-710.

47. Khanna D, Ranganath VK, Fitzgerald J, et al. Increased radiographic damage scores at the onset of seropositive rheumatoid arthritis in older patients are associated with osteoarthritis of the hands, but not with more rapid progression of damage. *Arthritis Rheum*. 2005;52(8):2284-2292.

48. Aletaha D, Alasti F, Smolen JS. Rheumatoid factor determines structural progression of rheumatoid arthritis dependent and independent of disease activity. *Ann Rheum Dis*. 2013;72(6):875-880.

49. Landewe R, van der Heijde D. Presentation and analysis of radiographic data in clinical trials and observational studies. *Ann Rheum Dis*. 2005;64 Suppl 4:iv48-iv51.

50. Forslind K, Ahlmén M, Eberhardt K, Hafström I, Svensson B. Prediction of radiological outcome in early rheumatoid arthritis in clinical practice: Role of antibodies to citrullinated peptides (anti-CCP). *Ann Rheum Dis*. 2004;63(9):1090-1095.

51. Vallbracht J, Rieber J, Oppermann M, Förger F, Siebert U, Helmke K. Diagnostic and clinical value of anti-cyclic citrullinated peptide antibodies compared with rheumatoid factor isotypes in rheumatoid arthritis. *Ann Rheum Dis*. 2004;63(9):1079-1084.

Tables and Figures

Table 1 - Summary Statistics for each Cohort

	ERAS		Total Missing (%)	ERAN		Total Missing (%)	**ERAS+ERAN**	
	Total	With SvdH		Total	With SvdH		**Total**	**With SvdH**
Demographics								
Year of Recruitment	1986-2001	1986-2001	0 (0)	2002-2013	2002-2013	0 (0)	1986-2013	1986-2013
Age at Onset (Mean (SD))	55.3 (14.6)	54.9 (14.5)	0 (0)	57.1 (14)	58 (13.5)	0 (0)	56.1 (14.4)	55.7 (14.3)
Female (%)	66	66	0 (0)	68	65	0 (0)	67	65
Clinical Markers								
Seropositive (%)	63	64	9 (0.1)	60	61	142 (11)	62	63
Baseline DAS (Mean (SD))	6.32 (1.33)	6.32 (1.33)	13 (0.1)	4.53 (1.58)	4.5 (1.64)	46 (4)	5.51 (1.7)	5.84 (1.62)
Baseline ESR (Median (IQR))	37 (44)	38 (44)	7 (0.1)	24 (29)	21 (28)	183 (15)	30 (39)	34 (41)
Baseline HAQ (Mean (SD))	1.15 (0.8)	1.15 (0.8)	5 (0.1)	1.08 (0.8)	1.03 (0.8)	37 (3)	1.12 (0.8)	1.12 (0.8)
Baseline Anaemia (%)	41	42	5 (0.1)	28	24	32 (3)	35	37
Months from symptom onset to First Visit (Median (IQR))	6 (7)	6.5 (7)	0 (0)	6 (8)	6 (8)	91 (7)	6 (8)	6 (8)
Observations	1465	1216	1216	1236	446	1216	2701	1662

Numbers represent means (SD), medians (IQR) and proportions were used where indicated.
SvdH = Sharp/van der Heijde, DAS=Disease Activity Score-28, ESR=Erythrocyte Sedimentation Rate, HAQ = Health Assessment Questionnaire.

Table 2. Mean and relative difference in baseline level and annual rate of progression for Total SvdH, JSN and erosion scores between ERAS and ERAN. Estimates based on fixing the values of the covariates to the sample means. Controlling covariates = age, sex, baseline DAS28, baseline HAQ, low Hb (<12/13) at baseline, months from symptom onset to first rheumatology visit, steroid use prior to first assessment and DMARD use

Estimated means	ERAS	ERAN	Absolute Difference	Relative Difference (IRR) [95% CI]	P-Value
Total SvdH at baseline	10.5	6.2	4.3	**0.59 [0.50-0.70]**	**<0.001**
Total SvdH annual rate	6.9	2.5	4.5	**0.35 [0.24-0.47]**	**<0.001**
JSN score at baseline	7.4	3.6	3.8	**0.49 [0.41-0.58]**	**<0.001**
JSN score annual rate	3.9	1.2	2.7	**0.31 [0.21-0.42]**	**<0.001**
Erosion score at baseline	1.8	1.7	0.1	0.94 [0.73-1.19]	0.593
Erosion score annual rate	1.9	0.8	1.1	**0.43 [0.25-0.61]**	**<0.001**

Table 3. Mean and relative difference in baseline level and annual rate of progression for Total SvdH based on seropostive (RF+) status between ERAS and ERAN. Estimates based on fixing the values of the covariates to the sample means

		RF-	RF+	Difference	Relative Difference (IRR) [95% CI]	P-Value
ERAS	Total SvdH at baseline	9.5	11	1.5	1.16 [1.00-1.35]	0.056
	Total SvdH Annual rate	5.1	8.6	3.6	**1.70 [1.42-1.97]**	**<0.001**
ERAN	Total SvdH at baseline	6.0	6.2	0.2	1.04 [0.76-1.42]	0.811
	Total SvdH Annual rate	1.9	2.0	0.2	1.09 [0.51-1.67]	0.855

Figure 1. 12-month Cumulative Incidence of DMARD use for ERAS and ERAN

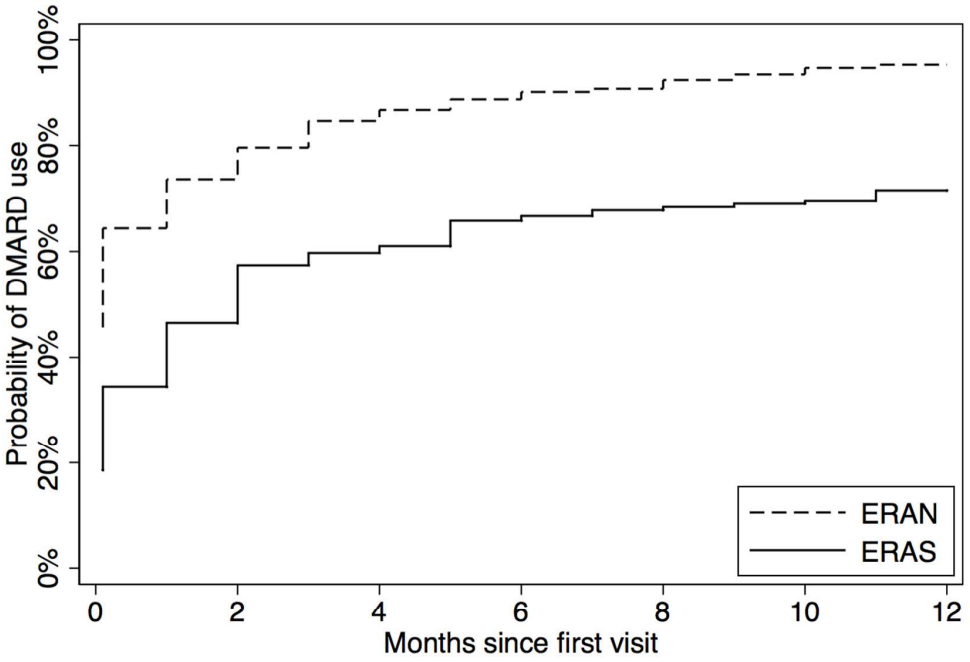


Figure 2 - Progression of A) Total SvdH, B) JSN and C) Erosion score for ERAS and ERAN

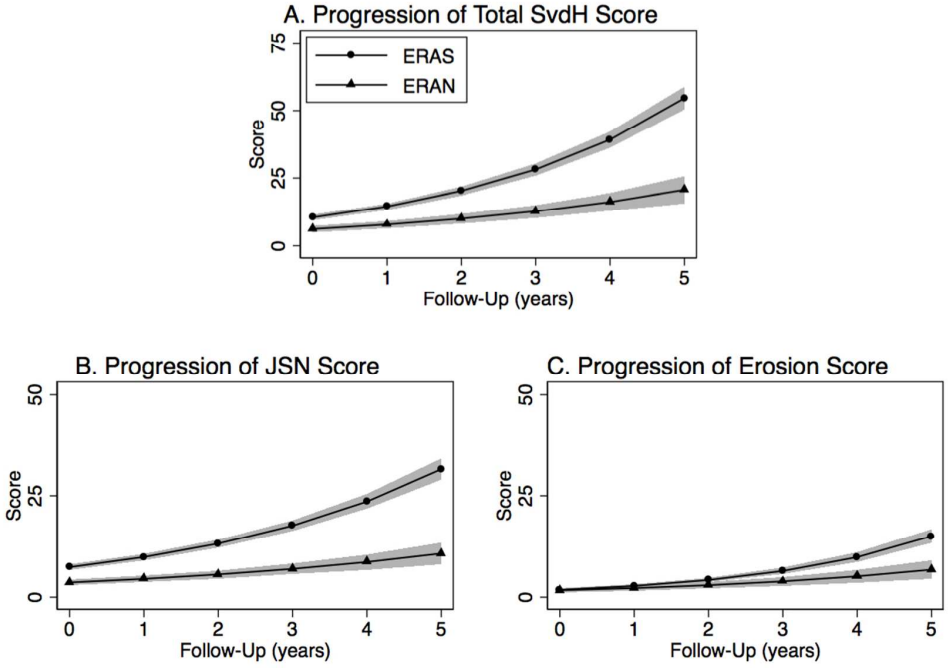
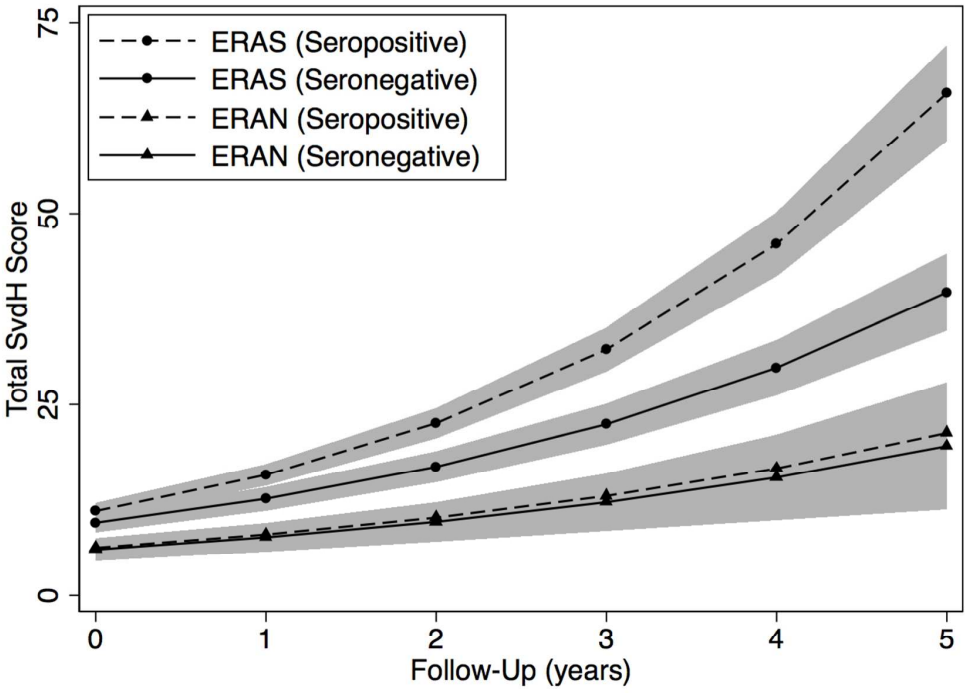


Figure 3 - Progression of Total SvdH score for ERAS and ERAN stratified by seropositivity



Supplementary Material 1

Association of seropositivity with joint space narrowing (JSN) and erosions in ERAS and ERAN

The influence of seropositive RA on the separate joint space narrowing (JSN) compared to seronegative rheumatoid arthritis (RA) indicated a similar pattern to the total Sharp/van der Heijde (SvdH) score (Supplementary Table 1 & Supplementary Figure 1A). Seropositive RA was not associated with increased JSN scores at baseline for either ERAS or ERAN, but was associated with a relative increase in the annual rate of progression. When compared to seronegative RA, seropositive RA indicating an increase of 58% and 19% for ERAS and ERAN respectively. The relative difference in the increased annual rate of progression of JSN scores for seropositive RA in ERAS compared to ERAN was not significantly different (IRR 0.72 [95% CI 0.26-1.18], p=650).

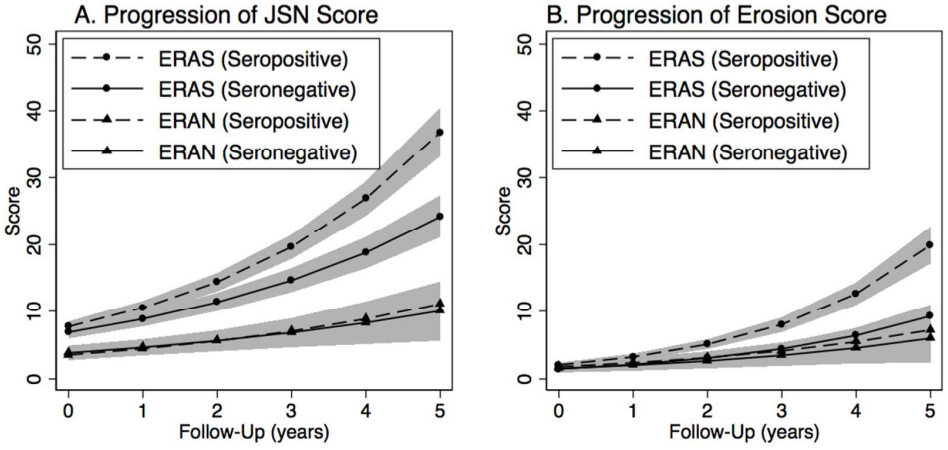
Differing from the JSN score, seropositive RA was associated with a 40% increased erosion score at baseline, compared to seronegative RA, in ERAS. The relative difference was reduced and non-significant for ERAN. Seropositive RA was associated with relative increases in the annual rate of progression, compared to seronegative RA, of 107% and 21% for ERAS and ERAN, respectively (Supplementary Table 1 & Supplementary Figure 1B). Again, though the relative impact of seronegative RA was reduced in ERAN compared to ERAS the difference was non-significant (IRR 0.58 [95% CI 0.17-0.98], p=0.234).

Supplementary Table and Figure

Supplementary Table 1 - Mean and relative difference in baseline level and annual rate of progression for JSN and Erosion scores based on seropostive (RF+) status between ERAS and ERAN. Estimates based on fixing the values of the covariates to the sample means

		RF-	RF+	Absolute Difference	Relative Difference (IRR) [95% CI]	P-Value
ERAS	Total JSN at baseline	6.9	7.7	0.8	1.12 [0.96-1.31]	0.152
	Total JSN Annual rate	2.9	4.6	1.7	**1.58 [1.32-1.84]**	**<0.001**
ERAN	Total JSN at baseline	3.8	3.5	0.3	0.92 [0.67-1.27]	0.619
	Total JSN Annual rate	1.1	1.3	0.2	1.19 [0.35-2.02]	0.480
ERAS	Total Erosion at baseline	1.5	2.0	0.6	**1.40 [1.12-1.75]**	**0.003**
	Total Erosion Annual rate	1.2	2.4	1.3	**2.07 [1.61-2.52]**	**<0.001**
ERAN	Total Erosion at baseline	1.5	1.8	0.3	1.17 [0.74-1.85]	0.495
	Total Erosion Annual rate	0.7	0.9	0.2	1.21 [0.28-2.14]	0.929

Supplementary Figure 1 – Estimated sample means of A) JSN and B) Erosion score for seropositive and seronegative patients in both ERAS and ERAN



STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **[Included in Title on page 1]**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **[See objectives and conclusion in abstract on page 2]**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **[Page 3 & 4]**
Objectives	3	State specific objectives, including any prespecified hypotheses **[Page 2]**
Methods		
Study design	4	Present key elements of study design early in the paper **[Page 4]**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **[Page 4 & 5]**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **[Page 4]**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **[N/A]**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **[Page 4 & 5]**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **[Page 5]**
Bias	9	Describe any efforts to address potential sources of bias **[Page 6]**
Study size	10	Explain how the study size was arrived at **[Page 4]**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **[Page 5 & 6]**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **[Page 5 & 6]**
		(*b*) Describe any methods used to examine subgroups and interactions **[Page 6]**
		(*c*) Explain how missing data were addressed **[Page 6]**
		(*d*) If applicable, explain how loss to follow-up was addressed **[Page 6]**
		(*e*) Describe any sensitivity analyses **[N\A]**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **[Page 7]**
		(b) Give reasons for non-participation at each stage **[Page 7]**
		(c) Consider use of a flow diagram **[N/A]**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **[Page 7]**
		(b) Indicate number of participants with missing data for each variable of interest **[Page 7]**
		(c) Summarise follow-up time (eg, average and total amount) **[Page 8]**
Outcome data	15*	Report numbers of outcome events or summary measures over time **[Page 8]**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were

		adjusted for and why they were included **[Page 8 & 9 & 10]**
		(*b*) Report category boundaries when continuous variables were categorized **[N/A]**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **[Page 8 & 9 & 10]**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **[N/A]**
Discussion		
Key results	18	Summarise key results with reference to study objectives **[Page 9]**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **[Page 12 & 13]**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **[Page 11 & 12]**
Generalisability	21	Discuss the generalisability (external validity) of the study results **[Page 12]**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **[N/A]**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

Appendix H

Other Publications and Abstracts

Publications

Carpenter, L., Nikiphorou, E., Sharpe, R., Norton, S., Rennie, K., Bunn, F., … Young, A. (2016). Have radiographic progression rates in early rheumatoid arthritis changed? A systematic review and meta-analysis of long-term cohorts. Rheumatology (Oxford, England), kew004.

Nikiphorou, E., Norton, S., Young, A., **Carpenter, L.**, Dixey, J., Walsh, D. A., & Kiely, P. (2016). Association between rheumatoid arthritis disease activity, progression of functional limitation and long-term risk of orthopaedic surgery: combined analysis of two prospective cohorts supports EULAR treat to target DAS thresholds. Annals of the Rheumatic Diseases.

Nikiphorou, E., **Carpenter, L.**, Morris, S., MacGregor, A. J., Dixey, J., Kiely, P., … Young, A. (2014). Hand & foot surgery rates in RA have declined from 1986-2011, but large joint replacements remain unchanged. Results from two UK inception cohorts. Arthritis & Rheumatology.

Carpenter, L., Nikiphorou, E., & Young, A. (2013). Importance of registries in informing clinical practice for arthritis. Clinical Practice, 10(6), 723–736.

Abstracts

BSR 2016 (April2016)

Carpenter L., Norton., S., Nikiphorou., E., Jayakumar., K., McWilliams, D., Dixey J., Keily, P., Walsh., DA., & Young A. How has the progression of Erosions and Joint space narrowing changed in early rheumatoid arthrits over the last 3 decades? Evidence from the ERAS and ERAN cohorts. Accepted as Oral presentation at Glasgow BSR 2016. Rheumatology 2016.

EULAR 2015 (June2015)

Nikiphorou E., **Carpenter, L.**, Norton S., Kiely, P., Dixey, J., Young A. Difference levels of moderate disease in rheumatoid arthritis (RA) are associated with varying risk for joint destruction and failure. Time to update DAS cut-offs for biologic DMARD use? Accepted as oral presentation and press release OP179 at EULAR Rome 12th June 2015. Annals of the Rheumatic Diseases 2015, 74:2 June 2015

BSR 2015 (April2015)

Carpenter, L., Sharpe, R., Nikiphorou E., Norton S., Bunn, F., Scott, D. L., Young. A. Radiographic progression rates over the first 10 years in patients with early rheumatoid arthritis: A systematic review. Accepted as poster presentation O81 at BSR Manchester 28th April 2015. Rheumatology 2015, 54:1 April 2015

Norton S., Nikiphorou E., **Carpenter L.**, Walsh D., Kiely P., Dixey J., Young. A. Impact of conventional disease modifying therapy on mortality risk in two

UK rheumatoid arthritis cohorts. Accepted as oral presentation O05 at BSR Manchester 28th April 2015. Rheumatology 2015, 54:1 April 2015

ACR 2014 (November2014)

Carpenter L., Nikiphorou., E., Norton., S., Jayakumar., K., Dixey J., Young A. Patients With Moderate Disease Activity In The First 5 Years Of Rheumatoid Arthritis Still Progress Radiographically Despite Conventional Disease Modifying Therapy. Accepted as poster 1381 at ACR Boston 17th Nov 2014. Arthritis & Rheumatism 2014; 66: S10 Oct 2014 (DOI: 10.1002/art.38914).

Nikiphorou E., **Carpenter L.**, Norton S., Young A. Eventual joint failure and surgery rates in rheumatoid arthritis remain high in patients with moderate disease activity in the first 5years of disease.. Accepted as oral presentation 1816 at ACR Boston 17th Nov 2014. Arthritis & Rheumatism 2014; 66: S10 Oct 2014 (DOI: 10.1002/art.38914).

Norton S., Nikiphorou E., **Carpenter L.**, Walsh D., Kiely P., Dixey J., Young. A. Reduced mortality risk in rheumatoid arthritis: findings from two UK inception cohorts. Accepted as oral presentation 2764 at ACR Boston 18th Nov 2014. Arthritis & Rheumatism 2014; 66: S10 Oct 2014 (DOI: 10.1002/art.38914).

BSR 2014 (April2014)

Carpenter,L., Jayakumar. K., Nikiphorou, E., Norton, S., Dixey, J. & Young, A. Can early, clinically significant radiographic progression in the first year of disease predict orthopaedic surgery in patients with rheumatoid arthritis? Accepted as Oral presentation at BSR Liverpool 29th April 2014. Rheumatology 2014, 53:1 April 2014