PHD

An investigation into the predictors and frequency of sustained remission in patients with rheumatoid arthritis undergoing treatment with anti-tumour necrosis factor therapy using the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis

Hamann, Philip

Award date:
2018

Awarding institution:
University of Bath

An investigation into the predictors and frequency of sustained remission in patients with rheumatoid arthritis undergoing treatment with anti-tumour necrosis factor therapy using the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis

Philip David Hugo Hamann

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Pharmacy and Pharmacology

September 2017

Table of contents

Table of Contents

Table of tables

Table of figures

Table of equations

Outputs from this thesis

Peer-reviewed publication

Hamann, P., Holland, R., Hyrich, K., Pauling, J. D., Shaddick, G., Nightingale, A., & McHugh, N. (2017). Factors associated with sustained remission in rheumatoid arthritis in patients treated with anti-tumor necrosis factor. Arthritis Care & Research, 69(6), 783–793. http://doi.org/10.1002/acr.23016

Oral presentations

BSR Annual Congress 2016, Glasgow. Differences in DAS28-CRP and DAS28-ESR influence disease activity stratification in rheumatoid arthritis and could influence use of biologics, treatment efficacy evaluations and decisions regarding treat-to-target: an analysis using the BSRBR-RA.

EULAR Annual Congress 2016, London. The development of the modified DAS28-CRP to improve agreement with DAS28-ESR and ensure appropriate disease activity stratification in RA.

Poster presentations

BSR Annual Congress 2016, Glasgow. Factors associated with sustained remission in rheumatoid arthritis in patients treated with anti-tumour necrosis factor (anti-TNF): a systematic review.

EULAR Annual Congress 2016, London. Differences in DAS28-CRP and DAS28-ESR influence disease activity stratification in rheumatoid arthritis and could influence use of biologics, treatment efficacy evaluations and decisions regarding treat-to-target: an analysis using the BSRBR-RA.

Invited presentations

BSR Annual Congress 2017, Birmingham. Development of the mDAS28 Scoring System.

BSR Annual Congress 2015, Manchester. Discussion and outline of BSR Fellowship Plans

Online tool

Calculator for the mDAS28-CRP. *https://mdas28.shinyapps.io/MDAS/*

TEDx Talk

Can we predict how we respond to medicines? *https://youtu.be/0_OVgREPnZE*

Blog posts

Blankety-Blank- Understanding Missing Data. *https://epibath.wordpress.com*. Posted January 18, 2017.

How long is a piece of string? The importance of consistent measurement in Rheumatoid Arthritis. *https://epibath.wordpress.com*. Posted July 28, 2016.

Personalised treatment is what it's all about nowadays in medicine. *https://epibath.wordpress.com.* Posted November 20, 2014.

Magazine Article

Featured in: 'RA remission more likely when anti-TNF combined with methotrexate' By Jennifer Davies. Arthritis Today. 2016.

Acknowledgments

There are many people I would like to thank. Firstly, I would like to thank my supervisors, Professor Neil McHugh, Dr John Pauling, Professor Gavin Shaddick and Professor Kimme Hyrich for their inspiration, guidance and supervision. Without their enthusiasm, constructive and helpful feedback, this work would not have been possible. Thank you Neil for being my educational supervisor for seven years, and guiding me through my clinical and academic training to be in a position to now achieve my ambition of being a clinical academic. John - thank you for supporting me in applying for the fellowship that enabled me to complete this work, as well as providing so much constructive feedback, and always ensuring I focussed on the clinical relevance of my work in the wider clinical and research environment. Gavin – many thanks for all your help and for such clear explanations of the complex statistical methods that I was able to utilise in this thesis. Thank you for inspiring me to develop my knowledge and use of Bayesian statistics further. Kimme, thank you for always making time for me. Your expertise and knowledge of epidemiology and the BSRBR-RA has been invaluable in enabling me to undertake this work.

Secondly, I would like to thank the BSRBR-RA team in Manchester. From the very beginning, everyone has been so welcoming and helpful. From showing me how the register runs on a day to day basis, the extraction of the data and inviting me to meetings, facilitating my remote attendance at meetings, and generally making me feel part of the team. In particular, I am grateful to Kath Watson and Katie McGrother who have always been on hand to help with any questions I had about the register and in extracting the data from the BSRBR-RA. I am also grateful to all the patients and clinicians who have contributed to the BSRBR-RA to make it the incredible resource that it is today.

I would also like to thank Dr Richard Holland and Dr Alison Nightingale who both cheerfully helped with the large task of dual screening the abstracts for the systematic review, and provided technical help in using Access™. I am also grateful to Amelia

Jobling for always being on hand to answer my many questions about the R software package.

I am endlessly thankful to my wife Hannah, who has been a constant source of advice, support and encouragement throughout this thesis and been so understanding during my long periods of absence at conferences, meetings and in writing-up this thesis, and to my daughter Ella, who always provides a ray of sunshine. I am also grateful to my family, particularly my father, and parents-in-law for their support and interest in my research.

Finally, I would like to thank the British Society for Rheumatology for supporting me to undertake this work. I am grateful for their generosity and belief in me to fully fund my work for three years. I am also grateful to have had me such academic freedom to pursue my ideas. Thank you.

I would like to dedicate this thesis to my late mother.

Declaration

I declare that the work undertaken in this thesis is my own and I take full responsibility for the opinions and content. I was responsible for the development of hypotheses, planning and execution of analyses as well as interpretation and discussion of results.

The work undertaken in Chapter 7 has been published (doi: 10.1002/acr.23016), and the manuscript was completed with assistance from the co-authors. Dr Holland, Dr Pauling and Dr Nightingale also assisted with the dual screening of the abstracts.

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Philip Hamann

Abstract

Background. Response to anti-tumour necrosis factor (anti-TNF) therapy in patients with rheumatoid arthritis (RA) varies between patients. Incidence of sustained remission in the UK is not known, and factors contributing to its achievement are poorly understood. Prior knowledge of response would enable better targeting of anti-TNF therapy, leading to better outcomes and reduced morbidity.

Aims. This thesis aims to identify incidence of sustained remission and low disease activity (LDA) in patients with rheumatoid arthritis (RA) taking anti-TNF therapy. Clinical and demographic factors associated with sustained remission and LDA were identified.

Methods. I undertook a systematic literature review of the incidence of, and factors associated with, sustained remission in patients with RA taking anti-TNF therapy. Results informed a subsequent analysis of data extracted from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis (BSRBR-RA). I used two approaches to examine sustained remission and LDA. Firstly, pre-defined DAS28 thresholds were used to identify individuals in sustained remission and LDA. Secondly, a data-driven approach used latent class mixed modelling (LCMM) to identify independent trajectories of response within the data.

Results. Sustained remission and LDA occurred infrequently in the literature review (range 4.2 – 38.1% sustained remission) and was uncommon in the BSRBR-RA (14.9% and 26.3% respectively), but had improved significantly over time. Significant associations were identified between the candidate variables and sustained remission and LDA, both using pre-defined thresholds and LCMM analyses. LCMM analyses identified response at six months to be a good indicator of long-term outcomes.

Conclusions. Sustained remission and LDA remains uncommon, although outcomes are improving. Clinical and demographic features are associated with achieving these

outcomes, suggesting it may be possible to use phenotypic features to guide therapy. Additionally, the clear response trajectories identified at six months, suggest it may be possible to identify non-responders to anti-TNF therapy earlier than six months.

List of abbreviations

ACPA – Anti-citrullinated peptide antibody

ACR - American College of Rheumatology

AIC - Akaike information criterion

ANOVA - Analysis of variance

ARA - American Rheumatism Association

BIC - Bayesian information criterion

BMI – Body Mass Index

BSR - British Society for Rheumatology

BRAGGSS - Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate

BSRBR-RA – British Society for Rheumatology Biologics Register for Rheumatoid Arthritis

CDAI – Clinical disease activity index

CI – Confidence interval

CRP – C-reactive protein

CT - Computed tomography

DAG – Directed cyclic graph

DAS- Disease activity score

DAS28 – Disease activity score of 28 joints

DMARD - Disease modifying anti rheumatic drug

EBV –Epstein-Barr virus

EMB - Expectation-maximisation and bootstrapping

EQ-5D - European quality of life questionnaire including five domains

EULAR – European League Against Rheumatism

ESR – Erythrocyte sedimentation rate

ESI - Events of special interest

FDA - Food and Drug Agency

GAM - Generalised additive models

GLM - Generalised linear models

GM-CSF - Granulocyte–macrophage colony-stimulating factor

HAQ - Health assessment questionnaire

HDA – High disease activity

ICER - Incremental cost effectiveness ratio

IFN - Interferon

IL – Interleukin

JAK – Janus kinase

LCMM - Latent class mixed modelling

LDA – Low disease activity

LOCF – Last observation carried forward

MAPK – Mitogen activated protein kinase

MAR – Missing at random

MCAR – Missing completely at random

MCMC - Markov-chain Monte-Carlo

MDA – Moderate disease activity

MD-HAQ - Modified health assessment questionnaire

MHC - Major histocompatibility complex

MCP – Metacarpophalangeal

MICE - Multiple imputation using chained equations

MNAR – Missing not at random

MRI - Magnetic resonance imaging

MTP – Metatarsophalangeal

NF-kB – Nuclear factor kB

NHS – National Health Service

NICE – National Institute for Health and Care Excellence

NOAR - Norfolk Arthritis Register

NSAID – Non-steroidal anti-inflammatory

OLS - Ordinary-least-squares

OMERACT – Outcome measures in rheumatology

OR – Odds ratio

ONS – Office for National Statistics

PI3K - Phosphatidylinositol 3-kinase

PIP - Proximal interphalangeal

PAD - Peptidylarginine deiminase enzyme

PAS - Patient Activity Score

PAS-II - Patient Activity Score–II

PGA – Patient global assessment of disease activity

QALY - Quality adjusted life year

RA – Rheumatoid arthritis

RCT - Randomised controlled trial

RF – Rheumatoid factor

RAI – Ritchie Articular Index

RANKL - Receptor activator of NF-κB ligand

RAPID3 - Routine Assessment of Patient Index Data 3

RADAI - Rheumatoid Arthritis Disease Activity Index

RAID - Rheumatoid Arthritis Impact of Disease

RCT – Randomised controlled trial

RMSE – Root mean squared error

SD – Standard deviation

SDAI – Simplified disease activity index

SF- 36 - 36-Item Short-Form health survey

SJC - Swollen joint count

STAT - Signal transducer and activator of transcription

S:TJR – Swollen:Tender joint count ratio

TJC - Tender joint count

TNF – Tumour necrosis factor

UIP - Usual interstitial pneumonia

VAS – Visual analogue score

VEGF – Vascular endothelial growth factor

VIF - Variance inflation factor

VPN – Virtual Private Network

Chapter 1

1 Introduction

The role of the physician has always been to alleviate pain and suffering, and ideally cure patients. Enshrined within all versions of the Hippocratic oath, and summarised by Hippocrates himself as; *'cure sometimes, treat often, comfort always'*. The highest aim of medical treatment continues to be the goal of 'curing' disease. Defined by the Oxford English Dictionary (1), a medical cure is:

> '*Successful medical treatment; the action or process of healing a wound, a disease, or a sick person; restoration to health.'*

This definition highlights that a successful medical treatment is one that cures, and restores health, with no time specification for how long the restoration of health should last; the implied assumption being that to 'cure' is to restore to a natural and permanent state of 'health'.

Rheumatoid arthritis (RA) is a chronic inflammatory joint disease, causing joint pain, swelling and stiffness. It remains an incurable disease, with the capacity to cause significant pain, disability, suffering and, if left untreated, premature death. There are now a wide range of therapeutic options available for the treatment of the condition with good evidence for their efficacy. However, there remains a significant challenge when transposing evidence-based findings generated at a cohort or population level to the individual patient level. Seeking greater personalisation of treatment, using putative molecular and serum biomarkers to guide therapy, is an attractive and laudable aim. The work undertaken in this thesis shall seek to contribute to this challenge; attempting to use real-life data from one of the world's largest RA clinical registries to define clinical and demographic features that might guide stratification of therapy for RA.

To fully understand the challenges in managing, measuring and assessing response to treatment in RA, an appreciation of the clinical symptoms and treatment of the condition, and the evolution of our knowledge in this regard over the last 200 years, is essential.

1.1 A brief history of RA

RA is often thought of as a 'young' disease. The first description of the disease in modern medicine was in 1800 by Augustin Landre-Baeuvais, working at Saltpêtrière asylum, with his thesis on "Primary Asthenic Gout" (2). Landre-Baevais described a constellation of symptoms that are now identified as RA, although, as the title of his thesis purports, the condition was, at that stage, considered a forme fruste of gout. His work paved the way for others to investigate this previously uncharacterised condition, and in 1859, Alfred Garrod published his work 'Treatise on the Nature of Gout and Rheumatic Gout' (3). Garrod had identified that a build-up of excess crystals in the blood was the primary cause of gout. However, he also identified a group of patients who had symptoms similar to gout, but who did not have any crystals in their blood. He called this condition 'rheumatic gout'. The separation of what went on to become known as RA from gout was the first step in properly defining the condition. It was Garrod's son, Archibald who first coined the term 'rheumatoid arthritis' with his thesis 'Treatise on Rheumatism and Rheumatoid Arthritis' in 1890 (4).

Whether RA really is a 'new' condition that has developed in the post-industrial revolution era remains unclear. One of the arguments for RA being a 'young' condition is due to the relative lack of medical records from historical times that describe the condition, and absence of paleopathlogical evidence demonstrating the classical erosive skeletal damage commonly associated with the condition (5). The lack of medical record evidence may have been due to the fact that historically, while gout was associated with wealth and prosperity, RA was a disease that, according to Landré-Beauvais, 'resides in the home of the indigent' (2). The reason why RA was more prevalent within lower socioeconomic classes of society is unclear. However, it is possible that the general poorer nutrition and health, particularly dental health (now

proposed as a key driver of anti-citrullinated peptide antibody (ACPA) positive RA (6)), may have played a role. Financial barriers to accessing medical care for chronic conditions amongst poorer people might account for under-documentation of RA by medical practitioners of the time (7). Another reason why RA may not have been prevalent in historical records is that first signs of RA usually don't develop until the late 30's or 40's, which was beyond the life expectancy of many in historical societies (including the UK); mean life expectancy did not extend beyond 50 years of age in the UK until 1907 (8). Therefore, the rise in prevalence of RA in Europe may also have been as a result of the rise in life expectancy in society (7).

Examination of archaeological remains is also challenging, because primary joints affected by RA are predominantly in the small joints of the hands and feet. The nature of such bones is that they are often incomplete or damaged in skeletal remains, making it challenging to identify definitively the pathognomonic changes associated with the disease (7). Although there have been some descriptions of possible rheumatoid deformities in exhumed skeletons in Europe and north Africa (9,10), 'Old World' paleopathological evidence remains sparse, and often disputed. Interestingly, there is putative evidence of the existence of RA in the Americas for over four Millennia with the identification of likely RA in 3500-year-old skeletal remains from Alabama, USA (11).

Although RA was only formally identified as a distinct condition in the late 19th century, and archaeological evidence is contentious, Hippocrates recorded a constellation of symptoms in a patient in the 4th century BC that would certainly be in keeping with a modern diagnosis of RA, describing:

'In the arthritis which generally shows itself about the age of thirty-five there is frequently no great interval between the affection of the hands and feet; both these becoming similar in nature, slender, with little flesh...For the most part their arthritis passeth from the feet to the hands, next the elbows and knees, after these the hip joint. It is incredible how fast the mischief spreads' (7).

In addition to Hippocrates, Thomas Sydenham also described symptoms that would be in keeping with RA under the heading *'Rheumatismus'* in his collection of case histories *'Medical Observations Concerning the History and Cure of Acute Diseases '* in 1676 (12). Sydenham described:

> *'a sharp pain, now in this, now in that joint, (but most in the wrists, shoulders, and knees) shifts about, leaving redness and swelling in the different parts as it takes them in turn.'*

Another reason that RA may have become more prevalent in European societies from the 19th century onwards may have been in part due to two major historical events that significantly altered environmental exposures to the peoples of Europe in the 16th and 18th centuries, and have both subsequently been identified to be integral to the pathophysiology of RA.

The first of these events was the importation of tobacco from the 'New World' in the sixteenth century. Tobacco had been smoked by native Americans for thousands of years (13), and as mentioned previously, there is more historical evidence of the condition than in Europe. Initially, following its importation to Europe, tobacco use was reserved for medicinal purposes, however, in the 17th and 18th century, it's use became more widespread with smoking for pleasure rather than for medical purposes (14). Consumption of tobacco increased steadily until the 1950's when the negative health implications of smoking began to be appreciated (15), although the impact of smoking on RA occurred later in the 1990s and beyond (16).

The second event thought to have contributed to the rise in number of cases of 'polyarticular gout' and appearance of rheumatoid as a recognised condition, was the importation of sugar from the West Indies to Europe, which led to increased rates of periodontitis (17) in the population. Both smoking and periodontitis have been linked with the production of ACPAs which are strongly associated with RA (Chapter 1; 1.3.3.1 and 1.3.3.2). It is likely that these events acted as environmental pressures that added to pre-existing genetic susceptibilities within the population, such as the HLA-DRB1 and PTPN22 alleles (18,19). These factors, combined with increasing life expectancy

(meaning more people lived to an age where RA was likely to manifest itself), are all likely to have contributed to the rise in the incidence and prevalence of the disease in European society.

1.2 Pathogenesis of RA

The joint symptoms in RA are primarily driven by inflammation of the synovium, which causes swelling, pain and stiffness. The exact pathogenesis of RA remains unknown, but dysregulation of cytokines and infiltration of lymphocytes leads to an inflammatory cascade with increased synovial thickness and vascularity.

Following initial inflammatory attack on the synovium, synovial cells become hyperplastic and thickened, with the recruitment of fibroblast-like synoviocytes, neutrophils, macrophages and development of new blood vessels to form pannus (20). Neutrophils and macrophages secrete pro-inflammatory cytokines which cause localised damage at the synovial/cartilage junction whereupon, the underlying cartilage and bone can be targeted by the inflammatory process (usually by matric metalloproteinases) which act to degrade the cartilage by disassembling type 2 collagen. Chondrocyte apoptosis limits further the regenerative capacity of cartilage with cartilaginous thinning and subsequent joint space narrowing (21) (visible radiologically). This cartilaginous and bone damage form the beginnings of permanent damage of the inflammatory process. Prolonged inflammatory attack causes the development of periarticular erosions (mediated by macrophage-colony stimulating factor and receptor activator of NF-κB ligand (RANKL)), which promote activation of osteoclasts which erode and thin the bone around the joint (22) (visible on plain x-ray as erosions and peri-articular osteopenia respectively).

1.2.1 Pro-inflammatory cascade

Within the synovium macrophages secrete a wide range of cytokines. These include: TNF; IL-6, granulocyte–macrophage colony-stimulating factor (GM-CSF); interferons

(INF), IL-15, IL-18, IL-32. Pro-angiogenic factors are also secreted, including, vascular endothelial growth factor (VEGF), TNF, IL-17 and IL-1. These promote the release of matrix enzymes from chondrocytes and fibroblast-like synoviocytes which degrade collagen, and RANKL, TNF, IL-1 and other cytokines promote the resorption of bone and development of erosions by osteoclasts (21).

TNF is centrally involved in the inflammatory cascade and activates leukocytes and synovial fibroblasts. It activates endothelial cells which promotes the adhesion and entry of lymphocytes from the vascular system to the synovium. TNF also has a prominent role in suppressing regulatory T-cell function and activating osteoclasts (23). The primary homeostatic functions of TNF are in the defence against pathogens and inhibition of tumorigenesis, however it plays a central role in the induction of inflammatory mediators central to RA (including IL-6 and NF-κB). Activation of NF-κB promotes the transcription of INF-β, which in turn activates janus kinase (JAK) and signal transducer and activator of transcription (STAT) pathways which promote the production of pro-inflammatory cytokines including CXCL-9 and CXCL-10, which promotes T-cell migration to inflamed synovium (24).

IL-6 has a pleotropic array of actions including the induction of hepatic acute-phase proteins that are important in the trafficking if inflammatory cells. It also plays an important role in B-cell differentiation that produces autoantibodies in RA as well as promoting and sustaining the differentiation of Th17 cells - important in sustaining chronic immune responses. One of the key actions of IL-6 is through the JAK-STAT pathway which can activate mitogen activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K) pathways, all of which are central to the propagation of an inflammatory response (25). IL-6 also has an important role in regulating lipid metabolism, anaemia of chronic disease and is associated with fatigue (26).

IL-1 also causes leukocyte activation and promotes migration of inflammatory cells to the synovium where it plays a role in activating matrix metalloproteinases. IL-1 has important biological effects in infection, although chronic activation of the cytokine is deleterious. IL-1 enhances the recruitment of inflammatory cells locally by up-

regulating the expression of adhesion molecules, allowing inflammatory cells to exit the circulatory system and enter the synovium. It also induces the cyclo-oxygenase pathway, as well as the production of prostaglandins and thromboxanes, and acts to promote the genesis of Th17 cells (27). IL-1 also stimulates the production of matrix metalloproteinases which are important in the degradation of cartilage and erosions. Systemically IL-1 mediates fever and glucose metabolism (21).

The role of the IL-17 cytokine was discovered relatively recently, but appears to play an important role in shifting the inflammatory response towards a pro-inflammatory state and may be important in explaining why many inflammatory conditions persist when treatment is withdrawn (28).

The multiple pathways that drive RA may also go some way to explaining why effective treatment with one, or indeed many disease modifying agents still fails to 'cure' the condition, and relapse often occurs.

The increasing understanding of the role of cytokines, and the development of techniques that facilitated mass-production of humanised monoclonal antibodies that could specifically target these cytokines, enabled the generation of a new class of therapeutic agents for the treatment of RA (29).

1.2.2 T-cells

RA is generally considered to be a T-cell driven autoimmune condition, with antigen presentation from dendritic cells promoting the differentiation of Th0 helper cells to Th1 T-cells in lymphoid germinal centres. Over the last few years, an additional subtype of T-cells (Th17) has been identified that is important in the maintenance of the inflammatory cascade and produces a IL-17, a cytokine with pleiotropic effects that can promote and maintain the inflammatory response. Th17 cells engage in cross-talk with immature B-cells which subsequently differentiate to plasma cells and plasmablasts which are responsible for the production of ACPA and rheumatoid factor (RF) antibodies (19,21).

1.2.3 B-cells

The role of B-cells in the pathogenesis of RA remains elusive. The presence of B-lymphocyte stimulators, and the B-cell proliferating ligand APRIL, all point to an important role of humoral immunity. The efficacy of B-cell depleting agents such as rituximab in the treatment of RA (particularly ACPA positive RA) attest to the importance of this cell lineage in the propagation and maintenance of the condition (30).

1.2.4 Autoantibodies

One of the earliest identified biomarkers of RA was the rheumatoid factor antibody. First identified in association with RA concurrently by Harry Rose and Erik Waaler, RF is still the most widely used biomarker in RA. Despite relatively poor sensitivity and specificity, RF antibodies are associated with nodular and erosive disease. More recently, antibodies against citrullinated peptides (ACPA) have also been identified and have much greater sensitivity and specificity for RA, and have also been implicated in the pathogenesis of RA (Chapter 1; 1.3.3.1 and 1.3.3.2). Peptide citrullination is driven by the activation of PADs which substitute negatively-charged arginine with neutrally-charged citrulline amino acids on self-proteins. PAD activation is thought to occur during cellular stress, by the influx of calcium (31). The substitution of citrulline for arginine is thought to strengthen the binding that occurs in the peptide-binding groove of HLA-DRβ1, and enhance the possibility of auto-reactivity against 'self' proteins. The presence of both ACPA and RF antibodies in the serum are very strongly linked to the presence of RA, and can be predictive of future onset of RA, even if no overt symptoms are present at the time of detection (32).

1.3 Epidemiology of RA

Today, RA is the most common autoimmune inflammatory arthritis in the UK and affects approximately 1% of the adult population in the UK (33). Symptoms typically present in the fifth to seventh decade of life, although it can occur at any age.

1.3.1 Genes and epigenetics

RA has a significant genetic composition, which may account for up to 50% of the risk of developing the condition (19). There is strong epidemiological evidence of a significant genetic component of the disease. Prevalence in European and North-American Caucasian populations is reported at 0.5 – 1.0% of the population (34). However, by comparison, some Native American tribes (such as the Chippawa and Pima) have significantly higher prevalence of RA (6.8% and 5.3% respectively (35)). By comparison, the prevalence of RA in far eastern countries such as China and Japan are much lower (0.2 - 0.3% respectively) suggesting a genetic or environmental component to the disease. Prevalence is also noted to be very low across sub-Saharan Africa (35), although low prevalence in such countries may be partially attributable to lower life-expectancies and limited universal healthcare coverage which may mean that the disease is less likely to be identified by research studies.

Whilst there is a familial risk associated with RA, it appears to be lower than other autoimmune conditions such as Type 1 diabetes and multiple sclerosis and whilst studies vary, it is estimated that there is approximately a twofold increase in risk in first-degree relatives of individuals with RA (36).

Genome-wide studies have identified multiple associations between RA and genes associated with immunologic processes (37). Genetic variation at HLA-DRB1 and PTPN22 alleles are most commonly associated with the condition, although multiple other associations have been identified (36,38). Variations in the HLA-DRB1 allele have been identified with varying strengths of association (35) with RA, thought to be due

to the 'shared epitope'- a five amino-acid sequence motif in residues of the HLA-DRβ chain that are located in the peptide-binding groove and responsible for antigen presentation. The reducing cost and increasing computing power have led to a dramatic increase in studies into genetic risk factors for RA, as well as using genetics platforms to explore the pathogenesis of the condition. Both T and B-cell signalling pathways have been identified, and signalling pathways involving tumour necrosis factor (TNF), Nuclear factor kB (NF-kB) and antigen presentation have also been associated with RA (39).

The role of epigenetics in the development of RA has also been increasingly identified as a possible route for the integration of environmental and genetic factors that may lead to the development of the clinical condition recognised as RA. Methlylation and acetylation of the genome acts to makes sections of the genome more or less susceptible to transcription, and environmental factors such as smoking have been shown to be potent inducers of acetylation and methylation, highlighting how environmental factors could influence genetic transcription without requiring mutation at a nucleotide level (40).

1.3.2 Gender and hormonal factors

The increased risk of RA amongst women, particularly those after menopause suggests that that the balance of hormones may play a role in the development of RA. The reduction in disease severity during pregnancy and exacerbation of disease activity when breastfeeding was noticed by Philip Hench as early at 1938 when he reported a case-series of 20 women with RA whose symptoms remitted during pregnancy (41). Multiple studies have also suggested that exposure to the oral contraceptive pill is associated with a reduced risk of developing RA, although there have also been conflicting findings and no overall association was identified in a recent meta-analysis (42) .

Breastfeeding is thought to influence the onset and progression of RA, and has been associated with a significantly increased risk of development of RA (43), relapse and

more severe disease (44) which has been attributed to both the pro-inflammatory nature of prolactin (45). However, the role of breastfeeding in the development and progression of RA continues to be contentious, with two more recent studies identifying a protective effect (46,47). Most recently, increased prolactin receptor expression has been identified in macrophages in synovial tissue of patients with RA and psoriatic arthritis compared with controls suggesting a possible mechanism by which prolactin might act in inducing or exacerbating RA in breastfeeding women (48).

1.3.3 Environmental risk factors

As alluded to earlier, environmental risk factors are also thought to play a role in the development of RA, as well as being important in modulating disease severity.

1.3.3.1 Smoking

The strongest environmental association with RA remains tobacco smoking, which may cause up to 35% of the risk in seropositive RA (49). It has been implicated in both the pathogenesis of the disease, and more aggressive erosive forms of the disease through the promotion or production of ACPA (50). Smoking is strongly associated with the development of seropositive (specifically ACPA positive) RA in individuals with the HLA-DRB1 shared epitope (51), and a recent meta-analysis has shown that current smokers have a significantly increased risk of developing seropositive RA compared to non-smokers (OR 1.64). This risk is exacerbated in males where the risk is increased to nearly four-fold (OR 3.91) (52). It is thought that smoking causes cellular stress in the lung parenchyma leading to the activation of peptidylarginine deiminase enzymes (PADs) which lead to the substitution of arginine for citrulline amino acids. Such citrullination of self-peptides is thought to interact strongly with the peptide binding groove of the HLA-DRβ1 protein, particularly in individuals with the shared epitope. The binding of self-peptides to the peptide binding groove leads to auto-antigen presentation by antigen presenting cells to both T- and B-cells, which breaks immune self-tolerance and leads to production of antibodies and an associated inflammatory response within synovial joints (53,54).

1.3.3.2 Periodontitis

Periodontitis has been associated with an increased prevalence of RA (55), and RA patients with severe periodontitis have been shown to have higher disease activity score of 28 joints (DAS28) when compared to RA patients who have no or moderate periodontitis (56). The putative mechanism by which periodontitis might be associated with RA is through the observation that dental infections with *Porphyromonas gingivalis* (*P. gingivalis*) are associated with increased levels of detection of ACPA in the gingival crevicular fluid surrounding infected gums (57). Additionally, ACPA positive patients with RA have also been identified as having higher titres of anti-*P.gingivalis* antibodies in serum (58).

P. gingivalis has been identified to have native PADs which have been demonstrated to citrullinate host peptides, leading to the development of ACPAs (6). It is postulated that these ACPAs then act systemically to break immune tolerance and invoke an inflammatory arthritis in a similar manner to that proposed for smoking (Chapter 1; 1.3.3.1), although no studies have demonstrated a clear cause/effect relationship. However, this association requires further investigation, and appears to be less clear-cut than that identified for smoking, with a recent meta-analysis identifying that the association between RA and periodontitis is lost when compared against control patients who had osteoarthritis rather than healthy controls (59).

1.3.3.3 Infections

Along with *P. gingivalis*, additional bacterial and viral infective triggers of RA have been proposed including: Epstein-Barr virus (EBV); cytomegalovirus; parvovirus; chikungunya; proteus and mycobacteria (60). The most common mechanism by which infective agents are thought to initiate RA is by molecular mimicry with loss of immune self-tolerance, but numerous other mechanisms (including microbial super-antigens, neo-antigen development, bystander immune activation) have been investigated (60). Whilst associations between infective triggers and RA have been identified, and infection with many of the postulated agents does cause arthralgia, definitive evidence

of causality is lacking. Reasons for such a lack of definitive proof of causality are multiple (61). The relatively late-onset of RA in life means that there is a long-exposure time to many factors (infective and otherwise) that are challenging to identify and control for. Moreover, an extended pre-clinical phase of the disease before the appearance of overt synovitis has been acknowledged by the observed emergence of positive RF and ACPA several years prior to diagnosis (62). Induction of inflammatory arthritis in animal models provides a limited approximation of human disease mechanisms and does not consider the complex human genetic risk factors previously discussed (Chapter 1; 1.3.1), and the lack of cross-species genetic conservation, particularly within the immune system.

1.3.3.4 Diet and gut microbiome

More recently, the role of the gut flora has been investigated in its role as a potential contributory factor to the development and maintenance of autoimmune conditions such as RA. Early studies have investigated the hypothesis that gut dysbiosis may both contribute to and modulate the severity of RA, and certain bacteria (*Prevotella copri* in particular) appear to be more prevalent in the stool of patients with RA, as well as promoting interleukin (IL) 17 responses in vitro when exposed (55). The hypothesis that intestinal flora influence systemic immune system homeostasis certainly seems a promising avenue for investigation, and results from the many ongoing studies will be of great interest.

1.3.3.5 Obesity

Obesity is one of the greatest public health challenges affecting many Western countries, and is increasingly being identified as having profound influence on the inflammatory profile of individuals. Adipose tissue has a diverse range of actions on the immune system which have been specifically identified as influencing disease activity in RA including leptin, adiponectin, visfatin and others (63).

Numerous studies have investigated if obesity increases the risk of developing RA, and although results are mixed, a recent meta-analysis suggests that obesity may increase the risk of developing RA compared to non-obese individuals, with a dose-response relationship with increasing body mass index (BMI) (64). In addition to investigating the association between obesity and risk of developing RA, many studies have examined how increased body mass influences outcomes in RA, and there appears to be a negative relationship between increasing BMI and optimal outcomes using composite outcome measures (65).

1.3.3.6 Other environmental triggers of RA

In addition to the aforementioned risk factors, other possible environmental triggers of RA have been identified including air pollution (66), silica (67,68) and textile dust (69). The findings that exposure to increased air pollution, dust, poor dentition, infections and smoking are all associated with an increased risk of developing RA may go some way to explaining Landre-Bauvais's original observation that RA appeared to be more prevalent amongst individuals in lower socio-economic groups – individuals that it would be considered likely would have a greater exposure to such stimuli.

1.3.4 Changing epidemiology

Whilst there are many challenges in identifying the true incidence of RA within a population, it is evident that the incidence of RA is dynamic, suggesting an interplay between environmental, infectious, genetic and other risk factors that influences the rate of the disease in society. A systematic review in 2006 noted variation in incidence rates globally (particularly between southern European and northern European/American studies), although there was a notable absence of evidence for large parts of the world (70).

Until the turn of the century, there appeared to be consistent evidence of a declining incidence of RA in developed nations (from where the majority of data are available) with many studies noting a declining incidence, along with an increasing age of onset of the disease (71). This had been attributed to a birth cohort effect, where successive

generations were deemed less likely to develop RA (possibly by a reduction in exposure to a precipitant of the condition), with a resultant increase in age-of-onset of the condition, as members of earlier birth cohorts have relative greater risk of developing the condition. The study of the Olmstead County population in Minnesota, USA had noted a consistent decade-by-decade decrease in incidence of RA from 61.2/100,000 population in 1955 to 32.7/100,000 population in 1995 (72). However, the latest update to the analysis of this population has identified an increase in incidence for the first time in 50 years to 40.9/100,000 population, driven by a significant increase in incidence in women from 39.9 to 53.1/100,000 population between 1995 and 2007. The cause of the latest increase in incidence in the Olmstead County study is unknown, although the authors suggest declining oestrogen content of the oral contraceptive pill, increasing BMI and a plateauing in smoking cessation rates amongst women could all be contributing factors (73).

Furthermore, a recent systematic review has suggested that rates of RA remain constant, although a decrease in disease severity appears to be evident, likely driven by improving identification and treatment of the condition (34).

Such conflicting findings suggest multiple factors at play, including changing life-expectancies, and environmental exposures, as well as better identification, characterisation and treatment of the disease.

1.4 Clinical signs and symptoms of RA

RA is a multisystem autoimmune disease but the major tissue target is the synovium, leading to inflammation within the synovial articulations. Most commonly affected are the small joints in the hands and feet - usually the metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints of the hands, and metatarsophalangeal (MTP) joints of the feet.

1.4.1 Symptoms

1.4.1.1 Joint pain and stiffness

Symptoms typically comprise pain, stiffness and swelling at the affected joints. A characteristic diurnal variation to symptoms, with pain and stiffness being worse in the morning, before easing off during the middle portion of the day, and often worsening towards the latter part of the day is typical. Movement of the affected joints usually eases symptoms of stiffness, and prolonged periods of inactivity often causes symptoms to worsen. If left untreated, the inflammatory attack on the joints causes localised joint destruction by damaging synovium, and in turn the bone, leading to restriction of joint movement, joint erosion and subsequent disability.

1.4.1.2 Fatigue

Fatigue often accompanies symptoms of active joint disease and can be as profound and as impactful as the joint symptoms themselves (74). RA patients report fatigue as one of the most important symptoms related to their disease (75), although it is often poorly addressed by clinicians (76). Often, with appropriate suppression of inflammation, fatigue can improve. However, fatigue often persists despite achievement of optimal outcomes, such as remission (77).

Numerous different factors are thought to contribute to the development and maintenance of fatigue in RA patients. At a biological level, increased markers of inflammation (including TNF, IL-6 and C-reactive protein; CRP) have been associated with fatigue in a number of studies (78), however it is difficult to isolate the effect of inflammatory markers from multiple other confounders such as pain, disrupted sleep, increased disability, medication side effects etc. which also contribute to fatigue (79). Improvements in fatigue (although incomplete) are often noted with successful immunomodulatory treatment of RA, suggesting inflammation does play a role (80).

Non-biological factors that have been associated with fatigue include depression and low mood, although fatigue itself is noted to cause depression, so identifying the

direction of causality is difficult (79). Poor psychological coping strategies, learned helplessness, difficult social circumstances and relationship difficulties have all also been associated with the level of fatigue experienced by RA patients (79,81).

Reduced physical activity levels are also associated with fatigue, and exercise programmes have been noted to improve fatigue in RA patients. Obesity also appears to be associated with increased levels of fatigue, although whether this is acting via increased inflammatory cytokines released by adipose tissue, increased effort of movement, or other mechanisms is unclear. RA patients with multiple comorbidity also experience greater levels of fatigue (82).

One of the most basic challenges in understanding fatigue is effectively and reliably quantifying it (82). Currently many different scoring tools are available, which all assess fatigue in different ways (e.g. impact of fatigue on ability to do tasks, mood, absence of fatigue), as well as using different mechanisms for quantifying effect (continuous scores, categorical scores, categorised continuous scales) (83). Most composite disease activity scores (such as the DAS28) do not formally assess fatigue other than via the patient global assessment of disease activity, meaning fatigue may disproportionately affect the disease activity score independent of the inflammatory component of the disease (84). Given this unreliable relationship, particular challenges may arise in assessing the specific effect of immunomodulatory therapies in RA if fatigue is prominent.

1.4.1.3 Pain

Pain is often highlighted as the most significant problem for patients with RA (85) and impacts both on the physical and psychological wellbeing of patients. The aetiology of pain in RA is complex and multifaceted, and has significant interactions between both pre-morbid factors (such as comorbidities, genetics and psychological aspects) and factors related to RA itself (including inflammation, psychological distress, secondary joint damage and altered pain processing) (86).

The processing of nociceptive stimuli can be broadly broken down into peripheral and central processing of painful stimuli. At the simplest level, pain from RA may be directly attributed to the immune-mediated attack at the affected joints causing peripheral nociceptive fibres to be activated, with a subsequent pain response. The pain may be due to direct damage or swelling at the joint, and is often described using terms such as aching/throbbing/tender. In addition to pain caused by immune mediated attack, peripheral pain may be induced by the articulation of previously damaged joints, or stretching of a joint capsule that surrounds a swollen joint. Both Aδ (fast) and C (slow) pain fibres can be activated, leading to a multimodal pain response. In addition to joint pain, soft-tissue swelling may cause compression of nerves (such as carpal tunnel syndrome) and may lead to a more neuropathic-type pain response (often described as 'shooting' or 'stabbing'). Neuropathic-like pain can also occur without an overt cause and can be associated with a clinical spectrum that includes fibromyalgia (86).

The central processing of pain plays a significant role in modulating the experience of pain responses. The gate-control theory (87) highlights the role of the central nervous system in the inhibition of painful stimuli. However, in chronic pain scenarios such as RA, these central inhibitory pathways may become dysfunctional, leading to reduction in pain thresholds and pain amplification rather than reduction (88).

Higher cortical function is also essential in modulating the perception of pain. Activation of spinothalamic tracts in chronic pain can activate the sensory cortex and thalamus as well as the limbic system – leading to disruption of sleep, and mood. Furthermore, individuals with concurrent or past depression appear to have more ready activation of descending pain pathways. Because painful stimuli are unpleasant, individuals with chronic pain have higher rates of depression, which in turn can affect sleep-wake cycles and alter circadian rhythms, which further amplify the negative adaptive response to pain, and may play a role in modulating systemic inflammatory responses in RA (86).

Because of these complex interactions, treatment of pain in RA is one of the most challenging aspects of management. Treatment of RA pain needs to be targeted at both abrogating the inflammatory component of pain, as well as providing appropriate

symptom control for both peripheral and central neurological aspects of chronic pain. The multifaceted nature of pain in RA, makes symptom quantification challenging, particularly when assessing disease activity of RA. Composite measures of disease activity (such as the DAS28 and simplified disease activity index; SDAI) incorporate the quantification of pain into the respective scores using a single visual analogue scale (VAS). Tender and swollen joint counts, as well as an inflammatory marker (the erythrocyte sedimentation rate (ESR) or CRP) are used to attempt to provide more objective clinical assessments of disease activity. However, reduction in central nervous system inhibition of nociceptive stimuli may result in pain being reported at sites that otherwise would not be uncomfortable, meaning that the tender joint count may be less effective at discriminating sites of active inflammation (89,90). All these aspects conspire to make the clinical assessment of active (inflammatory) RA very challenging. Understanding the degree to which inflammation is causing symptoms is essential in the management of RA, as it suggests that further immunomodulation may be appropriate and may alleviate symptoms. However, if pain is as a result of non-inflammatory causes, analgesic treatment strategies should be instigated.

Whilst the composite measures of disease activity are essential in capturing the patient perspective of disease activity, they may be less effective at accurately quantifying the degree to which active inflammation is playing a part in causing pain, compared to existing joint damage or altered pain perception. Disease activity scores are the mainstay of clinical assessment of drug efficacy in RA, and such difficulties have a significant impact of the ability to assess the true impact and efficacy of a disease modifying drug in RA. These considerations are of great relevance to the work undertaken in this thesis and will be discussed in more depth in later chapters (Chapters 7, 9 and 10).

1.4.2 Signs

1.4.2.1 Hands and feet

In early disease, signs of RA may be minimal. Signs of early RA include swelling around the joints (which may be subtle). Joints that are typically affected are the MCP, PIP and various carpal joints in the wrist. In the feet, the MTP joints are most commonly affected (91). Inflammation at these joints can cause the plantar fat pads to move, with the resultant symptom of 'walking on pebbles' that patients often describe. On occasion, there may be no overt clinical swelling of the joints and ultrasound imaging can be of assistance in assessing for subclinical synovitis.

If the inflammatory process of RA is not suppressed, the initial early subtle signs of RA progress to more overt, less reversible ones. Initial subtle swelling at the joints may increase as more inflammatory cells infiltrate the synovium, and increased vascular permeability and blood flow cause extravasation of fluid which causes further swelling of the synovium. Increased production of intra-articular fluid causes the joint to become further swollen and (due to the inflexibility of the joint capsule) the range of movement of the joint becomes reduced. The increased intra-articular pressure caused by these multiple factors causes further pain. If joint swelling persists, laxity of the joint capsule and peri-articular ligaments occurs and allows the joints to assume characteristic deformities defined by competing mechanical forces in play around affected joint groups.

Within the hands, a number of deformities occur including volar subluxation of the radiocarpal joint (resulting in guttering of the extensor tendons), subluxation of the ulnar styloid (causing a 'piano key' deformity), ulnar drift at the MCP joints causes imbalance of forces between the flexor and extensor tendons of the hands and can cause the fingers to drift laterally to the ulnar side of the hand (ulnar deviation). Damage to the PIP joints can cause swan-neck and boutonnières deformities through a combination of imbalanced forces between flexor and extensor tendons, joint capsule laxity and damage at the A1, A2 and A3 pulleys in the finger (91).

Muscle atrophy of the intrinsic muscle of the hands compounds instability at the joints and secondary degenerative changes can occur within affected joint groups. The loss of muscle strength and instability at joints which are not the target of inflammatory attack can exacerbate otherwise quiescent degenerative changes, and leads to loss of grip strength and fine motor movement.

1.4.2.2 Other joints

In addition to the joints in the hands and feet, any synovial joint in the body can be affected, however, the most common joints to be targeted apart from those in the hands and feet include the atlanto-axial (C1/C2) joint in the cervical spine, shoulders, elbows and knees. With the exception of the atlanto-axial joint, the spine is usually relatively spared from attack in RA, as are the femoro-acetabular joints in the hips. The same process of inflammatory attack at the joints as described for the hands and feet occurs at the afore mentioned joints, with similar effects of joint swelling, erosion and damage (91).

Of major clinical significance is erosive disease at the atlano-axial joint. Swelling and pannus formation at this joint can cause compression of either the C1 and C2 spinal nerves which can lead to referred occipital pain. Alternatively, if the spinal cord is compressed, referred pain can occur anywhere in the body (although due to spinal cord anatomy the upper limbs are usually affected first). Erosive disease at the atlanto-axial joint can cause instability which can be life-threatening. Instability at this joint can be caused by erosion of the odontoid peg, or rupture of the transverse ligament which prevents the odontoid peg from compressing the spinal cord. Symptoms of instability at the atlanto-axial joint include localised tenderness, referred occipital pain and upper limb neurological signs (brisk reflexes, dermatomal tingling sensations, muscle weakness). Symptoms can be uni- or bilateral and are often instigated or exacerbated by dynamic movements at the cervical spine. Flexion and extension lateral plain films views of the cervical spine can help identify if there is significant movement of C2 relative to C1, however if clinically suspected, magnetic resonance imaging (MRI)

imaging of the cervical spine is the gold standard imaging modality. Because of the life-threatening nature of this manifestation of RA, it is a surgical emergency and an urgent referral to a neurosurgeon is required to identify if surgical stabilisation is necessary (91).

1.4.2.3 Extra-articular features

In addition to joint pathology, RA can affect almost any part of the body. One of the most common extra-articular manifestations of RA is the formation of nodules. Associated with RF- and ACPA- seropositive RA, nodules can occur almost anywhere in the body. The most common sites for nodule formation are on the fingers and elbows, however, they can also occur in the lungs. Rheumatoid nodules in the lungs can be indistinguishable from malignancy on most imaging, and may only be confidently diagnosed on biopsy. Rheumatoid nodules are usually not harmful in themselves, although can be physically painful and inconvenient, depending on the anatomical location of the nodule (91).

The cardiovascular system is also affected by RA with a significantly increased risk of mortality. Chronic inflammation leads to accelerated atherosclerosis and increased risk of myocardial infarction and death compared with general population rates (92-94). In addition to disease modifying treatment of RA, appropriate surveillance and aggressive primary and secondary preventative measures are necessary to minimise cardiovascular risk (95).

The lungs can also be affected by RA and can be a significant cause of morbidity and mortality. Inflammatory attack on the lungs can cause pulmonary fibrosis, most commonly usual interstitial pneumonia (UIP) with honeycombing and traction bronchiectasis seen on high resolution chest computed tomography (CT) scanning (96).

Aggressive, untreated RA can also lead to systemic vasculitis, which can in turn affect skin (causing vasculitic ulcers), nervous system (causing mononeuritis multiplex) and

renal system. Rarely the eyes can be affected by RA, with scleral thinning and risk of globe rupture if not managed appropriately. Amyloidosis can also occur (typically AA type), which can also lead to renal and neurological impairment. Fortunately, due to earlier treatment and improving therapies, extra-articular manifestations are becoming increasingly uncommon (91).

1.5 Imaging in RA

The use of imaging in the diagnosis and management of RA is increasing as further efforts are made to diagnose the condition earlier, and to ensure inflammation is suppressed maximally.

1.5.1 X-rays

Plain film x-rays remain the main imaging technique used in the diagnosis and management of RA. The plain film changes associated with RA are well described and typically comprise of peri-articular erosions, joint space narrowing, and periarticular osteopenia. Soft-tissue swelling can be also identified, but typically only significant swelling can be identified (97).

Despite being the oldest imaging technique, x-rays have a number of significant advantages over more recent imaging modalities. X-rays are cheap and quick to undertake and have relatively little radiation exposure to the patient. With digitisation, images can be examined quickly and precise measurements, zooming and contrast changes can be made to the image.

Because standard positions are adopted when the image is taken, changes over time can be accurately mapped. Furthermore, changes on plain imaging can usually be clearly attributed to a pathology (e.g. erosions). The downsides of the x-rays however are that the changes seen on plain film are typically irreversible, often take a while to

form and are usually associated with greater disease durations which have had long-term disease activity. As such, it is less useful in early disease.

1.5.2 Ultrasound

Ultrasound is increasingly being utilised in both diagnosis and management of RA. The lack of radiation, and low cost associated with ultrasound imaging make it an extremely attractive imaging modality. Its limitations are that it requires specifically trained operators to undertake the scan who are able to interpret the scan results both relative to the site of imaging, as well as having an understanding of the pathology. Because the scan is dynamic, it is difficult to obtain standardised images which makes ultrasound images less reliable for longitudinal assessment of joint changes for clinical and research purposes.

Ultrasound imaging can identify subclinical synovitis, increased synovial thickness, joint swelling (including effusions) and increased vascularity. It is also able to identify early erosions that would not be visible on plain film x-rays (98).

1.5.3 Magnetic resonance imaging

MRI is the most recent imaging modality and is able to examine all the aspects of joint swelling and synovitis identified by ultrasound. In addition, MRI scanning can identify intra-osseous changes (including bone oedema) that are not visible on ultrasound or x-ray. With advanced machines, 3D reconstructions can be created. The difficulty with MRI images is that the technique can identify such subtle changes, it can be difficult to know the clinical significance of such changes, although this will likely change as the evidence base supporting this technique increases.

Whilst an excellent imaging modality, it is limited by the cost of equipment, the requirement for highly skilled radiologists to interpret scans, the size of the equipment, and requirement for magnetic shielding that often needs to be built into the fabric of

the building housing the scanner. It also has high running costs and cannot provide dynamic images (99).

1.6 Classification criteria for RA

The first classification criteria for RA were defined by the American Rheumatism Association (ARA; subsequently the American College of Rheumatology; ACR) in 1956 and revised in 1958. These criteria contained eleven criteria which included clinical (joint pain, morning stiffness, joint swelling, symmetry of swelling, rheumatoid nodules), laboratory (serum RF, synovial biopsy and nodule biopsy) and radiographic criteria (plain film evidence of RA e.g. erosions). In 1966, simplified classification criteria were proposed. The New York criteria had only four criteria (tender and swollen joints, serum RF and radiographic findings), but was more cumbersome and never gained widespread use. In 1987, the ARA updated the classification criteria for RA to be more streamlined than the original 1958 format (100) and are outlined in Table 1. The criteria required at least four of the seven criteria to be satisfied, with clinical symptoms being present for at least six weeks.

Criterion	Definition
1. Morning stiffness	Morning stiffness in and around the joints, lasting at least one hour before maximal improvement
2. Arthritis of three or more joint areas	At least three 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 one 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 juxta-articular regions, observed by a physician
6. Serum RF	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 RA 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 1. 1987 ACR Classification criteria for RA (100)

These criteria were updated in 2010, when a combined ACR/EULAR taskforce was formed to make classification criteria more applicable to earlier onset disease (i.e. before radiographic and nodular changes) in line with the changes in treatment paradigms since 1987 (101). The score was designed to be used in patients who have at least one joint with definite clinical synovitis which was not better explained by another disease. The 2010 update also included ACPA antibody status, and does not require radiographic or nodular changes to be present.

Classification criteria for RA (score-based algorithm: add score of categories A–D; a score of 6/10 is needed for classification of a patient as having definite RA)	
A. Joint involvement	
1 large joint	0
2-10 large joints	1
1-3 small joints (with or without involvement of large joints)	2
4-10 small joints (with or without involvement of large joints)	3
10 joints (at least 1 small joint)	5
B. Serology (at least 1 test result is needed for classification)	
Negative RF and negative ACPA	0
Low-positive RF or low-positive ACPA	2
High-positive RF or high-positive ACPA	3
C. Acute-phase reactants (at least 1 test result is needed for classification)	
Normal CRP and normal ESR	0
Abnormal CRP or abnormal ESR	1
D. Duration of symptoms	
<6 weeks	0
≥6 weeks	1

Table 2. Updated 2010 ACR/EULAR classification criteria for RA (102)

1.6.1 Pre-clinical phase of RA

In recent years much work has focussed on the development of RA and what is now identified as 'pre-clinical RA' (103). Pre-clinical RA is split into three phases A-C, which includes genetic risk factors (phase A), environmental (phase B) and systemic autoimmunity without arthritis (phase C). The increasing understanding of the different pre-clinical (described previously; 1.3.1, 1.3.3, 1.2.4) phases has led to hopes of identifying factors that could prevent the disease from ever manifesting (55).

1.7 The development of treatments for RA

1.7.1 The pre-pharmacologic era

Prior to 1899, there were very limited treatments for RA. Thomas Sydenham recommended extended bloodletting, purgatives and dietary restrictions (12) and willow or poplar bark have been used since ancient Greek times (subsequently identified as a natural source of salicylic acid and the active component of aspirin) (104). Another popular treatment was spa or balneotherapy. Certain spas in Europe, specifically catered for individuals with musculoskeletal conditions such as RA (105). Renoir, who suffered with aggressive RA from the age of 50, used to regularly attend the spa at Vichy, Bourbonne-les-Bains, and Aix-les-Bains and would often attend the spas for months on end, travelling with his family, staff, piano and pet parrot (106).

1.7.2 Aspirin

The development of Aspirin by Bayer plc. in 1899 was the first effective treatment available for individuals with RA. Touted as the 'miracle cure' for rheumatism, it was widely used, often with high doses and associated side effects such as ototoxicity and gastrointestinal bleeds.

The active component of aspirin is salicyclic acid and its anti-inflammatory effect is mediated through its ability to prevent the formation of prostaglandins and thromboxanes via the cyclooxygenase 1 and 2 pathways. Whilst the anti-inflammatory effect can be efficacious in reducing symptoms related to RA, it does not modulate the key biochemical pathways (outlined previously) that drive RA, and hence is not a disease modifying drug in RA.

1.7.3 Gold

Gold was the first disease modifying treatment used in the treatment of rheumatoid. Its use for RA was first suggested by Jacques Forestier, a rheumatologist at Aix-les-Bains (the spa that Renoir frequented for treatment of his RA) in the 1920s, and was in some cases, efficacious in inducing remission. However, it has significant side effects which become more common as the cumulative dose to the individual increased. Side effects included, renal and liver impairment, as well as ocular toxicity. Systemic build-up of gold salts in the skin over prolonged periods of time also led to a slate-grey appearance of the skin. Sudden onset acute pulmonary distress was also noted to occur sporadically, occasionally with devastating consequences (104)

The pharmacological action of gold (sodium aurothiomalate) is not fully understood, but it is known to inhibit the synthesis of prostaglandins and major histocompatibility complex (MHC) II peptide interactions as well as interfering with B-cell function (107).

1.7.4 Corticosteroids

It was another 20 years before the next major breakthrough in the treatment of RA with the discovery of corticosteroids by Dr Philip Hensch and colleagues at the Mayo Clinic in 1948 (108). Hensch and his collaborators had noticed jaundice, along with other conditions such as pregnancy, infection and the post-surgical state, all appeared to temporarily improve the symptoms of RA. This lead Hensch to develop the hypothesis that adrenal hormones may be the cause of this clinical improvement. Investigation into corticosteroids was initially delayed due to the Second World War, however, in 1948 the first use of cortisone (known at the time as 'Substance E') for RA was given to a 29-year-old patient who had been chair-bound due to her disease. After four days of treatment, the response was so profound that the patient was able to walk out of hospital. Hensch was keen to validate his findings further, however, despite only initially reporting his findings at the routine Mayo clinic weekly Wednesday evening physicians meeting in 1949, the results of the trial became widely publicised, with corticosteroid therapy being widely adopted. In 1950, Hensch and his collaborator

Edward Kendall, shared the Nobel prize for Medicine or Physiology (109). However, within a few years of being widely adopted, the long-term adverse events associated with chronic steroid therapy became evident. Steroid exposure is now minimised as much as possible, to short-term therapy while establishing longer term disease modifying therapy, or as short term flare management (19).

The anti-inflammatory effect of corticosteroids acts by binding with intracellular glucocorticoid receptors within the cytoplasm of cells. Binding of corticosteroids to these receptors facilitates translocation to the nucleus where it interacts with glucocorticoid response elements which both blocks the transcription of inflammatory genes whilst promoting the transcription of anti-inflammatory proteins.

1.7.5 Methotrexate

Shortly after the success of cortisone, methotrexate was first trialled in RA in 1951 by Gubner et. al. (110). Despite Gubner et. al. and others (111-114) demonstrating clinical effectiveness, uptake in clinical practice was slow, partly due to a reluctance to use an anti-cancer drug in a condition regarded as 'benign' (115). Additionally, steroid treatments were seen as so efficacious (and the risk of long-term corticosteroid therapy had yet to be identified), there was little appetite for using a potentially risky anti-cancer drug in treating rheumatoid. However, use of methotrexate became more widespread following randomised controlled clinical trials through the 1980s (116-119), and head-to-head studies in the 1980s and 1990s (120-123), finally replacing intramuscular gold as the standard of care for RA in the 1990s. Demonstration of success of combination therapy (124) and earlier aggressive treatment with DMARDs resulted in a reversal of the treatment paradigm of 'start low, go slow' to an aggressive early intervention with combination therapy and a 'treat to target' strategy (125).

Methotrexate is an anti-folate agent and has a number of mechanisms of action that are thought to be central to its disease modifying properties in RA. Methotrexate suppresses T-cell activation as well as down-regulating B-cell function and inhibiting the binding of IL1β to its receptor. It is also a potent inhibitor of dihydrofolate

reductase which catalyses the conversion of dihyrofolate to the active form, tetrahydrofolate. Folic acid is essential in the production of the nucleic acid thymidine, so inhibition of this impairs DNA, RNA and protein synthesis. Evidence also suggests that methotrexate may be involved in promoting adenosine release. Adenosine appears to have mixed pro- and inflammatory actions, depending on receptor and cell type, but methotrexate appears to promote the anti-inflammatory actions (126).

1.7.6 Sulfasalazine

Parallel to the development of methotrexate, sulfasalazine was first licenced by the US Food and Drug Agency (FDA) in 1950 following first trials into the treatment of RA in 1942 (127). Sulfasalazine has demonstrated efficacy in modifying disease activity in RA and remains a widely used first-line treatment and in combination therapy with methotrexate. It is a pro-drug that is broken down into its active components in the gut (sulfapyridine and mesalazine). Sulfapyridine is then subsequently absorbed systemically, while the majority of the mesalazine remains in the colon. In vitro studies have demonstrated suppression of expression of IL-1, IL-2, IL-6, IL-12 and TNF, although exactly how sulfasalazine acts is unknown. Sulfasalazine also reduces synovial hyperplasia and chemotaxis of inflammatory cells to the joints. Numerous studies have shown sulfasalazine to be efficacious as a disease modifying drug in RA, both independently and in combination with other disease modifying anti-rheumatic drugs (DMARDs; usually methotrexate) (125,128). It has advantages over methotrexate in that it is safe in pregnancy and breastfeeding, although it can cause azoospermia in men.

1.7.7 Hydroxychloroquine

Hydroxychloroquine, and its sister-drug chloroquine, are antimalarial agents that have particular efficacy in RA. As with many of the older synthetic disease modifying agents, its exact mechanism of action remains unknown, however it has been shown to raise the pH within lysosomes which reduces the efficacy of antigen presentation, and interferes with toll-like receptors. They are clinically efficacious in RA, although their

impact on radiographic progression is less impressive than other DMARDs (such as methotrexate and sulfasalazine). As such, hydroxychloroquine monotherapy is rarely used in all but the most mild cases of RA, although it is frequently used in combination DMARD therapy. It is a generally well-tolerated drug and, like sulfasalazine, it is safe for use in pregnancy and breastfeeding. Its most serious side-effect is that of retinal toxicity, which although rare, is associated with life-time cumulative dose, so vigilance of individuals on long-term hydroxychloroquine therapy is essential (129).

1.7.8 Other synthetic disease modifying therapy and novel synthetic DMARDS

In addition to the aforementioned drugs, other DMARDs have been used in the treatment of RA, including leflunomide, azathioprine, cyclosporine and others. However, methotrexate, sulfasalazine and hydroxychloroquine remain the most widely used DMARDs in first-line pre-biologic therapy for RA in the UK. More recently, novel targeted small molecule DMARDs have been developed that target JAK enzymes, important in the inflammatory pathway associated with RA.

1.7.9 Anti-TNF

The next transformative step in the treatment of RA was the development of anti-TNF. As previously discussed (1.2), TNF plays a key role in many of the key pathogenic pathways of RA. Blockade of TNF signalling has a pleiotropic array of actions both directly on cells via the TNF receptor, as well as indirectly, through subsequent down regulation of key inflammatory cytokines IL-6, IL-1 and many others. Extensive work in the 1980s and 1990s identified the importance of the cytokine in RA and led to it being identified as a potential target for therapy. In addition, the development of molecular techniques that enabled the generation of specifically targeted monoclonal antibodies allowed the possibility to ‘design’ and manufacture monoclonal antibodies that could harness the findings from research into clinical treatments (130).

The first clinical trial of anti-TNF in 20 patients with RA occurred in 1992 at Charing Cross Hospital, London. Although open label, with no placebo control and lasting only eight weeks, the results were profound (131), with all 20 patients demonstrating an improvement in their arthritis. Following the success of this initial trial, further investigations were undertaken, resulting in a multicentre placebo-controlled, double-blind, randomised trial which demonstrated prolonged efficacy of treatment with repeated administration of drug (132). The demonstration of clinical effectiveness of anti-TNF blockade in RA led to rapid development of other monoclonal antibody therapies targeted at TNF (133). Initial, and subsequent clinical trials (132,134) demonstrated, not only an improvement in clinical signs and symptoms, but also an apparent halting of radiographic progression of damage, something that had not been previously identified in studies of methotrexate (135). However, even in early studies, it was noted that response rates to anti-TNF were between 60-80% of patients, and exact reasons for non-response were not clear (29). In addition to variations in response, the potential for immunogenicity by anti-TNF was recognised, and high anti-TNF dosages, and combination with methotrexate was noted to limit this response (136).

1.7.10 Other biological therapies

As further advances in understanding of RA have identified other key cytokines involved in the pathogenesis of RA, additional targeted monoclonal antibodies have been developed, including anti-IL6, anti-CD20 and more recently, anti-IL17.

1.8 Evolution of treatment paradigms for RA

With the increasing array of effective disease modifying treatments for RA, and a greater understanding of the epidemiology, immunopathology and progression of the disease, the approach to managing the condition has changed dramatically over the past 30 years.

Historically, treatment of RA was often reactive in nature – with treatment escalation or intervention initiated following evidence of disease progression. Use of therapeutic agents was often cautious and escalation of drug doses slow, often described as 'start low, go slow'. However, as studies of more aggressive treatment strategies, particularly the COBRA (137), BeST (138) and TICORA studies (125), demonstrated superior outcomes to traditional more cautious dosing strategies, a new consensus emerged amongst the rheumatology community of the importance of earlier diagnosis and instigation of disease modifying drugs, before joint damage was evident. With the wider array of therapeutic agents available, better outcomes for more patients seemed more attainable. The parallel development of widely accepted outcome measures (such as the DAS28; covered in more depth in Chapter 2) also enabled treatment targets to become more standardised, and allowed treatment goals to be set, with progress towards these goals monitored more easily. With the publication updated ACR/European League against Rheumatism (EULAR) guidance (139), the 'treat-to-target' approach has become the standard of care, and the wide array of both synthetic and biological DMARDs now available, remission has become the *de facto* initial target for all newly diagnosed RA, and the majority of existing RA patients (140).

1.8.1 The National Institute for Health and Care Excellence (NICE)

In the UK, the use of anti-TNF for the treatment of rheumatoid arthritis is directed by guidelines laid down by NICE. NICE became a legal entity in 1999 and its primary remit was to assess the evidence-base for treatments (including drugs, procedures, and since 2010 also standards of care) offered by the NHS. Although the cost-benefit of treatments was not initially included as a necessary step in the process for approvals of treatments, it was soon added. The initial NICE approval for the use of anti-TNF in RA required patients to have persistent high disease activity (defined as a DAS28 score of >5.1 on two occasions a month apart) despite treatment with at least two synthetic DMARDs (one of which should be methotrexate) at the maximum tolerated dose. Whilst there have been subtle modifications to the recommendations for the use of anti-TNF as additional biological treatment agents have become available, the principal of anti-TNF being a second-line agent for use in high disease activity remains. This has

important implications for the recruitment of patients in the BSRBR-RA (discussed in chapter 4) and means that there are very few patients with a baseline DAS28 score of less than 5.1, or who have had less than one synthetic DMARD prior to starting anti-TNF.

1.8.2 Sustainability of remission

Whilst remission remains the target of treatment, most studies focus on the achievement of outcomes at a predetermined point in time, often at 6 or 12 months after initiating therapy. This is understandable as it fits with a standard clinical trial design paradigm for testing *a priori* hypotheses. However, the outcomes of trials at such single time points do not represent a permanent state. Patients and clinicians know that RA is a condition that waxes and wanes with flares of disease activity. A single cross-sectional measurement of remission rates does not give a clear picture of outcomes over time, something that is of keen interest to both patients and clinicians. However, before sustained remission can be measured and investigated, a key hurdle needs to be overcome: how long should sustained remission be? Ideally, it should be sustained for life – "*the restoration of health" – Oxford English Dictionary* (1). However, such an end point is both an extremely rare occurrence and impractical to measure, given it necessitates following up patients until death. Deciding the duration of sustained remission also requires decisions to be made on the frequency of measurements of disease activity. Therefore, how often should disease activity be measured to define a patient as being in 'sustained remission'? These questions are discussed in more depth in chapters 2, 7, 9 and 10.

1.8.3 Development of the BSRBR-RA

The British Society for Rheumatology Biologics Register for RA (BSRBR-RA) was established in 2001 as part of a Europe-wide initiative to monitor the safety of anti-TNF treatments for RA (141-143). Initially planned to last five years, the BSRBR-RA has evolved as more anti-TNF medications have come to market, and has now expanded to include all biologic-class medications.

The BSRBR-RA is now one of the largest biologics registry in the world, with over 25,000 individuals followed up over a period of time spanning 15 years. The wealth of data collected in the BSRBR-RA allows a broad spectrum of studies to be undertaken. The work undertaken using the data from the BSRBR-RA has been hugely influential in developing the evidence base for the safety and efficacy of biologic agents, and has been used as a model globally for other biologics registries and for other diseases. The length of data collection now also allows a comprehensive longitudinal assessment of disease activity and sustained remission to be undertaken. The register (including the background to its inception and methods) is described in more detail in Chapter 3.

1.9 Chapter summary

The diagnosis and management of RA has changed remarkably in a relatively short time. In the space of 100 years, RA has evolved from being a condition that was barely recognised by most clinicians with no effective treatment options; to being a condition with a panoply of treatment options; evidence–based national and international standards that recommend that clinicians identify the condition and commence disease modifying treatments within three months, aiming for complete elimination of symptoms. However, despite excellent progress, not every patient achieves remission. Some patients achieve remission only temporarily. Some have prolonged period of efficacy from their disease modifying agent before temporarily, or permanently, losing therapeutic benefit from the drug. Whilst clinical trials point to great improvements in outcome with use of many drugs (anti-TNF in particular), it is not clear to what extent these outcomes are experienced by patients outside the randomised controlled trial (RCT) setting. Given the wide choice of drugs available for patients, clinicians face a new challenge: which drug is best for each patient?

This thesis seeks to begin to address these questions for the most widely used class of biologics collected by the BSRBR-RA; anti-TNFs. It will explore how frequent and sustainable optimal outcomes are outside of the clinical trial setting. However, as with most aspects of RA, even the definitions of ‘optimal outcomes’, ‘remission’, and

'sustained remission' are heterogeneous, and require further consideration before it can be examined as an outcome. This will be explored in greater depth in Chapter 2.

Chapter 2

2 Returning to remission – A review of disease activity scores in RA

The advance in available drugs and the treatment paradigms have dramatically improved outcomes in RA. However, whilst the cure remains elusive, remission is increasingly becoming attainable for an increasing proportion of patients with RA. To begin to understand what remission is, it is helpful to return to the Oxford English Dictionary (1), where the definition of remission in a medical context is defined as the:

> *'Lessening of the severity of a disease or symptom; disappearance of symptoms or cessation of the activity of a disease for a period; an instance of this.'*

At first glance, such a definition of remission appears to delineate a clear state that could be identified and applied to RA. However, on closer inspection, the definition of remission begins to lay out the many difficulties in both achieving and identifying such a state. A myriad of questions arises including: How to identify the severity of the disease or symptoms? Is severe disease one that causes disability, joint pain or swelling, fatigue, or one of the many other symptoms that can be caused by a multisystem disease such as RA? What symptoms should dissipate? How long should the cessation of disease last to count as remission?

Therefore, before starting to investigate remission in RA, three overarching questions first require answering:

1. How to measure disease activity in RA in a valid and reproducible manner?

2. What constitutes remission when using a validated measure i.e. what is the threshold at which remission should be set?

3. How long should remission last for when it does occur i.e. what is the minimum time remission should be sustained for?

The first challenge lies in identifying a robust measure of disease activity for RA which allows quantification of severity of the disease. Different approaches to this have evolved over time, and the development of these scores have been essential to build an evidence base by which to treat patients in the most efficacious manner.

2.1 Measuring disease activty – the development of composite disease activity scores

Due to the systemic nature of RA and its propensity to affect a wide range of synovial joints, as well as symptoms of early morning stiffness and fatigue and pain, establishing a quantified assessment of 'disease activity' has been difficult. Early development of a disease activity index included the Ritchie Articular Index (RAI) (144) in 1968. However, this only measured joint pain and did not assess joint swelling or inflammation, both hallmark features of RA. Whilst intra-rater reliability was good, inter-rater reliability was poor. This posed problems in achieving a consistent measure of disease activity, not only in clinical practice, but also in clinical trials, where a lack of consensus made quantifying and comparing drug efficacy between trials especially challenging. There also remained an absence of any international consensus on which outcomes should be measured in RA, and how they should be quantified. This lack of consensus led to difficulties in comparing outcomes across clinical trials and slowed development of evidenced-based treatment, and in turn led to the formation of an international group of rheumatologists to define a core set of outcome measures to be used in RA clinical trials, subsequently known as OMERACT. In 1992, the first meeting of OMERACT led to the agreement of the first core set of data to be collected in RA clinical trials (145).

In 1990, van der Heijde *et al*. had developed a composite disease activity score (DAS) (146) which was subsequently modified to the disease activity score in 28 joints (DAS28) (147). These composite scores became the most widely accepted measures of disease activity in RA, had good inter- and intra-rater reliability and were sensitive to change. They also had the advantage of combining key measures from the OMERACT core outcome set into the composite score. Widespread adoption of the core data set specified by OMERACT and standardised outcome measures meant that for the first-time comparisons between multiple clinical trials became feasible. Over subsequent years, as limitations with the DAS28 were identified, new composite scores were developed, including the SDAI (148) and the Clinical Disease Activity Index (CDAI) (149). Increasingly, the evidence base for fully patient reported outcome measures has also evolved, with the development of numerous scores, including the Routine Assessment of Patient Index Data 3 (RAPID3), Rheumatoid Arthritis Disease Activity Index (RADAI), Patient Activity Score (PAS), Patient Activity Score–II (PAS-II), Patient-based Disease Activity Score without ESR (PDAS-2) and Rheumatoid Arthritis Impact of Disease (RAID) (150).

2.1.1 The disease activity score (DAS)

The DAS (146) was the first composite outcome measure to include physician, patient and laboratory assessment of RA disease activity into a single score. This involved combining the RAI measurement of tenderness in 44 joints, with an assessment of joint swelling, a patient global assessment of health and an inflammatory marker (the ESR). To establish content validity of the score, analysis of clinical records was used; if clinicians increased drug doses or changed treatments this was classified as active disease, and low disease activity was identified if drug doses were unchanged for one year or reduced. The components of the score were given differing weights:

Tender joint count (RAI)> Swollen joint count > ESR > Global health assessment

And was defined by the equation:

$$DAS = 0.53938*sqrt(RAI) + 0.06465*(swollen\ joint\ count) + 0.330*ln(ESR) + 0.00722*(general\ health)$$

The threshold for remission with the DAS was set at ≤1.6, corresponding to the attainment of the ARA 1981 remission criteria (151) (discussed in Chapter 2; 2.3.1), although of note, no time element was specified in the DAS definition of remission, meaning that the DAS record of remission represented a 'snapshot' of the clinical scenario, and gave no indication of the durability of the clinical state of remission.

2.1.2 The Disease Activity Score of 28 Joints (DAS28)

Five years after the development of the DAS, Prevoo *et al.* (152) modified the score to include 28 joints (omitting all the joints of the feet) rather than the original 44 joints assessed by the DAS. Whilst the DAS was a reproducible and valid score, the assessment of 44 joints was deemed to be cumbersome in clinical practice and had limited uptake in routine practice. In addition, instead of grading tenderness at each of the 44 included joints (between 0 – 3), the DAS28 opted for a binary assessment (yes/no) of tenderness at each joint. The streamlining of the DAS made the global measurement of the activity of RA more achievable in routine clinical care and it became the standard method of recording disease activity in clinical trials. Validation demonstrated robust construct, criterion, concurrent and content validity of the score, with the inflammatory burden of the disease quantified by the ESR. In 2003, Fransen *et al*. (153) presented evidence that the CRP could be used as a substitute for the ESR, with similar scores that could be used interchangeably with the DAS28-ESR. Subsequent work by Wells *et al.* in 2009 (154) demonstrated sensitivity to change and criterion and construct validity of the DAS28-CRP, although the study identified that the DAS28-CRP generally generated scores lower than the DAS28-ESR. Over the years, the use of DAS28-CRP has become widespread, as the cost of measuring CRP has fallen. The adoption of the DAS28-CRP has also been seen in clinical trials due to the ability to transport and measure CRP at centralised laboratory facilities, making inter-site

variability less of an issue when compared with the ESR (which must be determined locally, and is difficult to standardise in the same way as CRP measurement). However, robust evaluation of the validity of DAS28-CRP as a direct substitute for DAS28-ESR has only recently started to be analysed, and has been found to generally provide a slightly lower score when compared with the DAS28-ESR (155-157). However, despite the emerging differences between these scores, the ESR and CRP iterations of the DAS28 are still used interchangeably in clinical and epidemiological practice, as well as in national guidelines (158).

The equations used to calculate the DAS28-ESR and DAS28-CRP are outlined below:

*DAS28-ESR = 0.56*sqrt(28 tender joint count) + 0.28*sqrt(28 swollen joint count) + 0.70*Ln(ESR) + 0.014*(global health score)*

*DAS28-CRP = 0.56*sqrt(28 tender joint count) + 0.28*sqrt(28 swollen joint count) + 0.36*ln(CRP+1) + 0.014*(global health score)+ 0.96*

The disease activity thresholds for the DAS28 were independently derived from the original DAS (159), and achievement of remission according to the ARA 1981 remission criteria (160), (161). The thresholds in use are:

Remission < 2.6
Low disease activity (LDA) 2.6 – 3.2
Moderate disease activity (MDA) >3.2- 5.1
High disease activity (HDA) >5.1

2.1.3 Simplified Disease Activity Score (SDAI)

Despite the widespread adoption of the DAS28, limitations were noted in several areas. One of these was in the classification of 'remission'. The DAS28, and its precursor, the DAS, were developed at a time when the concept of complete remission was a rarity. However, as treatment of RA became more successful (due to more aggressive

treatment and new drugs), it became increasingly apparent that the definition of remission provided by the DAS28 was relatively lenient. With the weighting used in the calculation of the DAS28, it is possible for a patient to have a number of tender and swollen joints and still be classified as being 'in remission'.

To counter this, the SDAI was developed in 2003 by Smolen *et al.* (148) to provide a more stringent measure of disease activity, particularly at lower levels of disease activity, which were increasingly being achieved. Like the DAS28, it combined the domains of tender and swollen joints (using the assessment of the same 28 joints as the DAS28), a patient self-report of overall disease activity (the patient global health assessment; PGA) and an inflammatory marker (the CRP). However, in addition to these, a physician rating of disease activity was included to encompass a physician's general feeling of how active the disease was. These measures are simply added together with no formal statistical weighting, making the calculation of the score easier.

The equation for calculating the SDAI is:

SDAI = Tender Joint Count (0 – 28) + Swollen Joint Count (0 – 28) + Patient Global Score (0 – 10) + Clinician Global Score (0 – 10) + CRP in mg/dl (0 – 10)

Disease activity thresholds for the SDAI are:

Remission 0.0 – 3.3
Low disease activity 3.4 – 11.0
Moderate disease activity 11.1 – 26.0
High disease activity 26.1 – 86.0

2.1.4 Clinical Disease Activity Score (CDAI)

The CDAI was a more streamlined version of the SDAI developed by Aletaha *et al.* (149) to allow disease activity scoring to be undertaken without the need for any laboratory

test results. It involves all the disease activity parameters included in the SDAI without the inflammatory marker parameter (CRP). Again, as with the SDAI no complex computation is required. The score is more readily used in clinical decision-making because scores are instantly available without having to wait for laboratory results.

Disease activity thresholds for the CDAI are:

Remission 0.0 – 2.8
Low disease activity 2.9 – 10.0
Moderate disease activity 10.1 – 22.0
High disease activity 22.1 – 76.0

However, despite the more rigorous remission threshold, no longitudinal time element was included in the SDAI and CDAI definition of remission.

2.1.5 Heterogeneity in practice

Global use of these outcome measures varies, but in the UK the DAS28 (ESR or CRP version) is still the most widely used score, partly because a classification of 'high disease activity' (according to the DAS28) is required for patients to be eligible for anti-TNF treatment (162).

Overall, despite differences in weighting of components of the score, there are many similarities between all four of the widely disease activity scores, as summarised in Table 3. However, despite all scores having a set threshold for remission, *none of them specify a duration for which remission should be maintained*.

Disease Activity Score	No. Tender Joints	No. Swollen Joints	Inflammatory marker	Patient global score (PGA)	Physician global score (PhGA)	Score range	Disease activity thresholds	Score Derivation
Disease Activity Score of 44 Joints (DAS)	44 (scored as the RAI)	44	ESR	Yes – Visual Analogue Score (0 - 10cm)	No	0 -10	Remission <1.6 Low ≤ 2.4 Moderate >2.4 - High >3.7	DAS = 0.53938*sqrt(RAI) + 0.06465*(SJ) + 0.330*ln(ESR) + 0.00722*(PGA)
Disease Activity Score of 28 Joints (DAS28)	28	28	ESR or CRP	Yes - Visual Analogue Score (0 - 100mm)	No	0 - 9.4	Remission <2.6 Low 2.6 - 3.2 Moderate >3.2 - 5.1 High >5.1	DAS28-ESR = 0.56*sqrt(TJ) + 0.28*sqrt(SJ) + 0.70*Ln(ESR) + 0.014*PGA DAS28-CRP = 0.56*sqrt(TJ) + 0.28*sqrt(SJ) + 0.36*ln(CRP+1) + 0.014*PGA + 0.96
Simplified Disease Activity Score (SDAI)	28	28	CRP	Yes- Visual Analogue Score (0 - 10cm)	Yes - Visual Analogue Score (0 - 10cm)	0 - 86	Remission 0 – 3.3 Low 3.4 – 11.0 Moderate 11.1 – 26.0 High 26.1 – 86.0	SDAI = SJ + TJ + PGA + PhGA + CRP (0 – 10)
Clinical Disease Activity Score (CDAI)	28	28	None	Yes Visual Analogue Score (0 - 10cm)	Yes - Visual Analogue Score (0 - 10cm)	0 -76	Remission ≤ 2.8 Low ≥2.8 - 10.0 Moderate 10.1 - 22.0 High 22.1 - 76.0	CDAI = SJ + TJ + PGA + PhGA

Table 3. Validated Disease Activity Scoring Criteria for RA

2.2 Patient reported outcome measures

In addition to the composite outcome measures which combine physician and patient perspectives on the activity of RA, entirely patient-reported outcome (PRO) measures are also increasingly being seen as accurate and reproducible measures of RA disease activity (150). PRO instruments provide information on health status known only to the individual, and not influenced by observers or objective testing. The most widely used entirely PRO measures in RA include:

- RAPID – There have been multiple iterations of the RAPID score, however, the RAPID3 version is the most widely used. RAPID3 is a composite outcome measure which combines the Stanford modified health assessment questionnaire (MD-HAQ) (163) with three additional patient-reported outcomes; physical function, pain and patient global assessment. The score has a range of 0 – 30 and has been shown to have a good approximation to outcome measures such as the CDAI and DAS28 (164).

- RADAI - The RADAI and RADAI-5 are both five item questionnaires which encompass questions about how active an individual's arthritis has been over the past six months as well as over the current day including pain and stiffness. The RADAI-5 has a 0-10 scale and has thresholds for mild, moderate and high disease activity. Its agreement has been assessed against the DAS28, CDAI and SDAI and found to have overall moderate agreement with composite scores (165)

- RAID - Developed in collaboration between OMERACT and EULAR, the RAID is a patient reported outcome measure which seeks to measure the impact of RA on an individual. It covers seven key domains of pain, function, fatigue, sleep, coping, and physical and emotional well-being. It has been shown to correlate well with other patient reported outcome measures such as the HAQ, RADAI, 36-Item Short-Form Health survey (SF-36) and European quality of life questionnaire including five domains (EQ-5D) (166), and physician/patient composite outcome measures such as the DAS28 (167).

- P-DAS – the P-DAS incorporates the Stanford health assessment questionnaire (HAQ) questionnaire as well as a patient global assessment of disease activity. However, in contrast to the aforementioned patient reported disease activity scores, the P-DAS has two versions; incorporating a patient self-assessment of tender joint counts and ESR into one version (P-DAS1); and swollen joint counts and early morning joint stiffness into the other version of the score (P-DAS2) (168). As with the other scores, the two versions of the P-DAS have been validated and show sensitivity to change, and have moderate agreement with the DAS28 and CDAI scores (169).

However, despite the increasing use of entirely patient reported outcomes, the relative novelty of many of the scores means that they were not incorporated in the original structure of data collection within the BSRBR-RA, and as such will not be examined in this thesis.

2.3 Defining 'remission' in RA

As outlined previously, the definition of remission requires focussing to be of any meaningful use in the context of RA. The first effort to formally define remission in RA was undertaken in the early 1980's (151). Subsequently developed disease activity scores all had individual thresholds of disease activity that were identified as remission, although when comparing between disease activity scores, the level of disease activity required to meet the criteria of remission are different. The inconsistency in defining remission led to an international collaboration between the ACR and EULAR to construct an internationally accepted definition of remission in 2011 (170).

2.3.1 ARA 1981 Criteria

When remission was first formally quantified in 1981 by Robert Pinals *et al.* (151) for the ARA (later the ACR), it was a rarely achieved goal due to limited treatment options.

The 1981 remission criteria were: 1. Duration of morning stiffness not exceeding 15 minutes; 2. No Fatigue; 3. No joint pain; 4. No joint pain on palpation or movement; 5. No soft tissue swelling of joints or tendons; 6. ESR less than 30 mm/hr for women and 20 mm/hr for men; all for two consecutive months.

In 2011, the criteria to define remission were updated by the ACR and EULAR (170).

2.3.2 ACR/EULAR 2011 Criteria

The ACR/EULAR joint statement on remission (170) utilised a stringent Boolean criteria including; ≤1 tender or swollen joint; CRP≤1 mg/dl; and a patient visual analogue score of ≤1 on a 0-10 scale; or an SDAI score of ≤3.3. No longitudinal time element was specified, and the ESR was replaced by the CRP as the measure of inflammation. Due to the issues with the leniency of the DAS28 definition of remission, this was not included in the ACR/EULAR 2011 criteria. Whilst the ACR/EULAR criteria provided a robust and stringent definition of what remission was, unlike the 1981 definition it replaced, it did not define the minimum duration that remission should last.

2.4 Further challenges in assessing disease activity

In parallel with the desire to quantify the activity and impact of a multisystem disease such as RA on individuals with the condition in a holistic manner (something which all composited disease activity measures attempt to do), a parallel challenge arises; trying to differentiate the impact of treatment on the inflammatory aspect of the condition (which disease modifying immunomodulatory treatments target) from the non-inflammatory pain and fatigue that can accompany the condition (and are not necessarily influenced by disease modifying immunomodulatory therapy).

For disease modification treatment to be effective, there should be an inflammatory component to the disease. In many cases, the pain, fatigue and stiffness experienced by

individuals with RA is a likely result of an active inflammatory component of the disease. In this situation, targeting inflammatory disease activity should reduce pain, stiffness and fatigue. However, where there is a disconnect between inflammatory disease activity and an individual's experience of pain, fatigue and stiffness, suppressing inflammation may not have the desired therapeutic of reducing overall symptoms. The effect of chronic pain on amplifying the perception of pain has been well documented (171), as has the difficulty for individuals to discriminate between inflammatory pain and chronic non-inflammatory pain (172). This demonstrates the dual strength and weakness of using a composite outcome measure. On the one hand, the generic nature of the components of the composite outcome measure (such as the visual analogue score) ensures that the true impact of the disease on the individual is represented; however, using such generic measures does not enable the specific cause of the symptoms to be identified (e.g. inflammatory or non-inflammatory) (84).

Different composite disease activity scores have sought to focus on assessing the global effect of inflammatory RA disease activity more precisely by incorporating objective measures of inflammation (ESR and CRP), physician global scales, and objective measures of clinical inflammation (e.g. swollen joint count), as well as weighting different aspect of the components. The DAS28 in particular gives greater weight to the patient reported components than physician or laboratory determined measures (such as the CRP or ESR). Research has also been undertaken on potential methods of delineating between inflammatory and non-inflammatory pain, as a method of identifying which patients may have a greater or lesser chance of responding to disease modifying treatment such as anti-TNF. Of particular note is work by Kristensen *et al*. (173) who identified that patients with a high swollen to tender joint count ratio (S:TJR) were 2- to 3- times more likely to achieve a good response (defined as an ACR50 response[1]) at 6 months when compared with patients who had a lower S:TJR, highlighting a possible measure that may identify instances where there is a disconnect between inflammatory activity and symptom severity.

[1] ACR50 response: defined as at least a 50% improvement in the number of tender and swollen joints, and a 50% improvement in at least 3 of the following: the patient's global assessment of disease status; the patient's assessment of pain; the patient's assessment of function (measured using the Stanford Health Assessment Questionnaire); the physician's global assessment of disease status; and serum C-reactive protein levels.

To further complicate matters, there is evidence that synovitis can persist despite clinical remission, including ACR/EULAR Boolean remission (i.e. when CRP is ≤1) (174). Anti-TNF has also been noted to have a role in modulating pain pathways rather than solely targeting inflammatory pathways (175,176). This raises the question as to whether even inflammatory markers are sufficiently sensitive or targeted enough to detect all RA disease activity, or indeed if improvements in disease activity in patients with normal inflammatory markers represents a degree of placebo response, or other non-inflammatory mechanism of action (Chapter 1; 1.4.1.3).

Whilst the answers to all these questions are beyond the scope of this thesis, they are important considerations when understanding remission and predictors of such a clinical state.

2.5 The patient perspective on remission

Arguably the most important perspectives on what remission is in RA should embrace the patient perspective. Development of the ACR/EULAR remission definition involved six patient expert advisors, although the only patient-directed component considered for inclusion in the definition was the patient global assessment of disease activity (PGA; measured by the VAS) and patient reported pain (170). Visual analogue scales encompass the otherwise unmeasurable multiple facets of disease in a single scale. However, such a tool is a blunt one, and a recent international qualitative study of remission in RA (172), demonstrated the extremely diverse patient perspective on what remission meant to them. Interestingly, the thematic analysis split the patient perspective into the 'symptoms' of remission (such as joint pain and stiffness) and the 'impact' of remission on daily life (such as independence, ability to work and participate in family life). Analysis highlighted that whilst absence of symptoms such as joint stiffness, and swelling (the focus of most clinically driven outcome measures) was of importance to patients, the ability to plan future events knowing the likely degree of severity of symptoms, was of central importance to many patients, highlighting the absolute importance of sustainability of response to patients.

2.5.1 Absence of an agreed definition of 'sustained remission'

With the exception of the ARA 1981 (151) criteria, none of the remission thresholds include any longitudinal time element, meaning that their scores represent remission at a single time point ('point remission') rather than a sustained period of remission. More recently, OMERACT has begun to address this issue and suggested six months as an acceptable length of time to consider remission to be 'sustained' (177).

2.6 Importance of sustained remission in RA

Whilst there is extensive evidence supporting the use of anti-TNF from clinical trials and registry data (178,179), most studies report outcomes at only a single time point. Studies of longitudinal sustained response rates to anti-TNF are sparse and evidence suggesting how often, and in which patients, sustained remission is achieved is lacking.

In addition, studies frequently use the term 'sustained remission' when what is reported is sequential point remission. This differentiation is important, as sequential point remission rates often do not confirm that those patients identified as being in remission at time point one are the same individuals as those in remission at time point two; meaning that sustained remission reported by these studies may refer to a stable proportion of a cohort remaining in remission, rather than the more clinically meaningful outcome of the proportion of individual patients remaining in remission over time.

One of the key issues surrounding reporting sustained remission is in defining how long 'sustained' should be. Ideally, a sustained response would be either permanent, or at least for many years. However, establishing a robust evidence base for remission lasting many years requires data collection on a scale larger than any of the current registries in existence, and beyond the scope of phase four clinical trials.

Whilst it is debateable if six months (as suggested by OMERACT) can really be classed as 'sustained' in the context of a lifelong condition such as RA, selecting longer

durations of remission presents increasing challenges with ever decreasing numbers of patients with sufficient data to classify having achieved increasingly stringent criteria.

In addition to defining how long sustained remission should last, the frequency of measurement of disease activity is another consideration. Measuring disease activity on an annual basis, potentially allows for a patient to have multiple flares of disease during the year, which would be missed by infrequent data capture. Whilst measuring disease activity on a very frequent basis (such as weekly or monthly), would provide a very rigorous assessment of longitudinal disease activity, and could provide reassurance that remission was indeed sustained, the resources required to carry this out for thousands of patients over many years (in the case of registries), makes this currently unfeasible. However, a six-monthly disease activity measurement is frequent enough to allow for reasonable confidence that there are no persistent or recurrent flares of disease between assessments, whilst making data capture feasible in a real-world setting.

2.6.1 Limitations of current evidence and challenges in defining sustained remission

The final parameter to consider when defining sustained remission is the outcome measure to use. Although newer composite outcome measures (such as the SDAI (148) CDAI (149)) and the ACR/EULAR definition of remission (101) have proposed more stringent criteria for remission, their relatively recent development markedly reduces how much longitudinal data are available, an important consideration when assessing sustained remission outcomes. In addition, and relevant to this thesis, is the fact that the BSRBR-RA has collected data using the DAS28 score. Therefore, for the purposes of the analyses in this thesis, the DAS28 will be used to assess outcomes. Despite having a less stringent criteria for remission, the DAS28 remains one of the most widely used outcome measures in clinical and research practice globally, as well as providing the reference score for NICE guidance on access to biologics in the UK (162). Therefore,

findings from research using this outcome measure have the potential to have the greatest impact on clinical and research practice.

2.6.2 Pragmatism in clinical practice

As outlined previously, all the outcome measures a have a threshold score by which remission is defined. However, using thresholds can present their own difficulties in real-life clinical practice, and are a relatively blunt tool. Use of thresholds artificially dichotomises a continuous outcome scale into a categorical outcome (e.g. remission = yes/no). Thresholds are necessary to define remission in a reproducible manner, and therefore essential when undertaking comparative efficacy analyses between different groups or studies. However, application of disease activity thresholds at an individual level, and indeed targeting clinical thresholds, without consideration for other clinical factors, is unrealistic and unrepresentative of real-world clinical practice. For example, in the case of the DAS28, an individual with a DAS28 score of 2.59 at one time point would be classed as in remission. If the score was repeated six months later and was 2.61, this would be classed as being in LDA, and identified as a 'worsening of disease activity' and loss of remission. However, the difference between these two scores can be attributed to a 1mm difference on a visual analogue scale, with all other components of the DAS28 being identical. In clinical practice, such a difference in scores would not represent a clinical deterioration in disease activity, but by using disease activity thresholds to categorise scores, it would be recorded as so (a failure to maintain remission). In fact, using disease activity thresholds in this circumstance would mean a DAS28 score of 2.61 is classed in the same group as a DAS28 score of 3.2, a disease activity score that can only be achieved by substantial worsening of multiple components of the disease activity score. Therefore, whilst sustained remission is the goal of treatment, and treat-to-target recommendations suggest modifying and escalating treatment until a state of remission is achieved (180), clinical treatment decisions are often much more pragmatic and are poorly modelled by applying strict disease activity thresholds. Such real-world pragmatism represents a challenge when applying findings from clinical research (where trial protocols generally attempt to minimise such 'pragmatism') to the clinic. Taking this, and the fact that the test-retest

reliability for the DAS28 is reported as 0.19 (181), as well as the inter-score differences between the two versions of the DAS28, (further discussed in Chapter 7), the issues with using a set threshold for defining remission in RA become evident.

This leads to the question: are the traditional cut-off or threshold values for defining remission, low, moderate or high disease activity, the most clinically realistic method to identify individuals who would be described as achieving a good sustained response to their drug at an individual level?

2.6.3 New models to identify sustained good response

With increasing availability of computing power, more flexible and clinically realistic analyses of registry data are increasingly possible. One such method is trajectory analysis (also known as latent class mixed modelling; LCMM). An analysis of the mean trajectory of response provides a more realistic model of the situation encountered in the real-world setting. In clinical practice, the overall trend of disease activity scores is considered by the clinician when assessing a patient's response to therapy, rather than each disease activity score in isolation. To model this situation more accurately, trajectory mapping considers previous and subsequent disease activity scores when plotting an individual's response to drug over time, and groups individuals into common trajectories of response. The overall trend of results allows for 'blips' in scores or isolated disease flares that may cross a disease activity threshold. Each individual's trajectory of response is mapped, before group trajectories are identified to form common trajectories, independent of pre-determined disease activity thresholds. This trajectory mapping provides a much more accurate picture of clinical practice, and has the potential to identify associations that are more applicable to the real-world clinical environment and to the individual patient. Such an approach has previously been utilised effectively by a Dutch early arthritis cohort (182) which identified three distinct response trajectories following initiation of a treat-to-target strategy.

Once trajectories of common responses have been identified, the next step is to trace common trajectories back to baseline data, before commencement of therapy, with the

aim of identifying commonalties between individuals within each group, as well as differences between groups. If associations between clinical or demographic features, and long-term responses can be identified they may be much more relevant than associations with good threshold-defined responses at single time points.

2.6.4 Strengths of using registry data

Because clinical trials pre-select patients based on a rigorous set of inclusion criteria to create a homogeneous population, it may be challenging to identify possible clinical or demographic features that may predict good response to a drug. However, registry data allow heterogeneity in the population, which not only allows for analysis of predictors of sustained remission, but also more closely represents what occurs in the real-world usage of such drugs, offering an ideal setting to explore responses to drugs over time.

2.6.5 Considering the population

A consideration when investigating remission is the population under investigation. In the UK, the National Institute for Health and Care Excellence (NICE) has stringent rules regarding commencing anti-TNF treatment; patients must have sustained high active disease (defined as a DAS28 score of >5.1 on two occasions one month apart), and have failed at least two traditional synthetic DMARDs, one of which should be methotrexate (162). As a result, the cohort of patients on anti-TNF in the UK (and therefore collected by the BSRBR-RA) are a selected group of patients with persistent high disease activity (at least one month), resistant to multiple treatments, and have had their disease for at least six months before starting anti-TNF treatment. This practice differs markedly from many other European and North American practices, where access to anti-TNF is much less tightly controlled, and is likely to influence the demographics of those individuals included in biologic registries. This in turn will influence the associations identified in study findings, and should be considered in the analysis of results.

2.7 What might the implications for clinical practice be?

2.7.1 How will understanding predictors of sustained remission help in the management of RA?

Anti-TNFs now provide a central plank in the management of RA. With the advent of the post-patent biosimilar era, the role of anti-TNF in management of RA is likely to increase as cost barriers to use are reduced. In addition to anti-TNF, numerous other targeted biological medications are now available, as well as a new generation of synthetic small molecule disease modifying medications for the treatment of RA. In 30 years, the challenge faced by rheumatologists has shifted from finding new drugs to treat RA, to one of knowing which of the diverse array of available drugs to choose for the individual patient. Most disease modifying drugs (both biologic and conventional) used in RA take a minimum of 2-3 months to demonstrate efficacy in a patient, meaning a sequential, one-size-fits-all protocolised approach to selecting drugs for RA has the potential to introduce prolonged periods of poor disease control and joint damage if a patient does not respond to the 'first-line' drugs in the standard treatment template offered. In addition, if a patient has had multiple failed trials of disease modifying drugs, they may resort to accepting a modest response from a drug, rather than being appropriately guided to a treatment that may have a higher likelihood of achieving remission.

In addition to the clinical problems, the cost implications of failed trials of drugs in patients is significant. By developing a personalised 'likelihood of response' to different drugs, based on large-scale, real-world data, failed trials of drugs can be minimised, with improved clinical outcomes and reduced cost. By actively selecting patients for drugs that they are most likely to respond to, the cost per quality adjusted life year (QALY) added and incremental cost effectiveness ratio (ICER) could be reduced, which has the potential to reduce cost-burden of RA on individuals and healthcare systems.

By using findings from large real-world longitudinal data, it should be possible to identify associations that could enable the use of a patient's clinical and demographic profile to target therapy and maximise the chance of responding optimally to a drug.

2.7.2 Towards personalised care: can we predict who will achieve sustainable remission on anti-TNF?

Through the work outlined above, a clearer understanding of the long-term real-world efficacy of anti-TNFs used in RA can be elucidated. Firstly, through understanding how frequently sustained remission occurs, and secondarily, understanding which patients are most likely to respond optimally to anti-TNF treatment.

Anti-TNF medications are the most frequently used biologic agents, and the work undertaken in this thesis aims to provide the evidence to enable personalised recommendations for anti-TNF treatments. Further work will be required to focus on the other biologic and novel synthetic drugs. Currently, due to the strong evidence of efficacy of methotrexate in RA (183), coupled with low cost, initial treatment with methotrexate looks set to remain the main first-line drug in the treatment of RA. However, the challenge begins in choosing follow-on therapy if methotrexate fails.

Chapter 3

3 Thesis aims and objectives

The treatment options and outcomes for RA have advanced dramatically in recent years. If anything, the problems of 30 years ago have been inverted, and rather than having a paucity, we now have a plethora of evidence and treatments for RA. However, there remain significant barriers to achieving the aim of optimal treatment for every patient. One of the most basic questions that remains unanswered is what proportion of patients actually achieve a sustainable optimal response to their treatment in a 'real-world' setting? Are the measures we currently use to quantify disease activity really working to capture the true picture of longitudinal response to treatments? Is it possible to identify which patient is likely to respond optimally to which treatment either prospectively or early in the initiation phase of treatment? These questions apply to all modalities of treatments and outcome measures, but for this thesis, the focus will be on the use of anti-TNFs for RA, and the most widely used composite outcome measure used in the UK- the DAS28.

This thesis aims to address some of the problems outlined above by undertaking the following work.

3.1 Aim

The aim of this thesis is to identify the frequency and predictors of sustained remission in individuals with rheumatoid arthritis treated with anti-TNF. The study population includes individuals with RA, registered with the BSRBR-RA who were treated with their first anti-TNF. Remission is defined as a DAS28 score of less than 2.6.

3.2 Objectives

1. To collate existing evidence for the factors that are associated with sustained remission in RA through undertaking a systematic review of the literature.

2. To use BSRBR-RA data to quantify the discrepancy between the DAS28-ESR and DAS28-CRP scores that are used in the clinical assessment of patients taking anti-TNF and establish if the two scores can be used interchangeably. Establishing this is essential to ensure that the subsequent analysis of factors associated with sustained remission does not identify spurious associations, or miss important associations, due to inaccuracies in the disease activity scoring systems used.

3. To use the BSRBR-RA to investigate the predictors of sustained remission for patients on anti-TNF, both by using the set disease activity thresholds, and by using trajectory analysis to identify groups with good, adequate, poor and absent response over time, and if there are any factors associated with membership of these groups.

Understanding which clinical and demographic features are associated with sustained remission, and the degree to which they can predict response to anti-TNF will help in developing predictive algorithms which could be used to work towards the goal of truly personalised therapy for individuals with RA.

Chapter 4

4 BSRBR-RA methods

4.1 Background and history of the BSRBR-RA

Since TNF was first identified as a possible therapeutic target in the treatment of RA, there were concerns about possible safety implications of modulating the biological pathways of such a key cytokine. Early clinical trials demonstrated appropriate safety profiles for the medications (29), however, both case reports and continued theoretical concerns about the drug meant that potential safety issues around possible increased rates of infection, demyelination, autoimmune disease and malignancy (particularly lymphoproliferative disease) continued to be a significant consideration post-licencing (143,184).

Due to the clinical trial evidence of demonstrated clinical efficacy of anti-TNF in the treatment of RA, it was evident that the class of drugs was likely to significantly alter the clinical treatment landscape of in the UK. However, existing strategies for post-marketing surveillance of the drugs were identified to have significant limitations (185), particularly for rare events and those with a long latent period that may not occur within the timeframe of an RCT (186). Because RCTs are powered to demonstrate efficacy, rather than potential adverse events, the required sample sizes, and trial durations are insufficient to examine questions about drug safety. Additionally, RCTs typically select patients with as few co-morbidities as possible which is unrepresentative of the population of patients that are likely to receive these drugs in regular clinical practice.

Given these concerns, it was apparent that a long-term registry of individuals with RA treated with anti-TNF was needed (187,188).

4.2 Study design

The BSRBR-RA is an observational cohort study of RA patients starting biological agents, with a control arm of biologic-naïve patients as a comparator group. Initially the BSRBR-RA collected follow-up data on the use of biologic agents in all rheumatic conditions. However, a specific register for patients with ankylosing spondylitis taking biologic agents was established in 2012, and a specific register for patients with psoriatic arthritis is currently being established. After the foundation of the ankylosing spondylitis biologics register, the BSRBR was re-named the BSRBR-RA (initially referred to only as the BSRBR) to recognise that the majority of patients enrolled were patients with RA. The register now focusses on the collection of data on biologic agent use in RA patients only.

The initial study was powered to detect a two-fold rise in the risk of lymphoma following treatment with anti-TNF therapies compared to non-biologic treated patients, and it was calculated that 20,000 person-years of follow-up would be required in both cohorts to identify such an increase with a power of 80%. This equated to approximately 4000 individuals in the exposed cohort for each drug, and required a similar number in the unexposed cohort, each followed for an average of 5 years. Because individuals with RA have an increased risk of infection and malignancy compared with the background rate in a population, the comparator population selected for the BSRBR was drawn from a group of patients with active RA, who had only ever received conventional synthetic DMARDS.

4.2.1 Recruitment

Recruitment to the BSRBR-RA began in October 2001. Initial recruitment for patients starting the first anti-TNFs infliximab (Remicade™) and etanercept (Enbrel™) was successful with recruitment targets being met by 2007 and 2005 respectively (Table 4). However, due in part to the large-scale uptake of anti-TNFs, recruitment to the non-exposed cohort was more challenging and the target of 4000 was not achieved, with the cohort being closed at 3771 individuals in 2009. New anti-TNFs and other

biological agents were added to the registry as they became available. Adalimumab (Humira™) was added in 2002 and target recruitment (n=4000) was achieved by 2007. No new anti-TNF agents were added to the registry until 2010 when certolizumab pegol (Cimzia™) was added, and a new comparator anti-TNF cohort (comprising Remicade™, Enbrel™ and Humira™) was opened in 2012. Recruitment is ongoing to these groups, with targets of 2000 and 4000 participants respectively.

More recently, as biosimilar anti-TNF drugs have become available, these have also been added to the register, along with other biological agents used for the treatment of RA (Table 4) and will continue to provide essential long-term safety information on these agents.

Drug	Cohort Details	Recruitment Dates	Indication	Target sample size
DMARD	csDMARDS	2002-2009	RA	4000
Anti-TNF	Enbrel™ (etanercept)	2001-2005	RA	4000
	Remicade™ (infliximab)	2001-2007	RA	4000
	Humira™ (adalimumab)	2003-2008	RA	4000
	Cimzia™ (certlizumab pegol)	2010 onwards	RA	2000
	Anti-TNF comparator cohort	2012 onwards	RA	4000
	Remsima™ (infliximab)	2015 onwards	RA	500
	Inflectra™ (infliximab)	2015 onwards	RA	500
	Benepali™ (etanercept)	2016 onwards	RA	2000
	Flixabi™ (infliximab)	2016 onwards	RA	500
Anti-IL1	Kineret™ (anakinra)	2001-2009	RA	n/a (not NICE approved for RA)
Anti-CD19	Mabthera™ (rituximab)	2006-2011	RA	1100
Anti-IL6	RoActemra™ (tocilizumab)	2011 onwards	RA	850+

Table 4. BSRBR-RA recruitment dates and target cohort sizes.

4.3 Recruitment Centres

Recruitment to the BSRBR-RA has been from over 250 rheumatology clinics across the UK with a total of 29 rheumatology centres involved in recruiting patients to the

DMARD control cohort. This UK-wide recruitment means that the data on the BSRBR-RA provides an accurate representation of treatment and outcomes in the UK as well as covering a wide socio-demographic range.

4.4 Ethical approval

Ethical approval for the BSRBR was awarded by the North West Medical Research and Ethics Committee and initially included recruitment of patients starting on Enbrel™ and Remicade™ as well as the DMARD control cohort, although over time, additional biologic agents have been added. The register has ethical approval in place to continue until at least 2028.

4.5 Funding

The BSRBR is funded by a consortium of pharmaceutical companies and forms part of their obligations for post-marketing surveillance in the UK. Funding is distributed and managed by the British Society for Rheumatology, and the registry is managed and run by the Arthritis Research UK Centre for Epidemiology at the University of Manchester. Analyses have been undertaken by researchers from a range of academic institutions across the UK and the data are available for use by external institutions by contacting the BSR in London. There is complete academic freedom in the analysis and dissemination of findings from the BSRBR-RA. Contributing pharmaceutical companies can review any work prior to publication and can make suggestions, but do not have the power to amend or veto any publication.

4.6 Inclusion criteria

The BSRBR-RA has very broad eligibility criteria for recruitment to the register. To be included, patients must:

- Have a diagnosis of RA confirmed by a consultant, or one that satisfies 1987 ACR classification criteria (100) for RA.
- Be commencing on a biologic agent for the treatment of RA (not required for recruitment into the DMARD control arm).
- Be recruited within six months of starting the anti-TNF.
- Be aged above 16 years of age at the time of recruitment.
- Be able to give informed consent to participate in the study.

4.7 Data collection

Data are collected at time of recruitment (baseline) and six-monthly for the first three years from both the consultant and patient. Thereafter data are collected on an annual basis from the consultant only. Events of special interest (including adverse events) can be reported at follow-up time-points or on an *ad hoc* basis between follow-up time points.

4.7.1 Baseline information

Following consent, baseline information about the patient is collected from consultants using a paper-based questionnaire. Information on key areas is collected (Table 5; Copy of form appendix 12.1). In most cases, this task is delegated to a rheumatology or research nurse affiliated with the consultant. Once the consent form and consultant baseline information is received by the BSRBR-RA team, a separate questionnaire is sent out to the patient which includes further additional questions regarding drug therapy, adverse events, and past medical history (summarised in Table 5).

Data	Clinical Baseline	Clinical Follow-up	Patient Follow-up	Patient-held diary	ESI	Details	Baseline	Follow-up	Ad hoc
Demographics	✓	✓	×	×	✓	Date of birth, Study ID, NHS number	Yes	No	No
Gender	✓	×	×	×	×		Yes	No	No
RA details	✓	×	×	×	×	Antibody status (ACPA/RF), 1987 ACR Criteria, year diagnosed, year seen by rheumatologist	Yes	No	No
Systemic features	✓	×	×	×	×	Sicca syndrome, serosal and eye involvement, systemic or nailfold vasculitis, pulmonary fibrosis, other	Yes	No	No
Joint replacements	✓	✓	×	✓	×	Clinical baseline - Location and type Clinical follow-up and patient-held diary - details and duration of hospital admission	Yes	Yes	Yes
Disease activity	✓	✓	×	×	×	Clinical baseline and follow-up - 28 tender and swollen joint count, ESR or CRP, PGA (quantified by VAS), DAS28 score and date of assessment	Yes	Yes	No
Concomitant drug therapy	✓	✓	✓	✓	✓	Patient follow-up - number of new drugs Patient-held diary - name of drug and reason Clinical baseline - name, dose, indication (free-text)	Yes	Yes	No
Biologic information	✓	✓	×	×	✓	Clinical baseline and follow-up - biologic type, date started, dose, drug delivery route, frequency, previous biologics (including names, date started/stopped and reason for stopping)	Yes	Yes	No
DMARD therapy	✓	✓	×	×	✓	Clinical baseline - concomitant and previous (including data on: name, date started/stopped/ongoing/frequency, dose, route of administration) Clinical follow-up – current DMARD (including data on: name, date started/stopped/ongoing/frequency, dose, route of administration)	Yes	Yes	No
Comorbidity	✓	✓	×	✓	✓	Clinical baseline - hypertension, angina, heart attack liver disease, renal disease, tuberculosis history etc. Clinical follow-up - new illnesses Patient-held diary - new consultant referrals and reasons ESI – Captures specific information on respective ESI	Yes	Yes	Yes
Smoking status	✓	×	✓	×	×	Clinical baseline - current/ex/never smoker Patient follow-up - detailed smoking information (number per smoked per day;).	Yes	No	No
Steroid Use	✓	✓	×	✓	×	Clinical baseline and follow-up: use (Yes/No) and route of administration IV/PO/SC Patient-held diary – Free text on new drugs started (not steroid specific)	Yes	Yes	No
Diabetic	✓	✓	×	×	×	Clinical baseline - data on diet/tablet/insulin treatment Clinical follow-up and patient-held diary - new diagnoses of diabetes collected	Yes	No	No

Data	Clinical Baseline	Clinical Follow-up	Patient Follow-up	Patient-held Diary	ESI	Details	Baseline	Follow-up	Ad hoc
HAQ	✓	×	✓	×	×	Data collected for first three years only.	Yes	Yes	No
Quality of life	✓	×	✓	×	×	Clinical baseline and patient follow-up - SF-36 (pre- 2013), EQ-5Q (post- 2013). Data collected for first three years only.	Yes	Yes	No
Biologic switching	✓	✓	×	×	×	Clinical baseline and follow-up: information on switching to biosimilar drugs (name, reason for switching)	Yes	Yes	No
Other	✓	×	×	×	×	Blood pressure, height, weight, chest X-ray, quantiferon/elispot details, herpes zoster vaccine details	Yes	No	No
Other referrals/hospitalisations	×	✓	✓	✓	×	Patient-held diary - other new referrals/reasons/speciality Patient follow-up - number of new referrals/hospitalisations Clinical follow-up- hospitalisations/referrals collected if deemed an adverse event	No	Yes	Yes
Adverse Events (inc. ESI)	✓	✓	×	✓	✓	Clinical baseline – prior cancer diagnosis, date and site Patient-held diary - any hospital admissions/referrals recorded (inc. joint replacements, serious infections). Clinical follow-up - any adverse events + if defined as 'serious' + details of adverse event and if biologic/DMARD was thought to be the cause and if biologic/DMARD was discontinued temporarily or permanently. Death (including details on: if patient receiving biologic at time of death, which biologic, date of first and last dose, date of discontinuation of drug). If ESI identified – clinician is asked to complete specific ESI form. Data linkage - malignancy (National Cancer Registry), death (ONS)	Yes	Yes	Yes
ESI – Events of special interest, HAQ- Health assessment questionnaire. Specific ESIs include - Aplastic anaemia, Pancytopaenia, Serious Neutropenia, Serious Congestive heart failure, Cerebrovascular accident, Demyelination/Optic neuritis, Lymphoproliferative Malignancy, Malignancy, Myocardial Infarction/Serious Acute Coronary Syndrome, Pregnancy, Pulmonary Embolism, Serious Infection, Tuberculosis, Serious Lupus/Lupus-like illness, Hepatitis B Reactivation, Serious Haemorrhage, Serious skin reaction, Serious lower GI ulcer/bleed/perforation, Serious hepatic dysfunction/failure, Serious hypersensitivity reaction.									

Table 5. Summary of data collected by the BSRBR-RA

4.7.2 Follow-up information

Data are collected over many domains including ongoing disease activity and drug use (Table 5). All adverse events are captured, although there are adverse events of special interest (ESI). Adverse events can be reported between follow-up points by both patients and clinicians and reminders are included in routine follow-up questionnaires. Adverse event reports are compiled regularly to the pharmaceutical companies that contribute to the running of the registry, specifically including the rates of certain adverse events of interest as summarised in Table 5.

The BSRBR-BR also has established ongoing data linkages with the Office for National Statistics (ONS) which includes data on deaths and the National Cancer Registry which enables additional capture as well as cross-validation of data.

4.7.3 Losses to follow-up and study withdrawals

Throughout the duration of the BSRBR-RA, the proportion of patients withdrawn from the study has been extremely low (1.6%) and return of questionnaires by both consultants and patients has been high (88% and 68% respectively).

4.8 Transmission of BSRBR-RA data and selection of data cohort for analysis

In accordance with good research practice robust processes exist around the access to, and use of BSRBR-RA data. The process is briefly described here.

- A project proposal to use BSRBR-RA data is written, including aims, objectives and hypotheses to be tested. This proposal is submitted to the British Society for Rheumatology (BSR) registers committee.

- The research proposal is considered by the BSR registers committee which includes members of all three national registers (the BSRBR-RA, -AS and -PsA) as well as representatives of the BSR. The proposal is evaluated for its scientific rigour and to ensure that the data available from the register are sufficient to evaluate the proposed hypothesis. The proposal is also checked against ongoing studies to avoid duplication of research efforts. If approved, the proposal is recommended to proceed.

- Following BSR approval, the researcher contacts the team at the BSRBR-RA with the specific dataset required for the approved proposal. This dataset is extracted from the static dataset and transmitted to the researcher using the methods outlined below. The researcher provides regular updates to the BSRBR-RA on the progress of the analysis, and is invited to present interim and final results at BSRBR-RA analysis meetings which occur regularly at the University of Manchester.

Before file transmission, the researcher, and any other individuals previously identified who may use the data must agree to and sign the BSRBR-RA data sharing agreement (Appendix 12.2). Transmission of the extracted dataset can then occur in two ways:

1. Data can be stored internally within the University of Manchester secure servers, and can be accessed by the researcher locally using the University networks; or remotely, using a secure virtual private network (VPN) connection.

2. Data can be transferred via the University of Manchester secure data transfer service (ZendTo™) as a password-protected and encrypted zip file. The researcher is contacted prior to upload of the file to confirm the provided email address is correct and to provide a secondary email address. The requested datafile is uploaded to the data transfer service where it is available for 14 days. A time-limited link is sent to the researcher's primary email address to access the zip-file. The password required to open the zip-file is sent separately to the secondary email address that was previously confirmed. The researcher then confirms receipt of the

data and stores it in accordance with the previously signed data sharing agreement. For work undertaken in this thesis, the data were stored on an encrypted and password protected .dmg file within a restricted access folder on University of Bath servers.

4.8.1 Selecting the dataset to extract

As previously outlined, the BSRBR-RA contains data on a wide range of biologic agents and conventional DMARDs (Table 4), as well as patients who have switched between different biologic agents. The analyses in this thesis have focussed on two distinct cohorts – patients with paired ESR and CRP data, including all biologic-treated patients (except Ro-Actemra™- treated patients (Chapter 8), and patients starting their first anti-TNF (Chapters 9 and 10). As a result, the complete BSRBR-RA dataset was not required and specific subsets of the data were extracted from the BSRBR-RA for these analyses. This occurs following a two-step process:

1. Every six months, a static datafile is extracted from the live BSRBR-RA database. Because the live BSRBR-RA database is continually being updated with new registration and follow-up data, it requires continuous and rigorous data cross-checking (e.g. to follow-up details of any SAE or deaths reported) and data cleaning (ensuring reported information is accurate). As such, it is not appropriate to use the live dataset for analyses as it changes daily and may have data that have not been validated. However, it is important that analyses occur using the most contemporaneous and complete dataset possible. To balance these two needs, static datafiles that have been cross-checked and cleaned are extracted from the live BSRBR-RA database on a six-monthly basis. This allows researchers to access to data that is both contemporaneous, and cross-checked with minimal data inaccuracies. This means that data can be version-tracked, does not change unexpectedly mid-way through analyses, and ensures effective work within teams-minimising the risk of different members of a team working on different datasets.

2. If a specific subset of patients from the static datafile are required, these can be extracted from the latest static datafile and a subset provided to the researcher. For

analyses undertaken in this thesis, the specific criteria for the required subsets of data from the BSRBR-RA was emailed to the database coordinator who liaised with the BSRBR-RA team to arrange for this specific subset of the data to be extracted.

4.9 Strengths and weakness of cohort studies relevant for this thesis

4.9.1 Problems with registry data in assessing response

Whilst the BSRBR-RA offers a unique insight into the real-world clinical use of anti-TNF and response to the drug, there are limitations of such a data resource that are essential to acknowledge. A key point is that because the BSRBR-RA is a cohort study and not an RCT, causality cannot be attributed to any associations identified. This is because, without randomisation, it is not possible to exclude confounding factors which may influence the outcome (discussed in Chapter 9; 9.5.9).

It is also important to recognise that the study design and sample size of the BSRBR-RA was designed originally for establishing long-term drug safety, not establishing drug efficacy. Therefore, certain aspects of the study design may not be optimal for specifically examining longitudinal drug response profiles. For example, it may be interesting to examine exactly how early onset of response is for anti-TNF. However, with data collection points at baseline and six months, it is not possible to conclusively identify the time of onset of response earlier than six months, or whether patients achieve an early response, say within weeks, and then lose it by 6 months (secondary treatment failure).

Whilst there are robust data collection methods employed by the BSRBR-RA, as with all epidemiological studies, missing data remains a problem that must be addressed in all analyses. These methods are described in detail in Chapter 6.

4.9.2 Strengths of using a longitudinal registry for assessing long-term response

Despite the points outlined above, the BSRBR-RA has many strengths that makes it an appropriate resource to undertake longitudinal analysis of anti-TNF efficacy. Whilst RCTs are the gold-standard method for establishing a cause-effect relationship, the inclusion and exclusion criteria of such studies create a homogeneous population, which are not representative of the real-world population in which the drug is used. Follow-up protocols of RCTs are also unrepresentative of routine clinical practice, and rheumatology clinics that participate in RCT studies are more likely to be larger secondary or tertiary referral centres which are likely to differ in resource availability and clinical practice to the majority of rheumatology clinics that do not recruit to RCTs.

Registry studies reflect much more closely what is observed in clinical practice and can give useful insights into real-world efficacy of the drug. They are much less labour intensive for clinicians to recruit to and data are collected from routine practice. Because the BSRBR-RA has been ongoing for over 15 years, it is also possible to examine the evolution of clinical practice, and how the demographics of the patients using the drug have changed over time. The long-duration of follow-up enables investigation of outcomes (such as the long-term durability of response) that would not be possible in most RCTs. In addition, because the BSRBR-RA has recruited patients taking a range of anti-TNF medications, it is possible to investigate if drug choice influences outcomes (189).

Overall, the size and duration of the BSRBR-RA ensures that it is a powerful resource to test the hypothesis of sustained remission and guide future research into this area.

Chapter 5

5 Statistical Methods used in this thesis

This chapter outlines the key statistical methods used in the analyses in this thesis, their underlying concepts, assumptions and rationale for using them. The methods are described in the order in which they are used in this thesis and cover the following key areas:

- **Data preparation and missing values** (section 5.1). Prior to undertaking analysis, the dataset is prepared (Chapter 6). This involves examination of the dataset provided (described in 4; 4.8.1) for missing data, and subsequent imputation of missing values and analysis of multiply imputed datasets

- **Methods of assessing agreement**. The analysis undertaken in Chapter 8 involves comparison of agreement between the two versions of the DAS28. Section 5.2 of this chapter examines the different methods of assessing inter-score agreement

- **Non-linear modelling.** The development of a modified version of the DAS28-CRP (Chapter 8) requires use of generalised linear models as well as splines. This is described in section 5.3.

- **Regression methods.** Chapter 9 investigates sustained remission and LDA and the predictors associated with these outcomes. Logistic and stepwise regression modelling is used, as well as methods to investigate and quantify collinearity and are described in section 5.3

- **Latent class mixed modelling (LCMM)**. Finally, the trajectory analyses undertaken in Chapter 10 draw heavily on Bayesian statistic and LCMM. The theory behind these concepts is outlined in section 5.5.

5.1 Data preparation and missing values

5.1.1 Missing data

Analyses in Chapters 8, 9 and 10 all use the BSRBR-RA dataset which has missing data. Missing data are a feature in almost all datasets and their identification and management present challenges in ensuring bias is avoided when analysing data. There are three main patterns of "missingness", each of which will require different approaches when performing analysis:

1. Missing completely at random (MCAR)
2. Missing at random (MAR)
3. Missing not at random (MNAR)

There are no definitive statistical tests that can 'prove' any particular missing data pattern, so categorisation of the missing data should be undertaken by examination and understanding of the methods used to collate the data to enable a balanced decision to be made about the pattern of "missingness".

5.1.2 Missing completely at random

As the name implies, MCAR occurs when missing data points occur completely at random through a dataset. Missing data points have no relation to each other, the intervention or outcome. If missing data are present MCAR is optimal, as management of such missing data are least likely to introduce bias. Most imputation methods assume MCAR or MAR. In reality, MCAR very rarely happens in epidemiological

datasets as invariably there are relationships between missing data variables that make the assumption of MCAR false.

One way of examining if missing data are MCAR is to identify records with missing data; and then examine if the values for the remaining complete variables associated with the record fit the normal distribution of the complete dataset using a t-test. The null hypothesis for such a test is that there is no difference between records with missing data points and complete data records. Therefore, in this situation, to satisfy the assumption of MCAR, the null hypothesis should not be disproved. This method has significant problems which are described in depth by Little (190), however the fundamental problem with such an approach is that a separate t-test is required for each variable in the dataset resulting in issues with multiple testing if multiple variables are present. Little's test for MCAR (190) goes some way towards addressing this issue, and uses a χ^2 test across a matrix of observed means and missing data that either rejects the null hypothesis or not. The null hypothesis for Little's test is that there is no significant difference between the missing data variables and the available variables. Therefore, a p-value of >0.05 gives support for the assumption of MCAR. However, Little's test can still be prone to falsely rejecting a null hypothesis of MCAR, particularly for large datasets with large numbers of variables.

5.1.3 Missing at random and missing not at random

If MCAR cannot be confidently identified, data should be examined to identify if the missing data are MAR. MAR occurs when missing data points are related to a variable, but not the value of the variable. MAR can be explained by the following example. If a questionnaire includes a question on an individual's weight, it may be that women are less likely to complete this question, leading to missing data. This means that the "missingness" is related to another variable (gender - women are more likely to have missing weight data than men in this example), and so missing data are not MCAR. If women of all weights are equally likely to skip the question on weight, then the missing data for weight could be considered to be MAR. However, if heavier women are more likely to skip the question on weight than lighter women, then the "missingness" of the

data is now related to the actual value of the data and are MNAR. The distinction between MAR and MNAR is important because most imputation of missing data requires an assumption of MAR at a minimum (although some methods such as complete case analysis should satisfy MCAR).

Prior to analysis of the dataset, it is essential to establish which of the above missing data patterns is present in the data. This is undertaken using Little's test, as well as visualising the missing data patterns (Chapter 6).

5.1.4 Managing missing data

Missing data can cause major problems if not appropriately managed, impacting on the accuracy and power of the findings if data are inappropriately imputed or deleted respectively. There are numerous methods for managing missing data. The most commonly used methods include: complete-case analysis/list-wise deletion; separate category for missing; last-observation carried forward; hot-deck imputation; simple mean imputation and multiple imputation. These methods will be described briefly.

5.1.4.1 Complete-case analysis/list-wise deletion

The most basic method of managing missing data is to exclude all records with any missing data. However, this can result in substantial loss of power of the dataset in all but the most complete datasets. List-wise deletion can also introduce bias if there is a cause for "missingness" (191). Using the earlier example of a questionnaire with a weight question; if all missing data points were excluded, and women were more likely to skip a question on weight than men, the final dataset would have a disproportionate number of men in it and would be unrepresentative of the actual data collected, causing bias in the results. Because of this, complete-case analysis is rarely used.

5.1.4.2 Separate category for missing data

This method involves recoding those data that are missing as a separate category (e.g. NA or 'missing'), but keeping them in the analysis for all the available variables. This method maintains the precision of analysis, but still does not adjust for possible selection bias if there is a reason for "missingness" or data are MNAR. It also can lead to difficulties with continuous variables, whereby inclusion of a 'NA' or 'missing' category effectively creates a categorical variable in an otherwise continuous scale (191).

5.1.4.3 Last observation carried forward

In longitudinal datasets, another method for managing missing data is to carry forward the last recorded value when a missing data point arises (191). However, this method reduces the variance of any estimate and can cause auto-correlation (Figure 1). The impact of auto-correlation is amplified if records have multiple sequential missing data points. However, where the variance for the available values for the variable is small, and the likelihood of multiple sequential missing data points is unlikely, the impact of this type of imputation is reduced. The impact of such imputation can be explored by undertaking analysis with and without last observation carried forward (LOCF) imputation and examining the difference in results. If a significant difference is noted between the two analyses, further investigation as to the cause of such discrepancy is worthwhile. Another difficulty with LOCF is if there are missing baseline data, missing values may still be present even after LOCF imputation.

LOCF imputation was used for imputing age at time of DAS28 in the analysis in Chapter 8 where data were missing.

5.1.4.4 Hot-deck and simple mean imputation

Hot-deck imputation uses available data as a 'donor' to fill-in a missing data point from within the same variable. There are different methods for selecting which donor should

be used to fill-in the missing data point. It can be randomly selected from the available data, the nearest neighbouring value can be used, or it can be based on other parameters that try match the donor value to the missing data record value (192). All different methods have pros and cons, but whatever method is used will apply assumptions on the missing values, which may influence results.

Mean imputation uses the mean value from the available observations for the variable to complete any missing data points. However, mean imputation causes auto-correlation which can give overly precise estimates (Figure 1). Neither of these methods are used in this thesis.

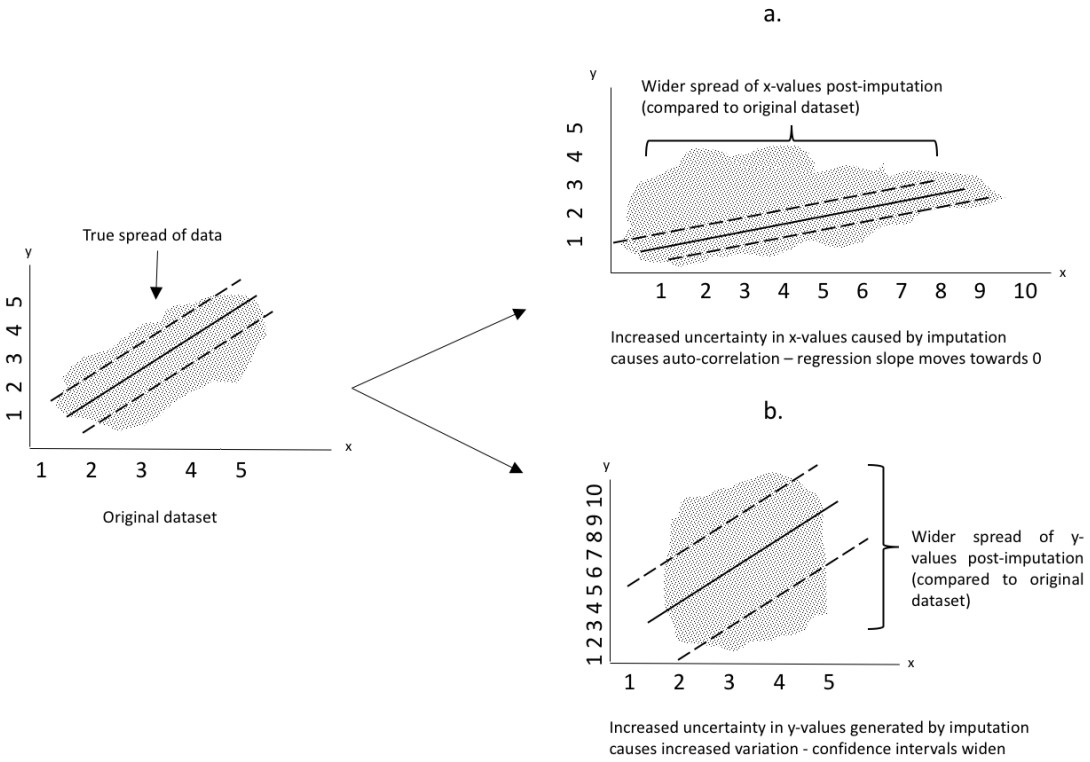


Figure 1. Demonstration of the impact of how varying values for x- and y-variables can affect regression relationships (auto-correlation (a) and attenuation (b) bias). If imputation has no boundaries to what x-value can be generated, then auto-regression can occur, which tends the slope of the regression line to 0. If there are no bounds applied to the y-component during imputation, then confidence intervals will tend to widen artificially

5.1.4.5 Multiple imputation

One increasingly used method of managing missing data is multiple imputation. One of the reasons for the limited use of this method previously was that it required a

significant amount of computational power to undertake. However, with the rapid increase in processing power available, this method can now be used for most datasets using standard personal computers. There are two main methods of multiple imputation - multiple imputation using chained equations (MICE), and expectation-maximisation and bootstrapping (EMB) algorithms. MICE algorithms use a Markov-chain Monte-Carlo (MCMC) method of simulating a dataset many times (usually thousands) with random numbers being selected for missing data points. After an initial 'burn-in' period (usually a few thousand simulations), the law of large numbers (193,194) means that values converge towards plausible ones and can be used to complete the missing data points. To ensure an accurate understanding of the uncertainty in the imputed data, multiple imputed datasets should be created (often five, but occasionally more) and Rubin's rules (Chapter 5; 5.1.5.1) used to accurately quantify uncertainty in estimates. However, because MICE requires thousands of simulations for each variable with missing data it becomes increasingly computationally intensive, and can take a long time to run. EMB algorithms are an alternative method that can be used, and were used to impute data in Chapter 6 which was used in the analyses in Chapters 9 and 10.

5.1.4.5.1 Bootstrapping & EMB algorithms

The Amelia II program in R (henceforth referred to as Amelia) was used to carry out multiple imputation using EMB algorithms for the analyses in this thesis (Figure 2). As with most multiple imputation methods, missing data are assumed to be MAR, meaning that the pattern of missing data are related to the observed data and not the missing data (as described in Chapter 5; 5.1.3). The other assumption is that the data (both missing and present) have a multivariate normal distribution, however there is evidence that Amelia works adequately for categorical as well as mixed datasets (195). Using the MAR assumption, Amelia identifies relationships between the variables included within the model and uses Bayesian inference to bootstrap parameters and distributions for each data point, before randomly drawing data from the normal distribution within these bootstrapped parameters. For example, if a dataset included data on height and weight of individuals, with data missing at random from the height

variable, Amelia would identify the distribution of values of the available height data, and the relationship between the height values and weight data to inform the parameters around each missing data point. This would mean that if an individual with a missing height data point weighed 100kg, Amelia would apply a different set of parameters than an individual who weighed 50kg who might also have missing height data. This is a particularly useful feature considering the analysis of outcomes in the BSRBR-RA, as there is likely to be a relationship between tender and swollen joint counts and the visual analogue scale, which can be identified using Amelia and allow generation of the most plausible values for each missing data point.

In addition to bootstrapping parameters for each variable, it is also possible to manually apply parameters to each variable, which means that expert knowledge can be used to optimise the model (something that is not possible using a MICE approach). In the case of imputing missing data for tender or swollen joint counts, it is possible to constrain results to within the allowed 0 - 28 range for this variable. This means that implausible values are not created, and improves the accuracy of the model, without having to run thousands of simulations.

These features of Amelia minimise auto-correlation and attenuation bias that can occur if inaccurate normal distributions or wildly extreme improbable data are generated (Figure 1). The algorithm runs in a step-wise manner as outlined in Figure 2, and each step refines the prior parameters of each value (the 'priors') (196). Therefore, the more variables that are included in the dataset, the better the accuracy of the imputed values.

As a default, Amelia generates five imputed datasets (although greater or fewer datasets can be generated if specified), which is usually sufficient to capture the variance and uncertainty of the missing data. Statistical methods can then be applied, (as if there were no missing data) to these five data sets and the results combined.

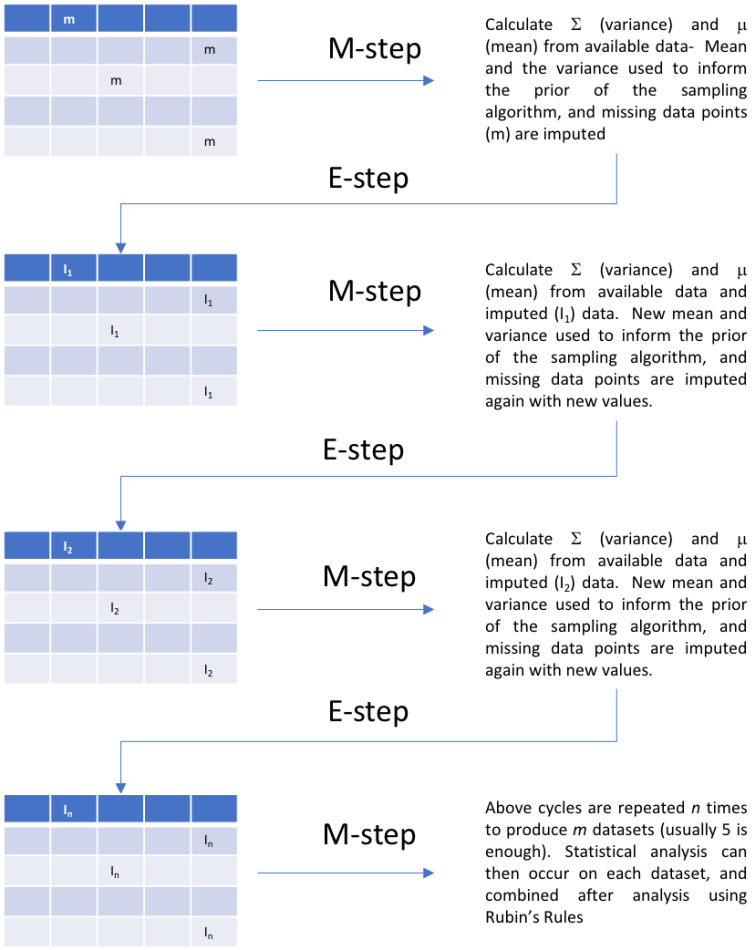


Figure 2. Imputation algorithm applied by the Amelia II program

5.1.5 Analysing multiply imputed datasets

5.1.5.1 Rubin's rules

Rubin's rules form the foundation on which multiple imputation analysis can be undertaken (197). When undertaking statistical analysis on multiply imputed datasets, it is important to undertake the proposed analysis on each of the imputed datasets separately, before combining the estimates into a final result. This enables the uncertainty both within each imputed dataset, as well as the uncertainly between each imputed dataset to be appropriately represented in the final estimate.

The point estimate ($\bar{Q}$) across the multiply imputed datasets is essentially the mean of the estimates across *m* datasets and is described as:

$$\bar{Q} = 1/m \sum_{t=1}^{m} \hat{Q}^{(t)}$$

Equation 1. Calculating the mean estimate ($\bar{Q}$) across m datasets

The within-dataset imputation variance (*W*) across all the multiply imputed datasets is calculated as the average variance (*U*) across all *m* imputed datasets:

$$W = 1/m \sum_{t=1}^{m} U^{(t)}$$

Equation 2. Calculating the mean within-dataset variance (W) across m datasets

The between-dataset imputation variance (*B*) is then calculated by calculating the difference between the point estimate for each dataset ($\hat{Q}^{(t)}$) with the mean point estimate across the all the imputed datasets ($\bar{Q}$):

$$B = \frac{1}{m-1} \sum_{t=1}^{m} (\hat{Q}^{(t)} - \bar{Q})^2$$

Equation 3. Calculating the between-dataset variance (B) across m datasets

The within and between imputed dataset variance (*T*) is given by combining estimates for *B* and *W* as:

$$T = W + \left(1 + \frac{1}{m}\right) B$$

Equation 4. Calculating the within- and between-dataset variance (T) across m datasets

These principles are incorporated in the software program Zelig which runs in the R environment and will be used in the analyses in this thesis.

5.1.5.2 A note on anomalies in date data due to imputation

When imputing time/date data (e.g. years) that may subsequently be subtracted from other time/date data that may also have been imputed, it is possible that anomalies may arise in the data that are theoretically impossible or implausible. Such an example from the dataset preparation (Chapter 6) is a negative disease duration. This arises because the disease duration is calculated by subtracting 'the year of starting on a biologic' from the 'year of diagnosis'. Because the two variables are imputed separately and disease duration calculated after imputation, in a small number of cases, the imputed value for year of diagnosis is greater than the year of biologic date, leading to a negative value. Similar occurrences can occur with 'age at diagnosis' and 'age when seen by a rheumatologist' fields which are calculated by subtracting one time/date value from another. One way of avoiding this would be to calculate these values before imputation, and then impute these missing values separately, applying bounds so that impossible or implausible values are not generated. Alternatively, implausible values could be manually converted to 0 after imputation. However, both of these methods could lead to impossible relationships, where the disease duration is related to the year of diagnosis for some data, but not others. This generates greater problems as it asymmetrically alters relationships within the dataset. It is also important to note that with multiple imputation, results are not reported from each individual imputed dataset, so individual implausible data are less directly relevant to the overall results. Instead, the multiple datasets are used to simulate variance and uncertainties in the imputed data and the individual values are used to generate composite estimates. Therefore, these implausible values contribute to the estimates of uncertainty quantified by the variance. In this analysis, five imputed baseline datasets are created and five imputed longitudinal datasets, meaning that a total of 25 datasets are created when combined. Analysis is then undertaken on all 25 datasets and estimates are combined using Zelig software. Manually altering the imputed data would impact on the quantification of uncertainty in the final estimate, and is less desirable than the few implausible results that may be occur otherwise. Therefore, in this analysis, implausible values post imputation have not been individually altered.

5.2 Methods of assessing agreement

5.2.1 Measures of agreement

Analysis undertaken in Chapter 8 involves comparison of inter-score agreement levels between different versions of the DAS28. Inter-score comparison requires use of specific statistical methods for continuous and categorical scores. Because the DAS28 is measured on a continuous scale with set disease activity thresholds, it has both continuous and categorical properties. Therefore, comparisons between the different versions of the score need to explore both the nature and degree of agreement for both types of variable.

5.2.2 Bland-Altman statistics

Bland-Altman plots and statistics are a validated measure of visualising and quantifying agreement between two measures that have a continuous scale (198). Whilst plotting the values of one version of the DAS28 against another can demonstrate correlation between the two scores, it is important to note that this does not demonstrate agreement. Intraclass correlation (ICC) statistics are used when comparisons are made using the same score (i.e. two observers scoring an x-ray for RA disease progression using the same score), however, if an inconsistent score is used, ICCs are not appropriate. The Bland-Altman plot is a technique that can be used to graphically demonstrate discrepancies between two scores that are on the continuous scale by plotting the difference between the mean of the two scores of interest and the individual values for the two scores. When comparing two scores (A and B), x and y values are calculated as follows:

$$y - axis\ values = A - B$$

$$x - axis\ values = \frac{A + B}{2}$$

Equation 5. Calculating values for Bland-Altman plot

In addition to demonstrating the discrepancies between the two scores, the 95% range of differences can be plotted, which will also demonstrate if such discrepancies vary with the size of the measurement. If the discrepancy between the two scores is associated with the size of the measurement then the Bland-Altman plot will be asymmetric (Figure 3).

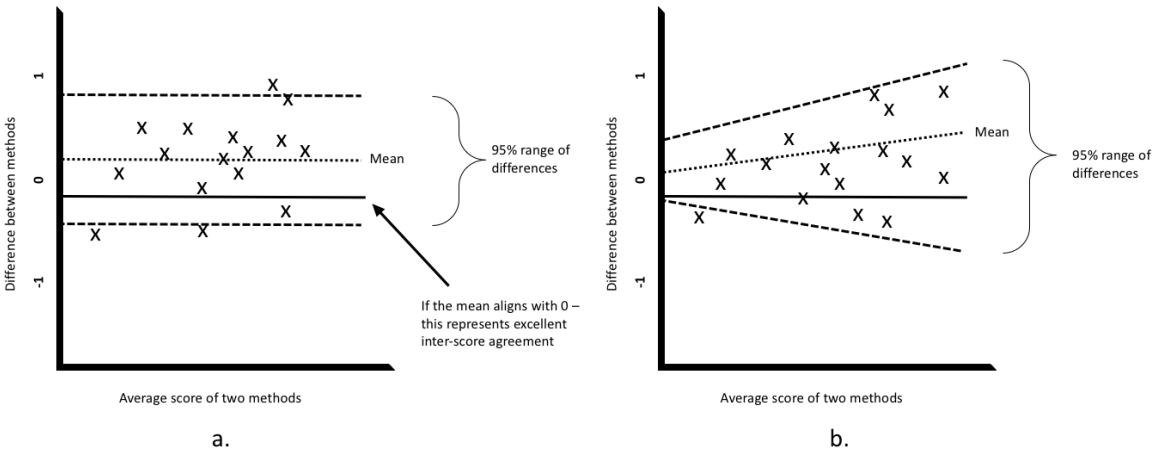


Figure 3. Example Bland-Altman plot with (a) symmetrical and (b) asymmetric interscore differences (which vary with the magnitude of the score)

5.2.3 Agreement matrices

The continuous version of the DAS28 can also be categorised into remission/LDA/MDA/HDA. To compare agreement between the categorical transformations of the different versions of the DAS28, agreement matrices are used (Figure 4). Such matrices allow comparison of how the two scores agree when identical disease activity thresholds are applied to each score. They are also useful for displaying the distribution of misclassifications (i.e. does one score routinely give scores higher than another; are disagreements dispersed across the spectrum of the available categories, or grouped closely around agreement).

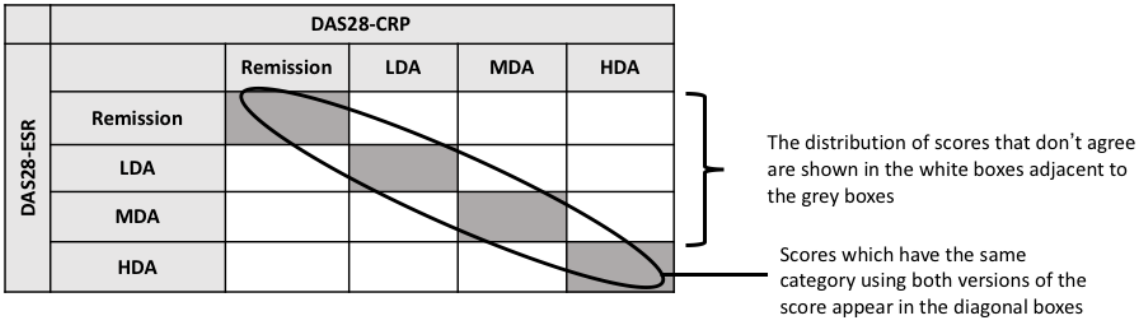


Figure 4. Agreement matrix for comparing categorical scores

5.2.4 Cohen's kappa for assessing agreement

Whilst agreement matrices are a useful method for visualising overall agreement between two scores, they do not consider levels of agreement that could have occurred by chance. Kappa (κ) statistics are one such way of examining this.

Cohen's κ (199) is an established methodology for investigating agreement between two scores, adjusting for agreements that could occur by chance. It takes the form:

$$\kappa = \frac{P_o - P_c}{1 - P_c}$$

Equation 6. Equation for calculating kappa

Where P_o is the observed agreement and P_c is the agreement that could occur by chance. Figure 5 shows how κ is calculated using example data from Chapter 8.

Observed agreement

		DAS28-CRP Classification				
		Remission	LDA	MDA	HDA	Total
DAS28-ESR Classification	Remission	4041				4857
	LDA		1303			3440
	MDA			8881		11810
	HDA				8563	10977
	Total	6063	4077	12056	8888	31084

$$P_0 = \frac{4041 + 1303 + 8881 + 8563}{31084}$$

Number of scores where both versions of the DAS28 were classified as HDA i.e. agreement

Chance agreement

		DAS28-CRP Classification				
		Remission	LDA	MDA	HDA	Total
DAS28-ESR Classification	Remission	947				4857
	LDA		451			3440
	MDA			4581		11810
	HDA				3139	10977
	Total	6063	4077	12056	8888	31084

$LDA_{esr} = DAS28 - ESR\ scores\ classed\ as\ LDA$

$LDA_c = LDA\ agreement\ by\ chance$

$LDA_c = \frac{LDA_{crp}}{Obs_T} * LDA_{esr}$

$LDA_{crp} = DAS28 - CRP\ scores\ classed\ as\ LDA$

$Obs_T = Total\ observations$

$$P_c = \frac{947 + 451 + 4581 + 3139}{31084}$$

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

Figure 5. Worked example for calculating Cohen's kappa

The maximum possible value that κ can take is +1 which indicates near-perfect agreement, 0 indicates agreement is due to chance, and -1 is the minimum possible value and indicates agreement worse than would occur by chance alone.

An underlying assumption of Cohen's κ is that that the two measures are independent, mutually exclusive (i.e. a single DAS28 score cannot be classified as being in two groups at the same time), and exhaustive (i.e. all scores will fall into a category). In the case of the DAS28 analysis, the assumption of independence is reasonable. Although the tender joint count, swollen joint count and PGA are shared between the different versions of the score, the inflammatory markers (CRP and ESR) are independent. In addition, the equations to calculate the final DAS28 score are different. Therefore, the assumption of independence of the two rating scales is justified. DAS28 scores classifications are mutually exclusive and exhaustive, so use of the test is valid.

There are no formally defined thresholds for what a 'good' or 'poor' value for κ is, however, the most commonly used are those proposed by Landis and Koch in 1977 (200). They are:

κ statistic	**Strength of agreement**
< 0.00	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost perfect

Table 6. Landis-Koch thresholds for kappa agreement

5.2.5 Root mean squared error (RMSE)

In addition to identifying the level of agreement using κ, it is important to quantify the magnitude of disagreement when it occurs. For example, a reference (y_i) and comparator ($\hat{y_i}$) score may have almost perfect agreement, but when the two scores do not agree, they may be slightly different or substantially different (i.e. when the two scores are not identical, there is a large error term). However, another comparator score ($\hat{z_i}$) may have a poorer κ value than $\hat{y_i}$ (i.e. have perfect agreement less often), but have very similar values to the reference score, albeit not identical (i.e. the error between $\hat{z_i}$ and the reference, y_i, when not identical is small). The RMSE allows such quantification and comparison of the overall error when a score is compared to a reference. It is calculated by the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}$$

Equation 7. Formula for calculating RMSE

For each individual, this involves subtracting the new score value ($\hat{y_i}$) from the reference score value (y_i) and squaring this difference. Squaring the difference ensures

that any negative values are converted to positive to prevent the possibility that positive and negative errors will cancel each other out. The mean is taken of all the squared errors and the square-root (of the mean) is taken. When the same reference is used for multiple scores which have the same scale, the RMSE can be compared to identify the score that has the smallest mean error compared to the reference score and a t-test can be used to identify if such a difference is between the two scores being compared is significant.

5.3 Linear and non-linear modelling

To model the relationship between the DAS28-ESR and DAS28-CRP (Chapter 8), various non-linear modelling methods were used.

5.3.1 Logistic and multinomial regression models

The regression models used in this thesis are mainly logistic. This is because analysis is between categorical outcomes (e.g. sustained remission or not). For those analyses where there are multiple categorical outcomes (LCMM analysis of 3 or more classes; Chapter 10), a multinomial logistic regression model is used. This compares a reference category (for example the poor-response class) with each of the other classes in turn. Before comparing results between different datasets or subgroups, it is essential to check that the same reference category is used for all analyses.

5.3.2 Multiple collinearity

Multiple collinearity can be a major problem when multiple variables are analysed in a regression model if the variables are not completely independent. This is because, if there is a relationship between variables within a regression model, the variance of the regression model is increased, resulting in variables that may have a significant association with the outcome not being identified as their effect is attributed to another

variable that is collinear with it in the model. For example, when a patient experiences an increase in the number of swollen joints, it is likely that they will also experience an increase in the tender joint count, PGA, and by default, an increase in DAS28. These variables are collinear- that is they are all likely to increase and decrease in a similar manner. If all these variables are included in a single model, the variance will be large. When attributing the effect of each variable individually, it is possible that the statistical model will attribute an equal weighting to each variable. This means that the overall effect is split between four variables, and it may appear that each individual variable may only have a small effect (which may or may not be true). Alternatively, the overall effect may be attributed to one or two of the variables only, making the effect of these variables appear larger as they 'inherit' the effect of the other collinear variables. However, with large variances, it can be difficult to be sure that the overall effect has been apportioned to the correct variables. One way to deal with collinearity is to minimise the number of variables in a regression model, and not choose any variables that may have a collinear relationship. The difficulty that arises here is that there may be justified interest in investigating the differential effect of collinear variables (such as the swollen and tender joint count on an outcome such as sustained remission). A way to address this issue is to identify the extent of collinearity within a full model specification (including all variables), remove as many variables as possible using stepwise regression, and then recheck to ensure collinearity is reduced, resulting in a model that has the optimal balance between model fit and minimal collinearity.

5.3.3 Variance inflation factors (VIF)

One way of identifying the extent of collinearity is to quantify the amount of variance associated with each variable. This can be done by measuring the VIF. As the name suggests, VIF estimates how much the variance is inflated for each variable, and is undertaken by regressing each predictor within the model against all other predictors within the model to give an R^2 value for each predictor variable (201). This R^2 value is then included in the following equation to calculate the VIF:

$$VIF = \frac{1}{1 - R_i^2}$$

Equation 8. Calculating the variance inflation factor for a variable

Where *i* is the variable of interest.

A VIF of 1 means there is no collinearity. Values between 1 and 5 suggest moderate collinearity and values greater than 5 are highly correlated. There is no formal definition as to what level of collinearity is acceptable but it should be minimised if possible.

5.3.4 Stepwise regression

Stepwise regression is an automated process that can be used to reduce variables in a regression model to its minimum number of most significant components. There are two main methods of stepwise regression that can be used, backwards and forwards stepwise regression. The former works by starting with the full regression model with all variables included, before systematically excluding each variable one at a time from the model to identify which variable contributes the least to the model fit, and can be deleted without significantly altering the model fit. If a variable which satisfies this criterion is identified, it is removed from the model and the process is repeated with the remaining variables, until no further variables can be deleted without significantly affecting the model fit. Forwards stepwise regression takes the same concept as backwards stepwise regression, but instead starts with no variables, and then sequentially adds variables that significantly improve model fit. Different criteria can be used to decide the criteria for inclusion (i.e. what is 'significant'), but for the purposes of this analysis, Akaike information criterion (AIC) (202) is used for assessing the model fit. Both forwards and backwards stepwise regression is undertaken and the model with the lowest AIC is chosen as the best model fit with the fewest number of variables.

There are several limitations with stepwise regression to be aware of. Firstly, because the process is automated, it removes the ability to use expert knowledge to specify a model. However, it can be very useful if there is limited information to help with selecting a limited set of variables from an extensive list of *a priori* specified variables. Another important limitation is that once a variable is removed from the model, it cannot be added back in at a later step. This means that for multiple collinear variables, it may become arbitrary which variables are deleted, and the order of variable removal may influence the final model selected. This can be minimised if multiple iterations of stepwise regression are used and both forwards and backwards stepwise regression is used to find the best model fit, as is used in the analysis undertaken in Chapter 9.

5.3.5 Generalised linear models (GLMs)

GLMs have a similar underlying function to a linear model. However, certain aspects of a linear model are amended to allow more flexible mapping of data. A linear model with one variable and *i* observations (x_{1i}) has the generic form:

$$Y = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

Equation 9. Generic linear model

Where β_0 represents the y-intercept, and β_1 the gradient, or the slope, associated with variable x_1. ε is an independently normally distributed random error term. Linear models fit a regression line using an ordinary-least-squares (OLS) model, where the line of best fit that is chosen has the least squared distance from all the data points and the line.

5.3.6 Generalised additive models (GAMs)

GAMs are a useful method of modelling non-linear relationships and offer a less restrictive approach than using polynomial terms within a regression model. In practice, relationships are commonly non-linear and, in such cases, performing linear

regression is likely to result in unrealistic outcomes. Quadratic or cubic functions (or higher order polynomials) can be used to model non-linear patterns in data, however their use often can result in issues at the extremes of the data, as can be seen in Figure 6. This is especially the case if there is a ceiling/maximum value for one variable (such as the ESR; see Chapter 8).

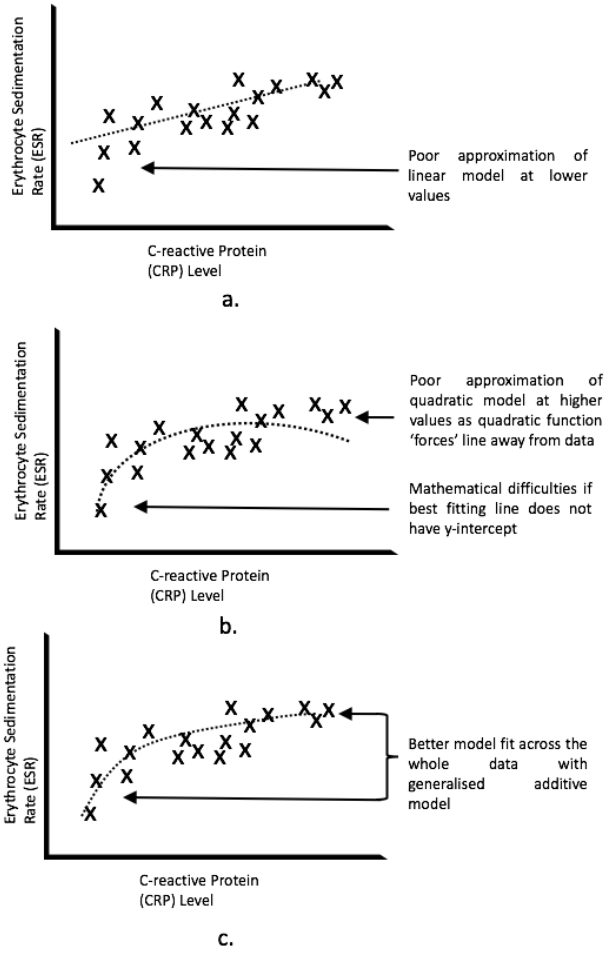


Figure 6. Comparison of (a) linear, (b) quadratic and (c) generalised additive models for non-linear data. (a) and (b) demonstrate how the properties of fixed linear models can coerce regression lines to clearly sub-optimal fits for the extremes of the data

GAMs have a similar structure to generalised linear models (Equation 10), but instead of having a fixed β for the regression slope, a smooth function $f(x)$ is used.

$$Y = \beta_0 + f(x)_{1i} + \varepsilon_i$$

Equation 10. Generic GAM equation

5.3.7 Smoothers

As the name suggests, smoothers fit a 'smooth' line through the data. One of the simplest type of smoother is a 'running mean', where the value of the line at any point is the mean of a set number of points around the point in question. If the number of points that is included in calculating the mean values is too small, these can over-fit the data, producing lines that can be unstable, and can be strongly influenced by outlying data points. If all the data points are used to create a single mean, then a straight line is generated. There are many smoothing functions that can be added to a GAM. In Chapter 8, a monotonically constrained spline is used, which is a smoothing function which restricts the GAM from having a negative gradient (meaning that as CRP increased, the corresponding predicted ESR value could not decrease).

There are many types of smoothing functions, and this provides additional flexibility to GAMs when fitting non-linear data.

5.3.8 Splines

Splines are useful in GAMs as varying curves can be plotted for different parts of a dataset. Different splines are joined by 'knots', which are the point at which one spline curve transitions to another (203).

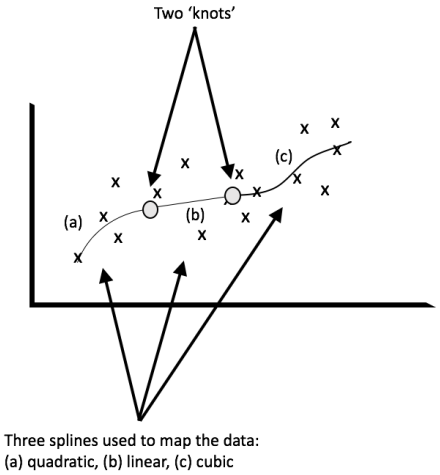


Figure 7. Example data fitted with a model containing two knots and three splines

Multiple 'knots' can be used, and the number can be manually specified or determined automatically based on the complexity of the data. By using multiple splines and 'knots' the best-fitting regression line can be fitted to data using the iterative process used in GAMs (204).

5.4 Bayesian statistics

Bayesian statistics are used throughout this thesis and are used for a number of different purposes. To understand the application of Bayesian methodology and how it differs from a more classical quantitative statistical methods approach, it is essential to understand the underlying concept of Bayes theorem, which forms the foundation for Bayesian statistics.

5.4.1 Bayes theorem

Bayes theorem is a powerful methodology that involves estimating the probability of a hypothesis (the prior) being true based on the probabilities of factors that may impact on the outcome. It is particularly powerful as it enables incorporation of uncertainty from multiple sources, expert opinion or experience, that can be conferred upon a separate dataset to estimate the probability of a hypothesis being true.

The theorem has been attributed to the Reverend Thomas Bayes (1701 – 1761), although his notes were actually published posthumously by a fellow preacher and mathematician Richard Price in 1763 in the Royal Society Journal *Philosophical Transactions* (205).

It is described by the following equation:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

Equation 11. Bayes Theorem

Where P(H|E) is the probability of the (prior) hypothesis (H) being true given the evidence (E).

To explain Bayes theorem, an example is helpful, and a hypothetical rheumatology clinic setting can be used to explain the process of updating a prior hypothesis with evidence.

5.4.1.1 Bayes theorem in practice: Example 1

Prior to the hypothetical rheumatology clinic commencing, we wish to calculate the probability that the first patient to be seen will have RA. This is the 'prior hypothesis'. A list of patients due to attend the clinic provides very little information other than the patient name and age. In this scenario, with limited information, the estimated probability that the person has RA is essentially a random guess. However, other sources of information can be used to influence the probability of the prior hypothesis before reviewing the patient. An understanding that RA is uncommon in the general population (~1% of the UK population (33)), may shift the probability of the prior hypothesis to be less likely. However, incorporating the evidence that the clinic is in a rheumatology department, and RA is the most common inflammatory arthritis seen by rheumatologists (say 1 in 3 rheumatology patients), the probability of the prior

hypothesis would be shifted to make the likelihood of the first patient in clinic having RA to be more likely. We know the first patient is a woman who is 48 years old and this adds more information. Given that RA is more common in women, and often presents in the fourth decade, the prior hypothesis can be updated again; further increasing the likelihood that the first patient has RA. This process continues, and is essentially the process used in clinical history taken and examination (although in an unquantified manner); more information is obtained allowing the prior to be updated until the patient's final diagnosis can be concluded with a high degree of confidence.

This example highlights the iterative nature of Bayes theorem and how new information can alter probabilities without having to undertake extensive sampling to establish a normal distribution or pre-specify confidence levels to accept or reject a null hypothesis.

5.4.1.2 Bayes theorem in practice: Example 2

The following example shows how Bayesian inference can help with interpretation of screening test results.

Consider a 30-year-old patient who has a screening blood test to detect a cancer. There is no prior concern that the individual has cancer. The test has a specificity of 99%. If the result comes back as positive it may appear that the probability that the patient has cancer is 0.99 (the prior hypothesis). However, the age of the patient is essential to inform the prior hypothesis if for example, the cancer detected by the blood test only occurs in 1 in 100 individuals under the age of 40 (a probability of 0.01). Using Bayes theorem, the prior hypothesis can be updated accordingly by multiplying the probabilities:

$$0.99 \times 0.01 \ = 0.0099$$
$$(i.e. less\ than\ 1\%\ chance\ of\ cancer)$$

Bayes theorem can therefore profoundly change the probabilities of the prior hypothesis and can be a very powerful tool. It has a wide range of uses and is used in the imputation methods used by Amelia (5.1.4.5.1).

In understanding Bayes Theorem, it is important to remember three key factors (206):

1. Have a clearly stated prior hypothesis.
2. Consider the likelihood of the hypothesis being true from all perspectives.
3. The prior hypothesis should be updated when new relevant information is available.

5.4.2 Bayesian model checking for selecting post-stepwise models

Bayesian model checking is utilised in its most simple form in this thesis when selecting which model to use following stepwise regression on multiple imputed datasets. As described earlier (Chapter 5; 5.1.4.5.1), multiple imputation creates multiple datasets, all with slightly different values. In Chapter 9, because five baseline and five longitudinal datasets are imputed separately and subsequently combined, there are a total of 25 imputed datasets. This is essential to quantify the uncertainty in the estimates that are generated from imputed data. However, having multiple datasets creates a potential problem when using stepwise regression. This is because the slight differences in values across the datasets influences the variables that are selected through the process of stepwise regression for the final model, meaning that a different set of variables could be selected for each imputed dataset. This could be challenging to deal with using a non-Bayesian, frequentist approach. However, a Bayesian approach allows a straightforward solution, whereby the prior hypothesis is set to be the full regression model with all variables included (i.e. before stepwise regression). Stepwise regression is then run on each of the imputed datasets and the final variables that are selected is noted for each of the 25 imputed datasets. A comparison is made across the datasets, and the post-stepwise model (i.e. the variables that are selected post-stepwise regression) that occurs the most frequently is selected as the most

probable representation for all 25 datasets (207). It is important to note that the post-stepwise regression models for each dataset are compared as a whole, and not broken-down into the constituent variables. This is because if the post-stepwise regression model is fragmented and only some of the variables selected from one model, that model no-longer represents the best-fit for that dataset.

5.4.3 Posterior probabilities

Posterior probabilities use the ability of Bayesian statistics to update or test a hypothesis, based on empirical data. Posterior probabilities are used in Chapter 10 as a method for examining the reliability of a model fit to the data. Broadly, posterior probabilities involve predicting the probability of an event occurring given the data or evidence used to generate the prediction. In the case of the analysis in Chapter 10, the posterior probabilities are generated by randomly sampling values from the distribution of the observed values to generate a new dataset (207). The findings from the analysis of the observed data are then applied to the randomly generated dataset to examine if the findings hold. An example from the analysis undertaken in Chapter 10 can help explain the relevance of such analysis to this thesis.

5.4.3.1 Posterior probabilities: Example

In Chapter 10, LCMM is used to map the trajectory of response of individual patients to anti-TNF. These trajectories are then grouped into a pre-specified number of common trajectories (or classes) according to the latent class model specifications. The accuracy of the model given the data can be checked using Bayesian model checking and posterior probabilities. In this case, a new dataset is generated using random draws from the distribution of the BSRBR-RA dataset used in the analysis. The latent class model generated using the original data is then checked against the new randomly generated data and the probability of the randomly generated dataset to be grouped into the same classes as the original data can be compared. Strong agreement between the original and simulated dataset lends support to the proposed latent class model.

5.4.4 Bayesian information criterion

BIC is a technique used to help prevent overfitting in mathematical models. BIC is used in Chapter 10 where it is used to help identify the most appropriate number of classes for latent class analysis. BIC was first proposed by Schwarz in 1978 (208) and is described by the following equation:

$$\mathrm{BIC} = p\ln(n) - 2\ln(\hat{\theta})$$

Equation 12. Bayesian Information Criterion. Where p = number of parameters investigated (e.g. the number of variables or classes), n = number of data points (i.e. the number of observations), and $\hat{\theta}$ = the maximised value of the likelihood function

BIC can help select the best fitting and most parsimonious model. This avoids selecting too many variables in a model which may increase the model fit, but also increase the variance of the model. When comparing models, the lowest BIC represents the model with the best fit to the given data. An important assumption of BIC is that the sample size (n) should be much larger than the number of parameters (p) in the model. BIC is also not capable of managing complex collections of models or high dimensional statistics (such as stepwise variable selection).

5.5 Latent class mixed modelling (LCMM)

LCMM is used in analysis in Chapter 10 to model patients' disease activity whilst on anti-TNF over time. LCMM is an extension of linear mixed model theory which is used to model complex time-course data, and is particularly useful for epidemiological data (209).

The LCMM package (210) in R is used to undertake analyses, and both posterior probabilities and Bayesian Information Criterion (BIC; discussed in Chapter 5; 5.4.4) are used to identify the trajectory model with the best model fit.

When considering time-course data for multiple individuals, there are several factors that must be considered. Firstly, when considering the population as a whole, there is

likely to be variation in the spread of the data between participants (which may follow a bivariate normal distribution). At an individual participant level, with repeated measures over time, measurements would also have a degree of variation. Therefore, there is both between-subject variation, as well as within-subject variation. There may also be variation in the time between repeated measurements for participants (something that occurs ubiquitously in epidemiological data). To address this, linear mixed models use a two-step process to model both within- and between-participant data. Such models usually assume a normal distribution and are therefore primarily used for continuous variables (209). LCMM extends the linear mixed model to incorporate categorical, binary, ordinal and continuous but asymmetric data. In addition, LCMM can manage otherwise non-observed heterogeneity within a population, such as responders/non-responders (essential in modelling multiple trajectories of response). LCMM is also able to operate where longitudinal processes may be altered by one or multiple times-to-event (such as study drop-out, disease progression or drug switching). All these factors make LCMM an ideal modelling tool for mapping longitudinal response to anti-TNFs in the BSRBR-RA cohort.

The LCMM output can be visualised graphically as mean trajectories of response. This is achieved by plotting all the DAS28-ESR scores in the dataset. The LCMM algorithm then 'examines' each individual patient record (the sequential DAS28-ESR scores) over the duration of the dataset and 'joins the dots' of each of the sequential DAS28-ESR scores. This creates a trajectory for one patient. The process is repeated for each patient record to generate (in the case of the BSRBR-RA) thousands of trajectories. The individual trajectories are then grouped into the number of pre-determined classes and distinct 'mean trajectories' can be plotted that represent the trajectory that best fits the data. Patients are then assigned to the mean trajectory that their sequence of DAS28-ESR scores best fit.

5.6 Chapter summary

This chapter has outlined the key methodologies used in this thesis – in particular the longitudinal mapping of responses and Bayes theorem. The majority of methodologies

used in this thesis have been discussed. However, for the sake of brevity, there are some basic statistical methods (such as T-tests and Chi^2 tests) that have been used in analysis, but not specifically described in detail here.

Chapter 6

6 Dataset preparation

This chapter describes the specific preparation of the dataset that was used to enable the analyses in Chapter 9 and 10. Briefly, this includes:

- Application of inclusion criteria for the dataset prior to analysis.
- Examination of the patterns of missing data.
- Examination of the data (original and imputed data) after imputation.
- Identification of patients in sustained remission and LDA.

A different version of the BSRBR-RA dataset was used for the analysis described in Chapter 8. This is because all biologic class drugs (excluding Ro-Actemra™) were included in the analysis undertaken in Chapter 8, whereas only anti-TNF class drugs were included in the analyses undertaken in Chapters 9 and 10. As there was very little missing data identified for that analysis, specific details of the dataset and dataset preparation are described in that chapter. Further descriptions of 'the dataset' in this chapter refer to the version of the dataset that was extracted for use in the analyses undertaken in Chapters 9 and 10.

6.1 Inclusion criteria

The dataset was requested in line with the procedure outlined in Chapter 4 (4.8). As previously outlined, there is extensive data checking that is undertaken by the BSRBR-RA team, so very little data cleaning that is required. However, the following criteria were used to ensure dataset homogeneity required for the proposed analyses in Chapters 9 and 10. Data points outside the following boundaries were coded as missing and included in the subsequent missing data analysis and imputation.

- ESR values were limited to a maximum of 150 (the maximum value for the test).
- Visual analogue score values used to quantify the PGA were limited to less than 100 (the maximum value for the score).
- Patients not taking an anti-TNF at baseline were removed from the dataset.
- Only patients taking Enbrel™ (etanercept), Humira™ (adalimumab), Cimzia™ (certolizumab pegol) and Remicade™ (infliximab) were included.
- Individuals who were not bio-naïve at baseline were removed from the dataset.
- Individuals enrolled after September 2013 were removed from the dataset. Because this analysis focusses on data collected over the first three years of six-monthly data collection, individuals enrolled after September 2013 would not have been able to complete three years of follow-up. Therefore, to avoid classifying individuals as having missing data for events that have not yet happened, these records were removed. Likewise, because this analysis focuses on the first three years of data collection (the first six follow-up visits), data collected after the first six follow-ups was removed.
- Weight data > 200kg was coded as missing. Examination of the data revealed a few occurrences of implausible weight data. As such a pragmatic maximum limit of 200kg was chosen for weight data.
- Height data >250cm or <60cm was coded as missing. As with weight data, there were some implausible outlier data points for height data. Accordingly, pragmatic boundaries for height were selected.

6.2 Examining missing data patterns in the BSRBR-RA

The BSRBR-RA dataset used for the analyses in Chapters 9 and 10 is primarily stored in two formats. A wide-format dataset (one row per patient record) which contains all baseline data that is collected at registration (such as date of birth, height, weight and DAS28-ESR before starting a biologic). The longitudinal component of the database is stored in a long-format, where each subsequent follow-up is added as a new row in the database (many rows of data per patient record). Because the two components of the dataset are stored separately, the missing data are examined individually.

6.2.1 Missing data in the anti-TNF dataset

On initial inspection of missing data within the dataset, it is possible to see that there is very little missing data overall. Baseline data from registration has the least missing data (Table 7) compared with the longitudinal component of the BSRBR-RA dataset (Table 8). The greatest amount of missing data occurs in the DAS28-ESR scores collected in the longitudinal dataset (24.6%). Because the DAS28-ESR is generated by a calculation of its constituent parts (the tender and swollen joint count, visual analogue score and ESR), missing data in any of these variables make the calculation of the DAS28-ESR impossible, so missingness from each of the components is compounded when looking at the DAS28-ESR score.

No missing data	**1 - ≤5% missing**	**5 – 10% missing**	**>10% missing**
Patient ID Follow-up number Gender Age Anti-TNF type Age when starting anti-TNF	Year of onset of RA (1.0%) Smoking history (1.3%) Information on if first biologic (1.3%) Weight (2.9%) Tender Joint Count (3.3%) Swollen Joint Count (3.4%) VAS (4.0%) Date of form completion (4.3%)	HAQ (8.8%) ESR (9.8%)	DAS28-ESR (10.9%) Date of HAQ form completion (12.4%) Height (13.3%)
N = 14436			

Table 7. Baseline Missing Data

No missing data	**1 - ≤5% missing**	**>10% missing**
Patient ID Follow-up number Still on biologic Biologic Name	Change in biologic (1.0%)	Tender Joint Count (11.6%) Swollen Joint Count (11.6%) VAS (15.7%) ESR (19.7%) DAS28-ESR (24.6%)
N = 60031		

Table 8. Missing data in longitudinal dataset

The missing data pattern rejects the null hypothesis that the data are missing completely at random ($p < 0.001$) using Little's test for MCAR (190). This result is not surprising, as it is likely that if one item is incomplete on a BSRBR-RA form, there may be another related item missing on that form. Because data are not missing completely at random, it is necessary to visually examine the data set to ensure that the assumption of MAR can be justified.

6.2.2 Visual examination of missing data within the BSRBR-RA

Because the data are unlikely to have a missing data pattern that is MCAR, it is necessary to examine the dataset to see if there are any associations between missing and available data that might lead to the assumption of MAR being false. As outlined in Chapter 5 (5.1), there is no 'test' that proves or disproves MAR, so an examination of the dataset is required. An understanding of the method of data collection is required to ensure there are no study protocol-related issues that might cause a systematic bias. Chapter 4 outlines the study-specific methods of the BSRBR-RA. Overall, there are no methods that might lead to a systematic bias in the data collection. The next step is to examine the data to identify any patterns between the missing and complete data that might lead to suspicions of a MNAR data pattern. It is essential to be confident that data do not have a MNAR pattern before imputing data. One way of examining relationships between missing and available data is graphically. There are numerous ways of graphically representing missing data relationships, but two methods were primarily used in the examination of this dataset - hierarchical clustering of missingness and pairwise scattering of missingness. In depth comparisons of the data were undertaken, although for the sake of brevity, only representative plots are shown here.

6.2.3 Hierarchical clustering of missingness

Hierarchical clustering (211) (Figure 8) is a useful way of identifying associations between missing data in a dataset. By clustering missing data together, it is possible to see an association between records that have missing data from more than one variable.

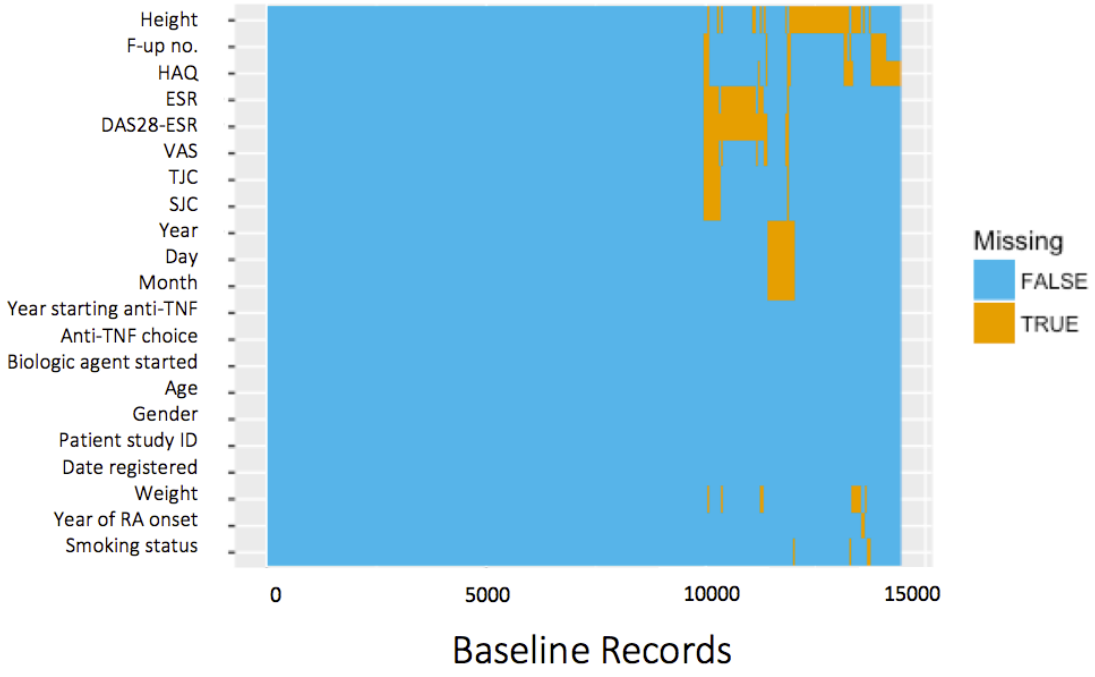


Figure 8. Missing data in baseline dataset clustered according to missingness. F-up no. = follow-up number, HAQ = health assessment questionnaire, VAS = visual analogue score, TJC = tender joint count, SJC = swollen joint count

Figure 8 demonstrates associations between missing data for tender and swollen joint counts, ESR and the PGA (quantified by the VAS) - suggesting that if one of the variables from tender or swollen joint count, ESR or VAS is missing, the other three variables are also likely to be missing. This association is not concerning, as at a practical level, if one aspect of a clinical examination is incomplete on a BSRBR-RA form (e.g. tender joint count), it is likely that other aspects of the same clinical assessment (e.g. swollen joint count) will be missing, and it is unlikely to represent a bias in missingness that might adversely affect analysis.

There is overlap with missingness for the DAS28-ESR and its components. This is because for this analysis, DAS28-ESR was calculated from the component parts collected in the dataset (rather than the reported DAS28 scores collected on the clinical questionnaires) and requires all four variables (tender and swollen joint counts, ESR, VAS) to generate a score.

There is also an association in those data that are missing one component of the date of assessment, being more likely to be missing the other components of the date data. Again, this is unsurprising as it is likely that someone who completes the form is likely to either complete the date, or not, rather than only complete a partial date. Other than these two areas of associated missingness, there is very little overlap in missing data that is of concern in the baseline dataset. However, these associations are likely to explain the reason that the null hypothesis of Little's test for MCAR was rejected. Overall, there do not appear to be unexpected associations between missing data that may influence the analysis of predictors of sustained remission. Exploration of the missing data in the longitudinal database was also undertaken. Little's test of MCAR rejected the null hypothesis of MCAR. However, similar missing data relationships to those seen in the baseline data were observed, and did not demonstrate any associations between missing data that rejected the assumption of MAR.

6.2.4 Pairwise scatter of missingness

Pairwise scatter plots of missingness were also used to examine the distribution of data in the missing and complete data to identify if there were any patterns in the missing data that might suggest MNAR (Figure 9).

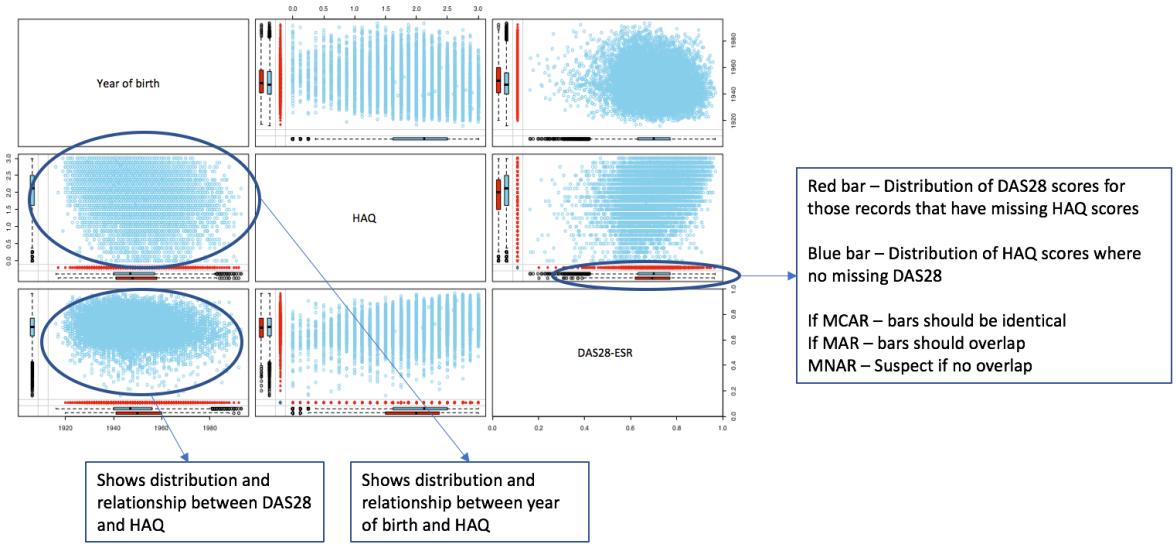


Figure 9. Visual exploration of missing data using pairwise scatter in the baseline dataset

Visual exploration of the missing data at baseline and in the longitudinal dataset was undertaken for all variables captured by the BSRBR-RA as outlined above, and did not demonstrate any concerning relationships that disproved the assumption of MAR. As the assumption of MAR is a reasonable one, multiple imputation can be undertaken. As described in Chapter 5 (5.1.4.5.1), multiple imputation was undertaken using the Amelia package in the R environment.

6.2.5 A note on manual interference with a dataset - remission and LDA at baseline

Examination of the baseline data demonstrated 66 individuals out of a total of 14436 (0.5% of the cohort) whose disease activity was recorded as LDA or less at baseline prior to commencing anti-TNF. Of these, 26 (0.2% of the cohort) were in remission at baseline before commencing anti-TNF. It may appear sensible to remove these records from the analysis as these individuals have achieved part of the outcome (remission/LDA) before starting anti-TNF. However, manually removing these individuals from the analysis would generate a new challenge in deciding which data to remove. This is because by removing these known data from the dataset, artificial assumptions are imposed on the data that are propagated through to influence the missing data which in turn influences imputation data. For example, in the analysis of sustained remission, should only those individuals who are in remission at baseline be excluded, leaving those who may be in LDA at baseline in the analysis? If so, then for the subsequent analysis of LDA, should the additional individuals who are in LDA but not remission at baseline be removed? If these individuals are removed, this makes the cohorts used in the sustained remission and sustained LDA analysis different, with different imputation parameters. Removing individuals who are in LDA or less at baseline from both the sustained remission and sustained LDA analysis also causes theoretical problems, as the difference between MDA and LDA is much smaller than the difference between MDA and remission, meaning that the differential effect of using LDA as a cut-off does not have equal effect on both disease activity classes, and hence subsequent analyses. In this case, minimal manual interference with the raw data is likely to cause least bias and given the very small number of individuals as a proportion

of the whole dataset (<0.5%), it is unlikely that these individuals would significantly alter the findings from this analysis.

6.2.6 Examining imputed data for baseline variables

Following imputation of the data in the BSRBR-RA (used in Chapters 9 and 10), the distribution of the values of the imputed variables was compared against the distribution of values of the original variables to ensure the new values are a reasonable approximation to the original dataset.

Examining the distribution of the original and imputed values for the baseline dataset demonstrates reasonable agreement in spread of data, providing confidence in the imputed values (Figure 10).

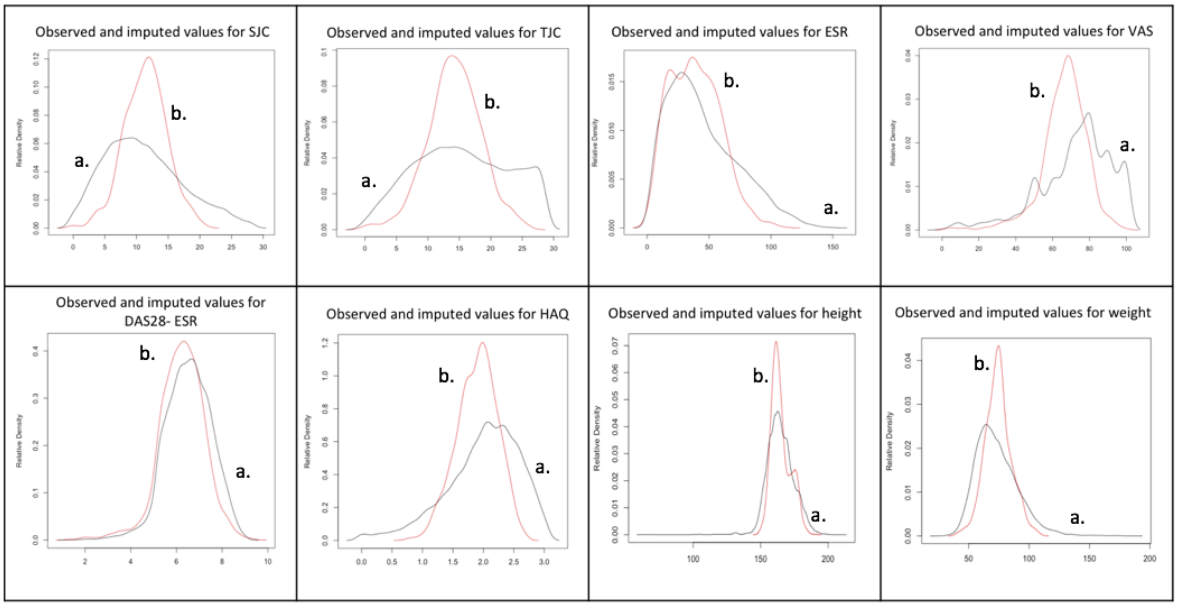


Figure 10. Comparison of original and imputed data for baseline dataset. Black line (a) - Original data, Red line (b) - Imputed data. SJC – Swollen joint count, TJC – Tender joint count, VAS – Visual analogue score. X-axis – range of values for specific variable (i.e. 0-28 for TJC). Y-axis – relative density

6.2.7 Examining imputed data for longitudinal variables

The distribution of the data for the longitudinal dataset is somewhat different to that observed in the baseline dataset (Figure 11). The distribution of the data for tender and swollen joint counts are left-skewed and the mean value of the imputed data are greater than that in the observed data. This is not surprising, as it would be expected that the overall number of swollen and tender joints would decrease with treatment, which might skew the data. An attempt was made to normalise the spread of the data using a log-transformation. Because zero is a plausible number of swollen or tender joints to be recorded, and the natural logarithm of zero is infinity, all tender and swollen values were increased by +1 (to move them onto a 1-29 scale). However, log-transformation of these values generated a very unusual data pattern for which the imputed dataset was less congruent than the original. As a result, the original data (not log-transformed) has been used in all analyses. Although the imputed values for the tender and swollen joint count are not closely aligned to the original values, the imputed values for the DAS28-ESR are much more closely approximated.

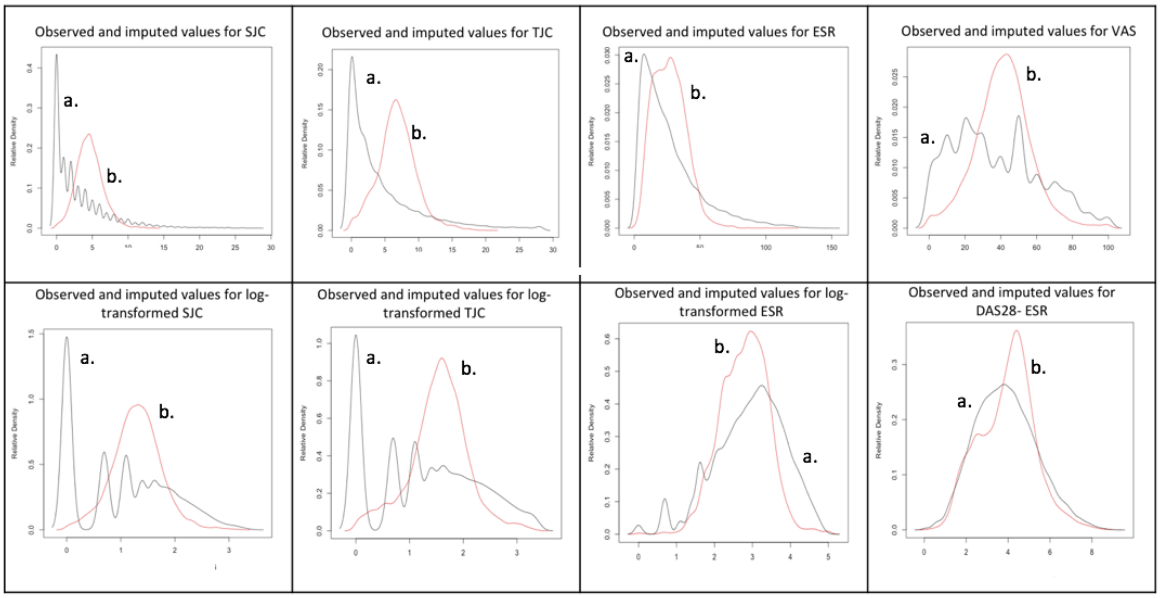


Figure 11. Comparison of original and imputed data for longitudinal dataset. Black line (a) - Original data, Red line (b) - Imputed data. SJC – Swollen joint count, TJC – Tender joint count, VAS – Visual analogue score. X-axis – range of values for specific variable (i.e. 0-28 for TJC). Y-axis – relative density

As mentioned previously (Chapter 5; 5.1.5.2), imputing data that are derived from components that are also imputed could lead to implausible relationships (i.e. an

imputed DAS28-ESR score may not be the same as the score calculated by its component parts for that record). Therefore, a secondary DAS28-ESR score column (DAS28-2) was created (post-imputation) and DAS28-ESR scores were calculated based on the (original and imputed) component parts of the DAS28-ESR. The distribution of these DAS28-2 data were then compared with distribution of the complete and imputed DAS28-ESR values and were not found to be significantly different. Therefore, to minimise any impossible relationships between the components of the DAS28-ESR and the score itself, the calculated (rather than the imputed) DAS28-ESR values were used in subsequent analyses.

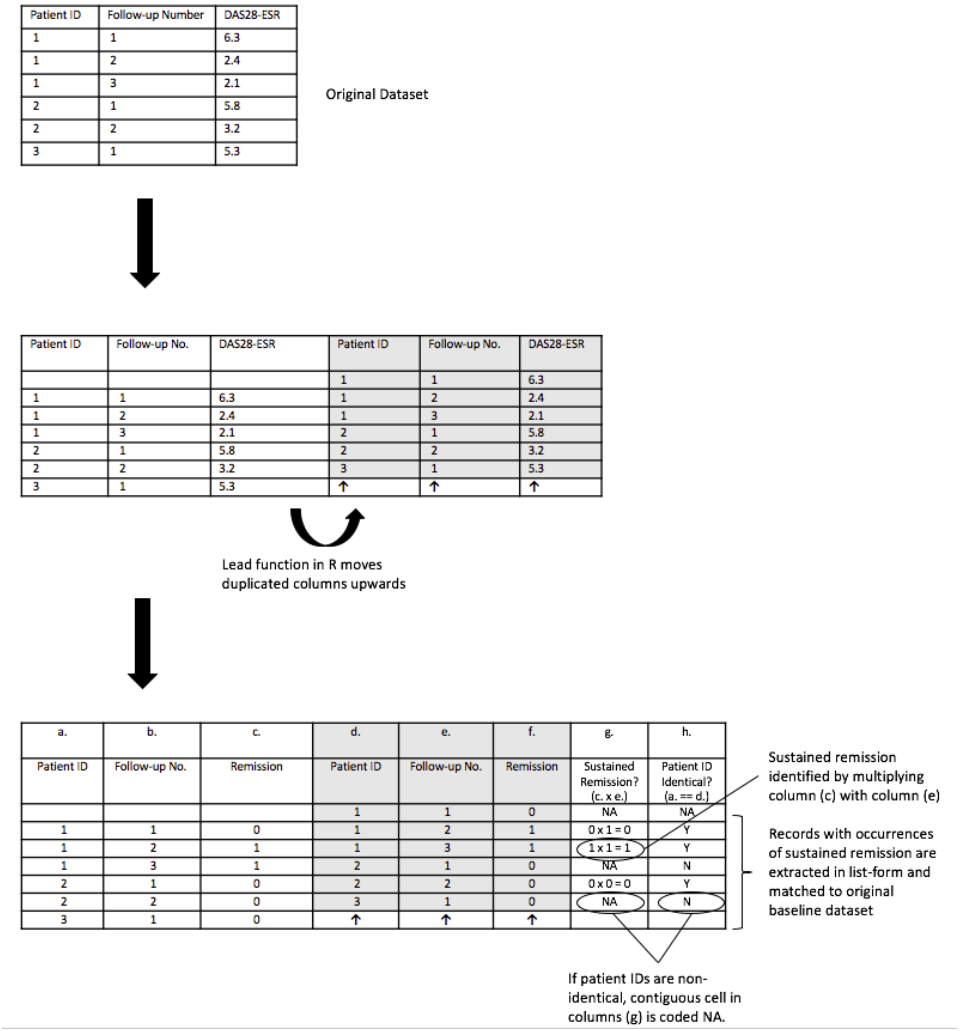
Although the imputed data for the longitudinal data had a poor match with the original data for the tender and swollen joint counts, the DAS28-ESR data was a good fit. Because identification of individuals in sustained remission and LDA will use the DAS28-ESR values rather than the components of the score, the imputed values are acceptable to proceed with analysis.

6.3 Identifying sustained remission and LDA

6.3.1 Binary vector multiplication

The analysis undertaken in Chapter 9 requires those individuals who are in a state of remission or LDA for two consecutive follow-ups to be identified. This presents a challenge as the longitudinal component of the BSRBR-RA is presented in 'long-format' (i.e. sequential follow-up visits are added as additional rows in the dataset, rather than additional columns). To identify an individual who has two sequential follow-ups at any point within the first six follow-up visits, a looping function could be created that 'reads' down the dataset and identifies any two rows that are in remission or LDA. However, this would require a complex loop command as it would also have to concurrently read the patient's ID number and restart the loop function for each new patient. It would also not be possible to record the output from such a looping function within the same dataset, which would cause further problems. A neater solution is to

vectorise the solution and then use binary multiplication (Figure 12) to identify and categorise sustained remission/LDA.



Patient ID	Follow-up Number	DAS28-ESR
1	1	6.3
1	2	2.4
1	3	2.1
2	1	5.8
2	2	3.2
3	1	5.3

Patient ID	Follow-up No.	DAS28-ESR	Patient ID	Follow-up No.	DAS28-ESR
			1	1	6.3
1	1	6.3	1	2	2.4
1	2	2.4	1	3	2.1
1	3	2.1	2	1	5.8
2	1	5.8	2	2	3.2
2	2	3.2	3	1	5.3
3	1	5.3	↑	↑	↑

a. Patient ID	b. Follow-up No.	c. Remission	d. Patient ID	e. Follow-up No.	f. Remission	g. Sustained Remission? (c. x e.)	h. Patient ID Identical? (a. == d.)
			1	1	0	NA	NA
1	1	0	1	2	1	0 x 1 = 0	Y
1	2	1	1	3	1	1 x 1 = 1	Y
1	3	1	2	1	0	NA	N
2	1	0	2	2	0	0 x 0 = 0	Y
2	2	0	3	1	0	NA	N
3	1	0	↑	↑	↑		

Figure 12. Method used to identify sustained remission from the longitudinal BSRBR-RA dataset

Binary vector multiplication involves duplicating the patient ID, follow-up number column and DAS28 score for each record. The 'lead' function in R can then be used to shift all the records in the newly duplicated columns up by one row. The means that when reading across the datasheet, the DAS28 score from the current and next follow-up are in neighbouring columns. The continuous DAS28 data is then categorised into 'remission' and coded as 1, or 'not in remission', and coded as 0. Binary multiplication across the datasheet is then used to identify those who are in remission on two successive points by multiplying the original (now categorised) DAS28 score column with the 'new' duplicated and categorised DAS28 column that has been advanced by 1

row, with the product being recorded in a subsequent new column (the sustained remission column). This means that if an individual is in remission on two successive occasions, they will have a categorised DAS28 score of 1 on both occasions. 1 x 1 = 1, so 1 will be recorded in the sustained remission column. If an individual is in remission at one follow-up visit, but not the next one (or vice versa) then the multiplications will be either 1 x 0 or 0 x 1, both of which equal 0, which is recorded in the sustained remission column. The sustained remission column will now have codes of 1 or 0 to identify if an individual was in sustained remission at each successive follow-up point.

To avoid one patient record overlapping into another when using the lead function in R, where the patient IDs are not identical, the contiguous sustained remission cell is recoded as NA. A subset of this data is taken for all records where sustained remission is identified. Finally, to avoid patients with more than one episode of sustained remission occurring during their first 6 follow-up visits being 'counted' more than once, the 'unique' function in R is used. This creates a separate list of patient IDs from the subsetted data, which is de-duplicated, so that even if a patient has more than one episode of sustained remission (identified by a duplicated patient ID), they are only counted once, and only appear in regression analyses once. This 'unique' list of patients who have ever achieved sustained remission in the first six follow-ups can then be matched to the baseline dataset (which is in wide format) using the match function in R (%in%). Patient records are given a binary code '1' to identify them as patients who have achieved sustained remission. The remaining patients can then be coded as '0' (no sustained remission), allowing logistic regression to be undertaken for the specified variables.

6.4 Chapter discussion

As was expected, missing data were identified in the BSRBR-RA dataset. These data were not MCAR. However, visual examination of the dataset and appraisal of the study methods has not demonstrated any significant concerns that the pattern of missingness is MNAR. Imputed data had a good match to the spread of the original dataset for the baseline dataset. Imputed data had a less good fit with the original data

for the longitudinal dataset, and log-transformations of the data did not improve this. The imputed DAS28-ESR data had a good approximation with the original data however, and as the DAS28-ESR data were used to identify patients in sustained remission/LDA, use of the imputed dataset was acceptable.

Identification of sustained remission and LDA was possible using binary vector multiplication, which will allow analysis of these outcomes to be subsequently examined (Chapters 9 and 10).

6.5 Key points

- Missing data from the BSRBR-RA dataset have a MAR pattern.

- Imputation of missing data provided values that are reasonably matched to the original data.

- Patients who have achieved either sustained remission and/or sustained LDA can be identified from the BSRBR-RA dataset.

Chapter 7

7 Factors associated with sustained remission in RA in patients treated with anti-TNF: reviewing the evidence

As previously outlined (Chapter 1; 1.8), aggressive treat-to-target strategies, alongside increased use of biologic agents such as anti-tumour necrosis factor antibody (anti-TNF)(125,138), have improved outcomes for patients with RA and the aim of achieving sustained remission is a realistic aspiration.

However, the response to anti-TNF is variable and requires further investigation prior to undertaking subsequent analyses. Understanding the existing evidence of how the demographic and clinical features of a patient may influence the likelihood of achieving sustained remission with anti-TNF will help inform subsequent analyses of the BSRBR-RA.

The majority of published studies report remission rates at a single time point, or sequential point remission rates. Previous systematic reviews have only investigated predictors of point remission in RA (212), however, given the chronicity of a condition such as RA and the long-term benefits of remission, a durable positive response to anti-TNF is a more clinically relevant outcome. Methods and results are combined in this chapter to allow cross-referencing in the subsequent chapters of this thesis.

7.1 Aim

To undertake a systematic review of the literature to evaluate the existing evidence for demographic and clinical factors associated with the achievement of sustained remission in individuals with RA treated with anti-TNF.

7.2 Patients and methods

The systematic review protocol was registered prospectively with the PROSPERO database (http://www.crd.york.ac.uk/PROSPERO/, reference CRD42015015983). PRISMA-P (213) and PRISMA (214) recommendations were followed in the development, implementation and reporting of the review.

7.3 Inclusion criteria

To be included in the review, papers had to meet the following criteria:

1. Phase three or four clinical trials, long-term extension trials or cohort studies reported as original research in the form of journal papers;

2. Adults (≥18 years of age) with RA according to ACR 1987(100) or ACR/EULAR 2010 (101) criteria;

3. Report on anti-TNF used for the treatment of RA;

4. Report on at least one measure of RA disease activity using DAS (146), DAS28 (147,215), CDAI (149), SDAI (148), ACR/EULAR remission (170) or ARA 1981 remission criteria (151);

5. Report on predictors of sustained remission (at least six months) (216).

7.4 Exclusion criteria

Studies where it was not possible to isolate the required data on patients in sustained remission, case-control, cross-sectional studies, case reports/series, phase one and

two/laboratory studies, qualitative studies, survey-based studies, narrative reviews, conference abstracts and editorials were excluded.

7.5 Search methods for identification of studies

EMBASE, Medline and the Cochrane Controlled Trials Register were searched using the Ovid platform to 4th September 2015. The full search strategy for Medline is provided in Table 9. No language restriction was applied to search results. Reference lists of included studies were searched for additional citations and all authors were contacted for additional information to assist with the review and meta-analysis. Additional data were kindly provided by Dr Barnabe, Dr Einarsson, Dr Balogh, and Professor Tanaka.

Diagnosis (KW)		Drug (OR, KW)		Outcome (OR, KW)
Rheumatoid arthritis (74933)	A N D	Tumor necrosis factors* (2066) Tumor necrosis factor-alpha* (116033) Antibodies, Monoclonal* (180019) Antibodies, Monoclonal, Humanized* (25271) Antibodies, Monoclonal, Murine-Derived* (9706) Receptors, Tumor Necrosis Factor* (16571) Certolizumab pegol. kw (373) Golimumab. kw (326) Infliximab. kw (8655) CT-P13* (Infliximab biosimilar). kw (3) Adalimumab. kw (3549) Etanercept. kw (4027) Anti-TNF .kw (6196) Tnf.kw (108867) TNFR-Fc fusion protein .kw (4201) Combined hits with OR 328541 Diagnosis AND Drug combined give 10829 hits	A N D	Remission, Spontaneous* (15099) Treatment outcome* (658026) Severity of illness index* (171810) Reproducibility of results* (281754) Remission induction* (32979) Induction chemotherapy* (5743) Recovery of function* (34252) Pharmacology, clinical* (2201) Drug evaluation* (80081) Drug utilization* (20410) Evaluation studies as topic* (119716) Maintenance chemotherapy* (1435) Quality of health care* (115945) Sustained remission (kw) (930) Maintained remission (kw) (130) Prolonged remission (kw) (611) Responder (kw) (10266) Disease control (kw) (46264) Disease activity (kw) (24636) DAS28 (kw) (1474) Clinical Disease Activity Index (kw) (184) Simplified Disease Activity Index (kw) (128) DAS (kw) (42944) Remission (kw) (109223) Combined above hits with OR 1676984

Table 9. Medline Search Criteria, *Denotes MeSH heading, (kw) denotes keyword, Search Performed using Ovid interface on 4th September 2015. All above articles were searched as keywords - Diagnosis AND Drug AND Outcome gives 3797 hits. Limited to humans, and adolescent, all adult, young adult, middle age, middle aged, all aged, aged- gives 2413 hits

7.6 Assessment of studies for inclusion into the review

All search results were dual screened with dual data extraction and quality scoring using a custom Access™ database. The quality of studies was assessed using the Newcastle-Ottawa Scale (217). A narrative review of studies with relevant quantitative data extraction was undertaken. Corresponding authors were contacted to obtain unadjusted data to enable meta-analysis where appropriate. Sources of heterogeneity were investigated through structured critical appraisal.

7.7 Meta-analysis

Statistical analysis was undertaken using Review Manager Software version 5.3 (218). Because the factors incorporated in calculating adjusted odds ratios (OR) were not consistent between studies, unadjusted OR were used in meta-analysis where data were available from at least three studies. A random effects model was used to allow for between study variation. Heterogeneity between studies was assessed using I^2 (219) and publication bias was assessed using funnel plots.

7.8 Results

7.8.1 Study identification

A total of 4438 papers were identified from the search strategy. 4220 records were excluded and 218 full text papers, including 50 randomised controlled trials, were assessed. Six papers met the inclusion criteria and were included in the review (220-225). One of these papers (224) had included one patient aged less than 18 years old in one subgroup (personal correspondence), however, the mean age of all the subgroups and the overall cohort was in line with the other included papers, and it was decided to include the study in the review. The screening process is summarised in Figure 13.

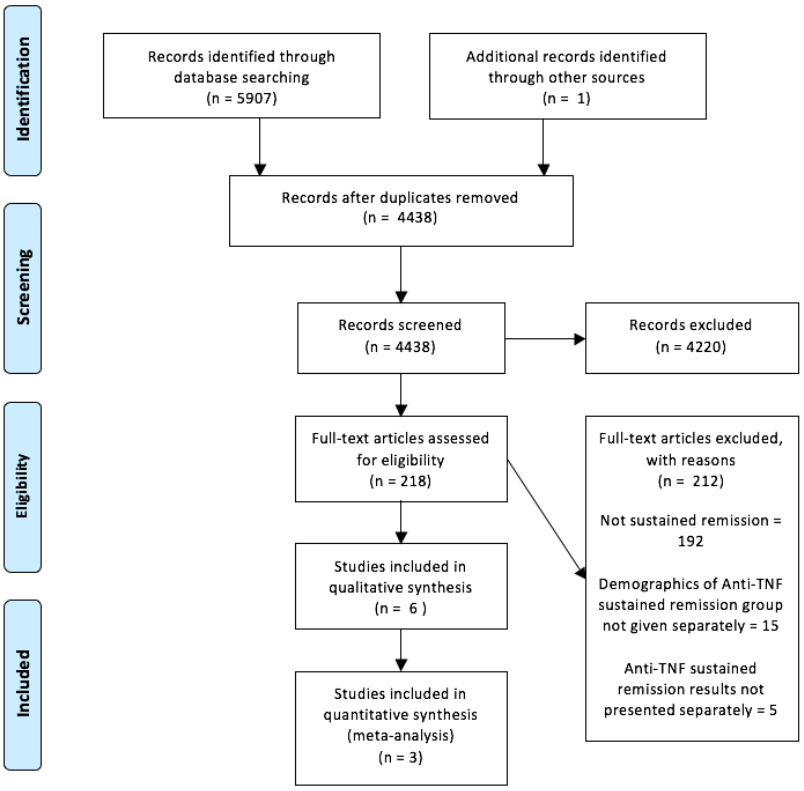


Figure 13. Results of screening process

7.8.2 Study design

The characteristics of the included studies (and quality assessed using Newcastle-Ottawa scores) are summarised in Table 10 and Table 11. Two of the included studies were multicentre studies (223,224) coordinated from one hospital, one of which was an open label, non-randomised trial (223). One study was a retrospective case note review (222) and three were registry studies (220,221,225).

[This page is intentionally blank]

Author, year	Country	Study duration & design	Sample size & gender (female)	Anti-TNF group inclusion criteria	Co-medication (included/ excluded)	Mean age & disease duration (yrs)	Disease activity measure	Mean +/- SD disease activity	HAQ score, mean or median +/- SD or (range)		Proportion of cohort in sustained remission (%), criteria	
Brocq 2009	France	7 years, retrospective cohort	304, 81.3[1]%	RA patients with failure of methotrexate and DAS28 ≥ 5.1. Previous anti-TNF treatment allowed.	Included Prednisolone (≤5mg) synthetic DMARD (stable dose) excluded NSAID users	58[1], 10 [1]	DAS28	6.4[1]	Not given		8.2% DAS28	
Furst 2011	USA	7 years, prospective cohort	3170, 77.4-79.8% (depending on subgroup)	RA patients starting on anti-TNF	Included synthetic DMARD (dose unspecified)	54.3, 2.5	DAS28	4.4 ± 1.4	M-HAQ values	0.5 (mean) ± 0.5 (SD)	8.9%, DAS28	Overall cohort sustained remission 7.6% by CDAI and 7.9% by DAS28
							CDAI	21.8 ± 13.7			9.7%, CDAI	
						54.1, 7.7	DAS28	4.5 ± 1.5		0.5 (mean) ± 0.5 (SD)	11.6%, DAS28	
							CDAI	21.0 ± 13.8			9.5%, CDAI	
						60.5, 20.3	DAS28	4.6 ± 1.5		0.6 (mean) ± 0.5 (SD)	4.9%, DAS28	
							CDAI	21.5 ± 13.1			4.2%, CDAI	
Balogh 2013	Ireland	1 year, Prospective Cohort	273, 74.4%	Biologic naïve RA patients with persistent disease activity starting on anti-TNF	None specified	59.9, 13.4	DAS28, ACR/EULAR	5.33 ± 1.07	Not given		9.9%, DAS28 (from ACR/EULAR subgroup)	
Barnabe 2014	Canada	7 years, prospective cohort	1116, 74.0%	Biologic naïve, RA refractory to parenteral methotrexate/ leflunamide, at least 2 study visits	None specified	54.4, 12.2	DAS28	6.03 ± 1.30	1.62 (mean), 0.68		21.5%, DAS28	
							CDAI	38.52 ± 13.59			9.9%, CDAI	
							SDAI	40.90 ± 14.66			4.5%, SDAI	
							ACR/EULAR 2011 (no CRP)	NA			10.8%, ACR/EULAR 2011 (minus CRP component)	
							ACR/EULAR	NA			6.8%, ACR/EULAR 2011	

Author, year	Country	Study duration & design	Sample size & gender (female)	Anti-TNF group inclusion criteria	Co-medication (included/ excluded)	Mean age & disease duration (yrs)	Disease activity measure	Mean +/- SD disease activity	HAQ score, mean or median +/- SD or (range)	Proportion of cohort in sustained remission (%), criteria
Tanaka 2015	Japan	3 years, open label, non-randomised study	197, 84.8%	RA patients with inadequate response (DAS28-ESR ≥3.2) to methotrexate and/or other non-biological DMARDs	Excluded steroid use NSAID and coxib use variable synthetic DMARD use	60.7, 8.9	DAS28-ESR	5.4	Not given	38.1%, DAS28-ESR
Einarsson 2015	Sweden	10 years, prospective cohort	2416, 77%	RA patients with active disease + ≥1 failed previous DMARD. Previous biologic allowed.	None specified	56.0, 11.8	DAS28	5.5	1.3	15.8%, DAS28

Table 10. Characteristics of included studies (continued). 1. Baseline data extracted from Brocq *et. al.* Joint Bone Spine 2007;74: 148–54

	Category	Acceptability criteria	Brocq 2009	Furst 2011	Balogh 2013	Barnabe 2014	Einarsson 2015	Tanaka 2015
Selection	Representativeness of the exposed cohort	Truly or somewhat representative of the average RA patient using anti-TNF in the community	Nil	*	Nil	*	*	Nil
	Representativeness of the non-exposed cohort	Drawn from the same community as the exposed cohort	*	*	*	*	*	*
	Ascertainment of exposure	Secure record or structured interview	*	*	*	*	*	*
	Demonstration that the outcome was not present at the start of the study	Documented that patients were not in remission at the time of entry	*	*	*	Nil	Nil	*
Comparability	Comparability of cohorts	Subgroups (exposed and non-exposed) were drawn from the same cohort (2 stars)	**	**	**	**	**	**
Outcome	Assessment of outcome	Independent blind assessment or record linkage	*	*	*	*	*	*
	Was follow-up long enough for outcome to occur	At least 6 months' follow-up after the first review following baseline visit	*	*	*	*	*	*
	Adequacy of follow-up	All subjects accounted for or <10% of patients lost to follow-up (unless detailed description that those lost to follow-up would not have introduced bias)	Nil	Nil	*	*	*	*
Summary		Selection (max 4 *)	***	****	***	***	***	***
		Comparability (max 2*)	**	**	**	**	**	**
		Outcome (max 3*)	**	**	***	***	***	***

Table 11. Newcastle-Ottawa quality scoring of included studies

7.8.3 Sources of heterogeneity

All studies included in this review were observational by design and therefore there was variation in the range of treatments and patients included. Neither of the two multicentre studies (223,224), or the retrospective case note review (222) described their selection criteria and therefore there is a potential for the introduction of bias. Three of the included papers were registry studies (220,221,225) and are more likely to be representative of the general RA population treated with anti-TNF.

7.8.4 Definitions of sustained remission utilised

The definitions of sustained remission varied across the included studies. The minimum length of time that different studies defined sustained remission as, varied from at least six months, to nine months, or 'two consecutive visits' (verified to be at least six months (226)); and a range of outcome measures (DAS28, CDAI, SDAI, ACR/EULAR criteria) were also used. Additionally, Einarsson et. al. (225) did not exclude patients who were in a state of sustained remission who had a single episode of increased disease activity. However, it was less clear how the other studies handled these cases.

7.8.5 Missing data

The extent of missing patient data in sustained remission was not clear in the study by Balogh et al. (224); and detail on 25% (seven patients) of the cohort in sustained remission was missing from the study by Brocq et al. (222) which meant that data from these studies could not be included in meta-analysis. A total of 46 patients (1.9%) were lost to follow-up in the study undertaken by Einarsson et al. (225), and last observation carried forward and LUNDEX correction was used to account for incomplete follow-up visits (227). Barnabe et al. (220) and Furst et al. (221) did not use imputation, but did not give any information on missing data. LOCF was used to impute missing data in the HONOR Study (223).

7.8.6 Achievement of sustained remission

There was wide variation in rates of sustained remission. The highest rate of DAS28 sustained remission was observed in the HONOR study (223) (38.1%), and the lowest rate of DAS28 sustained remission was noted in the CORRONA population (221) (7.9%).

7.8.7 Anti-TNFs and concomitant medications studied

The studies identified in this systematic review include a range of anti-TNF medications. Some studies (220,222,223,225) specifically reported which anti-TNFs were studied, whereas the studies by Furst et al. (221) and Balogh et al. (224) did not. Very little data was available for patients using the newer anti-TNF medications (certolizumab pegol and golimumab), and no data were available for biosimilar anti-TNF medications. Additionally, the use of concomitant allowable drug use (such as prednisolone and NSAIDs) differed between included studies (Table 10).

Despite these differences, there were many similarities in the baseline demographics, including mean age, gender, and concomitant synthetic DMARD use (Table 10) suggesting that although there is likely to be heterogeneity between studies, there are sufficient similarities to allow comparison between studies.

Study (n)	Predictor	Outcome measure used	Effect size	Association with sustained remission
Barnabe 2014 (1116)	Baseline physician global (High)	DAS28	OR 0.80, 95% CI 0.66 – 0.99	Negative
	Obesity	ACR/EULAR Boolean (excluding CRP)	OR 0.30, 95% CI 0.10 – 0.90	
	Tender Joint Count (high)	CDAI	OR 0.96, 95% CI 0.92 – 1.00	
	Early response to treatment (<16 weeks)	DAS28	OR 1.88, 95% CI 1.27 – 2.78	Positive
Furst 2011 (3179)	Higher baseline disease activity	DAS28	OR 0.37, 95% CI 0.19 – 0.73	Negative
		CDAI	OR 0.57, 95% CI 0.39 – 0.84	
	Disability	CDAI	OR 0.25, 95% CI 0.08 – 0.79	
	Increased disease duration (5-yearly increments)	CDAI (but not by DAS28)	OR 0.85, 95% CI 0.75 – 0.97	
	Age	DAS28 (but not by CDAI)	OR 0.79, 95% CI 0.63 – 1.00	
	Female gender	DAS28 (but not by CDAI)	OR 0.43, 95% CI 0.23 – 0.82	
	Concomitant prednisolone	CDAI (but not by DAS28)	OR 0.69, 95% CI 0.47 – 1.00	
	Prior anti-TNF use	CDAI	OR 0.98, 95% CI, 0.96 – 1.00	
		DAS28	OR 0.72, 95% CI, 0.54 – 0.94	
	Concomitant methotrexate	CDAI	OR 1.55, 95% CI 1.00 – 2.42	Positive
		DAS28	OR 2.83, 95% CI 1.18 – 6.80	
Balogh, 2013 (273)	Tender Joint Count	DAS28 (as a subgroup of ACR/EULAR criteria)	OR 0.910, p 0.031	Negative
	Increasing age		0.942, p<0.0001	
Tanaka 2015 (197)	Lower Patient Global Score	DAS28-ESR	41.3 vs. 54.9 mm, p=0.0004	Positive
	Shorter disease duration		7.5 vs. 9.6 yrs, p= 0.005	
	Lower baseline HAQ-DI score		0.96 vs. 1.42, p<0.0001	
	Lower baseline ESR		44.1mm/hr vs. 53.0 mm/hr, p=0.0374	
	Lower baseline DAS28-ESR		5.11 vs. 5.70, p=0.005	
Einarsson 2015 (2416)	Female gender	DAS28	OR 0.57, 95% CI 0.44 – 0.75	Negative
	Higher baseline disease activity		OR 0.62, 95% CI 0.55 – 0.70	
	Earlier calendar year of starting anti-TNF		OR 0.89, 95% CI 0.85 – 0.93	
	Higher HAQ		OR 0.39, 95% CI 0.31 – 0.49	
	Increasing age		OR 0.98, 95% CI 0.97 – 0.99	
	Concomitant methotrexate		OR 2.02, 95% CI 1.51 – 2.71	Positive

Table 12. Predictors of sustained remission

7.8.8 Impact of patient demographics on sustained remission

7.8.8.1 Gender

Female gender was negatively associated with DAS28 sustained remission in two of the studies (221,225) (Table 12). In contrast, Barnabe et al. did not find female gender to be significantly associated with sustained remission by DAS28, ACR/EULAR 2011, or SDAI criteria using multivariate modelling, although univariate analysis (personal communication; univariate analysis used in meta-analysis) did suggest an association. No association between sustained remission and gender was identified by Tanaka et al. (223), and was not reported by the remaining studies (222,224). Meta-analysis demonstrated a reduced likelihood of achieving sustained remission in females compared with males with low data heterogeneity and a low likelihood of publication bias (OR 0.53, 95% CI 0.44 – 0.63; Figure 14 and Figure 15).

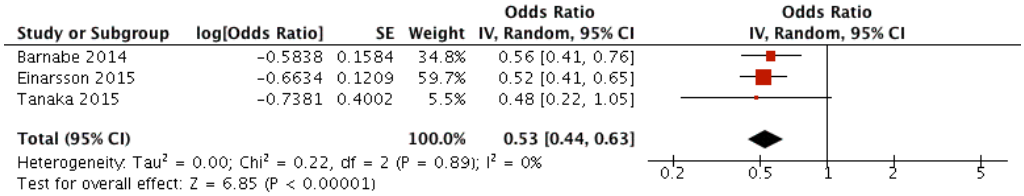


Figure 14. Effect of gender on achieving sustained remission

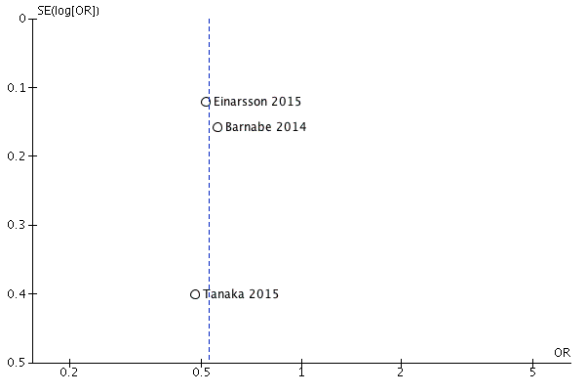


Figure 15. Publication bias of effect of gender on achieving sustained remission

7.8.8.2 Age

Increasing age was negatively associated with sustained remission by DAS28 in three of the studies (221,224,225) but was not associated with sustained remission in the study by Tanaka et al.(223) and was not reported by the remaining studies (220,222). Uniform unadjusted data were not available for this variable to enable meta-analysis.

7.8.8.3 Obesity

The only study to report the relationship between obesity and sustained remission identified a negative association according to ACR/EULAR Boolean criteria excluding the CRP, but not by the other remission criteria included in the study (220).

7.8.9 Impact of disease characteristics on sustained remission

7.8.9.1 Baseline disease activity

Higher baseline disease activity was associated with a reduced likelihood of achieving sustained remission using the DAS28 score in three studies (221,223,225) and CDAI in one (221). No association was noted between baseline disease activity and subsequent sustained remission in the multivariate analysis by Barnabe et al.(220). The association between baseline disease activity and attainment of sustained remission was not reported in the remaining studies (222,224).

7.8.9.2 Patient global score

A lower baseline patient global score was associated with sustained remission in the HONOR study (223). However, the only other study to include patient global scores did not identify any association (220).

7.8.9.3 Acute phase reactants

An elevated ESR was negatively associated with sustained remission in the HONOR study (223), but not in the study by Barnabe et al.(220). The CRP was not associated with sustained remission in the two studies where it was reported (220,225).

7.8.9.4 Number of tender and swollen joints

A greater number of tender joints negatively predicted sustained remission by CDAI criteria (220) and DAS28 criteria (224). However, no association was identified by Tanaka et al.(223). A higher swollen joint count was not identified as being associated with sustained remission in any of the four studies that reported this data (220,221,223,224).

7.8.9.5 Functional impairment

Higher rates of patient-reported functional impairment at baseline (assessed using the Stanford Health Assessment Questionnaire; HAQ) were consistently associated with lower rates of sustained remission (221,223,225). Only one study did not identify an association between baseline HAQ score and sustained remission on multivariate analysis (220). The effect of baseline functional impairment on remission status was not reported in the remaining studies (222,224).

7.8.9.6 Disease duration

One study identified that increased disease duration (stratified into five-yearly increments) was associated with a decreased likelihood of achieving sustained remission with the CDAI but not DAS28 criteria (221). Shorter disease duration was associated with an increased likelihood of achieving sustained DAS28 remission in one study (223). The remaining studies either found no association (220,225) or did not

report on the association between disease duration and sustained remission (222,224).

7.8.9.7 Early response to treatment

Response at 16 weeks after treatment was only reported by one study (220) and was associated with an increased likelihood of achieving of sustained remission by DAS28 criteria.

7.8.9.8 Concurrent and past medication use

Concomitant methotrexate use was positively associated with sustained remission by DAS28 criteria (221,225) and CDAI (221). However, Tanaka *et al.* (223) did not find any significant difference in baseline methotrexate dose between the sustained and non-sustained remission groups.

7.8.9.9 Prednisolone

Prednisolone use was negatively associated with sustained CDAI but not sustained DAS28 remission in one study (221). However, no association was identified in the study by Einarsson *et al*. (225). Prednisolone use was restricted to a stable dose of less than 5mg in one study (222) and patients taking concomitant corticosteroids were excluded from two studies (220,223). The remaining study did not report corticosteroid use (224).

7.8.9.10 Prior anti-TNF use and efficacy

Furst *et al*. (221) was the only study to report data on prior anti-TNF use and found that this was negatively associated with sustained remission in both DAS28 and CDAI measurements. Einarsson *et al.* (225) investigated time to sustained remission for each anti-TNF and found that etanercept was associated with an increased likelihood of

achieving sustained remission within the first twelve months on treatment when compared with infliximab.

7.9 Chapter discussion

Despite the variability in both the definition of sustained remission, and the predictive factors reported by each study, some common themes have emerged. One of the most striking findings was the paucity of evidence available for sustained remission as an outcome. From over 4000 possible manuscripts identified in the search, only six studies were identified which met the inclusion criteria, all of which were observational. With the exception of one study (223), the proportion of patients achieving sustained remission ranged from only 4.5% to 15.8%.

The search was updated to include papers published to the end of 2017 (Figure 16). However, no additional manuscripts were identified that met the search criteria.

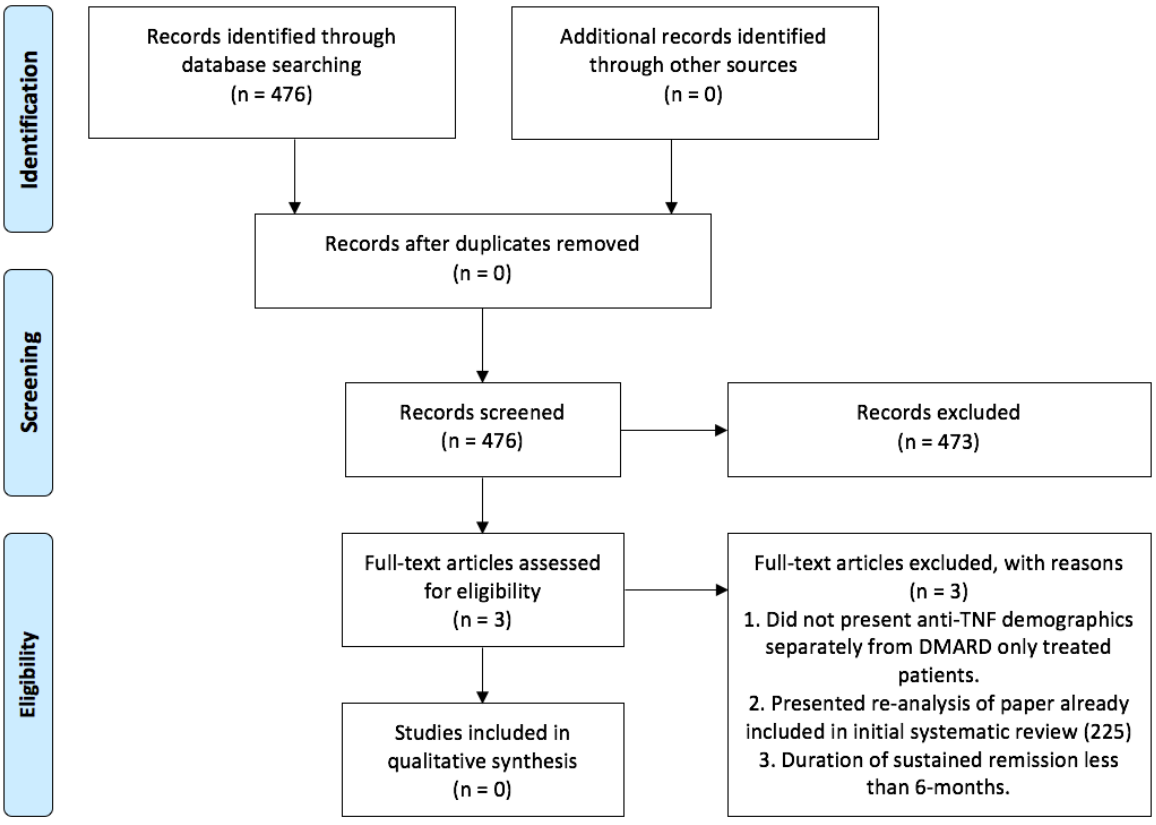


Figure 16. Updated systematic review search including additional manuscripts up to the end of 2017

A number of clinical factors including increased disease duration, higher baseline disease activity score, increased baseline tender joint count, and a greater baseline functional impairment are associated with a reduced likelihood of achieving sustained remission in individual studies. Demographic factors that appear to be negatively associated with sustained remission include female gender and increasing age. Only one clinical factor (methotrexate co-prescription) was associated with an increased likelihood of achieving sustained remission in more than one study. Supporting these findings, Katchamart *et al.* also identified these factors as predictors of point remission in a systematic review (212). Interestingly, the rates of sustained DAS28 remission identified in this review (7.9 - 38.1%) compare favourably with the range of point DAS28 remission rates (5 - 40%) identified by Katchamart *et. al.* However, the studies identified by Katchamart *et al.* included both biologic and synthetic DMARD treated RA patients enrolled in studies between 1999 and 2008. In comparison, this review focussed on anti-TNF treated patients only, and the oldest study in this review dates from 2009.

This review of the literature identified that female gender appears to be strongly associated with a reduced likelihood of achieving sustained remission in two of the included studies. However, female gender has been associated with a higher baseline ESR in a normal 'healthy' population compared with males, and prevalence is also known to increase with age (228). Given ESR is a component in the DAS28-ESR, it is possible that variations in gender and age may be confounders in the interpretation of the score which does not have different thresholds for these factors, rather than being true predictors of poor response. This may explain why Furst *et al*.(221) identified that female gender was associated with a lower likelihood of achieving sustained remission when using DAS28-ESR criteria, but not when using CDAI (which does not include an inflammatory marker component).

The finding that both increasing age and longer disease duration are both associated with a reduced likelihood of achieving sustained remission is unsurprising, and likely that one may be confounding the other. Further studies are required to ascertain the independence of these effects.

An important finding is the impact of baseline functional impairment on likelihood of achieving sustained remission. However, it remains uncertain whether worse functional impairment is a true predictor of response, or acting as a proxy marker of recalcitrant higher disease activity, irreversible joint damage, pain or fatigue, which may not be responsive to anti-TNF.

The only intervention that was associated with an increased likelihood of achieving sustained remission was methotrexate co-prescription. Whilst there may be alternative causal pathways that are responsible for this association (e.g. selection bias for those patients who are able to tolerate methotrexate), and observational data is not the optimal method for investigating the cause and effect relationships of medications, this association does appear to support the practice of co-prescription of methotrexate with anti-TNF where possible (229).

Of particular interest was the absence of any comorbidity data. Furst *et. al.* (221) and Einarsson *et al.* (225) both describe collection of comorbidity data, however no analysis was reported. None of the other studies included any reporting on comorbidity data. The association between RA and increased cardiovascular risk is well documented (94), as is the apparent risk reduction in RA patients successfully treated with anti-TNF (230). However, this review did not identify any evidence on cardiovascular outcomes in RA patients achieving sustained remission with anti-TNF.

Interaction between predictors and outcomes is challenging when using composite score outcome measures, particularly when variables included in the score are also identified as a predictor of that score. All the disease activity outcome measures used in RA are composite measures, and some of the predictors identified in this review, and the review by Katchamart *et al.* (212), are also components of these scores. An example of this is the association between higher baseline tender joint count and reduced likelihood of achieving sustained remission. It is unknown if having more tender joints prior to starting anti-TNF is a negative predictor of achieving sustained remission, or whether there is interaction with the composite outcome measure, within which tender joint count comprises a component. This review also identified that higher

baseline disease activity was negatively associated with the likelihood of achieving sustained remission, although it is possible that this association is on the causal pathway in the relationship between tender joint count and sustained remission. As previously discussed (Chapter 2), it is possible that composite disease activity scores may not be efficient in measuring reduction in inflammatory burden. Due to the multifaceted nature of a composite outcome measure, non-inflammatory components (such as the global health measure) may reduce the sensitivity in detecting the change in inflammatory activity achieved by anti-TNFs.

A surprising finding was that no objective clinical measure (such as the swollen joint count or inflammatory marker) was associated with sustained remission. It may be that improvements in objective measures of disease activity (such as the swollen joint count) improve more uniformly in response to anti-TNF in the majority of patients (including those in remission, low and moderate disease activity), whereas the more subjective components of the disease activity score (the tender joint count and global health measure) and patient directed outcome measures (such as the HAQ) are more variable in their response to anti-TNF. The World Health Organisation International Classification of Functioning, disability and health (WHO-ICF) (231) recognises the multi-faceted nature of an individual's perception of health, disability and functioning; and may provide some insight as to why there appears to be no association between sustained remission and objective measures of disease activity. In the WHO-ICF model, the actual health condition (such as RA) only accounts for one of six dimensions that contribute to an individual's perception of health and functioning. Participation in life situations, limitations on ability to undertake activities, impairment to body functions, environmental aids or barriers and personal factors all interact in an individual's perception of health. All the composite outcome measures included in this review contain measures that indirectly measure these non-disease dimensions (such as the global health measure). Personal factors also contribute to a patient's reporting of these subjective components. Therefore, the reduction in inflammatory burden could be offset by a lack of effect in subjective components of the score, which may not be directly related to disease activity. This poses wider questions for the use of composite outcome measures to quantify therapeutic efficacy of a targeted drug such as anti-TNF. Whilst blockade of the TNF pathway reduces joint damage, inflammation and swelling

of active RA, it may be that a patient's pain and fatigue is driven by other factors, unrelated to their RA. In these cases, classifying the patient as a 'non-responder' to anti-TNF is inappropriate; targeting external factors (such as low mood, altered pain perception or challenging life situations) may be more efficacious in achieving 'remission' than stopping or switching drug. As previously described, the non-specific nature of the global health measure allows for many aspects of health and quality of life to be encompassed in a composite score, and is both its strength and weakness (Chapter 1; 1.4.1.2 and 1.4.1.3). Specific quantification and reporting quality of life is challenging, and is a problem encountered in most chronic conditions. For example, fatigue is a significant problem in most chronic conditions, particularly those with an inflammatory component. A recent Cochrane review has shown an improvement in fatigue in response to anti-TNF and other biologic therapy, although modest (232). However, when using the more stringent outcome of fatigue remission, a cross-sectional evaluation of the BSRBR-RA demonstrated that only 37% of those individuals who achieved DAS28 remission at six months achieved a corresponding remission of their fatigue (77).

Composite outcome measures are essential in understanding the global impact of a complex multisystem disease such as RA. However, with increasingly aggressive treat-to-target treatment regimes, it may be more appropriate to consider which components of the composite measure are driving a patient's 'failure to respond' to anti-TNF rather than solely relying on the final composite score.

7.10 Conclusions

Despite the clinical relevance, sustained remission is a poorly reported outcome. Reporting the number of patients in a sustained state of remission or LDA would not require any additional data collection than currently occurs in most clinical studies, and would greatly assist in assessing the real-world clinical benefit of these treatments to patients. Furthermore, understanding of how individual components of composite outcome measures (such as the DAS28) vary in response to disease modifying treatment is needed in order to appropriately tailor treatment to the individual.

The associations identified in this chapter will be used as the foundations for the *a priori* variables specified in the analysis of sustained remission in the BSRBR-RA (Chapter 9). Additionally, consideration will be given to addressing the issue of multiple collinearity between the predictors of response, used in the final model.

It is evident from reviewing the existing evidence that the predictors of sustained remission vary considerably depending on the outcome measure used. As previously discussed (Chapter 2; 2.1.2) there is evidence that suggests the two versions of the DAS28 may not in fact be interchangeable, although they are used side-by-side in clinical practice (155-157). The BSRBR-RA does not specify which version of the DAS28 should be used, and as such, the composite DAS28 score recorded contains a mixture of DAS28-ESR and DAS28-CRP values. Before using these two outcomes interchangeably in the analysis of sustained remission, it is appropriate to use the data to examine the degree of discrepancy between the two versions of the DAS28 scores. This will be addressed in Chapter 8.

7.11 Key points from this chapter

- Sustained remission is an uncommonly reported outcome.
- Demographic and clinical features can help to predict sustained remission with anti-TNF.
- Female gender is associated with a reduced likelihood of achieving sustained remission.
- Methotrexate co-prescription with anti-TNF is associated with an increased likelihood of achieving sustained remission.
- Choice of outcome measure affects which associations are identified.

Chapter 8

8 Investigating agreement between the DAS28-ESR and the DAS28-CRP

8.1 Introduction

As outlined in Chapter 2, disease activity in RA is commonly measured using the DAS28. The original DAS28 was developed and validated using the erythrocyte sedimentation rate (ESR) (147,233). The development of the DAS28-CRP followed assessment of paired samples obtained from a relatively small cohort of 334 patients with subsequent wide adoption in clinical practice and trial settings (215). The DAS28-CRP and DAS28-ESR are typically used interchangeably, with identical disease activity stratification thresholds adopted in assessment of disease activity, treatment response and treat-to-target approaches.

However, a number of studies have highlighted consistently lower DAS28-CRP scores compared to DAS28-ESR scores, particularly at lower levels of disease activity that form the focus of treat-to-target management (155-157,234,235). This disparity in DAS28-CRP and DAS28-ESR values is important as it could influence patient management, both in terms of where high-cost drug reimbursement is only permitted if specific disease activity thresholds are reached (such as in the UK, and of importance for the BSRBR-RA cohort), but also when identifying episodes of sustained remission. Such disparity is of direct relevance when attempting to identify predictors of sustained remission, as differences in the methods utilised to measure the disease activity could disguise or falsely identify associations. The BSRBR-RA does not specify which version of the DAS28 is used. As such, both versions are recorded. Therefore, before attempting to identify the frequency and predictors of sustained remission and LDA in the BSRBR-RA cohort, it is essential to be confident that the scoring systems being used are reliable and accurate.

8.2 Aims

The aims of this analysis are:

1. To investigate the magnitude of discrepancy between the scores and its impact on disease activity stratification.

2. To investigate if common demographic factors influence the level of inter-score agreement.

3. To use paired DAS28-ESR and DAS28-CRP data on the BSRBR to investigate if a modified version of the DAS28-CRP could improve inter-score agreement.

8.3 Null hypothesis

The null hypothesis for this analysis is that the DAS28-ESR and DAS28-CRP are identical for all patients.

8.4 Methods

8.4.1 Subject selection and data collection

The methods of the BSRBR-RA have been described previously (Chapter 4). Patients treated with any biologic therapy with concurrent measures of ESR and CRP were identified, enabling paired calculation of DAS28-ESR and DAS28-CRP using existing formulae (236). Data obtained at baseline and following treatment with biologic agents, were used in the initial cohort analysis. Data from patients taking Ro-Actemra™ (tocilizumab) were excluded due to specific effects of IL-6 on serum CRP levels (237).

Patients taking only DMARDs were removed from the initial dataset to be used as an internal validation DMARD cohort.

8.4.2 Statistical analysis

The impact of age, BMI and gender on concordance between DAS28-ESR and DAS28-CRP was assessed by dichotomising the group for age (≥or< 50yrs) and gender and stratifying BMI according to World Health Organization thresholds (238). A random effects model was used to allow for the possibility that ESR and CRP were not measured from the same blood sample.

Agreement between the scores was compared using Bland-Altman statistics (198) (described in detail in Chapter 5; 5.2.2). Descriptive analysis was applied to compare disease stratification within accepted DAS28 disease activity thresholds.

8.5 Development of the modified DAS28-CRP

A nonlinear generalized additive model with monotonically constrained increasing regression spline (Figure 17, described in Chapter 5; 5.3.8) was used to model the relationship between paired ESR and CRP values, generating a predicted ESR from CRP (204).

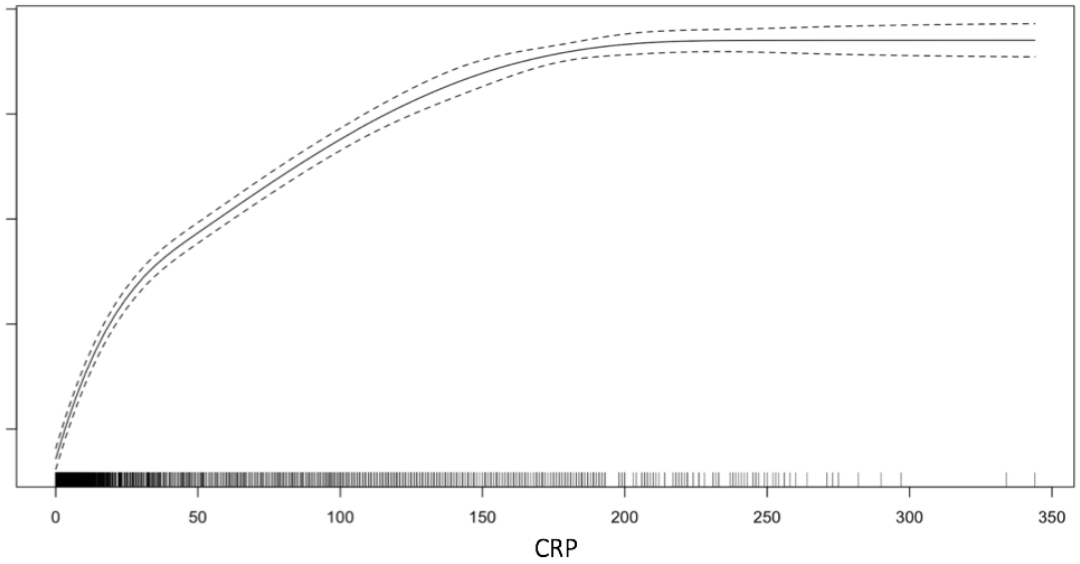


Figure 17. Monotonically constrained increasing regression spline used to model the predicted relationship between ESR and CRP as the CRP increases, allowing calculation of predicted ESR used in the calculation of the mDAS28-CRP (see Figure 19 for R code)

Resultant predicted ESR values were used in the existing DAS28-ESR formula to calculate the estimated disease activity, or modified DAS28-CRP (mDAS28-CRP). Kappa values and root mean squared error (RMSE; Chapter 4) calculated the mean error of DAS28-CRP and mDAS28-CRP (200). The differences in errors between the DAS28-ESR and DAS28-CRP or mDAS28-CRP were compared using the Wilcoxon signed rank test. Age at enrolment to the BSRBR-RA was used where it was not possible to calculate the age at time when the DAS28 score was measured. A cohort of biologic-naïve patients treated with DMARDs with paired DAS28-ESR and DAS28-CRP readings obtained from the BSRBR-RA was used to undertake internal validation of the mDAS28-CRP model.

8.6 Results

8.6.1 Subject characteristics

Paired ESR and CRP values were available for 8,509 subjects, with 31,074 paired assessments in the biologic cohort. The majority of subjects were female (76%), with

mean age of 57.3 years (standard deviation (SD) 12.2) and a mean baseline disease duration of 12.7 years (SD 9.6).

Baseline data table	
Number of paired readings	31074
Number of patients	8509
Mean Baseline Age (yrs, SD)	57.3 (12.2)
Gender (% Female)	76%
Mean Baseline Disease Duration (yrs, SD)	12.7 (9.6)
BMI (SD)	26.9 (6.0)
Mean DAS28-ESR (SD)	4.44 (1.73)
Mean DAS28-CRP (SD)	4.13 (1.60)

Table 13. Baseline demographics of cohort

8.6.2 Missing data

There were no missing data at baseline for age, gender, and height. There were missing data on the date of DAS28 measurement for 18% (5602) of total readings (Table 14). Comparison of missing and complete datasets respective to age, gender, disease duration and BMI did not reveal any significant differences between the two groups, suggesting MAR was a reasonable assumption, and LOCF imputation using age at enrolment to the BSRBR-RA was used.

	Complete data set	Missing data set	P value
Number of paired readings	25472	5602	NA
Mean Baseline Age (yrs, SD)	55.5 (12.1)	55.7 (12.1)	0.5
Gender (% Female)	76.4	75.6	0.3
Mean Baseline Disease Duration (yrs, SD)	12.8 (9.5)	12.8 (9.4)	0.7
BMI	26.9 (6.0)	26.8 (6.0)	0.4

Table 14. Demographics of complete and missing datasets

8.6.3 Discordance between DAS28-ESR and DAS28-CRP

Comparing differences between the two scores revealed that the DAS28-CRP was on average 0.3 points lower than the corresponding DAS28-ESR for the whole cohort (Table 15). When stratifying by age and gender, differences between the two scores were more pronounced for women and patients aged over 50 although the mean inter-score differences did not alter when categorised by baseline BMI (Table 15).

Impact of patient demographics on DAS28-ESR & DAS28-CRP concordance						
		Mean DAS28-ESR	**Mean DAS28- CRP**	**Mean Difference**	**95% Confidence Interval**	**n**
Overall		4.44	4.13	0.30	0.30 to 0.31	31074
Male		4.17	4.02	0.15	0.13 to 0.16	7380
Female		4.52	4.17	0.35	0.35 to 0.36	23694
Age	**<50**	4.27	4.09	0.17	0.16 to 0.19	7786
	>50	4.50	4.15	0.35	0.34 to 0.35	23288
Underweight (<18.5)		4.51	4.19	0.32	0.29 to 0.35	1054
Normal (18.5 - < 25)		4.34	4.04	0.30	0.29 to 0.31	12348
Overweight (25 - < 30)		4.43	4.15	0.29	0.28 to 0.30	9988
Obese (≥30)		4.59	4.27	0.33	0.32 to 0.34	7684

Table 15. Comparative mean difference between DAS28-CRP and DAS28-ESR (by Bland-Altman statistics), and effect of gender, age and BMI. n = number of paired scores

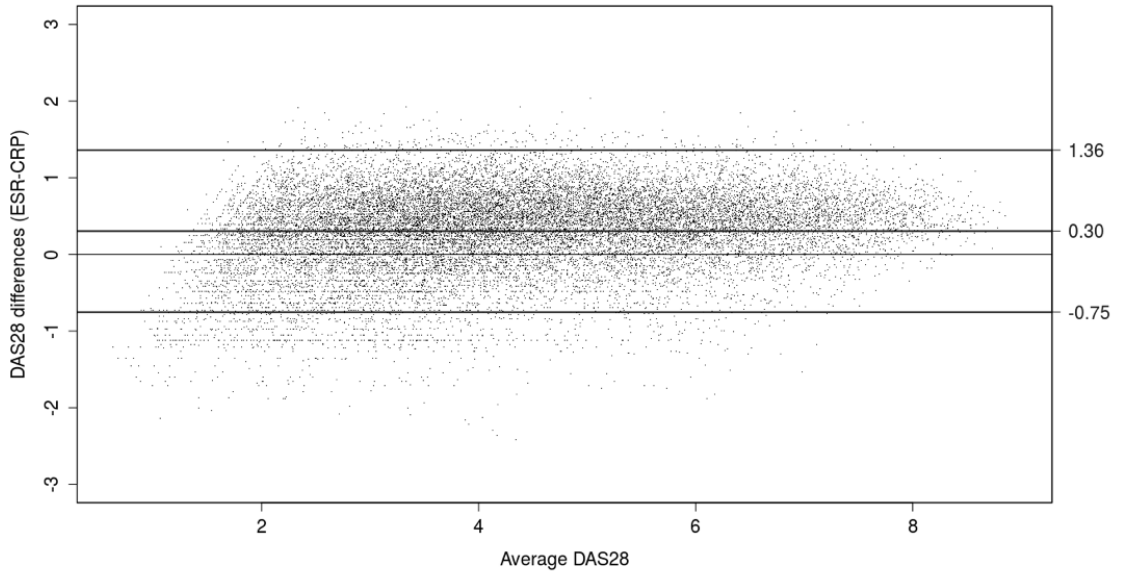


Figure 18. Bland-Altman plot of data using DAS28-CRP and DAS28-ESR showing mean difference (0.30) and spread of 95% of scores (-0.75 to 1.36)

8.6.4 Impact of disparity between DAS28-ESR and CRP on disease activity stratification

Disparity between the DAS28-ESR and DAS28-CRP had a significant impact on disease stratification, particularly within the low disease activity (LDA) category where the two scores only agreed in 32.0% of cases. The DAS28-ESR classified fewer patients in remission compared with the DAS28-CRP, and more in high disease activity (Table 16).

		DAS28-ESR			
		Remission (<2.6) 15.6% (n=4856)	**LDA (2.6 - ≤3.2) 11.1% (n=3438)**	**MDA (3.2 - ≤5.1) 38.0% (n=11807)**	**HDA >5.1 35.3% (n=10973)**
DAS28-CRP	**Remission (<2.6) 19.5% (n=6062)**	66.6% (4040)	25.4% (1541)	7.9% (481)	0
	LDA (2.6 - ≤3.2) 13.1% (n=4074)	15.9% (646)	32.0% (1302)	52.2% (2126)	0
	MDA (3.2 - ≤5.1) 38.8% (n=12053)	1.4% (170)	4.9% (590)	73.7% (8880)	20.0% (2413)
	HDA >5.1 28.6% (n=8885)	0	0.1% (5)	3.6% (320)	96.3% (8560)
* Total percentages not equal to 100% due to rounding					

Table 16. DAS28-CRP/DAS28-ESR agreement matrix showing distribution of inter-score misclassification. Numbers in brackets are number of scores falling into each disease activity category. Percentages are calculated by dividing the number of scores that agreed with the DAS28-ESR score classification by the total number of mDAS28-CRP scores that were in that disease activity category (numbers in brackets in left column) to give the proportion of mDAS28-CRP scores that agreed with the DAS28-ESR score

8.6.5 Development of mDAS28-CRP and impact on agreement with DAS28-ESR and disease activity stratification

When applying the mDAS28-CRP (Figure 19) the difference between mDAS28-CRP and DAS28-ESR scores was significantly reduced compared to scores generated by the DAS28-CRP ($p < 0.001$), particularly for women and patients over 50 years of age (Table 18).

```
require(scam)

esr_crp_relationship = scam(esr~s(crp, bs="micv"), data=df)

predicted_esr = predict(esr_crp_relationship)

mdas28_crp =
0.56*sqrt(tender_joint_count)+(0.28*sqrt(swollen_joint_count))+(0.70*log(predicted_esr))+(0.0
14*global_VAS)

Where:

df = BSRBR-RA dataset
crp = CRP value
esr = ESR value
predicted_esr = predicted ESR value generated using CRP from esr_crp_relationship
tender_joint_count = tender joint count
swollen_joint_count  = swollen joint count
global_VAS = Visual analogue score
```

Figure 19. R-code used to calculate mDAS28-CRP

Inter-score agreement with the DAS28-ESR improved across remission, low and moderate disease activity categories, at the expense of a small reduction in agreement at HDA (Table 18). Kappa statistics showed an improvement in agreement from 0.62 to 0.65 and RMSE reduced from 0.61 to 0.56. Although there was variation in mean differences between subgroups for the mDAS28-CRP, none of these subgroups had a greater mean difference than was observed between the DAS28-ESR and DAS28-CRP overall. Adjustment of analysis using a random effects model allowing for the possibility that CRP or ESR were measured at different times did not alter results, and was therefore not used in the final mDAS28-CRP model.

Impact of patient demographics on DAS28-ESR & mDAS28-CRP concordance						
		Mean DAS28-ESR	**Mean mDAS28-CRP**	**Mean Difference**	**95% Confidence Interval**	**n**
Overall		4.44	4.61	-0.17	-0.18 to -0.17	31074
Male		4.17	4.49	-0.34	-0.34 to -0.31	7380
Female		4.52	4.65	-0.12	-0.13 to -0.12	23694
Age	**<50**	4.27	4.57	-0.30	-0.32 to -0.29	7786
	>50	4.50	4.62	-0.13	-0.13 to -0.12	23288
Underweight (<18.5)		4.51	4.66	-0.16	-0.18 to -0.12	1054
Normal (18.5 - < 25)		4.34	4.51	-0.18	-0.19 to -0.17	12348
Overweight (25 - < 30)		4.43	4.62	-0.19	-0.20 to -0.18	9988
Obese (≥30)		4.59	4.74	-0.15	-0.16 to -0.14	7684

Table 17. Comparative mean difference between mDAS28-CRP and DAS28-ESR (by Bland-Altman statistics), and effect of gender, age and BMI. n = number of paired scores

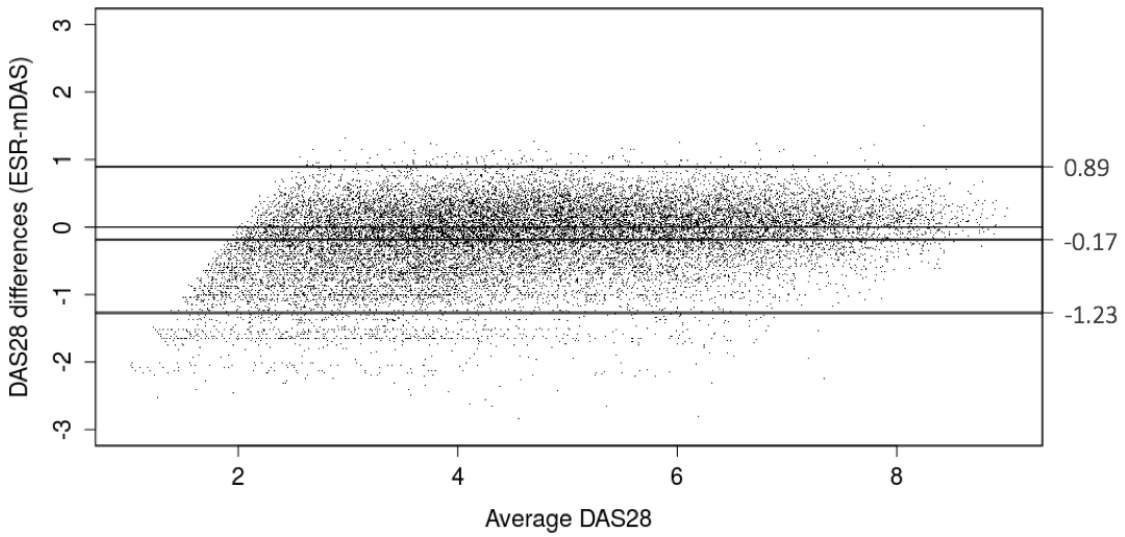


Figure 20. Bland-Altman plot of data using mDAS28-CRP and DAS28-ESR showing mean difference (-0.17) and spread of 95% of scores (-1.23 to 0.80)

mDAS28-CRP/DAS28-ESR agreement matrix*					
		DAS28-ESR			
		Remission (<2.6) 15.6% (n=4856)	**LDA (2.6 - ≤3.2) 11.1% (n=3438)**	**MDA (3.2 - ≤5.1) 38.0% (n=11807)**	**HDA (>5.1) 35.3% (n=10973)**
mDAS28-CRP	**Remission (<2.6) 10.3% (n=3208)**	84.5% (2711)	13.9% (446)	1.6% (51)	0
	LDA (2.6 - ≤3.2) 11.3% (n=3519)	42.3% (1490)	39.2% (1380)	18.4% (649)	0
	MDA (3.2 - ≤5.1) 41.9% (n=13031)	5.0% (655)	12.3% (1603)	76.2% (9932)	6.5% (841)
	HDA >5.1 36.4% (n=11316)	0	0.1% (9)	10.4% (1175)	89.5% (10132)
* Total percentages not equal to 100% due to rounding					

Table 18. mDAS28-CRP/DAS28-ESR agreement matrix showing distribution of inter-score misclassification. Numbers in brackets are number of scores falling into each disease activity category. Percentages are calculated by dividing the number of scores that agreed with the DAS28-ESR score classification by the total number of mDAS28-CRP scores that were in that disease activity category (numbers in brackets in left column) to give the proportion of mDAS28-CRP scores that agreed with the DAS28-ESR score

The DMARD control cohort from the BSRBR-RA was used as an internal validation cohort. Patients in the DMARD control cohort were slightly older than the biologics cohort (58.2 vs 57.3 years respectively), and there was a lower proportion of women compared with the biologics cohort (73% vs 76%). The overall number of paired readings was also considerably smaller than the biologics cohort (748 vs 31,074), which limited the opportunities for subgroup analysis by age, gender and BMI.

For the DMARD control cohort, the mean differences between the DAS28-CRP and the DAS28-ESR remained the same as identified in the biologic cohort and the mDAS28-CRP reduced this mean difference. Kappa agreement was also improved using the mDAS28-CRP, and RMSE was reduced. Overall, the reduction in error was statistically significant (Table 19).

DMARD internal control cohort		
	DAS28-CRP	**mDAS28-CRP**
Bland-Altman Mean Difference vs DAS28-ESR (points; 95% Confidence Interval)	0.31 (0.27 to 0.34)	-0.13 (-0.16 to -0.10)
Kappa	0.69	0.73
RMSE (points)	0.55	0.47
Wilcoxon Rank Sum error difference vs. DAS28-ESR (DAS28-CRP - mDAS28-CRP)	0.10***	

Table 19. Comparative mean difference, Kappa, RMSE and error difference between DAS28-CRP and mDAS28-CRP when compared to DAS28-ESR. *** p<0.001

When applying disease activity thresholds, the mDAS28-CRP performed in a similar manner to that identified in the biologic cohort, improving class agreement across remission, LDA and MDA, with a modest reduction in agreement at HDA (Table 20).

DMARD internal control cohort					
	% agreement (DAS28-CRP with DAS28-ESR)	**% agreement (mDAS28-CRP with DAS28-ESR)**	**DAS28-ESR**	**DAS28-CRP**	**mDAS28-CRP**
			n= number of paired DAS28 scores		
Remission (<2.6)	57.7	80.0	21	26	15
LDA (2.6 - ≤3.2)	51.7	52.9	31	29	17
MDA (3.2 - ≤5.1)	69.3	81.4	229	306	220
HDA (>5.1)	98.2	90.1	467	387	496
Total			748		

Table 20. Inter-score agreement between DAS28-ESR and DAS28-CRP when stratified by DAS28 thresholds for the DMARD internal validation cohort. LDA = Low disease activity, MDA = Moderate disease activity, HDA = High disease activity. n = number of DAS28 scores per category when disease activity thresholds applied

8.7 Chapter discussion

This analysis supports existing evidence suggesting DAS28-ESR and DAS28-CRP should not be viewed as interchangeable outcome measures, highlighting the limitations of doing so, with important consequences for clinical and research practice, and of particular relevance to subsequent analyses for this thesis (155-157,234,235).

The initial development of the DAS28-CRP by Fransen *et al*., which included data from 334 patients, used linear regression and high Pearson correlation coefficient to suggest equivalence with the DAS28-ESR. However, agreement analysis was not undertaken (215). Consequently, whilst correlation between the scores was high, equivalence and interchangeability cannot have been said to have been demonstrated and concern about the limited validation of the DAS28-CRP prior to the release of the score was raised by Aletaha *et al.* in 2005 (149). Disparity between DAS28-ESR and DAS28-CRP has led some to propose applying lower disease activity thresholds for the DAS28-CRP (155,157). However, this approach could prove problematic given many guidelines are based solely on DAS28 levels without specification of the inflammatory response marker used (162,233,239).

It is evident that variation in equivalence between the two scores is most pronounced for older patients and women (demographics representing the majority of the RA population) and at lower disease activity levels (the target for most treatment strategies (139)). This discordance precludes easy comparison of outcomes of studies that have adopted different versions of the DAS28.

To address this, I used real-world data from over 8500 patients from the BSRBR-RA to develop a modification of the DAS28-CRP (the mDAS28-CRP) which has demonstrated superior agreement with the DAS28-ESR compared to the original DAS28-CRP. The improvement in agreement of disease activity stratification at lower disease activity thresholds was achieved at the expense of a minor reduction in agreement at HDA. However, the effect of this in clinical practice would be to encourage more active treatment for patients with higher disease activity, in line with current treatment

paradigms. Younger and male patients have the lowest inter-score agreement using the mDAS28-CRP; however, neither group has worse agreement than that seen between the DAS28-CRP and DAS28-ESR overall.

A potential limitation of the mDAS28-CRP is that it was developed using a single cohort. The cohort is also mainly of Caucasian ethnicity, which may influence ESR and CRP relationships differently compared with other ethnicities (156,235). The main cohort used for the development of the mDAS28-CRP included patients on biological agents recruited to a registry which may introduce selection bias, although it is unlikely this would impact on ESR/CRP comparisons. Furthermore, patients included in the BSRBR-RA are enrolled from across the UK, representing a broad population and spectrum of RA management.

In support of the mDAS28-CRP, the internal validation DMARD cohort also demonstrated similar results. It is possible that unknown confounders may influence whether an individual has both an ESR and CRP test undertaken rather than only one, although this seems unlikely. There were some missing data, although there were no significant demographic differences between missing and complete groups (Table 14).

A further limitation is the relatively small internal validation cohort, cross-validation of the mDAS28-CRP in larger external cohorts is necessary before the mDAS28-CRP could be widely adopted. Other research groups from Canada (CATCH Registry) and the Netherlands (Dutch RA Quality Register) have expressed interest in participating in future validation studies which will enable testing of the mDAS28-CRP in a more diverse population of patients.

However, the key finding from the analysis undertaken in this chapter and particularly relevant to this thesis is that the DAS28-ESR and DAS28-CRP should not be used interchangeably, particularly when stratifying disease activity. As such, further analysis of sustained remission will focus on using only the DAS28-ESR, the originally validated version of the DAS28, and the form in which the majority of the data in the BSRBR-RA exists.

The mDAS28-CRP will not be used in place of the DAS28-CRP for subsequent analyses undertaken in this thesis for two reasons. As mentioned previously, the score does need further validation in cohorts outside of the BSRBR-RA to ensure the improved agreement between the DAS28-ESR and mDAS28-CRP is maintained. Secondarily, use of the mDAS28-CRP would not 'solve' the issue of missing data in the BSRBR-RA to be used in subsequent analyses. As such multiple imputation will still be required. Because the mDAS28-CRP has a slightly different relationship between its component parts and final score compared to the DAS28-ESR, this would likely impact on the bootstrapped imputation parameters used in the Amelia imputation algorithm (Chapter 5; 5.1.4.5.1). As such, use of both the mDAS28-CRP and DAS28-ESR may make the imputed values less reliable than using the DAS28-ESR alone.

8.8 Key points from this chapter

- The DAS28-ESR and DAS28-CRP do not generate interchangeable scores.

- Inter-score agreement is poorer for female and older patients.

- Inter-score discrepancies have a significant impact on disease activity stratification.

- The mDAS28-CRP has improved inter-score agreement with the DAS28-ESR compared with the DAS28-CRP.

Chapter 9

9 Predictors of sustained remission for RA patients on anti-TNF

As outlined previously, in addition to measuring the frequency of remission at a single pre-determined point in time, understanding how often sustained remission occurs is of importance to both clinicians and patients (Chapter 2; 2.6).

Evidence from the systematic literature review undertaken in Chapter 7 identified several factors associated with a reduced likelihood of achieving sustained remission including greater baseline disease activity, higher tender joint count, higher age, higher baseline disease activity, and female gender, with only methotrexate co-prescription being identified as associated with an increased likelihood of achieving sustained remission. It also identified the paucity of reporting of such a clinical outcome, with only six papers from nearly 4500 identified by the search criteria, highlighting the need for further research in this area.

This analysis will focus on identifying the frequency with which sustained remission occurs in the real-world using data obtained from the BSRBR-RA. Associations with *a priori* specified clinical and demographic features will be investigated. In addition to sustained remission, this chapter will examine the frequency of sustained low disease activity (LDA). Whilst sustained remission remains the aim of the treat-to-target paradigm, LDA is often used as a pragmatic target in clinical practice for patients that may be unlikely to achieve remission (240) (discussed in Chapter 2; 2.6.2). Therefore, this analysis will also examine the frequency of attainment of LDA, and the clinical and demographic features associated with this outcome.

With nearly 15,000 anti-TNF patients enrolled, the BSRBR-RA represents an ideal resource to investigate this outcome further.

9.1 Aim

This analysis aims to investigate the frequency of sustained remission and sustained LDA and the clinical and demographic factors associated with these outcomes using the BSRBR-RA.

9.2 Objectives of this chapter

- To examine how the demographics of the anti-TNF cohort of the BSRBR-RA have changed over time.
- To examine the frequency of sustained remission and sustained LDA in the BSRBR-RA defined using DAS28-ESR thresholds (Chapter 2; 2.1.2).
- To investigate if there are any clinical or demographic features that are associated with achieving sustained remission or sustained LDA.
- To examine how the frequency of sustained remission and LDA, and predictors of these outcomes may, have changed over time.

9.3 Null hypothesis

The null hypothesis of this analysis is that there are no associations between baseline clinical and demographic factors and the attainment of sustained remission or sustained LDA according to the DAS28-ESR.

9.4 Definition of sustained remission and sustained LDA

For the purposes of this analysis, sustained remission is defined as a DAS28-ESR score of ≤2.6 for at least 6 months while on anti-TNF treatment.

Sustained LDA is defined as a DAS28-ESR of ≤3.2 for at least 6 months while on anti-TNF treatment.

9.5 Analysis plan

9.5.1 Defining the study population for analysis

The BSRBR-RA contains data on a heterogeneous population of patients with RA taking a range of drugs for RA, including biological drugs and synthetic DMARDs. Due to the longitudinal nature of the registry, there are patients who have switched drug (due to adverse events or non-response). This may include switching within class (i.e. switching from one anti-TNF to another anti-TNF) or switching between classes of drug (i.e. switching from an anti-TNF to another class of action e.g. anti-CD20). Such switching creates difficulties in accurately identifying true associations after switching, due to overlap in effects of the different sequential exposure to drugs. Biological agents such as anti-TNF, act by modifying the cellular signalling pathways which in turn modify the makeup of the immunological response (21,29). This means that, after discontinuing a biological agent, the impact of the drug may continue beyond that time at which it is no longer detectable in the body. As such, the pharmacokinetic half-life of the drug is different to the biological half-life of the drug, and knowing when the actual effect of the drug ceases is difficult and likely to vary between individuals and drugs. To accurately identify associations between anti-TNF and sustained remission/LDA, this analysis will focus on a more homogeneous population, defined as adult patients with RA taking their first anti-TNF medication only. For those patients that switch anti-TNF (either to another anti-TNF or to a different class of biologic agent) or stop taking their first anti-TNF, data will be censored at the point of switching or stopping the drug.

9.5.2 Identifying biologic switching

The dataset used for this analysis includes only individuals starting on an anti-TNF as their first biologic agent. The process of requesting and receiving the dataset is described in Chapter 4 (4.8). The BSRBR-RA collects data on drug switching, so this will be used to identify any records where an individual has switched to another biologic

agent. In addition, binary vector multiplication (Chapter 6; 6.3.1) will be used to identify any change in biologic that is not specifically identified in the 'biologic switching' column. This could occur if the data on biologic switching is not completed by the clinician completing the BSRBR-RA forms and is not identified by the BSRBR-RA registry team at the time of data entry.

9.5.3 Identifying sustained LDA and remission

As outlined previously, for the purposes of this analysis, sustained LDA and remission will be defined as any patient achieving the required DAS28-ESR thresholds for at least six months. The BSRBR-RA collects DAS28-ESR outcomes on a six-monthly basis for the first three years that a patient is enrolled on the database. Thereafter, data collection reduces to an annual frequency. Therefore, a pragmatic approach to the operationalisation of the definition of sustained remission/LDA was applied to the BSRBR-RA dataset. To qualify as sustained remission or sustained LDA, a patient will need to have a DAS28-ESR score of less than 2.6 or 3.2 respectively, on two sequential follow-up visits. By this definition, sustained LDA includes sustained remission patients as well. The analysis will only include the first three years of data collection for each individual when study follow-up forms were completed every six months with all patients censored at three years (whereupon data collection frequency reduces to annually).

Although follow-up visits are supposed to occur at 6-monthly intervals in the first three years of the BSRBR-RA, there is variation in when the assessments actually occur. To allow coding of periods of sustained remission, follow-up number (i.e. first follow-up, second follow-up etc.), rather than the precise date that follow-up occurs on, will be used to enable identification of sequential follow-up visits and periods of sustained remission/LDA. Whilst there will be variation in when follow-ups occur between individuals, the alternative would be to artificially apply date/time thresholds within which follow-ups would need to occur to be counted as sequential. This leads to difficulties in defining what these time-windows should be. For example, if a baseline visit occurred on January 10th 2011, then a six-month follow-up visit should occur on

or around the 11th July 2011. However, in an epidemiological study such as the BSRBR-RA, follow-ups are unlikely to occur on an exact date, as data are collected from routine clinic appointments. As such, the six-month follow-up (the first follow-up time point), could occur before or after the 11th July 2011. If a three-month time-window either side of the defined follow-up date were used (i.e. half-way between sequential follow-up dates), follow-ups that occurred one-day outside of this artificial boundary would potentially be classified as being a different follow-up visit. However, if the next follow-up data were also available within the same follow-up time window (e.g. follow-up visit two at 12 months +/- three months), there would be two data points where only one could be counted, and an artificial missing data field generated for the previous follow-up visit. This would lead to difficulties in deciding which of the real data points to use for the second follow-up time point, and how to analyse artificially generated missing data fields generated by the analysis methods, which may not be at random.

Whilst accepting there will be variability in the time between follow-up visits is a limitation, the alternative method to correct this variability is likely to misrepresent the data by artificially generating missing data which may create a non-random missing pattern.

9.5.4 Identifying individuals in sustained remission/LDA

The method for identifying individuals in sustained remission or LDA uses binary vector multiplication and is described in Chapter 6 (6.3.1).

9.5.5 Defining clinical and demographic predictors to be included in the analysis

The clinical and demographic features used in this analysis are those identified in the systematic review of the topic (Chapter 7) and additional *a priori* specified variables, outlined in this chapter, that were identified during the course of reviewing the literature.

As discussed in Chapter 7 (7.9), one of the difficulties when examining associations with sustained remission and LDA as defined by a composite outcome such as the DAS28-ESR is that some of the key variables that would be of interest to investigate (such as the tender and swollen joint count) are also variables in the outcome of interest (the DAS28-ESR). Additionally, variables such as number of swollen and/or tender joints are likely to be strongly interrelated at an individual level. This, coupled with the multiple clinical and demographic predictors to be analysed, mean that collinearity between variables is likely to be a significant issue in interpreting results. The issues with collinearity are discussed in Chapter 5 (5.3.2). However, for this analysis, impact on the analysis will be assessed using variance inflation factor (VIF) analysis (Chapter 5; 5.3.3). A stepwise regression with Bayesian model checking will be used to establish the most stable model with lowest VIF (Chapter 5; 5.3.4 and 5.4.2).

The combination of *a priori* variable specification with subsequent stepwise regression and collinearity assessments uses a two–step Bayesian methodology (Chapter 5; 5.4). The first step uses existing evidence from multiple sources (i.e. different papers) to establish prior hypotheses (i.e. which variables are likely to be of importance in the model). The second step uses stepwise regression and collinearity modelling to assess if the number of variables in the model can be reduced. Confidence in the selected model will be increased if the direction of association in the subsequent reduced multivariable model are the same, and of similar magnitude to the non-reduced model.

The measure of collinearity establishes the likelihood that the associations identified may be attributed to other variables. If the VIF is reduced following stepwise regression, this adds confidence that the stepwise regression has occurred appropriately.

Univariable regression is often used as a method of variable selection. A variable is chosen if the associated p-value obtained from performing a univariable regression is smaller than a pre-determined significance threshold (commonly 0.05). After fitting a series of univariable regression models, the chosen variables are fitted together in a multivariable model. Whilst widely used, assessment of significance in a multivariable

model that constitutes the second stage of a two-stage should take into account the previous screening of variables based on significance in univariable analyses. There is debate as to if the subsequent p-values and confidence intervals truly reflect the relationships that are found, as the mechanism by which they are included in the model means that they must have achieved a certain level of significance. This thesis does not aim to state which method is 'best' and for clarity and for completeness, a univariable analysis has been undertaken, although subsequent analyses are based on the Bayesian two-step methodology outlined above.

A priori specified variables (identified in Chapter 7) to be included in the initial analysis include:

Gender (male/female)	Age (yrs)
BMI (kg/m^2)	Patient Global Assessment (VAS)
Tender Joint Count (0-28)	ESR (0 – 150mm)
Baseline HAQ-DI (0 – 3)	DAS28-ESR (0 – 10)
Disease Duration (yrs)	Year starting on biologic

Table 21. *A priori* specified variables identified in Chapter 7 included in initial analysis

Additional *a priori* variables identified from literature not included in the systematic review (Chapter 7) to be evaluated:

Swollen Joint Count (0 – 28)
Swollen:Tender Joint Count Ratio (Categorical variable: Low <0.5, Moderate ≥0.5 - ≤1.0, High >1.0)
Anti-TNF drug type (Categorical variable: 1 = Enbrel™, 2 = Remicade™, 3 = Cimzia™, 4 = Humira™)
Baseline smoking status (Current/Ex/Never)

Table 22. Additional *a priori* specified variables identified from the literature

9.5.6 Examining missingness

Evaluation of missing data patterns in the BSRBR-RA dataset are outlined in Chapter 6, Chapter 4 (4.2.1) outlines the BSRBR-RA methods used to recruit patients and collect data.

9.5.7 Multiple imputation

Multiple imputation (using the Amelia package in R) is used to maximise the data resource in the registry and minimise data loss due to missing data (Chapter 5; 5.1.4.5).

9.5.8 Examining associations

Imputation using Amelia generates five imputed datasets for each originator dataset. All statistical analysis is undertaken on separate datasets, before estimates are combined to generate a final single estimate using Rubin's Rules (Chapter 5; 5.1.5.1) using the Zelig package in R. Generating five datasets with slightly different estimates for the missing data points generates variability in the imputed datasets which allows accurate estimation of uncertainty generated by multiple imputation in the further analysis. Associations with *a priori* specified variables will be examined on each of the imputed datasets using stepwise regression and Bayesian model checking (Chapter 5; 5.3.4 and 5.4.2). Variables identified from the stepwise regression will then be used in a reduced logistic regression model using Zelig. Model fit will be compared between the original full *a priori* specified variables, and the truncated variable set following stepwise regression. The collinearity of the two different models will be examined by comparing the different VIF values for each variable in the regression model as well as analysis of variance (ANOVA) to examine if there is a significant difference in variance between the full and reduced models. The most parsimonious model with the best fit will be selected.

9.5.9 Confounding

A significant issue in observational studies is confounding. Confounding occurs when two or more factors may be combined in their influence on a single outcome (Figure 21). This can lead to false identification of causality between an exposure and an outcome. If confounding is not identified, the outcome may in fact be caused by another

unmeasured factor that is related to both the outcome and the exposure, but not included in the model. It is important to note that in an epidemiological study such as the BSRBR-RA, it is not possible to establish causality as this requires a study with a randomised controlled trial design. However, it is still essential to consider confounding in the design and interpretation of results from an epidemiological study as confounding may still influence the associations identified.

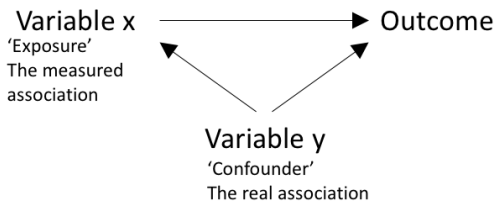


Figure 21. Directed Acyclic Graph (DAG) of effect of confounding on measured association

Evidence identified in the systematic review and discussed in Chapter 7 (7.9) highlighted some of the issues of confounding when using a composite outcome such at the DAS28 and predictors of outcome. In particular, a negative association between baseline ESR and rates of sustained remission may be confounded by the relationship between female gender and ESR (Figure 22). In this case, female gender has been associated with higher ESR values than in males (228). Because ESR is a component of the DAS28-ESR, increased average values for ESR may cause an artificial association to be identified between baseline ESR and poorer outcomes as measured by the DAS28-ESR, rather than female gender. The fact that Furst *et al*. (221) identified a negative relationship between female gender and sustained remission using the DAS28, but not using the CDAI (which does not include an inflammatory marker as a component), lends support to this hypothesis.

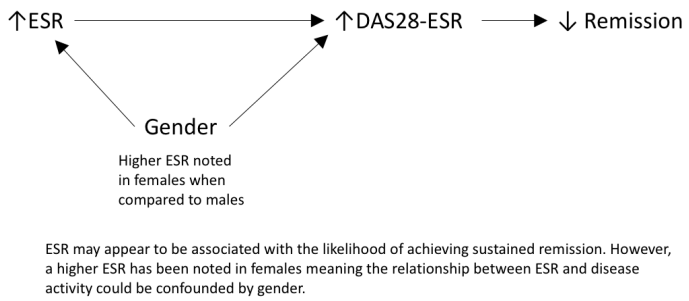


Figure 22. Confounding effect of gender when using the DAS28-ESR

Confounding could also occur between disease duration and baseline functional status measured by HAQ. Longer disease duration has been identified as being associated with a higher HAQ, with an average increase between 0.02 and 0.03 units per year (241). The HAQ is a composite score that measures both disease activity and damage. Because joint damage in RA is generally irreversible, and accumulates over time, this component of the HAQ makes up a greater proportion of the score over time, making the larger HAQ score (which is also associated with disease duration) less responsive to treatment (241). Therefore, any association between HAQ or disease duration and sustained remission alone could involve contributory factors from both HAQ or disease duration. In addition, an increased disease duration may reduce the likelihood of achieving sustained remission through other mechanisms unrelated to HAQ. Similarly, increased DAS28 may act as a confounder in the relationship between baseline HAQ and sustained remission.

One of the most complex relationships to establish is the choice of anti-TNF agent. Because the choice of anti-TNF agent is likely to be influenced by a multitude of factors, it is difficult to identify a true association between choice of anti-TNF and remission as an outcome (known as confounding by indication; Figure 23). However, it is still worthwhile including drug choice in the regression models as it is useful to understand if there is the possibility of drug choice influencing outcome, justifying further investigation. In this analysis it is examined as a categorical variable (Table 22).

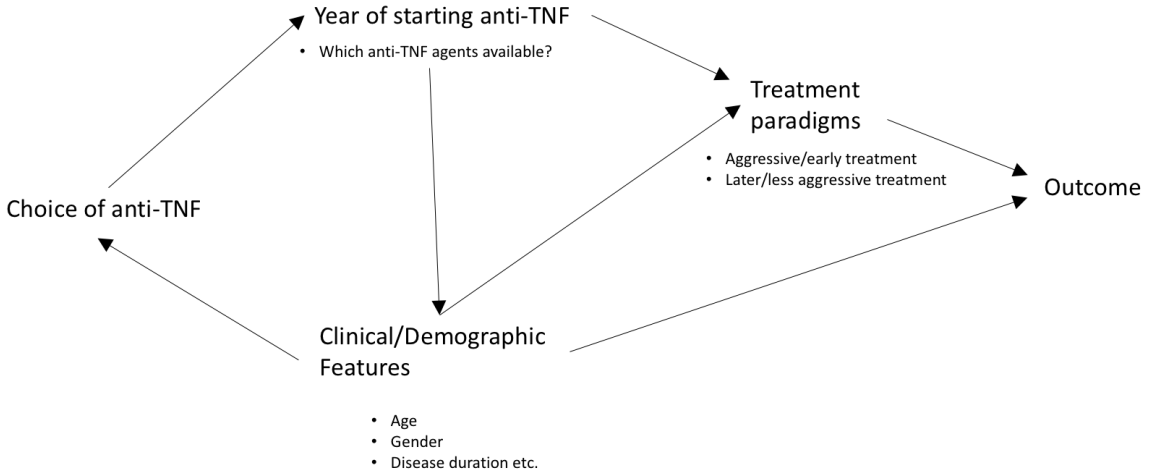


Figure 23. Effect of confounding by indication with anti-TNF choice

To mitigate against the confounding identified above, multiple variables are included in the full logistic regression models used. However, one of the drawbacks of incorporating multiple variables into any statistical model is the issue of overfitting and multiple collinearity as previously discussed (Chapter 9; 9.5.8). Therefore, findings from both the full and reduced regression models will be examined and compared.

9.5.10 Dataset preparation

Prior to undertaking analysis, the raw dataset requires preparation to ensure that it includes only the desired cohort of patients.

Chapter 4 outlines the process of requesting and extracting the raw dataset from the BSRBR-RA. The dataset preparation including multiple imputation, and inclusion criteria are described in Chapter 6.

As mentioned previously, the BSRBR-RA does not specify the version of DAS28 that needs to be used. However, this analysis includes only DAS28-ESR data. The rationale for this has previously discussed (Chapter 8; 8.7).

9.5.11 Calendar year effect

Since the first introduction of anti-TNF into routine clinical practice in 2000, clinical practice and the demographics of those individuals starting on anti-TNFs has changed (242). Therefore, after undertaking analysis on the complete cohort (from 2001 – 2013), it is split into two subgroups to identify how demographics of patients starting anti-TNF, remission and LDA rates have changed over time.

Recruitment for anti-TNF medications was paused from 2007 until 2010 (Table 4). In addition, updated EULAR guidelines for the treatment of RA were published in 2010 (243). The publication of such updated practice, and the natural pause in registers recruitment windows, makes the year 2010 an appropriate time to split the cohort, and allows further investigation into whether demographics or outcomes of patients on anti-TNF medications has changed over time. Although recruitment paused from 2007 – 2010, the earlier subgroup will be identified in this thesis as the '2001 – 2010 subgroup' to be clear that, although recruitment of new patients temporarily ceased, follow-up data were not excluded between 2007 and 2010.

9.6 Results

9.6.1 Comparing the cohorts baseline data

Following dataset preparation, comparison of the different datasets, prior to imputation, reveals significant changes in demographics and clinical factors over time (Table 23).

Variable	Whole cohort (2001 - 2013)	2001-2010 subgroup	2010-2013 subgroup	P-values (comparing subgroups)
Number	14436	13115	1321	NA
Female (%)	76.3	76.3	75.7	0.6
Age (yrs, SD)	56.0 (12.3)	56.0 (12.2)	56.3 (12.7)	0.4
DAS28-ESR (SD)	6.5 (1.0)	6.6 (1.0)	6.0 (1.0)	<0.01
Swollen Joint Count (SD)	11.1 (6.2)	11.4 (6.2)	8.7 (5.2)	<0.01
Tender Joint Count (SD)	15.5 (7.4)	15.6 (7.4)	14.6 (7.5)	<0.01
Patient Global Score (SD)	72.5 (19.8)	72.5 (19.8)	72.2 (19.5)	0.6
ESR (SD)	44.7(28.2)	46.0 (28.3)	29.6 (22.8)	<0.01
HAQ (SD)	2.0 (0.6)	2.0 (0.6)	1.6 (0.7)	<0.01
BMI (SD)	27.2 (8.1)	27.0 (6.8)	29.6 (17.1)	<0.01
Disease duration (yrs, mean, median (SD))	12.7, 11.0 (9.6)	13.0, 11.0 (9.6)	9.6, 6.0 (9.5)	<0.01
Time from first rheumatology consult to biologics (mean, median (SD))	12.0, 10.0 (9.0)	12.2, 10.0 (8.9)	9.5, 6.0 (9.0)	<0.01
Enbrel™, N (%)	4852 (33.6)	4449 (33.9)	376 (28.5)	NA
Remicade™, N (%)	4222 (29.2)	4196 (32.0)	26 (2.0)	NA
Cimzia™, N (%)	659 (4.6)	0.0	659 (49.9)	NA
Humira™, N (%)	4730 (32.8)	4471 (34.1)	260 (21.0)	NA
Current smokers, N (%)	3108 (21.8)	2861 (22.0)	247 (19.9)	0.03
Ever smoke, N (%)	5368 (37.7)	4922 (37.8)	446 (36.0)	
Never smoke, N (%)	5778 (40.5)	5232 (40.2)	546 (44.1)	

Table 23. Baseline demographics. Unless otherwise specified, mean is given

Comparing the different subgroups, the mean age at baseline has remained relatively constant (range 56.8 – 55.9 years), as has the gender make-up (75.5 – 76.3% female). However, mean baseline disability (measured by the HAQ) has improved significantly – from 2.03 to 1.60; and the mean disease duration has decreased from 13.0 years in the 2001 – 2010 subgroup to 9.6 years in the 2010 – 2013 subgroup. Baseline DAS28-ESR, tender joint count, swollen joint count and ESR have all decreased significantly, although the baseline PGA (measured by VAS) has not altered.

The BMI of patients starting on anti-TNF agents in the 2010-2013 subgroup is significantly greater than those starting in the 2001 – 2010 subgroup and there is also a declining trend in current and ex–smokers, with an increasing proportion of never smokers. The time from when a patient first sees a rheumatologist to commencement of biologics has also decreased significantly from a median of 10 to 6 years. There are

also differing profiles of biologic agents included in each subgroup due to when the different anti-TNF agents became available and differing recruitment dates for the BSRBR-RA.

9.6.2 Frequency of sustained remission: how often does sustained remission and LDA occur in the BSRBR?

Analysis of the imputed data demonstrates the infrequency of sustained remission in all subgroups. Only 2144 patients from the whole cohort of 14436 (14.9%) patients achieved one or more episodes of six-months remission (Table 24) on their first anti-TNF. Point remission was more common, but still infrequent, with 3175 (22.0%) of the whole cohort achieving this outcome. Comparison of the two subgroups (2001-2010 vs. 2010-2013) shows a significant increase in the proportion of patients achieving both sustained and point remission (Table 24).

Sustained LDA is also infrequent, with only 3802 patients (26.3%) identified in the whole cohort. Just over half of those patients (2144) who had achieved sustained LDA had also achieved sustained remission, and 1031 (27%) of the sustained LDA group achieved at least one episode of remission, although it was not recorded as being sustained. Only 627 (16%) of those patients who had achieved sustained LDA had no recorded episode remission in the first six follow-up visits recorded by the BSRBR-RA. The proportion of sustained LDA patients who had achieved at least one episode of non-sustained remission remained stable over the two subgroups examined (7.1 – 8.0%). However, the proportion of patients who have never achieved point or sustained remission in the sustained LDA group reduced significantly from 4.4% to 2.7%, although numbers are small for the 2010-2013 subgroup. Stratifying by time shows that the proportion of patients achieving sustained LDA increased significantly, although this is driven primarily by the significant increases in the proportion of patients achieving sustained remission.

Cohort Dataset		2001 - 2013	2001 - 2010	2010 - 2013	p-value
N		14436	13115	1321	NA
Sustained remission N (%)		2144 (14.9)	1875 (14.3)	285 (21.6)	<0.001
Point remission N (incl. sustained remission; %)		3175 (22.0)	2802 (21.4)	391 (29.6)	<0.001
Sustained LDA N (%)	**All LDA (incl. sustained remission)**	3802 (26.3)	3375 (25.7)	427 (32.3)	<0.001
	LDA excl. sustained remission (≥1 episode point remission ever)	1031 (7.1)	927 (7.1)	106 (8.0)	0.2
	LDA (no episodes of remission)	627 (4.3)	573 (4.4)	36 (2.7)	0.005

Table 24. Changes in sustained remission and LDA over time. Percentages are of cohort dataset total population

9.6.3 Predictors of sustained remission

Associations with the previously identified *a priori* variables (Table 21 and Table 22) with the full BSRBR-RA cohort (from 2001-2013), as well as with the two pre-defined subgroups (2001-2010 and 2010-2013) are now examined.

9.6.3.1 Univariable analysis

As outlined in 9.5.5, univariable regression analysis was undertaken.

Variable	OR (95% CI)	p-value
Gender (Female)	0.56 (0.50 - 0.61)	0.00
HAQ (per unit increase)	0.40 (0.37 - 0.43)	0.00
DAS28-ESR (per unit increase)	0.69 (0.66 - 0.72)	0.00
BMI (per kg/m^2 increase)	0.98 (0.97 - 0.99)	0.00
Swollen:tender joint count (ordered categorical: low, moderate, high)	1.12 (1.04 – 1.20)	0.00
Disease duration (per year increase)	0.98 (0.98 – 0.99)	0.00
Tender joint count (per unit increase)	0.97 (0.97 - 0.98)	0.00
Swollen joint count (per unit increase)	0.99 (0.98 – 1.00)	0.00
PGA (per mm increase)	1.00 (0.99 - 1.00)	0.01
ESR (per mm increase)	0.98 (0.98 - 0.98)	0.00
Ex-smoker (vs current)	0.97 (0.86 - 1.10)	0.64
Never smoker (vs current)	1.03 (0.91 - 1.16)	0.63
Age at starting biologic (per year increase)	0.98 (0.97 - 0.98)	0.00
Remicade™ (vs Enbrel™)	0.74 (0.65 - 0.84)	0.00
Cimzia ™ (vs Enbrel™)	1.76 (1.44 - 2.16)	0.00
Humira™ (vs Enbrel™)	1.50 (1.35 - 1.68)	0.00
Year starting biologic	1.09 (1.07 - 1.10)	0.00

Table 25. Univariable regression modelling of *a priori* chosen variables

All included variables meet the significance threshold of $p < 0.05$ except for smoking. However, there is strong evidence (discussed in Chapter 1; 1.3.3.1) that smoking may play a key role in the pathogenesis of RA. Accordingly, this variable will not be excluded from subsequent statistical models on the basis of this univariable regression modelling.

9.6.3.2 Full cohort (2001-2013)

The characteristics of the full cohort are described in 9.6.1 (summarised in Table 23). Examination of the full *a priori* specified variables identifies several predictors that are strongly associated with sustained remission within the whole cohort (Table 26).

Sustained remission. Full regression model									
	Whole cohort			2001 - 2010 subgroup			2010 - 2013 subgroup		
Variable	OR (95%CI)	VIF	P	OR (95% CI)	Mean VIF	P	OR (95%CI)	VIF	P
Gender (Female)	0.59 (0.53 - 0.65)	1.04	<0.001	0.54 (0.48 - 0.60)	1.04	<0.001	0.81 (0.59 - 1.11)	1.04	0.19
HAQ (per unit increase)	0.53 (0.49 - 0.58)	1.11	<0.001	0.53 (0.48 - 0.58)	1.09	<0.001	0.54 (0.44 - 0.66)	1.07	<0.001
DAS28-ESR (per unit increase)	0.88 (0.75 - 1.03)	3.31	0.11	0.83 (0.70 - 0.99)	3.38	0.04	0.95 (0.65 - 1.37)	2.87	0.77
BMI (per kg/m² increase)	0.98 (0.97 - 0.99)	1.02	<0.001	0.98 (0.97 - 0.99)	1.02	<0.001	0.99 (0.97 - 1.01)	1.02	0.18
Swollen:tender joint count (ordered categorical: low, moderate, high)	1.00 (0.88 - 1.14)	1.83	0.99	0.95 (0.83 - 1.09)	1.83	0.48	1.30 (0.90 - 1.87)	1.83	0.16
Disease duration (per year increase)	1.00 (0.99 - 1.00)	1.06	0.82	1.00 (1.00 - 1.01)	1.06	0.63	0.99 (0.97 - 1.00)	1.04	0.08
Tender joint count (per unit increase)	0.99 (0.97 - 1.01)	2.63	0.20	0.99 (0.97 - 1.01)	2.68	0.20	1.00 (0.95 - 1.04)	2.38	0.83
Swollen joint count (per unit increase)	1.02 (1.00 - 1.03)	2.01	0.03	1.02 (1.01 - 1.04)	2.03	0.01	1.00 (0.95 - 1.05)	1.87	0.98
PGA (per mm increase)	1.01 (1.00 - 1.01)	1.41	<0.001	1.01 (1.00 - 1.01)	1.43	<0.001	1.01 (1.00 - 1.01)	1.35	0.29
ESR (per mm increase)	0.99 (0.99 - 0.99)	1.89	<0.001	0.99 (0.99 - 0.99)	1.92	<0.001	0.99 (0.98 - 1.00)	1.70	0.12
Ex-smoker (vs current)	1.15 (1.01 - 1.32)	1.03	0.03	1.20 (1.04 - 1.38)	1.03	0.01	0.91 (0.62 - 1.33)	1.02	0.62
Never smoker (vs current)	1.10 (0.97 - 1.26)		0.14	1.16 (1.01 - 1.33)		0.04	0.83 (0.57 - 1.19)		0.31
Age at starting biologic (per year increase)	0.98 (0.98 - 0.99)	1.09	<0.001	0.98 (0.97 - 0.98)	1.09	<0.001	1.00 (0.99 - 1.01)	1.08	0.88
Remicade™ (vs Enbrel™)	0.79 (0.69 - 0.90)	1.10	<0.001	0.80 (0.70 - 0.92)	1.13	<0.001	0.48 (0.13 - 1.68)	1.03	0.25
Cimzia™ (vs Enbrel™)	0.97 (0.75 - 1.25)		0.80	NA*		NA	0.91 (0.65 - 1.29)		0.61
Humira™ (vs Enbrel™)	1.36 (1.21 - 1.53)		<0.001	1.26 (1.09 - 1.45)		<0.001	1.19 (0.80 - 1.77)		0.38
Year starting biologic	1.02 (1.00 - 1.04)	1.37	0.11	1.06 (1.02 - 1.10)	1.31	0.01	0.96 (0.79 - 1.16)	1.09	0.66

Table 26. Predictors of sustained remission. Multivariable regression model for whole cohort and subgroups including all a priori specified variables. PGA – Patient Global Assessment of health. *Cimzia™ was only licenced for RA after 2010, so there are no Cimzia™ patients in the 2001 – 2010 subgroup

Sustained remission. Abbreviated models									
	Whole cohort			2001 - 2010 subgroup			2010 - 2013 subgroup		
Variable	OR (95%CI)	VIF	P	OR (95%CI)	VIF	P	OR (95%CI)	VIF	P
Gender (female)	0.59 (0.53 - 0.65)	1.04	<0.001	0.54 (0.48 - 0.60)	1.04	<0.001			
HAQ (per unit increase)	0.53 (0.49 - 0.58)	1.10	<0.001	0.53 (0.48 - 0.58)	1.08	<0.001	0.53 (0.44 - 0.64)	1.06	<0.001
DAS28-ESR (per unit increase)	0.88 (0.75 - 1.03)	3.30	0.11	0.77 (0.70 - 0.84)	1.74	<0.001			
BMI (per kg/m^2 increase)	0.98 (0.97 - 0.99)	1.02	0.00	0.98 (0.97 - 0.99)	1.01	<0.001	0.99 (0.97 - 1.01)	1.03	0.18
Swollen:tender joint count (ordered categorical: low, moderate, high)							1.34 (1.09 - 1.63)	1.03	<0.001
Disease duration (per year increase)							0.98 (0.97 - 1.00)	1.02	0.05
Tender joint count (per unit increase)	0.99 (0.97 - 1.00)	2.19	0.13						
Swollen joint count (per unit increase)	1.02 (1.01 - 1.03)	1.41	<0.001	1.02 (1.01 - 1.03)	1.34	<0.001			
PGA (per mm increase)	1.01 (1.00 - 1.01)	1.41	<0.001	1.01 (1.00 - 1.01)	1.22	<0.001	1.00 (1.00 - 1.01)	1.04	0.33
ESR (per mm increase)	0.99 (0.99 - 0.99)	1.89	<0.001	0.99 (0.99 - 0.99)	1.25	<0.001	0.99 (0.98 - 1.00)	1.03	0.01
Ex-smoker (vs current)	1.15 (1.01 - 1.31)	1.03	0.04	1.20 (1.04 - 1.38)	1.03	0.01			
Never smoker (vs current)	1.10 (0.97 - 1.25)		0.14	1.16 (1.01 - 1.34)		0.03			
Age at starting biologic (per year increase)	0.98 (0.98 - 0.99)	1.06	<0.001	0.98 (0.98 - 0.98)	1.06	<0.001			
Remicade™ (vs Enbrel™)	0.79 (0.69 - 0.90)	1.10	<0.001	0.80 (0.70 - 0.92)	1.13	<0.001			
Cimzia™ (vs Enbrel™)	0.97 (0.75 - 1.25)		0.80	NA*		NA			
Humira™ (vs Enbrel™)	1.36 (1.21 - 1.53)		<0.001	1.26 (1.09 - 1.45)		<0.001			
Year starting biologic	1.02 (1.00 - 1.04)	1.37	0.10	1.05 (1.02 - 1.10)	1.30	0.01			
ANOVA of residual deviances (full vs reduced model)	0.90			0.73			0.76		

Table 27. Predictors of sustained remission. Reduced multivariable regression model for whole cohort and subgroups. Greyed-out boxes represent variables excluded by stepwise regression. PGA – Patient Global Assessment of health. *Cimzia™ was only licenced for RA after 2010, so there are no Cimzia™ patients in the 2001 – 2010 subgroup

As expected, the original model demonstrates multiple collinearity between variables, particularly with DAS28-ESR, which has a VIF of 3.31 (evidence of multiple collinearity). Neither the DAS28-ESR, S:TJR, disease duration, tender joint count or never smoking appear to be associated with sustained remission. However, there is considerable collinearity associated with the DAS28-ESR, S:TJR, and tender joint count variables (VIF range 1.83 – 3.31).

Application of stepwise regression identified variables it was possible to remove from the original full regression model, with no significant impact on the residual deviance between the two models (analysis of variance; ANOVA p-value 0.90; Table 27). Compared with the full regression model, none of the identified associations in the reduced regression model change in direction, and the tender joint count is identified as having a significant association with sustained remission in the reduced regression model. In further support of the model simplification, all the VIFs are reduced compared with the original full model.

As identified in the systematic review, female gender and poor baseline functional status (HAQ), are strongly associated with a reduced likelihood of achieving sustained remission in both the full and reduced regression models (Table 26 and Table 27). Elevated BMI, ESR and older age at starting anti-TNF are all also negatively associated with the likelihood of achieving sustained remission in the reduced variable regression model, as they are in the full regression model (Table 27). An increased swollen joint count and ex-smoker status (but not never smoker status) are associated with an improved chance of achieving sustained remission in both models. An increased PGA is also noted to be significantly associated with an increased likelihood of achieving sustained remission. The biologic agent of choice also appears to have a significant impact on the likelihood of achieving sustained remission (in both regression models). Enbrel™ and Remicade™ were both approved for use in RA in the UK at the same time, so Enbrel™ was chosen as the reference anti-TNF agent. Humira™ use appears to be significantly associated with an increased likelihood of attainment of sustained remission, and Remicade™ is significantly less likely to be associated with sustained remission when compared to Enbrel™. No difference was observed with Cimzia™.

9.6.3.3 2001-2010 cohort

Examining the 2001-2010 subgroup within the BSRBR, many of the significant predictors of sustained remission remain the same as the overall 2001 - 2013 cohort analysis (gender, HAQ, BMI, swollen joint count, PGA, ESR, age at starting anti-TNF, anti-TNF agent and year starting anti-TNF; Table 26).

However, there are some notable differences between the 2001-2010 subgroup and the whole cohort. An increasing baseline DAS28-ESR appears to be negatively associated with the attainment of sustained remission in both regression models for the 2001-2010 subgroup (Table 26 and Table 27), which was not identified in the regression model for the overall cohort. Stopping, or never smoking is positively associated with the likelihood of achieving sustained remission in the reduced regression model (which was not identified in the analysis of the overall cohort), although the relationship between never smokers and sustained remission was only just significant (p-value 0.04).

As with the whole cohort analysis, there was a non-significant change in deviance (ANOVA p-value 0.73) between the full and reduced variable regression model for the 2001-2010 subgroup analysis, and the VIF was reduced for all variables included in the reduced regression model (Table 27).

9.6.3.4 2010-2013 cohort

Analysis of the 2010-2013 subgroup reveals marked differences in predictors associated with sustained remission when compared to both the overall cohort and the 2001-2010 subgroup. Firstly, there are considerably fewer predictors associated with sustained remission for this subgroup, and the predictors identified are quite different to the earlier subgroup. The full regression model (Table 26) only identified HAQ as being associated with sustained remission, although there is collinearity between the other variables (as identified in previous models).

The reduced variable model (Table 27) includes considerably fewer variables compared to the full regression model and has substantially reduced collinearity. The reduced variable regression model identifies associations that were not identified in the full regression model, suggesting that collinearity between variables is affecting the associations identified in the full regression model.

When considering the reduced regression model, as with the full cohort analysis, and the 2001-2010 subgroup, higher baseline HAQ is negatively associated with the likelihood of achieving sustained remission. Likewise, higher ESR levels prior to commencing anti-TNF are also negatively associated with the likelihood of achieving sustained remission. An increasing disease duration also appears to be negatively associated with achieving sustained remission in the 2010 – 2013 subgroup, although this association was not identified in the analysis of the whole cohort and 2001-2010 subgroup analysis. In the 2010-2013 reduced regression model, an increasing S:TJR is associated with an increasing likelihood of achieving sustained remission, which was not identified in either the whole cohort analysis or in the 2001-2010 subgroup analysis. In addition, no association is identified between tender or swollen joint counts individually.

9.6.4 Summary of findings for sustained remission

The factors associated with sustained remission across the full regression models are summarised in Table 28.

There is considerable similarity between the associations identified in the regression models for the 2001 – 2010 subgroup and full cohort regression models. This is unsurprising given the 2001 – 2010 cohort contains 13115 of the 14436 patients included in the whole cohort, compared with 1321 patients in the 2010 - 2013 subgroup.

Variables	Whole cohort (2001 – 2013)	2001-2010 subgroup	2010-2013 subgroup
Gender	-	-	Nil
HAQ	-	-	-
DAS28-ESR	Nil	-	Nil
BMI	-	-	Nil
Swollen:tender joint count ratio	Nil	Nil	Nil
Disease duration	Nil	Nil	Nil
Tender joint count	Nil	Nil	Nil
Swollen joint count	+	+	Nil
PGA	+	+	Nil
ESR	-	-	Nil
Ex-smokers (vs. smokers)	+	+	
Never-smokers (vs. smokers)	Nil	+	Nil
Age at starting biologic	-	-	Nil
Remicade™ (vs Enbrel™)	-	-	Nil
Cimzia™ (vs Enbrel™)	Nil	Nil	Nil
Humira™ (vs Enbrel™)	+	+	Nil
Year starting biologic	Nil	+	Nil

Table 28. Summary of statistically significant (p ≤0.05) associations with sustained remission using full regression model. Positive sign = increases likelihood of sustained remission. Negative sign = reduces likelihood of sustained remission. Nil = non-significant association. NA = no data

9.6.5 Sustained LDA

As outlined in Table 24, just over half (56%) of those individuals who achieved sustained LDA also achieved sustained remission. When examining the 2010 -2013 subgroup, this increased to two thirds (67%) of the sustained LDA group, demonstrating a significant improvement in outcomes over time. Consequently, many of the predictors of sustained remission are likely to be identified as predictors of sustained LDA. However, approximately a quarter of patients who achieve LDA experience non-sustained remission, and a further 16% achieve sustained LDA, but never achieve remission (Table 24). Therefore, the sustained LDA group represents a more heterogeneous group of patients and some predictors of sustained LDA may be identified that were not seen in the analysis of sustained remission and vice versa.

9.6.5.1 Full cohort 2001-2013

Comparing the predictors identified for sustained remission with those identified for sustained LDA for the whole cohort, there are marked similarities (Table 29). There are a few minor differences however. An increasing tender joint count is associated with a reduced likelihood of achieving sustained LDA (in both full and reduced LDA regression models), and a more recent year of starting anti-TNF is significantly associated with an increased likelihood of achieving sustained LDA but not sustained remission. Cimzia™ use is associated with a reduced likelihood of achieving sustained LDA compared to Enbrel™ in both the full and reduced models for sustained LDA, something that was not identified in the analysis of sustained remission. The PGA does not appear to be associated with sustained LDA where it is associated with sustained remission in both the full and reduced regression models. The DAS28-ESR was not included as a variable in the reduced regression model for sustained LDA (unlike sustained remission).

Although all the remaining significant associations are the same as the equivalent sustained remission analyses, the ORs are all slightly reduced. Female gender and higher HAQ are negatively associated with the likelihood of achieving sustained LDA in both the full and reduced regression models.

Examining the degree of collinearity of the reduced regression model compared with the full regression model, the VIF is reduced or equivalent for all the variables included. ANOVA model checking between the full and reduced variable analysis shows no significant difference in deviance (ANOVA p-value 0.78; Table 30).

Sustained LDA. Full regression model	Whole cohort			2001 – 2010 subgroup			2010 – 2013 subgroup		
Variable	**OR (95%CI)**	**VIF**	**P**	**OR (95%CI)**	**VIF**	**P**	**OR (95%CI)**	**VIF**	**P**
Gender (female)	0.65 (0.59 - 0.71)	1.03	0.00	0.63 (0.58 - 0.69)	1.03	0.00	0.90 (0.68 - 1.21)	1.04	0.49
HAQ (per unit increase)	0.59 (0.55 - 0.64)	1.11	0.00	0.59 (0.54 - 0.63)	1.09	0.00	0.57 (0.47 - 0.68)	1.07	0.00
DAS28-ESR (per unit increase)	0.97 (0.85 - 1.10)	3.43	0.60	0.89 (0.77 - 1.03)	3.52	0.13	1.15 (0.82 - 1.61)	2.89	0.41
BMI (per kg/m² increase)	0.98 (0.97 - 0.99)	1.02	0.00	0.98 (0.97 - 0.99)	1.02	0.00	0.98 (0.97 - 1.00)	1.03	0.05
Swollen:tender joint count (ordered categorical: low, moderate, high)	0.96 (0.87 - 1.07)	1.84	0.47	0.92 (0.82 - 1.03)	1.84	0.14	1.46 (1.06 - 2.03)	1.84	0.02
Disease duration (per year increase)	1.00 (1.00 - 1.01)	1.06	0.62	1.00 (1.00 - 1.01)	1.06	0.25	0.99 (0.98 - 1.01)	1.05	0.30
Tender Joint Count (per unit increase)	0.98 (0.96 - 0.99)	2.68	0.00	0.98 (0.96 - 0.99)	2.75	0.01	1.00 (0.96 - 1.04)	2.38	0.99
Swollen Joint Count (per unit increase)	1.01 (1.00 - 1.03)	2.04	0.04	1.02 (1.01 - 1.03)	2.06	0.00	0.98 (0.94 - 1.03)	1.87	0.44
PGA (per mm increase)	1.00 (1.00 - 1.00)	1.44	0.17	1.00 (1.00 - 1.01)	1.46	0.02	1.00 (0.99 -1.01)	1.36	0.66
ESR (per mm increase)	0.99 (0.99 - 0.99)	1.94	0.00	0.99 (0.99 - 0.99)	1.97	0.00	0.99 (0.98 -1.00)	1.71	0.01
Ex-smoker (vs current)	1.15 (1.03 - 1.27)	1.03	0.01	1.14 (1.02 - 1.28)	1.03	0.02	1.00 (0.72 -1.41)	1.02	0.98
Never smoker (vs current)	1.09 (0.98 - 1.21)		0.12	1.09 (0.98 - 1.22)		0.12	0.95 (0.69 - 1.32)		0.77
Age at starting biologic (per year increase)	0.99 (0.98 - 0.99)	1.08	0.00	0.98 (0.98 - 0.99)	1.09	0.00	1.00 (0.99 - 1.01)	1.08	0.44
Remicade™ (vs Enbrel™)	0.81 (0.73 - 0.90)	1.10	0.00	0.83 (0.75 - 0.92)	1.13	0.00	0.22 (0.06 - 0.77)	1.03	0.02
Cimzia™ (vs Enbrel™)	0.71 (0.57 - 0.90)		0.00	NA*		NA	0.81 (0.60 - 1.09)		0.16
Humira™ (vs Enbrel™)	1.23 (1.12 - 1.35)		0.00	1.14 (1.01 - 1.27)		0.03	0.91 (0.64 - 1.29)		0.59
Year starting biologic	1.03 (1.01 - 1.04)	1.37	0.00	1.08 (1.05 - 1.11)	1.30	0.00	0.99 (0.84 - 1.18)	1.09	0.93

Table 29. Predictors of sustained LDA. Regression model for whole cohort and subgroups including all *a priori* specified variables.
PGA – Patient global assessment of disease activity. *Cimzia™ was only licenced for RA after 2010, so there are no Cimzia™ patients in the 2001 – 2010 subgroup

Sustained LDA. Reduced regression model	Whole cohort			2001 – 2010 subgroup			2010 – 2013 subgroup		
Variable	**OR (95%CI)**	**VIF**	**P**	**OR (95%CI)**	**VIF**	**P**	**OR (95%CI)**	**VIF**	**P**
Gender (female)	0.65 (0.59 - 0.71)	1.03	0.00	0.63 (0.58 - 0.70)	1.03	0.00			
HAQ (per unit increase)	0.59 (0.55 - 0.64)	1.10	0.00	0.59 (0.55 - 0.64)	1.08	0.00	0.57 (0.48 - 0.67)	1.02	0.00
DAS28-ESR (per unit increase)				0.89 (0.77 - 1.03)	3.52	0.13			
BMI (per kg/m² increase)	0.98 (0.97 - 0.99)	1.02	0.00	0.98 (0.97 - 0.99)	1.01	0.00	0.98 (0.97 - 1.00)	1.02	0.06
Swollen:tender joint count ratio (ordered categorical: low, moderate, high)				0.92 (0.83 - 1.03)	1.84	0.14	1.33 (1.12 - 1.59)	1.01	0.00
Disease duration (per year increase)									
Tender joint count (per unit increase)	0.98 (0.97 - 0.98)	1.15	0.00	0.98 (0.96 - 0.99)	2.75	0.01			
Swollen joint count (per unit increase)	1.01 (1.00 - 1.02)	1.12	0.02	1.02 (1.01 - 1.03)	2.06	0.00			
PGA (per mm increase)	1.00 (1.00 - 1.00)	1.05	0.17	1.00 (1.00 - 1.01)	1.46	0.02			
ESR (per mm increase)	0.99 (0.99 - 0.99)	1.05	0.00	0.99 (0.99 - 0.99)	1.97	0.00	0.99 (0.98 - 1.00)	1.01	0.00
Ex-smoker (vs current)	1.15 (1.03 - 1.28)	1.02	0.01	1.15 (1.03 - 1.28)	1.03	0.02			
Never smoker (vs current)	1.09 (0.98 - 1.21)		0.11	1.10 (0.99 - 1.23)		0.09			
Age at starting biologic (per year increase)	0.99 (0.98 - 0.99)	1.05	0.00	0.98 (0.98 - 0.99)	1.06	0.00			
Remicade™ (vs Enbrel™)	0.81 (0.73 - 0.90)	1.10	0.00	0.83 (0.75 - 0.92)	1.13	0.00	0.22 (0.06 - 0.76)	1.00	0.02
Cimzia™ (vs Enbrel™)	0.72 (0.57 - 0.90)		0.00	NA*		NA	0.80 (0.61 - 1.06)		0.12
Humira™ (vs Enbrel™)	1.23 (1.12 - 1.35)		0.00	1.13 (1.01 - 1.27)		0.03	0.90 (0.64 - 1.28)		0.56
Year starting biologic	1.03 (1.01 - 1.04)	1.36	0.00	1.08 (1.05 - 1.11)	1.30	0.00			
ANOVA of residual deviances (full vs reduced model)	0.78			0.25			0.91		

Table 30. Predictors of sustained LDA. Reduced regression model for whole cohort and subgroups. Greyed-out boxes represent variables excluded by stepwise regression. PGA – Patient global assessment of disease activity *Cimzia™ was only licenced for RA after 2010, so there are no Cimzia™ patients in the 2001 – 2010 subgroup

9.6.5.2 2001-2010 cohort

When considering the 2001 – 2010 subgroup, the predictors of sustained remission and LDA are similar, but not identical. Unlike sustained remission, DAS28-ESR is not associated with sustained LDA in the 2001-2010 subgroup, as it was for the equivalent sustained remission analysis. However, an increasing tender joint count appears to be associated with an increased likelihood of achieving sustained LDA in both full and reduced regression models for this subgroup (Table 29 and Table 30), but not remission (Table 26 and Table 27). The S:TJR was included in the reduced regression model for sustained LDA, unlike the equivalent sustained remission analysis, however the association was non-significant. As with the previous models, there appears to be a calendar-year effect, as patients prescribed anti-TNF agents which were commenced more recently appear to have a greater chance of achieving sustained remission. The difference in model fit between full and reduced regression models is not significant (ANOVA p-value 0.73; Table 30) and VIF is reduced for all included variables.

9.6.5.3 2010-2013 cohort

As with the sustained remission analysis for the 2010 – 2013 subgroup (Table 26 and Table 27) the number of associations are considerably reduced compared with the full cohort and 2001 – 2010 subgroup analyses. However, unlike the sustained remission analysis, the full and reduced regression models have identified similar associations, with the exception of BMI, which just reaches significance in the full regression model ($p = 0.05$) but is lost in the reduced LDA model ($p = 0.06$). An increasing HAQ and ESR are both negatively associated with the chance of achieving sustained LDA whilst an increasing S:TJR is associated with an increased likelihood of achieving sustained LDA (OR 1.33, 95% CI 1.12 – 1.59). Overall, the associations identified for the sustained LDA analysis are almost identical to the sustained remission analysis, with the exception of disease duration, which does not appear to be associated with the likelihood of achieving sustained LDA. Remicade™ use also appears to have a significantly reduced likelihood of attainment of sustained LDA compared to Enbrel™.

As with previous reduced regression models, collinearity is reduced with a non-significant effect on model fit (ANOVA p-value 0.91; Table 30).

9.6.6 Summary of findings for sustained LDA analysis

As with the sustained remission analyses, analyses were undertaken on the cohort overall (2001 – 2013) as well as subgroups of the overall cohort that were stratified by time (2001 – 2010 and 2010 – 2013).

There is an increase in the proportion of patients achieving sustained LDA over time. This is primarily driven by the increase in the number of patients who are achieving sustained remission (who are included in the LDA cohort). When considering the proportion of patients who achieve sustained LDA, but not sustained remission, this has decreased from 44% to 33% comparing the 2001 – 2010 and 2010 – 2013 subgroups respectively. Furthermore, the proportion of patients who achieve sustained LDA but never achieve any episode of remission has decreased significantly from 4.4% to 2.7% over time. However, compared to the cohort as a whole, the proportion of patients who have achieved sustained LDA but not sustained remission has remained stable over time at 11.4% and 10.7% (comparing the 2001 – 2010 and 2010 - 2013 subgroups; Table 24).

Higher HAQ and ESR are associated with a reduced likelihood of achieving sustained LDA for all analyses. Whilst female gender appears to have been negatively associated with sustained remission in the full cohort and in the 2001 – 2010 subgroup (and was identified in the systematic review; Chapter 7), this effect was not demonstrated in the 2010 – 2013 subgroup. There was also a significant positive association identified between the S:TJR and sustained LDA in the 2010 -2013 subgroup.

Variables	Whole cohort (2001 – 2013)	2001 - 2010 Subgroup	2010 - 2013 Subgroup
Gender	-	-	Nil
HAQ	-	-	-
DAS28-ESR	Nil	Nil	Nil
BMI	-	-	-
Swollen:tender joint count ratio	Nil	Nil	+
Disease duration	Nil	Nil	Nil
Tender joint count	-	-	Nil
Swollen joint count	+	+	Nil
PGA	Nil	+	Nil
ESR	-	-	-
Ex-smokers (vs. smokers)	+	+	Nil
Never-smokers (vs. smokers)	Nil	Nil	Nil
Age at starting biologic	-	-	Nil
Remicade™ (vs Enbrel™)	-	-	-
Cimzia™ (vs Enbrel™)	-	NA	Nil
Humira™ (vs Enbrel™)	+	+	Nil
Year starting biologic	+	+	Nil

Table 31. Summary of statistically significant (P ≤ 0.05) associations with sustained LDA using full regression model. Positive sign = increases likelihood of sustained remission. Negative sign = reduces likelihood of sustained remission. Nil = no association. NA = no data

9.7 Chapter discussion

9.7.1 Demographics of the BSRBR-RA

This analysis shows that the demographics and clinical features of patients treated with anti-TNF in the UK over a 12-year period have changed significantly (Table 23). Subgroup analysis has revealed that individuals are being treated earlier in their disease course, with a significant reduction in disease duration at time of commencement of anti-TNF (from a median of 11 to 6 years; Table 23). However, despite this improvement, it is surprising that in 2010-2013, there was still a median time from onset of disease to commencement of biologics of nearly 6 years despite routine access to anti-TNF as part of clinical care. A possible explanation for this may be that at a population level, earlier and more aggressive use of synthetic DMARD may

offer several years of effective treatment before disease activity increases to above the NICE mandated DAS28 score threshold of 5.1.

Whilst there has been a significant reduction in disease duration at time of starting anti-TNF, it is interesting to note that the mean age at starting anti-TNF has remained constant over time. This seems counterintuitive given that disease duration prior to starting anti-TNF is decreasing. Exact reasons for this are unclear, but there are several possible explanations for such a result. It is possible that older patients are more willing to consent to participating in contributing to the BSRBR-RA, which may represent a general recruitment bias, or clinicians may be more confident in using the drugs in older patients now. Alternatively, the disease presentation and course of RA may be changing, with later disease onset (discussed in Chapter 1; 1.3.4). This could be related to a birth cohort effect, or external environmental factors such as a reduction in the population prevalence of known exacerbating, and possibly precipitating, features of RA such as smoking. In support of this, the prevalence of both current and ex-smokers has reduced significantly between the two subgroups analysed. A further explanation may be that a selection bias may be occurring in that different biologic agents (not available in 2001-2010, and excluded from the analyses in this thesis) are being used in younger patients, although this seems less likely.

Disease activity at time of commencing anti-TNF has also reduced significantly over the last decade (from 6.56 to 6.00; Table 23). When examining the component parts of the DAS28-ESR prior to commencing anti-TNF, there has also been a significant reduction in the number of swollen joints observed in those starting anti-TNF by nearly three swollen joints (from 11.35 to 8.70; Table 23). Whilst the number of tender joints has reduced significantly as well, it has only reduced by 1 (from 15.6 to 14.6). What is striking however, is that although overall disease activity has reduced significantly, the patient perception of the overall disease impact, as quantified by the PGA, has not reduced at all between the two subgroups, and has remained almost identical over time at 72mm (Table 23). This may be because the patient perceived relationship between disease activity and visual analogue score is a non-linear one; meaning that the differential impact on quality of life between having 11 swollen joints and 9 swollen joints is minimal. Other psychological aspects may also play a part in the reporting of a

visual analogue score (84). There may also be a greater expectation of treatment in the patients included in the 2010 - 2013 subgroup, and hence greater motivation for patients to express the impact of symptoms on their lives, rather than adopting a 'stoical' acceptance and under-reporting of symptoms as may have been the case historically when limited treatment options were available. Another possible explanation may be that patients may be receiving better education as to the importance of quantifying symptom severity, either through better explanation of the purpose of the metric to patients by healthcare staff, or through previous regular use of metrics that are used more frequently in clinical practice today than previously - both potentially resulting in more honest reporting of symptom impact.

An important point to note in the interpretation of the changing disease activity at time of commencing biologic is that there is a 'floor-effect' associated with use of the DAS28 in the UK due to mandatory NICE guidance on access to biologics. This guidance specifies a minimum DAS28 score of 5.1 is required in order to commence a biologic agent (162). Therefore, by definition, those patients who have been commenced in anti-TNF and enrolled onto the BSRBR-RA will have to have met this minimum threshold, and it would be inconceivable that the mean DAS28 score could reduce below this mandatory threshold in the UK healthcare environment. This disease activity threshold has been constant since the introduction of biologics in the UK and is likely to have influenced the baseline characteristics in the BSRBR-RA over time.

In addition to the reduction in baseline disease activity over time, there has been a significant reduction in the baseline HAQ prior to commencing anti-TNF of 0.43, demonstrating that patients starting anti-TNF are significantly less disabled at the time of starting biologics in 2010-2013 compared with 2001-2010 (Table 23). This is of importance as evidence has shown that worsening HAQ scores are challenging to reverse (241) and have a significant impact upon employment and engagement in society (244,245). This finding is evidence of improved outcomes in both biologic and pre-biologic treatment.

9.7.2 Frequency of sustained remission and LDA

This analysis has identified significant changes in the proportion of patients taking their first anti-TNF who achieve sustained remission and sustained LDA over a 12-year period. Sustained remission has significantly improved from 14.3 to 21.6% over the 12-year period examined in this analysis (Table 24). Sustained LDA has also improved from 25.8 to 32.3%, and suggests that earlier treatment with anti-TNF prior to the onset of irreversible joint damage and disability, is likely to be having an impact on the attainment of sustained remission and LDA. Examining the makeup of those patients who achieve sustained LDA has highlighted that within this group too, outcomes have improved, and the proportion of patients achieving sustained LDA, but no episodes of remission has diminished significantly (from 4.4 to 2.7%; Table 24). This suggests that clinicians and patients are aiming to achieve remission in as many instances as possible and are having success in doing so. However, it is important to note that as the range of biologic agents available has increased, it is probable that patients are more likely to switch to alternative biologic agents if a suitable response is not obtained. Because this analysis did not include patients taking all biologics or those who had switched anti-TNF agent, it is not possible to identify if the rate of sustained LDA and remission for all patients taking biologic agents has increased or not.

However, despite the improvements in outcomes over the past decade, it is sobering to note that between 68% and 78% of patients do not achieve either sustained LDA or remission respectively (Table 24). This highlights scope for further improvement in outcomes. With the increasing array of treatment options available for the management of RA, targeting anti-TNF at those patients who are most likely to achieve sustained remission or LDA may offer a way of improving outcomes by identifying the right drug for the right patient. The analysis of predictors of sustained remission and LDA has highlighted some potential variables that could assist in predicting outcome with anti-TNF.

9.7.3 Predictors of sustained remission and LDA

In both the sustained LDA and sustained remission analyses, a greater number of predictors were identified in the 2001-2010 subgroup analysis compared with the 2010-2013 analysis. This may be due to the larger group size in the 2001 – 2010 analysis compared with the 2010 – 2013 analysis. However, it is also possible that the change in demographics of individuals starting anti-TNFs over time, and increasing standardisation in treatment paradigms may have resulted in the impact of predictors identified in the 2001 – 2010 subgroup becoming less over time.

9.7.3.1 Gender

One change that appears to have occurred over time is the loss of association between gender and sustained remission or LDA in the 2010-2013 subgroup analysis compared with the 2001 – 2010 subgroup. Female gender was one of the most clearly identified associations with sustained remission in the systematic review with a pooled OR of 0.53 (95% CI 0.44 – 0.63), and was also identified as being associated with both sustained remission and LDA when the cohort was analysed, in both the whole cohort, and in the 2001 - 2010 subgroup analyses (OR of 0.63, 95% CI 0.58 – 0.69, and OR 0.65, 95% CI 0.59 – 0.71, for the 2001-2010 and whole cohort sustained LDA analyses respectively). The association with gender was even more pronounced for the sustained remission analysis with ORs of 0.54 (95% CI 0.48 – 0.60) and 0.59 (95% CI 53 – 0.65) for 2001 - 2010 and whole cohort sustained remission analyses respectively. It is possible that there may be an unidentified differential selection bias that has been applied to women and not men in the 2010 – 2013 subgroup, although what this could be is not clear. It may be that men with lower disease activity are less likely to participate in a research study, with a resultant apparent increase in disease activity amongst men that negates an effect of gender. It may also be that over the duration of the study, men and women have become more similar in reporting disease activity from RA, leading to a loss of this apparent relationship.

9.7.3.2 Health assessment questionnaire (HAQ)

One consistent finding identified across all analyses was the significant association between both sustained remission and LDA and the baseline HAQ. In all analyses, a higher HAQ score was associated with a significantly reduced likelihood of achieving sustained remission with an OR of between 0.53 (95% CI 0.49 – 0.58) and 0.54 (95%CI 0.44 – 0.66) per unit increase in HAQ, and a range of OR from 0.57 (0.47 – 0.68) to 0.59 (0.55 – 0.64) for LDA. This adds further support to the hypothesis that earlier treatment, before significant disability develops, improves chance of achieving sustained remission. The fact that a strong relationship is identified in both the 2001-2010 and 2010-2013 subgroups suggests that the relationship is not influenced by a calendar year effect.

9.7.3.3 ESR

Higher ESR was also significantly associated with a reduced likelihood of achieving sustained remission and LDA with an OR of 0.99 (95% CI 0.99 – 0.99) per unit increase of ESR in both sustained remission and LDA analyses. Whilst an OR of 0.99 may appear to be only a weak predictor, the OR is per unit increase in ESR (i.e. 1 mm/hour). Given the range of ESR values is 0 – 150, the compound effect of changes in ESR values actually has a large effect on the absolute risk (or likelihood) of achieving sustained remission or LDA.

9.7.3.4 Tender and swollen joint count, S:TJR, PGA, and DAS28-ESR

An increasing swollen joint count was significantly associated with an increased chance of achieving sustained remission for both the cohort as a whole, and the 2001 – 2010 subgroup. This association was maintained for sustained LDA in the 2001 – 2010 subgroup. Whilst this relationship may appear counterintuitive at first, it could be explained by the fact that an increasing swollen joint count is strong evidence of an ongoing (potentially) reversible inflammatory disease activity (that anti-TNF is effective at targeting), rather than a less reversible chronic pain phenomenon (that is

less responsive to treatment with anti-TNF) that may manifest as a greater tender joint count. In support of this explanation, an increasing tender joint count was negatively associated with the likelihood of achieving sustained LDA in the whole cohort and the 2001 – 2010 subgroup models.

Another unexpected finding was that an increasing PGA was associated with an increased likelihood of achieving sustained remission for the whole cohort and 2001 – 2010 subgroup. It is possible that those patients with active inflammatory disease experience a more fluctuating level of pain that means that that when they have active disease, they rate their PGA as more severe (i.e. a higher value) than someone with more of a chronic pain component to their RA, who may have an elevated but more stable PGA.

A noteworthy finding is the association of the S:TJR with both attainment of sustained remission and sustained LDA in only the 2010-2013 subgroup analyses. Kristensen et al (173) identified this as predictor of sustained remission in an early arthritis cohort, and it is possible, that as anti-TNF treatment is being used earlier in the disease course of RA, the predictors of sustained remission in the BSRBR-RA cohort are becoming more similar to those that have been identified in earlier arthritis cohorts.

A notable absence from the identified associations was the DAS28-ESR, which was only associated with sustained remission in the 2001-2010 subgroup analysis. This is probably due to the high levels of multiple collinearity that occurred when including both the components of the DAS28 and the composite score itself in the regression models.

9.7.3.5 BMI

Higher baseline BMI was significantly associated with a reduced likelihood of achieving sustained remission and sustained LDA in the full cohort and 2001 – 2010 subgroup analysis, although this association was only identified in the full variable regression modelling of the sustained LDA analysis of the 2010 – 2013 subgroup. It is possible that

the association may have been lost due to the small cohort size of the 2010 – 2013 subgroup in comparison to the 2001 – 2013 subgroup and the cohort overall. It is also possible that the changing demographic characteristics of the BSRBR-RA has blunted this effect in the later subgroup, as comparisons of baseline characteristics between the 2001 – 2010 and 2010 – 2013 subgroups demonstrated a significant difference in mean BMI from 27.0 in the 2001 – 2010 subgroup to 29.6 kg/m^2 in the 2010 – 2013 subgroup (Table 23). Higher BMI has been noted to be associated with poorer outcomes both in early RA patients (246), and those with more established disease taking anti-TNF (247). These findings suggest that higher BMI does appear to be independently associated with reduced rates of sustained remission.

9.7.3.6 Smoking

Ex-smoking status was significantly associated with the likelihood of achieving sustained remission and LDA in both the whole cohort and 2001-2010 subgroup analyses. Whilst there was a trend towards benefit for patients who had never smoked, this did not reach significance. Whilst the association between stopping smoking and improved outcomes is in line with what would be expected, it is surprising that a significant association was not identified when comparing never smokers with current smokers. It may be that the smoking data used in this analysis (never/ever/current) were too crude to pick up such an association and more quantified measures of smoking (such as pack-years) may have yielded different results. Comprehensive smoking exposure models have been proposed as a more accurate method of modelling smoking exposure and risk in epidemiological studies (248) and have been used to examine the risk of vascular outcomes in patients with scleroderma. This identified a negative effect of smoking on vascular outcomes where a categorisation or pack-year approach failed to identify an association (249). Such an approach could be used in further studies to specifically examine the association between smoking and sustained remission or LDA. However, Barnabe *et al.*(220) were able to identify an association between smoking and sustained remission using never/ex/current categories for smoking, although in that analysis ex- and never-smoking were grouped together, leading to a dichotomous outcome of current/not current smoker being used.

9.7.3.7 Age at time of anti-TNF commencement

Age when commencing anti-TNF appears to be negatively associated with the likelihood of achieving sustained remission in both the whole cohort and the 2001-2010 subgroup analysis, but this association was not identified in the 2010-2013 analysis (Table 26 and Table 27). This is interesting given that the mean age of patients starting anti-TNFs has not changed over time. It is possible however that the 2010-2013 subgroup analysis may be underpowered to identify this association.

9.7.3.8 Calendar-year effect

There appears to have been a calendar-year effect on both likelihood of achieving sustained LDA and remission in the whole cohort and 2001-2010 subgroup analysis, with those patients who started anti-TNF more recently being more likely to achieve sustained remission and LDA (Table 26 and Table 27). This effect was not observed in the 2010 – 2013 subgroup analysis. However, given that the window of recruitment for the 2010-2013 analysis was only 3 years, there may not have been sufficient time for a calendar-year effect to emerge.

9.7.3.9 Anti-TNF agent

Finally, the choice of anti-TNF appears to be associated with attainment of sustained remission and LDA. Humira™ is associated with a significantly increased likelihood of sustained remission and LDA, and Remicade™ associated with a decreased likelihood, for all whole cohort and 2001 – 2010 subgroup analyses, when compared to Enbrel™. Remicade™ use was also associated with reduced likelihood of attainment of sustained LDA when considering the 2010 – 2013 subgroup analyses, and Cimzia™ use was associated with a reduced likelihood of attainment of sustained LDA for the whole cohort analyses, but none of the subgroup analyses.

The possible reasons for these are multiple. It may be Humira™ is indeed associated with better outcomes. However, Enbrel™ and Remicade™ were first in class anti-TNF

agents, and therefore will have been used to treat patients with the worst disease activity, greatest disability and longest disease duration when the drugs first became available in the UK; all factors independently associated with a lower likelihood of achieving remission (Chapter 7 and Katchamart *et al.*(212)). Whilst these factors were adjusted for in this analysis, it is possible that measures such as the HAQ and DAS28-ESR do not fully adjust for these confounders.

The other point to note is that the relationship between anti-TNF agent and sustained remission and LDA mirrors that of the association with calendar year, suggesting that the effect of calendar year may be manifest in the relationship between anti-TNF choice and sustained remission/LDA that is not fully adjusted by the regression models.

Overall, this relationship is complex and needs further investigation. The interactions between clinical phenotypes, calendar year and other confounding factors related to this association need to be factored into any study design (Figure 23). However, it does warrant further investigation, and a case-matched study design with the primary aim of examining differences between drugs may be a more appropriate method to examine this relationship with more certainty.

9.8 Strengths and weaknesses of this analysis

One of the strengths of this analysis is that it has used real-world data to examine one of the most challenging clinical targets of treatment: sustained remission. Use of real-world data to examine such an end-point is important, as clinical trial data are lacking. When considering point remission rates, clinical trials often achieve much better outcomes than are seen in routine clinical practice, where patients are not selected based on tight inclusion criteria (186,250). As such, 'real-world' analyses are essential for both clinicians and patients to understand what outcomes are achieved outside of the trial setting.

This analysis has investigated an outcome (sustained remission) that is infrequently examined, but highly clinically relevant in a chronic, incurable disease such as RA. In

addition, a clinically pragmatic target of sustained LDA was evaluated, which makes the results of this analysis useful to clinicians appraising the achievements in treating RA patients in the UK.

A thorough assessment of missing data was also undertaken prior to analysis, and a robust imputation methodology was used for missing data. A comprehensive range of clinically useful predictors was selected, with an evidence-based approach using results from the systematic review, and the issue of multiple collinearity between variables was quantified and addressed using a defined methodology (stepwise regression). The fact that comparisons between the full and reduced regression models did not highlight significant changes in direction of relationships lends further support to the relationships identified in the full regression models, and minimises the likelihood of identification of spurious relationships occurring due to chance.

A further strength is that this analysis has examined how the demographics and frequency of sustained remission and LDA have changed over time, and highlight how improvements in clinical practice and treatment strategies have influenced outcomes. By stratifying the analysis, these results provide a contemporaneous assessment of anti-TNF treatment outcomes in the UK today, which could help guide treatment strategies for the future.

Despite the strengths outlined above, there are weaknesses in this analysis. Although missing data were managed in a comprehensive manner, and there are no protocol-specific reasons why a selection bias should occur in the BSRBR-RA, any registry study will have a degree of selection bias by virtue of patients having to consent to participate, and clinicians being willing to contribute to the study. As such, it is not possible to fully quantify the theoretical missing data from patients who never consented, or were never approached, to take part in the study. Additionally, in the analysis of longitudinal outcomes, there will be variation in the date that a follow-up should have occurred and when it actually did occur (described in detail in 9.5.3), meaning that the time between sequential follow-ups may in some cases be more or less than six months.

Another weakness is that this analysis only evaluated remission using DAS28 criteria, which is known to be a more lenient definition of remission compared to more contemporaneous definitions (101,102). However, physician global assessment of disease activity is not collected by the BSRBR-RA, which precludes analysis of newer composite outcome measures such as the CDAI and SDAI.

The sample size of the two subgroups was also markedly different, which is likely to have reduced the power of the 2010 – 2013 subgroup compared to the 2001 – 2010 subgroup. However, using 2010 as the year to split the cohort was an evidence-based decision as described in Chapter 9; 9.5.11.

This analysis has not included all different anti-TNFs, including Simponi™ (which is not collected by the BSRBR-RA) and biosimilars (which have been clinically available since 2013 and were therefore excluded from the analysis). This analysis also did not include concomitant DMARDs, non-steroidal anti-inflammatories (NSAIDs) or steroid use as variables in the analysis. The BSRBR-RA only captures steroid use as a 'yes/no' outcome at each follow-up, which lacks the granularity to be able to explore it as a co-variate in this analysis.

NSAID data on the BSRBR-RA has a significant missing data component as NSAIDs available over-the-counter, and often taken for short intervals, meaning it is difficult to know with confidence if reported NSAID use is accurate. Additionally, because NSAIDs are often taken on an 'as required' basis for short periods of time, their impact on a long-term outcome is difficult quantify with confidence.

Although synthetic DMARD data use is collected by the BSRBR-RA, it was not included. This was due to two reasons that made the variable challenging to manage in modelling for this analysis. Firstly, if concomitant DMARD data were included, should this be DMARD use at baseline (i.e. before the anti-TNF is commenced), or as ongoing DMARD use? If only baseline use was considered, then a patient may be recorded as taking methotrexate at baseline, which may be subsequently discontinued once the anti-TNF is started. With a pharmacokinetic half-life of between five and eight hours, and loss of clinical effect usually seen before three months, it would be unlikely that methotrexate

taken for the first six months of treatment with anti-TNF would affect outcomes in year two or three of taking anti-TNF. If ongoing use of DMARDs was included, how should DMARD-switching be managed if the anti-TNF agent remained the same? In addition to the challenges with incorporation of a time-varying covariate into a longitudinal model, there are likely to be interactions between such a variable and the choice of anti-TNF, as some anti-TNF agents have better evidence for use as monotherapy than others (251).

For these reasons, concomitant medications were not included as independent variables in the analysis.

9.9 Conclusions

The proportion of patients achieving sustained remission and LDA has increased significantly over time, although patients who achieve this outcome remain in the minority. It is also evident that there are clinical and demographic factors that are associated with sustained remission and LDA that could be used to help predict the likelihood of achieving this outcome in the first three years of treatment with anti-TNF, which could help clinicians and patients decide on future treatment strategies.

9.10 Key points from this chapter

- The demographics of patients taking anti-TNF in the UK has changed significantly over time.
- Whilst sustained remission and LDA outcomes have improved significantly, those patients who achieve these outcomes remain in the minority.
- Baseline clinical and demographic factors are significantly associated with long-term outcomes.

Chapter 10

10 Modelling trajectories of response to anti-TNF in patients with RA

As identified in Chapter 9, certain demographic and clinical features have been identified that are associated with the likelihood of achieving sustained remission and LDA. However, as previously described, using set thresholds for defining any parameter (such as remission or LDA) creates difficulties with values that fall close to either side of the imposed threshold. Using thresholds also converts a continuous outcome such as the DAS28-ESR (with a scale 0-10) into a categorical outcome (remission/LDA/MDA/HDA). This means that DAS28-ESR values that may be only 0.1 points apart may be allocated to different groups and be allocated to the same category label as scores that are less close to them, but which fall within the same category, defined by the threshold. Whilst conversion of a continuous variable to a categorical variable may be necessary for purposes of analysis, it can have limitations when applying findings from the analysis back to the clinical setting where the DAS28-ESR score is used as a continuous score.

A more clinically realistic model of how disease activity scores are used is to map the trajectory of response of a patient to anti-TNF over time which allows for fluctuations in response to a drug such as anti-TNF. By mapping the trajectory of many individuals with similar group parameters (as is feasible using the BSRBR-RA), it is possible to identify groups which have a similar response trajectory, and map a class trajectory for the group. Such an approach creates groups based on patterns in the data and not by externally determined thresholds.

Another difficulty when using epidemiological time-series data is the discrepancy between when the data *should* have been collected, and when it *was* collected. Because the previous analysis in Chapter 9 required identification of remission/LDA at two sequential time points, it was necessary to use the follow-up number, rather than the

actual date of completion of the follow-up form when determining sustained remission/LDA. However, trajectory mapping using latent-class analysis is able to handle variable time differences between follow-up visits, so it is possible to use the actual date that the follow-up form was completed rather than the follow-up number. This means a more accurate representation of the spread of data and disease activity in a temporal manner can be analysed.

This analysis seeks to use latent class modelling to identify if different trajectories of response are present within the group of patients taking anti-TNF on the BSRBR-RA, and if so, if there are any common traits of individuals that are associated with trajectories that map out a good or poor outcome.

10.1 Aim

To identify different clinical response trajectories associated with the use of anti-TNF in the real-world clinical setting, and predictors associated with different drug response trajectories.

10.2 Objectives of this chapter

1. To examine the longitudinal disease activity data in the BSRBR-RA to identify if there are clearly identifiable trajectories of response to anti-TNF medications.
2. To identify any clinical or demographic features that are associated with the different response trajectories identified in 1.
3. To examine how trajectories and predictors of response may have changed over time.

10.3 Null hypothesis

The null hypothesis of this analysis is that there is only one common response trajectory associated with the use of anti-TNF, to which all patients belong.

10.4 Methods

10.4.1 Dataset preparation

The same dataset preparation, management of missing data and multiple imputation methods as outlined in Chapter 6 are used in this chapter.

10.4.2 Analysis plan

This analysis will use latent class mixed modelling (LCMM; explained in Chapter 5; 5.5) to identify if there are common response trajectories (or classes) that can be identified from the anti-TNF cohort of patients within the BSRBR-RA. If different trajectories are identified, the clinical and demographic features of patients who achieve these different outcomes will be examined to identify if there are any common traits that might help predict a sustained good response, or that may be associated with a poor response.

10.4.3 Cohort specification and variable selection

Predictors that will be used in regression modelling will be the same *a priori* variables specified in Chapter 9 (Table 32). As in Chapter 9, this analysis will be undertaken on three datasets - one 'full' cohort including all patients starting anti-TNF from 2001-2013, and two subgroups extracted from this dataset; one covering 2001-2010, and

one covering the period 2010-2013. This will enable comparison of the effect of different variables, and how they may have changed over time.

Gender	Age
BMI	Baseline PGA
Tender joint count	ESR
Baseline HAQ	DAS28-ESR
Disease duration	Year starting on biologic
Swollen joint count	Swollen:tender joint count ratio (S:TJR)
Anti-TNF drug type	Smoking

Table 32. Candidate variables used in full regression models

10.4.4 Collinearity

As with analysis undertaken in Chapter 9, collinearity is likely to have a significant impact in the analysis of the specified candidate variables. Therefore, in addition to the full list of candidate variables (Table 32), the most stable model of variables identified using VIF and stepwise regression in the full cohort (2001 – 2013) and each subgroup (2001 – 2010 and 2010 – 2013) from the analysis in Chapter 9 will also be used (summarised in Table 33). Using the same variables will also allow comparison between results from this and the previous chapter.

	Whole cohort (2001 – 2013)	**2001 – 2010 subgroup**	**2010 – 2013 subgroup**
Variables in model	Gender HAQ BMI Tender Joint Count Swollen Joint Count PGA ESR Age at starting anti-TNF Smoking status Anti-TNF choice Year starting biologic	Gender HAQ BMI Swollen Joint Count PGA ESR Age at starting anti-TNF Smoking status Anti-TNF choice Year starting biologic	HAQ BMI S:TJR PGA Disease duration ESR

Table 33. Candidate variables used in reduced regression models.

This analysis will explore if two, three or four distinct trajectories of response can be identified from the data. For analyses where there are more than two classes, multinomial logistic regression is used (explained in Chapter 5; 5.3.1).

10.4.5 Time-lapse standardisation

Because individuals have been recruited to the BSRBR-RA over the past 16 years, it is necessary to standardise the time elapsed since the baseline visit, rather than using actual date. This is calculated by subtracting the date of completion of the baseline enrolment form from the date of completion of the respective follow-up forms, to give the number of days elapsed since enrolment.

10.4.6 Latent class mixed modelling

LCMM will be used to identify and model different response trajectories to patients starting their first anti-TNF. LCMM is described in Chapter 5. Analysis will use the LCMM package in R and will include two, three and four classes. Data will be examined by graphical plotting and the lowest Bayesian information criterion (BIC; Chapter 5; 5.4.4) will be used to select the best-fitting class model for each specific dataset. A logistic or multinomial regression model will then be used to identify associations with the previously outlined variables (Table 32 and Table 33). For this analysis, baseline and longitudinal datasets have been merged, so a total of five multiply-imputed datasets will be used. As in Chapter 9, regression analysis will be undertaken on each imputed dataset, and estimates combined using Rubin's rules (Chapter 5; 5.1.5.1) to give an accurate quantification of associations and uncertainties related to the estimates.

10.5 Results

10.5.1 Initial trajectory analyses

After application of the LCMM package to the data, it is important to visualise the results of the LCMM analysis to understand how the different number of classes fit the dataset (two, three or four), prior to undertaking regression modelling. The process by

which the mean trajectories of response are generated is described in Chapter 5 (5.5). Visualising the results of the LCMM analysis is essential to ensure that class allocations are uniform across the different imputed datasets (i.e. is the label 'class 1' always assigned to the best performing trajectory?), and to visually inspect the stability of the trajectories (are they smooth, with small confidence intervals or unstable with wide overlapping confidence intervals?).

Because it is not possible to combine all the imputed data together into one plot, plots from one representative imputed dataset are shown here, although they were all examined individually to ensure uniform class allocation prior to regression. As previously mentioned, the dataset has been split into the three different groups (whole cohort, 2001 – 2010 subgroup and 2010 – 2013 subgroup) and two, three and four-class models are shown (Figure 24). Visual inspection of the trajectories shows significantly different trajectories of response emerging for all three models.

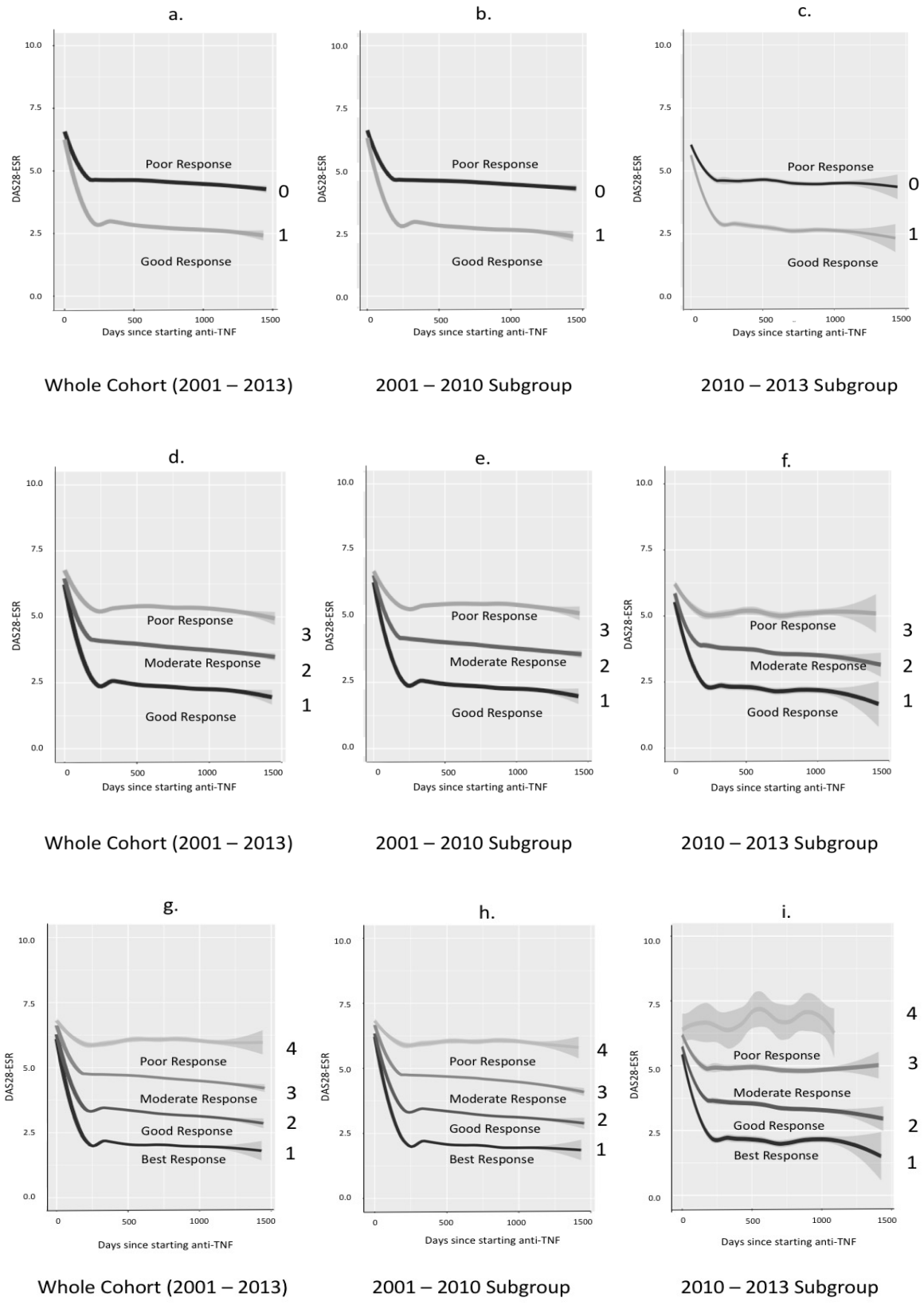


Figure 24. Mean group trajectories identified within one representative imputed dataset. Numbers indicate class trajectory. The dark - light lines (labelled 'best', 'good', 'moderate' and 'poor' respectively) show the mean response trajectories identified for each respective cohort/subgroup. Grey shading indicates 95% CI

Inspection of the data (after application of LCMM, but prior to regression modelling) clearly shows that there appear to be different trajectories of response within the whole cohort and different subgroups of the BSRBR-RA. It is possible to see that the baseline disease activity has reduced between the 2001-2010 and 2001-2013 subgroups for both trajectories in the two-class model (reduced y-intercept; Figure 24b, c), in keeping with results presented in Chapter 9 which showed a significant reduction in baseline disease activity (Table 23).

In the 2010-2013 subgroup, it appears that those individuals with a lower baseline disease activity score at the point of starting anti-TNF have a greater chance of achieving sustained good response, with good and poor response trajectories separated at baseline (reduced y-intercept for good response trajectory; Figure 24c, f and i). The greatest fall in disease activity occurs in the first six months from starting anti-TNF, for all trajectory analyses and across all datasets examined (Figure 24). It also appears that the differences in response are evident very early, possibly before six months. However, it should be noted that according to the study protocol, the BSRBR-RA data collection points are at baseline and six months (not between these times). This early separation may be due to some data that may have been collected earlier than six months, but may also be because of the smoothing function of the statistical analysis software and line of best fit.

What is evident however, is that the six-month response appears to be indicative of long-term response to the drug, and if remission or near-remission is not achieved by six to nine months, the data suggest that it is unlikely to subsequently occur within the first three years of anti-TNF use.

10.5.2 LCMM class size selection

The choice of which LCMM to best represent the groupings and their temporal patterns is made using the Bayesian Information Criterion (BIC; described in Chapter 5; 5.4.4). Because analysis is undertaken on five imputed datasets for the whole cohort and each

subgroup, the mean BIC from the five imputed datasets is shown (Table 34). The BIC is used to identify the optimally fitting model. The lowest BIC indicates the optimal balance between having the smallest number of classes and the model which best fits the data.

Mean BIC	2 Classes	3 Classes	4 Classes
Whole cohort	219136	219063	**219032**
2001-2010 subgroup	200511	200433	**200403**
2010-2013 subgroup	**18585**	18600	18617

Table 34. Comparison of model fit (using mean BIC across all multiply imputed datasets) for a selection of LCMMs with different numbers of classes

LCMM analysis for two, three and four-class analysis on the whole cohort and the two previously defined subgroups (2001-2010 and 2010-2013), shows the lowest BIC are seen for a four-class model for the whole cohort and 2001-2010 subgroup, and for a two-class model for the 2010-2013 subgroup. Graphical mapping of the different LCMM models in one of the imputed datasets (Figure 24) shows increasingly wide confidence intervals in the 2010-2013 subgroup as the number of classes increase, particularly in the no-response trajectory (latent class 4; Figure 24i), whereas tight confidence intervals are maintained in up to 4 classes in the whole cohort analysis and 2001-2010 subgroup (Figure 24).

10.5.3 Two class LCMMs

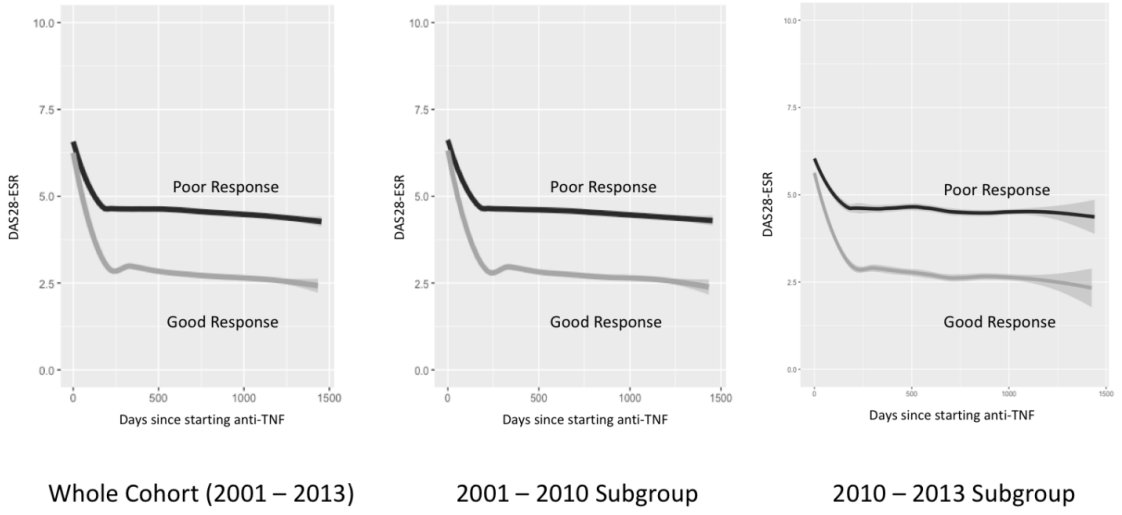


Figure 25. Mean response trajectories for a selection of two-class models: whole cohort, 2001 – 2010 and 2010 – 2013 subgroups. The light and dark lines (labelled 'good' and 'poor' respectively) show the mean response trajectories identified for each respective cohort/subgroup. Grey shading represents 95% CI

From the mean trajectories identified (Figure 25), it is evident that two trajectories of response to anti-TNF emerge in the whole cohort and two subgroups. Examining the class sizes associated with each trajectory, the proportion of patients achieving a long-term good response trajectory has increased significantly over the duration of BSRBR-RA (Table 35), similar to findings in Chapter 9 (Table 24).

Dataset	Poor response trajectory n (%)	Good response trajectory n (%)
Whole cohort (2001 – 2013)	9584 (66.4)	4852 (33.6)
2001 - 2010 subgroup	8733 (66.6)	4382 (33.4)
2010 -2013 subgroup	819 (62.0)	502 (38.0)
Chi^2 comparing 2001 – 2010 and 2010 – 2013 subgroups		<0.0001

Table 35. Change in good response category over time

In addition to the BIC, the posterior probability of misclassification can be examined to understand how well the data fit the two-class model. This quantifies the probability that the selected class for a patient is correct, as well as the probability that a patient has been allocated to the wrong class (described in Chapter 5; 5.4.3).

Response Trajectory	Poor	Good
Whole cohort (2001 – 2013)		
Poor	0.78	0.22
Good	0.22	0.78
2001-2010 cohort		
Poor	0.75	0.25
Good	0.21	0.79
2010-2013 cohort		
Poor	0.78	0.22
Good	0.22	0.78

Table 36. Posterior probability specificity associated with classification to the good and poor response trajectories for a two-class LCMM

As with the BIC, the mean posterior probabilities across all five imputed datasets are shown (Table 36). The data show that there is a good model specificity with 75 - 78% of individuals correctly classified to the correct trajectory and a correspondingly low probability of misclassification (Table 36).

10.5.3.1 Predictors of response for two classes

When applying the full set of candidate variables (Table 32) to the regression model for the whole cohort and the two subgroups, very few of the variables are significantly ($p \leq 0.5$) associated with differentiating between the two response trajectories. Table 37 summarises the significant predictors associated with achieving a good response trajectory when applying the full regression model (Table 32).

2001 – 2013 cohort – full regression model			
Variables	**OR**	**(95% CI)**	**P**
DAS28-ESR	0.69	(0.61 - 0.78)	<0.001
2001 – 2010 cohort – full model			
DAS28-ESR	0.67	(0.59 - 0.77)	<0.001
2010 - 2013 subgroup – full model			
Swollen joint count	0.95	(0.92 - 1.00)	0.03
PGA	0.99	(0.98 - 1.00)	0.02

Table 37. Significant ($p \leq 0.05$) predictors of response for two classes (full regression model)

Higher baseline DAS28-ESR scores (prior to anti-TNF) are associated with a reduced likelihood of achieving sustained good response for both the whole cohort (OR 0.69, 95% CI 0.61 – 0.78) and 2001-2010 subgroup (OR 0.67, 95% CI 0.59 – 0.77), a pattern that is reflected in the graphical plots of the data (Figure 25). However, this association is not maintained for the 2010-2013 subgroup. Instead, a higher swollen joint count and PGA is associated with a reduced likelihood of achieving sustained good response.

Summary of associations			
	Whole cohort	**2001 – 2010 subgroup**	**2010 – 2013 subgroup**
Good outcome			
Poor outcome	- High baseline DAS28-ESR	- High baseline DAS28-ESR	- High swollen joint count - Increasing PGA

Table 38. Summary of predictors associated with two-class trajectory model using the full regression model

When applying the variables selected from the reduced regression model (summarised in Table 33), the relationships identified are almost identical to those identified in the full regression model (Table 37).

Reduced regression model		
Variable	**OR (95% CI)**	**P**
Whole cohort (2001- 2013)		
DAS28-ESR	0.69 (0.61 - 0.77)	<0.001
2001- 2010 subgroup		
DAS28-ESR	0.75 (0.70 - 0.80)	<0.001
2010-2013 subgroup		
PGA	0.99 (0.98 - 0.99)	<0.001

Table 39. Predictors of response for two classes (reduced regression model)

Examining the whole cohort and the 2001 – 2010 subgroup, the relationships identified in the reduced regression model are maintained compared with the full regression model, although the effect of the baseline DAS28-ESR is reduced in the 2001 – 2010 subgroup. Only the baseline PGA is associated with sustained good response in the 2010 – 2013 subgroup reduced regression model, and the effect of the swollen joint count is lost as it was excluded following stepwise regression in Chapter 9 (9.6.3).

When examining the variables that were associated with a reduced likelihood of achieving sustained remission for the whole cohort and the 2001 – 2010 reduced regression model, the only variable identified in both LCMM and the previous chapter analysis, was the relationship between baseline DAS28-ESR and sustained remission for the 2001 – 2010 subgroup. Increasing HAQ and female gender, which were both associated with sustained remission in Chapter 9, are not identified as predictors of sustained good response according to the two-class LCMM.

The 2010-2013 subgroup LCMM analysis identified that a higher PGA and SJC were associated with a reduced likelihood of achieving sustained remission, relationships that were not identified in the sustained remission analysis. However, PGA was retained in the reduced regression model for sustained remission for the 2010 – 2013 subgroup after stepwise regression (Chapter 9, Table 27), although the relationship was not significant. The direction of association in the 2010 – 2013 subgroup LCMM analysis is opposite to that identified in the whole cohort and 2001 – 2010 subgroup analysis (Chapter 9, Table 27), with a higher PGA being associated with a reduced likelihood of achieving sustained good response.

However, the effect of increasing HAQ, S:TJR, disease duration and ESR, previously identified as being negatively associated with the likelihood of achieving sustained good response using the reduced regression model and the equivalent analysis in Chapter 9 (Table 26 and Table 27), were not identified in the respective LCMM analysis.

Summary of associations: sustained remission and two class LCMM						
	Whole cohort		**2001 – 2010 subgroup**		**2010 – 2013 subgroup**	
Variables in Model	Gender, HAQ, BMI, TJC, SJC, PGA, ESR, Age at starting anti-TNF, Smoking status, Anti-TNF choice, Year starting biologic		Gender, HAQ, BMI, SJC, PGA, ESR, Age at starting anti-TNF, Smoking status, Anti-TNF choice, Year starting biologic		HAQ, BMI, S:TJR, Disease duration, PGA, ESR	
	LCMM	**Threshold**	**LCMM**	**Threshold**	**LCMM**	**Threshold**
Associated with good outcome		• Ex-smoker status • Recent starting anti-TNF • ↑ SJC • ↑PGA • Humira™ use		• ↑ SJC • ↑ PGA • Ex-smoking • Never-smoking • Recent starting anti-TNF • Humira™ use		• ↑ S:TJR
Associated with poor outcome	• ↑ baseline DAS28	• Female Gender • ↑ HAQ • ↑BMI • ↑ ESR • ↑Age • Remicade™ use	• ↑ baseline DAS28	• Female Gender • ↑ HAQ • ↑ baseline DAS28 • ↑ BMI • ↑ ESR • ↑ Age • Remicade™ use	• ↑ PGA	• ↑ HAQ • ↑ Disease duration • ↑ ESR

Table 40. Comparison of associations between Chapters 8 and 9. Analysis using reduced regression models. SJC –swollen joint count, TJC – tender joint count, BMI – body mass index

10.5.4 Three class LCMM

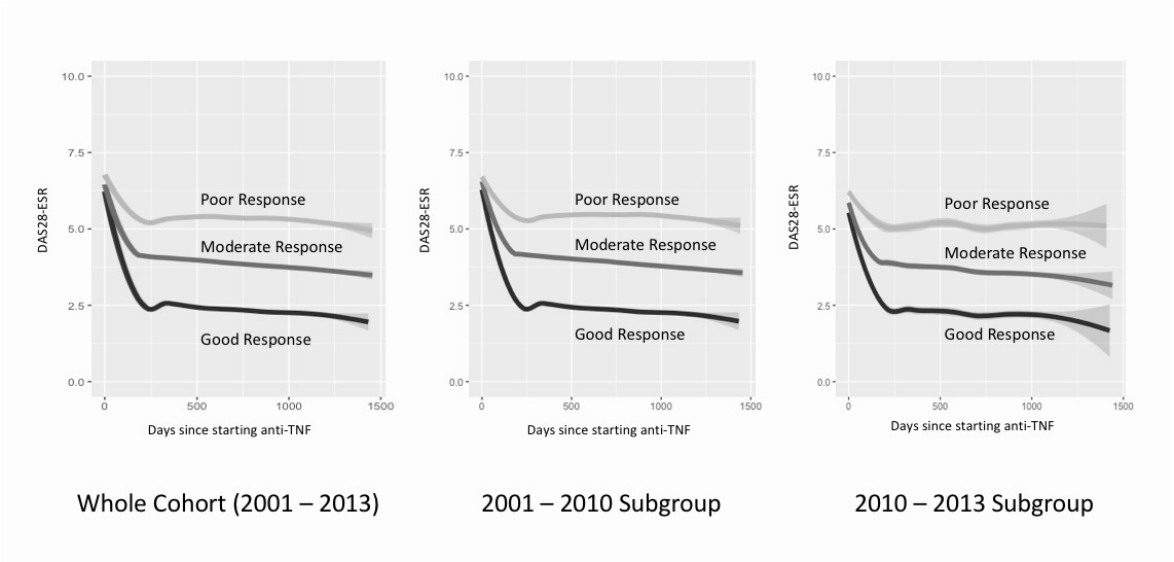


Figure 26. Mean response trajectories for a selection of three-class models: whole cohort, 2001 – 2010 and 2010 – 2013 subgroups. The dark - light lines (labelled 'good', 'moderate' and 'poor' respectively) show the mean response trajectories identified for each respective cohort/subgroup. Grey shading represents 95% CI

Examining a three-class LCMM shows an additional group of responses is identified, which increases model fit (defined by a reducing BIC) for the 2001-2013 and 2001-2010 datasets compared with a two-class model. There is small increase in mean BIC

for the 2010-2013 dataset (Table 34). However, confidence intervals around the mapped trajectories remain tight up to 1000 days (~3 years) after starting anti-TNF (Figure 26).

Comparing the response trajectories between the two- and three-class model, the good response trajectory maintains a very similar shape compared with the two-class model. However, what is apparent is that the poor response trajectory identified in the two-class model has split to reveal a moderate and poor response trajectory (Figure 26).

Dataset	Poor response, n (%)	Moderate response, n (%)	Good response, n (%)
Whole cohort	2709 (18.8)	9264 (64.2)	2463 (17.1)
2001 - 2010	2112 (16.1)	8749 (66.7)	2255 (17.2)
2010-2013	336 (25.4)	730 (55.2)	255 (19.3)

Table 41. Change in responses over time for three-class model

Comparison of the proportion of each cohort in each class over time shows that the majority of patients taking anti-TNF appear to fall into the long-term moderate response category (Table 41). Of note, there appears to be an increase in poor responders in the 2010-2013 cohort. However, this may be because of the increased uncertainty in the model and the increased chance of misclassification seen in the three-class model for the 2010 – 2013 subgroup (Table 42). The posterior probability of correct classification for the moderate response class in the 2010 – 2013 subgroup is 58% compared with 65% for the 2001 – 2010 subgroup. This increased uncertainty in the 2010 – 2013 subgroup is quantified by the increased BIC compared with the two-class model for the subgroup.

Response trajectory	Poor	Moderate	Good
Whole cohort (2001 – 2013)			
Poor	0.71	0.28	0.01
Moderate	0.22	0.63	0.15
Good response	0.01	0.23	0.77
2001 – 2010 subgroup			
Poor response	0.72	0.28	0.01
Moderate response	0.21	0.65	0.14
Good response	0.01	0.22	0.77
2010 – 2013 subgroup			
Poor response	0.69	0.28	0.03
Moderate response	0.24	0.58	0.18
Good response	0.01	0.24	0.75

Table 42. Posterior probability classification specificity for three-class LCMM

10.5.4.1 Predictors of response for three classes

Comparing the whole cohort and 2001 – 2010 subgroup, the DAS28-ESR is negatively associated with the likelihood of achieving the best response in both the full and reduced regression model (Table 43). A similar pattern with regards to the relationship with the DAS28-ESR and its components is seen in the differentiation between a moderate and poor response for the whole cohort analysis. However, the relationship between the baseline DAS28-ESR and response is not seen in the 2010 – 2013 subgroup although an independent effect of ESR identified. A higher baseline ESR is associated with a reduced likelihood of achieving in a moderate response trajectory for both full and reduced regression models (OR 0.998 for the full regression model), although rounding demonstrates an OR of 1.00 (95% CI 0.99 - 1.00).

There appears to be a calendar-year effect in both regression models for the whole cohort analysis (Table 43) and a more recent year of starting anti-TNF is associated with an increased likelihood of achieving sustained remission, for both best and moderate response trajectories. The choice of anti-TNF agent (Humira™) appears to be associated with achievement of a moderate, but not best response. Longer disease duration before starting anti-TNF is associated with a reduced likelihood of achieving

a sustained best response. This relationship is not seen in the reduced regression model as the variable is not included in the reduced regression model. However, an increasing swollen joint count is associated with a reduced likelihood of achieving sustained moderate response compared in the reduced regression model, a relationship that was not identified in the full regression model.

Whole cohort (2001 – 2013) full regression model				
Variables	**Best vs poor response**		**Moderate vs poor response**	
	OR (95% CI)	**P**	**OR (95% CI)**	**P**
DAS28-ESR	0.63 (0.52 - 0.77)	≤0.001	0.83 (0.71 - 0.99)	0.03
Disease duration	0.99 (0.99 - 1.00)	0.03	1.00 (0.99 - 1.00)	0.31
ESR	1.00 (0.99 - 1.00)	0.35	1.00 (0.99 - 1.00)	≤0.001
Humira™	1.06 (0.92 - 1.23)	0.39	1.19 (1.07 - 1.33)	≤0.001
Year starting biologic	1.05 (1.02 - 1.07)	0.00	1.02 (1.00 - 1.04)	0.07
Whole cohort (2001 – 2013) reduced regression model. Variables included in model: gender, HAQ, DAS28-ESR, BMI, TJC, SJC, PGA, ESR, anti-TNF agent, year starting biologic.				
DAS28-ESR	0.63 (0.52 - 0.77)	≤0.001	0.83 (0.70 - 0.98)	0.03
SJC	0.99 (0.98 - 1.01)	0.29	0.99 (0.98 - 1.00)	0.02
ESR	1.00 (0.99 - 1.00)	0.40	1.00 (0.99 - 1.00)	≤0.001
Humira™	1.07 (0.92 - 1.23)	0.38	1.19 (1.07 - 1.33)	≤0.001
Year starting biologic	1.05 (1.02 - 1.08)	≤0.001	1.02 (1.00 - 1.04)	0.06

Table 43. Predictors of response for three classes (whole cohort 2001 - 2013)

Examining the 2001 – 2010 subgroup (Table 44), similar associations are identified compared with the whole cohort analysis. An increasing baseline DAS28-ESR is associated with a reduced likelihood of achieving a best response in both the full and reduced regression model, and an increasing baseline DAS28-ESR is also negatively associated with the likelihood of achieving a moderate response compared with a poor response in the reduced regression model. As with the whole cohort analysis, there is a calendar-year effect associated with both trajectories of response. Unlike previous analyses however, increasing BMI appears to increase the likelihood of achieving the best response for both the full and reduced regression model, although this relationship only just meets the significance threshold for the full regression model. Ex-smoking status is associated with an increased likelihood of achieving a sustained best response for both the full and reduced regression models. An increasing age when starting anti-TNF is associated with a reduced likelihood of achieving sustained best response in the reduced regression model, although this is not identified in the full model.

2001 – 2010 subgroup full regression model				
Variables	**Best vs poor response**		**Moderate vs poor response**	
	OR (95% CI)	**P**	**OR (95% CI)**	**P**
DAS28-ESR	0.63 (0.50 - 0.79)	≤0.001	0.84 (0.69 - 1.02)	0.08
BMI	1.01 (1.00 - 1.02)	0.05	1.00 (1.00 - 1.01)	0.32
Smoking (ex vs current)	1.20 (1.02 - 1.41)	0.03	1.12 (0.98 - 1.27)	0.09
Year when starting anti-TNF	1.10 (1.05 - 1.15)	0.00	1.08 (1.04 - 1.12)	≤0.001
2001 – 2010 subgroup reduced regression model. Variables included in model: gender, HAQ, DAS28, BMI, SJC, PGA, ESR, smoking, age at starting biologic, anti-TNF choice, year starting biologic.				
DAS28-ESR	0.62 (0.55 - 0.69)	≤0.001	0.75 (0.68 - 0.82)	≤0.001
BMI	1.01 (1.00 - 1.02)	0.04	1.00 (1.00 - 1.01)	0.29
Smoking (ex vs current)	1.19 (1.01 - 1.40)	0.04	1.12 (0.98 - 1.27)	0.10
Age at starting biologic	0.99 (0.99 - 1.00)	0.04	1.00 (0.99 - 1.00)	0.10
Year when starting anti-TNF	1.10 (1.05 – 1.15)	≤0.001	1.08 (1.04 – 1.12)	≤0.001

Table 44. Predictors of response for three classes (2001 - 2010 subgroup)

As expected, the 2010 – 2013 subgroup analysis shows fewer associations than the other two datasets (Table 45). However, female gender appears to be associated with an increased likelihood of achievement of moderate response compared with a poor outcome, but has no association with the best response trajectory. The relationship with the baseline HAQ in the reduced regression model is in keeping with previous findings (Chapters 7 and 9) and suggests that an increasing HAQ is associated with a reduced likelihood of achieving the best response trajectory in the reduced regression model.

An increasing swollen joint count is associated with a reduced likelihood of achieving sustained best response in the full model, although the variable is excluded from the reduced regression model. An increasing PGA is also identified as being associated with a reduced likelihood of achieving sustained moderate response in the full, but not reduced regression model. The direction of relationship of both the swollen joint count and PGA is opposite to the response identified in the whole cohort and 2001 – 2010 subgroup analysis in chapter 9 for the same variables. Finally, as with previous analyses, an increasing baseline ESR is associated with a reduced likelihood of achieving sustained best and moderate response in the reduced regression model.

2010 – 2013 subgroup full regression model				
Variables	**Best vs poor response**		**Moderate vs poor response**	
	OR (95% CI)	**P**	**OR (95% CI)**	**P**
Gender	1.10 (0.74 - 1.64)	0.63	1.49 (1.07 - 2.07)	0.02
SJC	0.90 (0.85 - 0.96)	≤0.001	0.93 (0.89 - 0.98)	≤0.001
PGA	0.99 (0.98 - 1.00)	0.08	0.99 (0.98 - 1.00)	0.05
2010 – 2013 subgroup reduced regression model. Variables included in model: HAQ, BMI, S:TJR, disease duration, PGA, ESR.				
HAQ	0.71 (0.56 - 0.90)	≤0.001	0.86 (0.71 - 1.05)	0.15
ESR	0.99 (0.98 - 1.00)	0.01	0.99 (0.99 - 1.00)	0.01

Table 45. Predictors of response for three classes (2010 - 2013 subgroup)

10.5.5 Four class LCMM

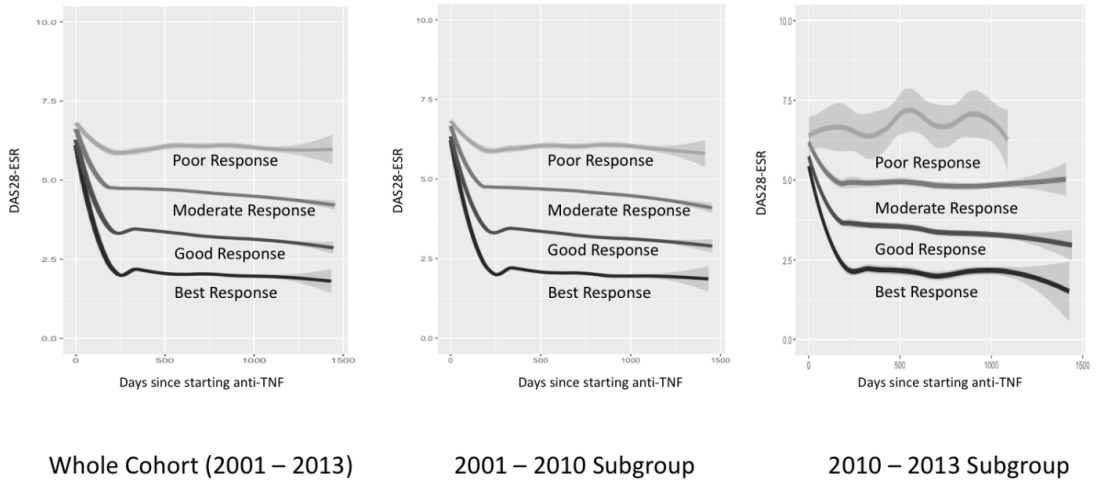


Figure 27. Mean response trajectories for a selection of four-class models: whole cohort, 2001 – 2010 and 2010 – 2013 subgroups. The dark - light lines (labelled 'best', 'good', 'moderate' and 'poor' respectively) show the mean response trajectories identified for each respective cohort/subgroup. Grey shading represents 95% confidence intervals

Examination of the whole cohort and the 2001-2010 subgroup shows four different trajectories of response to treatment emerge (Figure 27), and offers the best model fit for these two datasets, with the lowest BIC (Table 34). Increasing the number of classes from three to four has revealed a class of patients that appear to have a very minimal response (poor response trajectory). Whilst the four-class model does not represent the best-fitting model for the 2010-2013 subgroup, comparisons of class allocation

show a trend towards increasing proportions of individuals achieving sustained good responses over time, with an increasing proportion of moderate and best responders (Table 46). However, it should be noted that the worst response group is very small for the 2010 - 2013 subgroup with only 15 individuals included in it.

Dataset	Poor response, n (%)	Moderate response, n (%)	Good response, n (%)	Best response, n (%)
Whole cohort	465 (3.2)	7935 (55.0)	4792 (33.2)	1244 (8.6)
2001 – 2010 subgroup	517 (3.9)	7388 (56.3)	4163 (31.7)	1047 (8.0)
2010 – 2013 subgroup	15 (1.1)	556 (42.1)	485 (36.7)	265 (20.0)

Table 46. Change in response categories over time (4-class LCMM)

The probability of class misclassification between the different classes remains low for the whole cohort (Table 47) and 2001-2010 subgroup (Table 48), although classification specificity does reduce for the moderate and good response trajectories to 61-64%, suggesting a degree of overlap between these classes.

Whole cohort (2001 – 2013) Class specification (four classes)				
	Poor response	**Moderate response**	**Good response**	**Best response**
Poor response	0.75	0.24	0.01	0.00
Moderate response	0.15	0.64	0.20	0.01
Good response	0.03	0.23	0.62	0.12
Best response	0.00	0.00	0.28	0.72

Table 47. Posterior probability of misclassification for four-class LCMM (whole cohort)

2001 – 2010 Subgroup Class Specification (four classes)				
	Poor response	**Moderate response**	**Good response**	**Best response**
Poor response	0.75	0.25	0.01	0.00
Moderate response	0.15	0.64	0.21	0.01
Good response	0.03	0.23	0.61	0.14
Best response	0.00	0.00	0.28	0.71

Table 48. Posterior probability of misclassification for four-Class LCMM (2001-2010 subgroup)

10.5.5.1 Predictors of response 4 class LCMM

Response trajectory comparison	Best vs worst		Good vs worst		Moderate vs worst	
Variables	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**
	Whole cohort (2001-2013) full model					
DAS28-ESR	0.42 (0.27 - 0.65)	≤0.001	0.50 (0.33 - 0.75)	≤0.001	0.74 (0.49 - 1.12)	0.15
Humira™	1.25 (0.93 - 1.68)	0.14	1.43 (1.09 - 1.87)	0.01	1.49 (1.14 - 1.95)	≤0.001
Year starting anti-TNF	1.07 (1.01 - 1.14)	0.02	1.04 (0.99 - 1.10)	0.13	1.04 (0.98 - 1.10)	0.17
	Whole cohort (2001-2013) reduced model. Variables included in model: gender, HAQ, DAS28-ESR, BMI, TJC, SJC, PGA, ESR, smoking, age starting anti-TNF, anti-TNF agent, year starting anti-TNF					
DAS28-ESR	0.42 (0.27 - 0.65)	≤0.001	0.49 (0.33 - 0.75)	≤0.001	0.74 (0.49 - 1.11)	0.15
Humira™	1.25 (0.93 - 1.68)	0.14	1.43 (1.09 - 1.87)	0.01	1.49 (1.14 - 1.95)	≤0.001
Year starting anti-TNF	1.08 (1.01 - 1.14)	0.01	1.04 (0.99 - 1.10)	0.11	1.04 (0.98 - 1.10)	0.16

Table 49. Predictors of response for four classes (whole cohort 2001 - 2013)

Comparing the full and reduced variable models for the four-class LCMM when examining the whole cohort (Table 49), the predictors and magnitude of associations are almost identical. A higher baseline DAS28-ESR appears to be negatively associated with the likelihood of achieving the best and good response. The DAS28-ESR does not appear to differentiate between the poor and moderate response trajectories in the whole cohort full model.

The use of Humira™ is associated with differentiating between good and worst, and moderate and worst trajectory, but not with the likelihood of achieving a sustained best response in either the full or reduced regression model. A more recent year of starting anti-TNF is associated with an increased likelihood of achieving the best response (OR 1.07), but not with other trajectories.

Response trajectory comparison	Best vs poor		Good vs poor		Moderate vs poor	
Variables	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**
	2001 - 2010 subgroup full model					
DAS28-ESR	0.41 (0.25 - 0.65)	≤0.001	0.47 (0.30 - 0.73)	≤0.001	0.71 (0.46 - 1.10)	0.13
Year starting anti-TNF	1.13 (1.04 - 1.23)	≤0.001	1.10 (1.02 - 1.19)	0.01	1.09 (1.01 - 1.18)	0.02
	2001 - 2010 subgroup reduced model. Variables included in model: gender, HAQ, DAS28-ESR, BMI, SJC, PGA, ESR, smoking, age at starting biologic, anti-TNF choice, year starting biologic					
DAS28-ESR	0.56 (0.46 - 0.69)	≤0.001	0.64 (0.53 - 0.78)	≤0.001	0.88 (0.73 - 1.06)	0.19
Year starting anti-TNF	1.14 (1.04 - 1.24)	≤0.001	1.11 (1.03 - 1.20)	0.01	1.10 (1.02 - 1.18)	0.02

Table 50. Predictors of response for four classes (2001 - 2010 subgroup)

Undertaking the same analysis in the 2001-2010 subgroup (Table 50), the negative association of an increased baseline DAS28-ESR on the likelihood of achieving sustained best or good response is maintained in the both regression models. Anti-TNF choice does not appear to have any association with response in the 2001-2010 subgroup.

In addition, a calendar-year effect (year of starting anti-TNF variable) is also identified in the full and reduced model analysis of the 2001-2010 subgroup.

Given the very small group sizes, and higher BIC value in the four-class LCMM model for the 2010-2013, associations have not been presented as any associations identified are likely to be unreliable.

10.6 Chapter discussion

This analysis has revealed insights into how patients respond to anti-TNF over time, as well as the possible predictors that are associated with each trajectory. It is clear from the analysis that there are distinct trajectories of response to anti-TNF, and these trajectories disperse very early during treatment, with the majority of improvement in disease activity seen by six months. The fact that the trajectories diverge so clearly by six months suggest that it may be possible to identify those patients on different

trajectories earlier than six months, and the graphical representations of the different trajectories of response suggest that this may occur as early three months (Figure 24). However, as previously mentioned, there are sparse data before the six-month point, so the early separation of the different trajectories may be a function of smoothing of trajectories used by the statistical software. Evidence suggests that response to anti-TNF can be seen earlier than six months (178,252) and this data supports the clinical practice of assessing response at three and six months. Earlier identification of likely response to anti-TNF treatment would expedite clinical decision making about the chance of future success of continuing treatment.

Another interesting finding from this analysis is the absence of any 'U-shaped' response trajectories representing a secondary non-response trajectory. There is evidence that has demonstrated that secondary non-response occurs not infrequently (253). However, there does not appear to be such a trajectory identified in this analysis of the BSRBR-RA. There are several conceivable explanations for this. It is possible that such secondary non-response occurs insufficiently frequently to generate a distinct response trajectory that is identifiable in the LCMM analysis. It may also be that a six-month data collection frequency is insufficient to identify such a secondary non-response, and patients may lose response to anti-TNF, and switch to a different drug between the six-monthly follow-ups, resulting in these data being censored from this analysis. Examination of the different trajectory plots (Figure 24) does show a small rebound effect at approximately six months in the best response trajectories, and it is possible that this is generated by early secondary non-responders, who subsequently switch drug before the next data collection point. It also possible that there are more than four trajectories of response, and increasing the number of classes further may reveal a 'U-shaped' secondary non-response trajectory.

An alternative explanation is that because this analysis only included patients on their first anti-TNF, and did not include data from patients after a switch to a different anti-TNF, data which could form such a trajectory may have been excluded. Another possibility may be that most secondary non-response to anti-TNF occurs after three years of use, and hence would be outside the scope of this analysis.

When considering the poor response trajectories in these analyses, it may seem counterintuitive that such a trajectory would even exist, as one would assume that a patient and clinician would not continue with a treatment that is not demonstrating efficacy. This is particularly so in the 2010 – 2013 subgroup analysis, where there were multiple alternative therapeutic options that a non-responsive patient could try. The explanation for this is that the response trajectories presented are mean response trajectories for a group of patients. As such, at an individual level, patients classified on the poor response trajectory may have had some individual DAS28-ESR scores lower than the mean trajectory, but also may have some that were worse. Therefore, when all the scores over time are considered, the patient's mean trajectory would be poor. However, a clinician assessing a patient at a single point in time, may elect to continue with an anti-TNF despite a high previous DAS28-ESR if the current DAS28-ESR score is improving; or may consider previous single episodes of improved DAS28-ESR scores and may make a clinical decision that it would be worthwhile persisting with the current anti-TNF agent so see if a good response returns. The advantage of these LCMM analyses is that the probability of different future mean trajectories of response can be calculated, something that is challenging when only considering current or disease activity. This would be of use to assist clinicians when trying to decide on whether to continue or switch biologic in patients who have had a sporadic or intermediate response to anti-TNF.

There is a significant improvement in the proportion of patients achieving a sustained good response over time (Table 35), matching the changes in response observed in Chapter 9. This adds further support to the suggestion that earlier treatment, at lower DAS28-ESR scores and before development of significant disability, as evidenced by the changing demographics of the 2010-2013 subgroup (Chapter 9; Table 23) is leading to improved outcomes.

Compared with the analysis in Chapter 9, there are fewer predictors associated with sustained best and good response trajectories than when using a pre-determined threshold of remission or LDA. Of particular interest, is that female gender, which was strongly associated with a reduced likelihood of achieving sustained remission and LDA in Chapter 7 and Chapter 9, was not identified as a predictor of best response in

any of the LCMM analyses in this chapter. This may be because, as previously discussed, women tend to have higher ESR values compared to men (228), which may make them to be less likely to achieve sustained remission when using the DAS28-ESR as an outcome. Because the LCMM analysis does not use set thresholds, women with a sustained good response that may be just above the 2.6 threshold required for the DAS28 definition of remission, would still be grouped with a sustained good/best response trajectory, and this may explain why this variable was not identified in any of the analyses undertaken. Interestingly, female gender was identified as being associated with an increased likelihood of achieving sustained moderate response in the full regression model of the 2010 – 2013 subgroup.

The other surprising finding was that baseline function (measured by the HAQ), was only identified as being associated with a good response trajectory in the three-class 2010-2013 subgroup analysis using the reduced regression model (Table 45), where the association was reduced compared with the equivalent analysis in Chapter 9 (OR 0.71, 95% CI 0.56 – 0.90 and 0.53, 95% CI 0.44 – 0.64 respectively). The reason for this is not clear, as the HAQ was a variable that had the least collinearity in the analyses in Chapter 9, so is unlikely to be affected by the other covariates in the models. However, the baseline mean HAQ for the whole cohort and 2001 – 2010 subgroup was 2.0 (compared with 1.6 for the 2010 – 2013 subgroup) so it is possible that a ceiling effect may have occurred in the whole cohort and 2001 – 2010 subgroup analyses. This may have diminished the ability of the HAQ to be an effective discriminator of long-term response to anti-TNF in these analyses.

Comparing the associations identified in the two-class and four-class analysis for the whole cohort, the association of the baseline DAS28-ESR with the likelihood of achieving the best and good response compared with poor response (OR 0.42, 95% CI 0.27 – 0.65 and 0.50, 95% CI 0.33 – 0.75 respectively; Table 49) is greater than between good and poor trajectories identified in the two-class analysis (OR 0.69, 95%CI 0.61 – 0.78; Table 37), suggesting that in the four-class analysis, a more homogeneous population of patients who are achieving sustained remission is identified compared with the two-class analysis.

There is also evidence that the choice of anti-TNF may be associated with the long-term outcome. However, the relationship is identified in the four-class analysis of the whole cohort but not the 2001-2010 cohort. The relationship identified in the whole cohort four-class model was unusual. Unlike all the other predictors identified, the association appears to be strongest in differentiating between the moderate and poor response trajectories rather than between best and worst trajectories (Table 49), suggesting that there may be confounding by indication occurring with anti-TNF choice.

10.6.1 Which model to use?

From the data, it is apparent that there is no single trajectory model that fits both the whole cohort and both subgroups perfectly. This leads to the question: which model should be used? Using the best-fitting model for the whole cohort and 2001 – 2010 (four classes) is unreliable when considering the 2010 – 2013 subgroup (given the increased BIC, poorer posterior probabilities and evident instabilities seen on graphical mapping of the data). However, using a two-class model would not fit the whole cohort or 2001 – 2010 subgroups optimally (although it would fit the 2010 – 2013 subgroup well) and using three classes does not represent optimal fit for any of the datasets.

In this situation, it is helpful to take a pragmatic approach and think which model would be the most clinically useful. From a clinical standpoint, if the data were to be used in a clinical decision-support tool, the two-class model is the most useful, as it is useful to know if a patient is likely to respond optimally to a drug or not. Having additional 'in-between' response classes is less helpful in assisting clinical decision making than a binary outcome. Furthermore, if a clinician was considering using these findings to assist in making clinical decisions, it would be most appropriate to use data from a subgroup that is most similar to the patient being treated in clinic today (i.e. the 2010 – 2013 subgroup), rather than more historical cohorts (i.e. the 2001 – 2010 subgroup). When considering the 2010 – 2013 subgroup, the three- and four-class models were less stable compared to the two-class model.

However, if a historical description of responses achieved since anti-TNF was introduced (i.e. using the whole cohort 2001 – 2013 data), then a four-class model for the whole cohort would be an appropriate model.

The other question to address when considering which model to use is whoch regression model to use? Should the 'full' regression model (which include all *a priori* variables) or the 'reduced' regression model (using the reduced set of variables identified by the stepwise regression modelling identified in Chapter 9) be used? Examination of the data (in both Chapter 9 and 10) suggest that there is actually very little difference between the results of the two models, and the directions of associations identified by reduced regression models are similar to the full regression models, lending support to the associations identified by the latter and suggests that the associations identified by the full regression model are stable and not dramatically altered in magnitude or direction by minor changes in the variables included.

Because it is not possible to undertake stepwise regression on the higher-order multinomial regression models in this chapter, the variables identified by stepwise analysis were selected using the analysis from Chapter 9 which is not optimal. Therefore, whilst it was helpful to undertake analysis using the reduced regression models (to allow assessment of stability of associations within the full regression model, as well as allowing cross comparison with results from Chapter 9), the most appropriate regression model to select when examining possible predictors of sustained good response is the full regression model with all *a priori* variables included.

10.6.2 Strengths and weaknesses of this analysis

The strengths of this analysis are that it is data-driven, using clinical data to identify 'naturally' occurring trajectories within the data. By using registry data which includes data from across the whole of the UK, the response data represents a broad picture of clinical practice not only across the country, but also over time. The variables chosen for this analysis were selected based on evidence (Chapter 7), and responses were

mapped in real-time. The analysis used clearly defined methods to identify the most stable models based on multiple parameters (BIC, posterior probabilities and visual inspection of trajectories). As with Chapter 9, missing data were thoroughly investigated and imputation appropriately applied. Attempts were also made to minimise collinearity by using previously selected variables chosen by stepwise regression from Chapter 9.

However, there are also weakness. As discussed previously, the decision to exclude patients' data from the point of switching anti-TNF or to a different biologic, may have influenced the trajectories identified, and possibly made the identification of a U-shaped secondary non-response trajectory less likely. It is also possible that the six-monthly frequency of the data collection undertaken in the BSRBR-RA is not sufficient to identify all possible trajectories of response. However, collecting data more frequently is not possible without construction of a new cohort study which is impractical. An additional point to note is that although the follow-ups were limited to the first six (when data collection is undertaken at six monthly intervals), the actual number of days elapsed since registering extended up to nearly 1500 days (four years) for the last follow-up. This is because no time limit is imposed by the BSRBR-RA registry for delayed follow-ups. It is possible that this might have influenced results, although it seems unlikely that predictors of sustained response would be significantly different between three and four years. It does however highlight the importance of not enforcing externally defined follow-up time windows in the analyses of sustained remission and sustained LDA (Chapter 9; 9.5).

Another possible weakness in this analysis is that in exploring the relationships between many classes and in two sub-groups, multiple relationships have been explored, which poses a potential risk of multiple testing and type one error. However, the analyses and variables were specified *a priori* and were based on existing evidence.

A further possible weakness is the absence of specific co-morbidity variables from the regression models. It is possible that other underlying health conditions (such as heart disease, diabetes, depression etc.) may be associated with sustained remission. However, conditions such as depression, heart disease and diabetes have a very broad

spectrum of severity (e.g. diet controlled diabetes vs. recalcitrant diabetes requiring multiple medications and subcutaneous insulin), and it would be very challenging to evaluate with certainty the strength of association between even one co-morbidity and sustained remission without exceptionally granular detail on the underlying co-morbidity. The HAQ could be viewed as a global surrogate for 'other health conditions' that might result in functional impairment (e.g. heart disease may impair an individual's ability to undertake some of the activities of daily living as well as activity RA), and the association identified between HAQ and sustained remission in Chapter 9 could be viewed as evidence that this should be explored in more depth, and it is possible that predictors of response to anti-TNF may differ between subgroups of patients with different co-morbidities. Indeed, the relationship between co-morbidity and response to anti-TNF in individuals with RA could justify a further independent body of work in its own right. However, co-morbidity data were not included in analyses undertaken in this thesis for two reasons: Firstly, no specific co-morbidities were identified as being associated with sustained remission in the systematic review (Chapter 7) which was used as the mechanism for identifying the variables to be evaluated in this work; and secondly, co-morbidity data were not included in the BSRBR-RA dataset provided.

Finally, as mentioned previously, it is possible that the variables chosen for the reduced regression models (defined by the variables identified in Chapter 9) are not necessarily optimal for the multinomial regression models in this chapter. However, it is not possible to use stepwise regression models on a multinomial regression model, so it was not possible to generate specific optimised regression models for the LCMM analyses. Using the same variables between the analyses in chapter 9 and 10, has allowed comparison of results between analyses.

10.7 Conclusions

The analysis undertaken in this chapter has clearly identified that there are different response trajectories related to anti-TNF use, as well as identifying baseline variables that are associated with the attainment of these different trajectories. From a clinical

standpoint, the two-class model is probably the most helpful to aid decision-making as it gives a binary result (good/poor response), and the treatment trajectories and associations identified in this analysis could help clinicians identify likely good/poor response earlier in treatment. However, to accurately describe the population or patients taking anti-TNF enrolled on the BSRBR-RA, then the four-class model is more appropriate.

It is also clear that as well as improving outcomes, the variables which are associated with the different trajectories of response are changing over time, which is likely to be due to the changing population of patients that is being treated with anti-TNFs, and the changing clinical practice since anti-TNFs first became available in the early 2000s.

10.8 Key points from this chapter

- Different trajectories of response to anti-TNF are evident from the data in the BSRBR-RA.

- Attainment of sustained good/best response with anti-TNF is improving over time.

- Long term clinical response trajectories are evident following six months of treatment with anti-TNF.

- Clinical and demographic variables can help predict which response trajectory a patient is likely to achieve. Higher or greater baseline DAS28-ESR, swollen joint count, PGA, HAQ, disease duration, and age are associated with a reduced likelihood of achieving an optimal response. Starting anti-TNF more recently, stopping smoking, higher BMI, and Humira™ use are associated with an increased likelihood of achieving a better response to anti-TNF.

Chapter 11

11 Discussion

11.1 Thesis summary

The overall aim of this thesis was to investigate the frequency of sustained remission and LDA in the UK population of patients taking anti-TNF, and to establish if there were any clinical or demographic features that were associated with the attainment of sustained remission.

11.1.1 Establishing the evidence base

Before undertaking this work, it was necessary to establish the current evidence base for sustained remission, achieved by undertaking a systematic literature review of the subject (Chapter 7). Whilst not formally defined by ACR or EULAR, six months was selected as an appropriate length of time to define sustained remission. In the process of undertaking the systematic literature review, it became apparent that not only was the evidence for sustained remission scarce, but it was clouded by the multiple outcome measures in use. In particular, there was significant variation between versions of the DAS28 used. Therefore, prior to examining the frequency of sustained remission in the BSRBR-RA, it was necessary to quantify the potential discrepancy in scoring between the DAS28-ESR and DAS28-CRP (Chapter 8).

11.1.2 Quantifying DAS28 inter-score agreement

Analysis of paired DAS28-ESR and DAS28-CRP data obtained from the BSRBR-RA revealed that the differences between the two scores were indeed significant, and the discrepancy between the two scores varied according to age and gender, but not BMI.

This was a significant finding, as an inter-score discrepancy that varied dependant of demographic factures, had the potential to interfere with subsequent analysis on which clinical features may predict the attainment of sustained remission or LDA. The paired DAS28ESR and DAS28-CRP data available in the BSRBR-RA enabled development of a modified version of the DAS28-CRP (the mDAS28-CRP) that had improved agreement with the DAS28-ESR, whilst using the same components as the original score.

11.1.3 Investigating sustained remission and LDA

Findings from Chapters 7 and 8 informed the parameters for analyses undertaken in Chapter 9 and 10. In particular, the identification of significant inter-score differences between the DAS28-ESR and DAS28-CRP led to the decision to use one version of the DAS28 score in subsequent analysis (the DAS28-ESR). These final two results chapters sought to approach the problem of identifying and quantifying sustained remission from two perspectives: (1) using a pre-determined 'DAS28-ESR threshold' approach to define sustained remission, and (2) using a data-driven approach to identify sustained good response (and other) trajectories. *A priori* specified variables were used (identified from the systematic review undertaken in Chapter 7) to examine if any of the clinical and demographic data collected by the BSRBR-RA were associated with the likelihood of achieving sustained remission or sustained good response.

11.2 Principal findings

11.2.1 Sustained remission in the wider literature

11.2.1.1 Frequency

Examination of the frequency of sustained remission from the studies identified in Chapter 7 demonstrated that it occurred infrequently. With the exception of one study (223), the range of sustained remission was 4.5-15.8%, substantially lower than the

rate of point remission identified in RCTs (178). The study that was identified as having the greatest rate of sustained remission (223) (38.1%), whilst being observational in design, had a rigorous trial strategy, which may have influenced outcomes compared with less intensively managed registry and case-review observational studies which comprised the remaining studies.

11.2.1.2 Predictors

From the available evidence, female gender was strongly associated with a reduced likelihood of achieving sustained remission compared with males (meta-analysis OR 0.53, 95% CI 0.44 – 0.63). The reasons for this relationship were not identified in any of the studies. However, it was interesting to note that in one of the studies (221), the association was lost when an outcome measure that did not include the ESR (the CDAI) was investigated. It has been previously identified that healthy females have a higher ESR level than men (228), and it is possible that this may have influenced this relationship.

Other factors that appeared to negatively influence the likelihood of achieving sustained remission with anti-TNFs were: increasing age; longer disease duration; and higher HAQ. Methotrexate co-prescription appeared to improve the likelihood of achieving sustained remission. No objective clinical assessment parameter (such as the swollen joint count or inflammatory marker) was associated with the likelihood of achieving sustained remission. Possible explanations for this were not immediately apparent. However, these findings raised further questions including: Are composite outcome measures effective at measuring the anti-inflammatory effect of anti-TNFs? What is the best way to distinguish between the inflammatory and non-inflammatory components of the disease using a composite outcome measure such as the DAS28? And how much does confounding or collinearity between variables within a composite outcome measures influence the associations identified?

Another observation from the data extracted from the systematic review was that, in addition to the variation in the outcome measures used between the six included

studies, there was also variation in which version of DAS28 was used, and in some cases, the version used was not specified. Previous studies identified significant differences between the different versions of the DAS28 (155-157), therefore, it was possible that the inter-score differences between the different version of the DAS28 may have affected some studies ability to detect predictors of sustained remission.

11.2.2 Difficulties using the DAS28-ESR and –CRP

Because of the previously identified discrepancies between the two versions of the DAS28, it was appropriate to investigate this further. The BSRBR-RA does not specify which version of the DAS28 should be used in quantifying the disease activity on data collection pro-forma. As such, the DAS28 data recorded in the registry is a mix of both versions which have been recorded as the same outcome. However, in addition to the final composite score, the BSRBR-RA collects the component values of the score which allows calculation of composite scores using the raw component data. In some instances, both ESR and CRP were reported alongside the other components of the DAS28. This made it possible to calculate paired DAS28-ESR and DAS28-CRP scores and compare the scores generated.

Examination of the BSRBR-RA identified over 8000 individuals who had a total of 31,074 paired DAS28 scores. Comparisons between the two scores revealed significant differences between the two scores in line with previous research.

The discrepancy between the two scores was greatest for females, and older patients. This discrepancy between scores had a significant impact when applying the DAS28 thresholds for remission and LDA where agreement between the two scores was only 66.0% and 32.0% respectively.

Given that three of the four components between the two versions of the DAS28 were identical, it was possible to model the relationship between the paired ESR and CRP scores. A data driven approach, incorporating the use of non-linear modelling enabled quantification of this relationship, which subsequently enabled modification of the

DAS28-CRP to form a new mDAS28-CRP outcome measure. The mDAS28-CRP had greater agreement with the DAS28-ESR both at an individual score level (the mean difference was reduced from +0.3 to -0.17), and at a categorical level (remission and LDA agreement increased from 66.6% to 84.5% and 32.0% to 39.2% respectively).

Detailed consideration was given to the possibility of using the newly developed mDAS28-CRP with the DAS28-ESR in the subsequent analysis of sustained remission and LDA where the DAS28-ESR was not available, but ultimately it was not included in the subsequent analysis. The rationale for this was two-fold. Firstly, use of the mDAS28-CRP would not have solved the problem of missing data in the BSRBR-RA, meaning multiple imputation would still be required. Because the mDAS28-CRP and the DAS28-ESR have slightly different relationships with their respective component data (e.g. the tender and swollen joint count), imputing a combined DAS28 outcome (incorporating the mDAS28-CRP and DAS28-ESR) would have been less reliable than using the DAS28-ESR only, and the bootstrapped inference parameters (used to impute other non-DAS28 missing data) would have been different for the two scores. Secondly, because the mDAS28-CRP is a novel measure, it was prudent to use a more widely-accepted outcome measure (the DAS28-ESR) in subsequent analyses so that it could be readily understood by a wider research and clinical audience, without requiring readers to be familiar with the mDAS28-CRP.

The mDAS28-CRP was developed alongside the work undertaken in this thesis including an online calculator (available at: https://mdas28.shinyapps.io), and further validations studies are planned with collaborators in Canada and the Netherlands.

11.2.3 Sustained remission & LDA in the BSRBR-RA

Having identified significant potential issues with using the DAS28-ESR and DAS28-CRP interchangeably, the DAS28-ESR was chosen as the outcome measure for subsequent analysis. Results from the systematic review (Chapter 7) were used to guide the variables included in the regression model, and had highlighted the potential problems of confounding and collinearity. Therefore, in addition to undertaking

analysis using the full regression set of variables identified from Chapter 7, collinearity between variables in a regression model was examined and stepwise regression used to address this problem.

11.2.3.1 Frequency

Results showed that sustained remission and sustained LDA was uncommon in the dataset overall (just 14.9% and 26.3% respectively) and in line with results from the systematic review. However, it was evident that there had been significant improvement in these outcomes over time, with 21.6% and 32.3% of patients achieving sustained remission and LDA respectively in the 2010 – 2013 subgroup analysis. Detailed analysis of the make-up of individuals who achieved sustained LDA revealed that the majority either achieved sustained remission or remission at at least one point in time within the first three years of treatment with anti-TNF (Table 24).

11.2.3.2 Predictors

As expected, there was collinearity between the variables examined. However, predictors that were identified with sustained remission and LDA were identified. Female gender, increasing HAQ, BMI, ESR, Remicade™ use and older age at starting anti-TNF were all negatively associated with achieving sustained remission in the cohort when examined as a whole. Ex-smoker status, Humira™ use and increasing swollen joint count and PGA were all positively associated with the likelihood of achieving sustained remission. The subgroup analyses identified identical associations for the earlier subgroup (2001 – 2010), with the additional negative association of increasing DAS28-ESR at baseline with sustained remission, and positive association of non-smoker status vs. current smoking status and starting anti-TNF more recently. The 2010 -2013 subgroup only identified one negative association with sustained remission (the baseline HAQ). However, collinearity existed between the variables examined. Stepwise regression improved the model fit for all analyses. Post-stepwise regression, an additional three predictors were identified for the 2010 – 2013 subgroup; disease duration and ESR were associated with a reduced likelihood of

achieving sustained remission, and an increasing S:TJR was associated with an increasing likelihood of achieving sustained remission.

Stepwise regression improved model fit for both the whole cohort and both subgroups, but did not identify any additional predictors of sustained remission. Examination of predictors of sustained LDA demonstrated similar predictors as identified in the sustained remission analysis, although ORs were generally reduced, representing the more heterogeneous group of patients now included.

11.2.4 Response trajectories

11.2.4.1 Frequency

LCMM analysis reinforced the findings from Chapter 9, and demonstrated a significant improvement in the proportion of patients achieving a sustained good response over time, although as identified earlier, the proportion who achieve this optimal outcome remained in the minority (38%). Increasing the number of classes to three demonstrated that there was a large 'moderate response' group evident from the data for the whole cohort and two subgroup analyses. Increasing the number of classes to four was possible in the whole cohort and 2001 – 2010 subgroup, but was unstable in the 2010 – 2013 subgroup. The trajectory models used had a low posterior probability of misclassification for two-class models for all analysis (75-79% correct classification). The proportion of correct class allocation decreased as the number of classes increased, although the lowest level of correct classification was still 58% in the 2010 – 2013 subgroup analysis for three classes. Class misclassification was not examined for the four-class model of the 2010 – 2013 subgroup due to the poor model fit identified using BIC and visual examination (Figure 24).

11.2.4.2 Predictors

Similar predictors were identified for sustained good response as in Chapter 9, although generally fewer associations were identified using LCMM. One of the most surprising associations that was not identified was that female gender did not appear to be a predictor of response in any of the LCMM analyses, unlike results from the systematic review (Chapter 7) and Chapter 9. However, it does lend support to the hypothesis that an elevated baseline ESR in women compared to men may be confounding this relationship. Further work using the DAS28-CRP in place of DAS28-ESR would help identify if this relationship is independent of the ESR. In addition, HAQ did not appear to be associated with sustained good response in any of the LCMM analyses using the full variable regression models, and was only identified as being negatively associated with sustained good response in the reduced variable regression model for the three-class trajectory analysis for the 2010 - 2013 subgroup. The reason for the lack of association identified with the HAQ is not clear, although as described in Chapter 10, this may have been due to a 'ceiling-effect' associated with very high HAQ scores in the earlier subgroup. Increasing DAS28-ESR, or components of it (the swollen joint count and PGA) were associated with a reduced likelihood of achieving sustained good response in the whole cohort and both subgroup analyses in the two class analyses.

The choice of which of the different LCMMs was the most appropriate model to use was also discussed (Chapter 10; 10.6). The different class analyses offer an opportunity to map historical perspectives on how trajectories of response have changed over time, as well as identifying predictors that are associated with each outcome. However, from a clinical decision making standpoint, the two class model was deemed the most useful to help with the key decisions faced by clinicians, which are fundamentally binary in nature; 'should anti-TNF therapy be started or not?'; and 'should anti-TNF therapy be continued or not?'. Therefore, while there may be nuances in the make-up of the poor response class in a two-class model (which may contain a moderate response class); for a clinician following the current ACR/EULAR guidelines and aiming for remission (139), the two-class model is the most helpful.

Anti-TNF therapy has been available in the UK for over 15 years, so when considering a patient starting anti-TNF therapy today, it is less likely that they will have the clinical history associated with the patients who started anti-TNF when it first became available (e.g. chronic intractable disease, >3 failed synthetic DMARDS, high HAQ scores, etc.). Therefore, the 2010 – 2013 subgroup analysis is more likely to be representative of the patient population starting anti-TNF therapy in 2017, and gives a better indication of likely outcomes and predictors. Furthermore, the 2010 - 2013 subgroup three- and four-class LCMM analyses were less stable (by BIC, posterior probabilities and visual analysis), a further reason for selecting a two-class LCMM.

11.3 What do these findings add to the current evidence base?

11.3.1 The mDAS28-CRP

The work undertaken in this thesis adds to the current evidence base in a number of ways. Firstly, it reinforces the suggestion that the two versions of the DAS28 should not be used interchangeably. Evidence presented in this thesis shows that not only is there a discrepancy between the two scores, but that this also has a significant impact on which disease activity category the score is classified as. In a time where treatments are target driven, it essential to have a clearly defined and reliable target. Switching between using the two versions of the DAS28 is essentially 'moving the goalposts'. Furthermore, when making comparisons between trial outcomes from different studies, this evidence strongly suggests that results of studies using the same version of the DAS28 should be used for comparisons, and broad references to 'the DAS28' should be avoided for clarity. Whilst there have been moves to use a disease activity score with a more stringent definition of remission (170), within Europe, and in the UK in particular, the DAS28 remains the main outcome measure used in RA. As long as it remains the threshold used for access to biologic therapies, it is likely to continue to be the most widely used score. In addition, the huge wealth of existing data using the DAS28, will mean that the relevance of the score will continue for many years to come. The development of the mDAS28-CRP is a practical solution to the discrepancy between the two versions of the DAS28. While there remains a difference between the

mDAS28-CRP and DAS28-ESR, it is significantly reduced and has significantly better agreement than the original DAS28-CRP. However, further validation studies are required before such a modification to the score could be used widely.

11.3.2 Attainment of sustained remission and LDA

The examination of sustained remission and LDA (in Chapters 7, 9 and 10) adds to the existing evidence by demonstrating its relative infrequency, and the need for continued efforts to identify and improve the achievement of a sustainable optimal response to treatment in RA. The paucity of evidence in the systematic review, highlights the need for further investigation of this outcome, as well as a discussion about how best to incentivise investigation of such outcomes.

Whilst the attainment of sustained good response in patients starting anti-TNF for RA is infrequent, the analyses in Chapters 9 and 10 suggest that the achievement of these outcomes (including remission, LDA and 'good/best' trajectory) has improved significantly over time, possibly as a result of treating patients earlier and before significant disability develops. Supporting this, analysis of the changing demographics of patients included in the BSRBR-RA shows that patients starting anti-TNF in the 2010 – 2013 subgroup had significantly less disability, shorter disease durations and lower disease activity at baseline (Table 23).

The work undertaken to identify predictors of sustained remission and LDA has identified predictors that have a modifiable component (such as BMI and smoking), and non-modifiable components (including gender, baseline disease activity). Identification of modifiable risk factors could assist in encouraging clinicians to further encourage patients to make healthy life choices that may improve long-term outcomes for RA patients using anti-TNF therapies (and have other health benefits). However, it should be reiterated, that as the data used to undertake this analysis are observational, causality cannot be attributed to these factors. Even though causality has not been proven, recommending smoking cessation and weight loss in overweight/obese individuals is unlikely to be deleterious to a patient's health.

The LCMM analyses (Chapter 10) support the current practice of assessing anti-TNF response at six months, and demonstrate that response at six months after initiating anti-TNF therapy is a good indicator of longer-term outcomes. However, the LCMM analyses go further, and show that if a good/best response is not achieved by six months, it is unlikely that it will be achieved in the next three years. This could be of great use to clinicians in deciding what treatment decision to make at six months when following a treat-to-target strategy. If a patient has achieved a reduction in disease activity, but not LDA or remission by six months, the results presented in Chapter 10 suggest that subsequent attainment of either sustained LDA or remission is unlikely, and a clinician and patient should give serious consideration to modifying treatment.

11.4 Strengths and weaknesses

Specific strengths and weaknesses of each analysis are presented with the respective analyses, however, there are cross-cutting strengths and weaknesses that run through this thesis which are examined here.

The work undertaken in this thesis has a number of strengths. The over-riding research question is relevant and important to patients and clinicians alike. Anti-TNF agents and other biologic agents are used ubiquitously, yet there is still little personalisation of therapy, other than clinician 'gut feeling', so investigation of predictors of response to anti-TNFs (the most widely used class of biologics in RA) is important. Another strength of the work undertaken in this thesis is that analysis has been guided by evidence. Firstly, by establishing the current evidence base around the proposed question, before ensuring that the parameters used to identify the condition of remission/LDA were as robust as possible (by examining the DAS28-ESR and –CRP in detail). The identification of discrepancies between the two versions of the DAS28, ensured that a homogeneous measure of disease activity was used to avoid missing or spuriously identifying any predictors of response. Predictors of response used in the analyses were evidence-based and clinically relevant. Rigorous efforts were made to investigate and manage missing data. Finally, investigating sustained optimal response

from two perspectives; using the pre-specified thresholds for remission/LDA, and using a data driven trajectory analysis cross-validates findings about the frequency of the occurrence of the outcome, and predictors identified.

Nonetheless, despite the strengths outlined above, there are weaknesses. Most notably is that as these data are observational, causality cannot be attributed to the relationships identified in these analyses. Therefore, caution must be exercised in attesting to how modification of the identified relationships might influence the outcome of treatment.

Using 2010 as the date to split the cohorts could be viewed as a weakness as it meant that two cohorts were very different in size, and may have meant that some relationships with the later cohort were not identified due to sample size.

Whilst many variables were included in the analyses undertaken in Chapter 9 and 10, there were some notable variables that were not included; in particular co-use of DMARDs and steroids (discussed in more detail in Chapter 9; 9.8). The presence or absence of antibodies (RF and ACPA; discussed in Chapter 1; 1.2.4) would also have been interesting to examine, although only RF data are available from BSRBR-RA cohort from inception (ACPA was added later). Genetic data would have been very interesting to include, given the strong influence of genetics on the development of RA (Chapter 1; 1.3.1). However, genetic data are not available within the BSRBR-RA, although the linked Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate (BRAGGSS) registry does contain this data for a subgroup (≥2500) of BSRBR-RA patients.

A further weakness, is that this analysis has only mapped longitudinal disease activity over time using the DAS28-ESR. It would have been interesting to compare disease activity trajectories using other disease activity scores (such as the SDAI and CDAI), but the absence of the physician global score meant that this was not possible. It is also important to acknowledge that these analyses do not investigate other important outcomes such as radiographic progression or disability.

Finally, this analysis only includes patients taking their first anti-TNF for up to three years. Therefore, whilst there appear to be factors that strongly influence the likelihood of achieving sustained optimal response for anti-TNF therapy, it is currently not known if these predictors would be different for any secondary anti-TNF agent or other biologic agent.

11.5 Wider implications

This thesis has highlighted several important points. Despite the rapid expansion in therapeutic options available for the treatment of RA, and the focus on the treat-to-target paradigm, the rate of sustained remission remains disappointingly low. Even relaxing a treat-to-target paradigm to include LDA still only includes 32% of patients, leaving 68% of patients either only achieving LDA sporadically, or not at all. The wider lack of evidence using sustained remission or LDA as an outcome is also concerning. Whilst there is an understandable focus on RCTs with single point outcomes for efficacy studies required for licencing, the fact there were only six observational studies that investigated sustained remission as an outcome points to an area that requires more in-depth investigation.

Clinicians and patients know that treatment outcomes for RA need to be sustained in the long-term, not just at one point in time. Yet the available research does not reflect the importance of this clinical outcome. There may be a feeling that given the wide array of treatments available for RA, and documented success of aggressive treat-to-target interventions (125,137,138), that the management of RA has been adequately addressed. However, evidence presented in this thesis suggests that for the majority of RA patients, sustained remission is still a distant goal, only occasionally glimpsed and less frequently obtained. There is still far to go in optimising therapy.

It remains somewhat surprising that the temporal component of remission was specifically excluded from consideration for the updated combined ACR/EULAR definition of remission in RA (170). The stated reason for this was that *'the committee believed this [sustainability of remission] should be specified in each trial report'*.

However, this seems somewhat counterintuitive, given that the same argument could be used for almost any component of the assessment of remission.

The lack of a temporal aspect to the officially endorsed definition of remission fails to incentivise research into examining sustainable remission as an outcome. By their very nature, and as demonstrated in this thesis, sustained remission outcomes are more difficult to achieve than point-remission comparators. By failing to specify a minimum time for sustained remission in the ACR/EULAR definition, the studies that opt to report sustained remission outcomes would most likely have the appearances of 'poorer' outcomes than studies that only report point remission. The result is that there is little incentive for clinical trials, particularly those of novel therapeutics which may be trying to carve a niche in the (happily) crowded marketplace of rheumatological therapeutics, to report an outcome that would be more difficult to achieve than point remission. The lack of comparative outcomes in the literature would mean any study of a novel therapeutic demonstrating a realistic (but likely low) rate of sustainable remission would struggle to gain traction within the marketplace.

Some point to the fact that radiological outcomes may adequately quantify the temporal component of response to a drug (254), suggesting that absence of radiological progression of damage quantifies sustained good/optimal response in a clearly quantifiable way. However, the problem with using a radiological outcome to quantify sustained remission is that it does not represent the fluctuating temporal patient experience of RA, one of the most important aspects of treating any illness.

As discussed previously (Chapter 2), the DAS28 is by no means a perfect outcome measure, and the debate as to what 'remission' means, less still 'sustained remission' remains open. It is possible that if sustained normalisation of inflammatory response, or absence of swollen joints, was used to define remission, there may be more patients who achieve this over time. However, pain and fatigue are some of the most important outcomes from a patient perspective, irrespective of an ESR or CRP level, and cannot be completely ignored in the assessment of RA disease activity. Clinicians and patients will have different views as to what is important in defining remission; including what and how it should be measured. Ultimately, composite outcome measures such as the

DAS28 are not perfect, but do at least incorporate aspects important to patients and clinicians alike. Bearing this in mind, the question remains; is it realistic to expect everyone to reach remission as defined by any composite outcome measure?

Another argument against having a formalised definition of 'sustained remission' is that a threshold would have to be defined, both in terms of time (how long should sustained remission last?) and disease activity (ACR/EULAR, DAS28, CDAI, SDAI etc.). The application of thresholds is essential to enable analysis of outcomes in a research setting, however, wherever a threshold is set, there will be patients who don't quite achieve the threshold and become artificially grouped with patients who are more dissimilar to them than those 'just the other side' of the threshold (as previously discussed). However, this argument applies not only to sustained remission, but any categorisation of a continuous score (including DAS28, CDAI, SDAI and even BMI). By failing to have a standardised definition for what sustained remission should be, the little evidence that is generated is analysed using different definitions, making meta-analyses difficult and hampering the development of a robust evidence base just as was the case prior to the establishment of the OMERACT (216) core outcome set for RA trials.

A potential solution to the problem of applying artificial thresholds to individual patient data (a top-down approach) is to allow the 'data to speak for themselves' in the form of the trajectory analysis (a bottom-up approach). The analysis undertaken in Chapter 10 demonstrated that even in the smallest subgroup analysis (~1300 patients), it was possible to reliably identify at least two common trajectories of response, and the proportion of patients achieving the best outcome trajectory (38%) was similar to the proportion of patients achieving sustained LDA by threshold analysis (32%) in the most recent subgroup analysis (2010 – 2013). There were fewer 'statistically significant' associations identified (at a p-value ≤ 0.05) using this method compared with the threshold analysis (Chapter 9), which likely represents the increased heterogeneity of the patients included using a data-driven approach. Herein lies a problem with the frequentist approach to such research. An *a priori* hypothesis-driven approach to science (with significance thresholds) is the bedrock on which the majority of medical evidence rests, and is the most widely-accepted method of

conducting research. However, the problem with this approach is that it requires a single primary (and null) hypothesis to test. However, when considering the task of trying to predict response to a drug, it is clear from work in this thesis and existing evidence (Chapter 7), that in the case of anti-TNF there is no single predictor of optimal response for any patient. What has been demonstrated is that there are multiple variables that each play a part in predicting the likely response to anti-TNF. The weighting of each of these components is likely to be different for each patient, but will have an impact. In addition, even after taking into account all the variable identified in this thesis and others, there will be an unexplained component (genetics, epigenetics, environmental, health economic and societal factors etc.) that it will not be possible to explain, but could be quantified (e.g. as an 'error term').

With regards to whether significance thresholds (analysis in Chapter 9) or data-driven latent class modelling (Chapter 10) is the 'best' method of analysis really depends on the proposed use intended. Thresholds are undoubtedly useful in quantifying outcomes within a given population and can present clear and relatively unambiguous population-level results and associations (giving a clear 'snapshot' of outcomes at given time-points). However, it often requires dichotomisation of continuous variables, as well as applying unrealistic assumptions (i.e. everyone is followed-up at exactly 6-month intervals). Latent class modelling does not provide such easy-to-understand results (i.e. there are no 'remission' thresholds, and response trajectories are relative to other trajectories within that dataset, rather than between datasets). However, the data are presented in a way that is more akin to what really happened (i.e. exact follow-up times can be used and continuous scores are not artificially dichotomised). Therefore, results may be viewed as more representative of the 'true' picture of the real-world situation, and are more amenable to adaptation to develop predictive models that could provide real-time estimates of future outcomes based on an individual's past response (a Bayesian methodology). However, because thresholds are not enforced, fewer class-wide associations may identified (as was the case in Chapter 10), and results can be more difficult to conceptualise. Although each method has its strengths and weaknesses, an understanding of both perspectives is essential to achieve a detailed picture of outcomes of individuals with RA treated with anti-TNF.

Therefore, the task of trying to predict a patient's response to a drug such as anti-TNF, based on clinical and demographic features discussed in this thesis may seem an impossible task. However, if the problem is approached from a Bayesian perspective (described in Chapter 5; 5.4), it may be possible to incorporate the variable levels of uncertainties not only from one data-source (such as the BSRBR-RA), but others too; including epidemiological and genetic studies such as the Norfolk Arthritis Register (NOAR) and BRAGGSS; as well as existing clinical trial data, other registries etc. This combination of uncertainties from studies undertaken with different primary hypotheses may seem heretical from a frequentist standpoint, but tailoring and updating estimates of likelihood of a certain outcome for an individual, based on experience and evidence (from multiple sources) is exactly what a clinician does every day, and is the epitome of a Bayesian approach. The difficulty until recently, has been the ability to quantify this 'clinical acumen' or 'gut feeling', in a way that can be reproducibly used and at scale. However, with the significant increases in computational power available to researchers today, and rapid advance in understanding and use of machine learning, neural networks and artificial intelligence (AI), it is now possible to incorporate these varying estimates of probability from multiple sources into a single predictive estimate. Such an approach is already being used in a collaboration between clinicians and researchers at the Broad Institute, and IBM using the AI algorithm – Watson (255). In this collaboration, Watson is being used to analyse existing genomic and histopathological registry data to provide personalised treatment recommendations for individuals based on their personal genomic information. Watson is also being developed to be able to 'read' evidence on certain types of cancer (including unstructured data in the form of 'free-text' within research papers) from a wide range of resources including the internet, and provide up-to-the-minute evidence-based recommendations for treatment for individual patients, based on their clinical presentation and histological diagnosis. Trials are currently underway to use this technology to help guide clinicians' treatment recommendations (255).

This approach could be applied to the treatment of RA. Indeed, given the multiple longitudinal RA registries, clinical trial and genomic databases with structured datasets, the application of machine learning and AI algorithms to provide tailored

treatment recommendations to help clinicians treat patients may not be that far-fetched.

A great attraction of this approach is that it both leverages collective evidence available globally, and updates recommendations based on new evidence as it emerges (256). However, there are challenges in adopting this approach. Firstly, because AI uses self-generating algorithms, it is a truly 'black-box' approach and it is not possible to demonstrate exactly how an answer or recommendation was generated, meaning a human may not be able to cross-check the exact decision-making process used. This will understandably generate concern and convincing clinicians and patients to accept a computer-generated recommendation will require a step-change in how medicine is practiced. There are ethical issues too. If an adverse event occurs due to an AI generated recommendation, who is responsible? How can clinicians be sure that AI is making decisions in the best interests of the individual, and not in the best interests of society? Conversely, if outcomes from AI + clinician recommendations were far superior to decisions made without AI support, would it be ethical *not* to use them? These, and more questions will certainly arise when AI-assisted medicine arrives.

11.6 Next steps

With such implications (as discussed in 11.5), the initial follow-up work to this thesis would be undertake the analyses presented here in biologics with other mechanisms of action (such as rituximab or tocilizumab), as well as investigating if other longitudinal outcomes (such as disability) match those as measured by the DAS28. Furthermore, a specific analysis focusing on the effect of the use of DMARDs as a time-varying covariate in combination with anti-TNFs would also be worthwhile undertaking.

As alluded to previously, collaboration with studies such as BRAGGSS and NOAR, would enable a more comprehensive range of patients (including early arthritis patients), and predictors (including genetic markers) to be included in analyses. Furthermore, a pre-biologic patient population would be an ideal cohort in which to evaluate if predictors

of sustained response to biologics, and disease activity trajectories, are similar in a pre-biologic population.

Looking further forward, whilst AI medicine is still a way off, the use of Bayesian statistics to help predict likely sustained remission for anti-TNF and other biologics is a logical next step to the work started in this thesis. A Bayesian framework could enable the combination of existing RA treatment response data for multiple drugs, and from multiple sources (different registries, existing clinical trial data etc.) to build a predictive online data tool. This has already been achieved in the field of osteoporosis treatment, with the development, and widespread use of the FRAX tool (https://www.sheffield.ac.uk/FRAX/tool.jsp). Development of an online 'FRAX-style' tool for RA could recommend a specific treatment based on the likelihood of sustained remission within three years, based on various risk factors (such as identified in this thesis). LCMM-based 'trajectories' could then use real-time DAS28 scores (or SDAI, CDAI etc.) from patients as treatment is commenced, to provide updated likelihoods of achieving sustained remission as treatment progresses, allowing clinicians and patients the option to switch treatments earlier if desired outcomes are not likely to be met. Such a tool could be tested in a RCT e.g. standards care vs clinician + online prediction tool. If a trial of a relatively simple risk prediction tool were successful, more dynamic AI solutions could be incorporated to build a 'learning framework' that could constantly refine and improve its predictive capabilities as new research is generated.

Ignoring this avenue of medical research is not an option. AI is already being used to analyse exceptionally complex unstructured data, and will be used increasingly in the healthcare setting. As its utility and efficacy becomes more apparent, patients will expect their doctors to be proficient in the use such new tools, just as it is expected that doctors are able to use computers and the internet today.

As Isaac Asimov (author and scientist) said:

"It is change, continuing change, inevitable change, that is the dominant factor in society today. No sensible decision can be made any longer without taking into account not only the world as it is, but the world as it will be" (257).

Appendices

12 Appendices

12.1 Appendix 1: BSRBR-RA forms

ID

For office use only

BSR Biologics Register – Rheumatoid Arthritis
Clinical Baseline Form

Please complete the following PATIENT information

Gender: Male ☐ Female ☐

Date of birth: D D M M Y Y Y Y

Hospital Reg. No:

NHS No:

Consultant Rheumatologist:

Name of Hospital:

Preferred clinical contact email address:

Form completion date (today's date): D D M M Y Y Y Y

Title: Mr / Mrs / Miss / Ms Surname:

Forename/s:

Address: Postcode:

Telephone Number:

1. Does the patient have **Rheumatoid Arthritis**? ☐ Yes ☐ No

If **NO**, can you specify the other diagnosis? ☐

1a. Does the patient have **ACPA (anti-CCP) positive RA?** ☐ Yes ☐ No ☐ Don't Know

2a. What was the year of diagnosis? ☐

2b. What year was this patient first seen by a rheumatologist? ☐

3. ACR Criteria (please indicate which of the following apply to the patient):

Yes	No	Don't know	
☐	☐	☐	Morning stiffness >1 hour (ever)
☐	☐	☐	Arthritis or deformity/damage of three or more joint areas (PIP, MCP, wrist, elbow, knee, ankle, MTP) (now)
☐	☐	☐	Arthritis/deformity of hand/joint (now)
☐	☐	☐	Symmetry
☐	☐	☐	Nodules (ever)
☐	☐	☐	Rheumatoid factor positive (≥ 1/40) (ever)
☐	☐	☐	Erosions on hand or feet x-ray

4. Systemic features: Has the patient <u>ever</u> had any of the following?

Yes	No	Don't know	
☐	☐	☐	Sicca syndrome
☐	☐	☐	Serosal involvement (pleurisy/pericarditis)
☐	☐	☐	Eye involvement
☐	☐	☐	Systemic vasculitis
☐	☐	☐	Nailfold vasculitis
☐	☐	☐	Pulmonary fibrosis
☐	☐	☐	Other (please specify)

THIS SECTION IS INTENTIONALLY BLANK

5. Joint replacements/surgery: Has the patient <u>ever</u> had any of the following?

	Unilateral	Bilateral
Total knee replacement		
Total hip replacement		
Total shoulder replacement		
Total elbow replacement		
Wrist/hand/ankle/foot surgery		
Neck surgery		

6. Please indicate the current disease activity (i.e. at the time the patient started the new drug)

28 tender joint count ☐☐

28 swollen joint count ☐☐

ESR ☐☐☐

OR

CRP ☐☐☐

Patient global assessment (VAS) ☐☐☐ mm

(Out of 100)

Total DAS score (if known): _____

Date DAS28 taken:

____/_____/_____

DD MM YYYY

For patients starting a **biosimilar** - If DAS 28 is unavailable

According to the case notes, was the patient in low disease activity /remission at the point of switch to the biosimilar? Yes ☐ No ☐

7. Drug therapy: Please list all the patient's <u>current</u> treatment, for any indication

8. New Biologic/ Biosimilar Therapy (please use trade name)**:**

Which drug has the patient started?

Enbrel ☐	**Cimzia** ☐	**Inflectra** ☐			
Remicade ☐	**RoActemra** ☐	**Benepali** ☐	**Please specify trade name:**		
Humira ☐	**Remsima** ☐	**Other** ☐			

Please indicate the date of first therapy dose:

D	D	M	M	Y	Y	Y	Y

Please also indicate the average **dose and unit**: ☐ Frequency: ☐

Is this delivered intravenously or subcutaneously? ☐ IV ☐ SC

Is this the patient's first exposure to a biologic/ biosimilar agent? Yes ☐ No ☐ If **No**, please give details below

	Biologic therapy	DAS28 prior to starting	Start date	Stop date	Reason for stopping
1					
2					
3					
4					

Is the patient switching from an **originator** e.g. Remicade directly to a **biosimilar** of the **same** product, i.e. Inflectra or Remsima? Yes ☐ No ☐

If **yes**, please provide the reason for this switch and any comments below:

Comments:

☐ Clinical indication
☐ Patient choice
☐ Cost factors
☐ Other

Is the patient still on biologic/biosimilar therapy? Yes ☐ No ☐ **If NO, please give details on a separate sheet**

9. Is the patient currently receiving DMARD therapy? Yes ☐ No ☐

If **Yes**, please indicate which DMARD(s) and current dose.

DMARD Started (please tick)		mg	Frequency	Date Started D	D	M	M	Y	Y
Methotrexate									
Azathioprine									
Cyclophosphamide									
Cyclosporine									
Leflunomide									
Other **:**									

10. Previous second-line drug therapy:
Has the patient EVER had any of the following drugs?

	Yes	No	Don't know
IM Gold			
Auranofin			
Penicillamine			
Sulphasalazine			
Chlor/HCQ			
Steroids			

→ If currently receiving steroids, please indicate dose: ____

And route: ☐ IV ☐ SC ☐ Oral

We would now like to know more details about certain drugs:

	Yes	No	Don't know	1st Course: Date started: Month	1st Course: Date started: Year	1st Course: Date stopped: Month	1st Course: Date stopped: Year	2nd Course: Date started: Month	2nd Course: Date started: Year	2nd Course: Date stopped: Month	2nd Course: Date stopped: Year
Methotrexate											
Azathioprine											
Cyclophosphamide											
Cyclosporine											
Leflunomide											
Other, please specify											

If patient has started or stopped the same drug more than twice please give details on an additional sheet
(Do not include stopping a drug for less than three months)

11. Co-morbidity:

Has the patient <u>ever</u> had (i.e. required treatment for) any of the following illnesses? Please tick all that apply

	Yes	No	Don't know	Year of onset			
High blood pressure							
Angina							
Heart attack							
Stroke							
Epilepsy							
Asthma							
Chronic bronchitis/emphysema							
Peptic ulcer							
Liver disease							
Renal disease							
TB							
Demyelination							
Diabetes*							
Hyperthyroidism							
Depression							
Cancer‡							

Other co-morbidity not listed _____

_____

_____

‡If the patient has (or has ever had) cancer please specify date of diagnosis and site(s):

*If the patient is diabetic is (s)he:

Insulin dependent ☐ Tablet controlled ☐ Diet controlled ☐

12. Smoking status: Is the patient a:

☐ Current smoker ☐ Ex-smoker ☐ Never-smoked

13. Blood pressure: what is the patient's <u>current</u> (i.e. at the time that the biologic agent was started) blood pressure?

Systolic ☐☐☐ mm

Diastolic ☐☐☐ mm

14. Height and weight: what is the patient's <u>current</u> (i.e. at the time that the biologic agent was started) height and weight?

Weight ☐☐☐ kg

Height ☐☐☐ cm

15. Did the patient have a chest x-ray prior to starting the new therapy?

Yes ☐

No ☐

16. Has the patient had a QuantiFERON, ELISPOT (or other Gamma interferon based assays for TB) test?

Yes ☐

No ☐

Date/Details:

17. Has the patient received the Herpes zoster vaccine?

☐ Yes ☐ Date ☐ No ☐ Don't know

Thank you for completing this form!

This form should be accompanied by the following <u>pre-biologic therapy</u> patient-completed forms:

HAQ ☐

EQ-5D ☐

Please return to:

BSRBR-RA
Arthritis Research UK Centre for Epidemiology
Unit 4 Rutherford House
Manchester Science Park
40 Pencroft Way
Manchester
M15 6SZ

12.2 Appendix 2: BSRBR-RA Data Release Form

BSRBR RA

Biologic Studies Group

The University of Manchester
Rutherford House
Manchester Science Park
40 Pencroft Way
Manchester
M15 6SZ

Phone 0161-275
Fax 0161-275
E-mail: biologics.register@manchester.ac.uk
Website:

BSRBR-RA DATA RELEASE FORM

(1) DATA SECURITY AND CONFIDENTIALITY

Upon on receipt of the agreed BSRBR dataset, I Philip Hamann hereby of the University of Bath, undertake to adhere to the attached guidelines "*Information Governance in Health and Social Care Research*" which is a framework for handling information in a confidential and secure manner to appropriate ethical and quality standards at the University of Manchester. Of particular relevance are the sections on data protection

and research, data security (including the storage of data (hard-copy and computer data), the electronic transfer of data and the destruction of data).

x	I have read the patient information sheet, consent form and ethics approval documentation under which the BSRBR-RA data was collected.

BSRBR-RA Consent

http://www.inflammation-repair.manchester.ac.uk/Musculoskeletal/research/CfE/pharmacoepidemiology/bsrbr/healthprofessionals/forms/Registrationforms/AntiTNFCohortForms

BSRBR-RA Ethics

http://www.inflammation-repair.manchester.ac.uk/Musculoskeletal/research/CfE/pharmacoepidemiology/bsrbr/healthprofessionals/MRECDocuments/

x	I agree to be held fully responsible for this data adhering to the Data Protection Act 1998, once it is transferred from The University of Manchester. Once the transfer has taken place, the responsibility of the data is transferred from The University of Manchester to (your name) Philip Hamann.

x	I agree to ensure that all persons handling the BSRBR-RA data as part of this project (external to the BSRBR-RA team) have completed and returned the BSRBR-RA Data Confidentiality Form to the BSRBR-RA prior to the release of BSRBR-RA data.

(2) PUBLICATION

The University of Manchester is sub-contracted by the British Society for Rheumatology (BSR) to process the BSRBR-RA data. BSR has separate contracts with the participating pharmaceutical companies who support the BSRBR-RA. There is a publications policy relating to all material arising from all BSRBR-RA data (see below for details).

Timescales for Review of material

There are strict rules with regards to releasing BSRBR-RA material into the public domain and therefore all proposed abstracts/manuscripts/reports and presentations arising from this data must be circulated for review to all BSRBR-RA stakeholders prior to submission via the BSRBR-RA offices in Manchester. Therefore, I, Philip Hamann agree to adhere to the following timescales for circulation of data/results prior to submission, to enable the University of Manchester to fulfil its contractual obligations with the stakeholders:

(i) 15 calendar days:

- Abstracts for conference presentations
- Posters, presentation slides and papers that are going to be used at a medico-scientific conference or published elsewhere for which there has been no material change in the results/conclusions from those submitted in the abstract.

(ii) 30 calendar days:

- Papers that are to be submitted for full publication in a scientific/medical journal or used in a report to be submitted to an external body (outside BSR).
- Posters, presentation slides, papers and reports that are going to be used at a medico-scientific conference or published elsewhere for which there has been material change in the results/conclusions from those submitted in the abstract.

(3) DATA ARCHIVING

For archiving purposes, all relevant working analysis files should be returned to the BSRBR-RA at the end of the project, including the raw datasets, any programmed statistical analyses and the result sets. The files should be sufficient to allow the analyses to be re-run from the raw datasets by another researcher, giving the same final results. Please provide a copy of the final manuscript/abstract directly to the BSRBR-RA. Also provide details of any statistical software tools used and the final versions of the article/papers/abstracts.

Please return all files to the BSRBR-RA Database Manager by an agreed method. To discuss the most suitable method for your centre please contact katie.mcgrother@manchester.ac.uk or phone 0161 306 1893.

(4) BSRBR-RA DATA/STATISTICS ADVICE

The BSRBR team will be available to offer advice throughout the project:

(i) For BSRBR data issues, please contact the BSRBR database manager katie.mcgrother@manchester.ac.uk or phone 0161 306 1893.
(ii) For BSRBR analysis issues, please contact Kath.watson@manchester.ac.uk (phone 0161 306 1898) or Kimme.hyrich@manchester.ac.uk.

This form is valid for the period of **2 years** from the date shown below. Should the project take longer than 1 year, this will need to be discussed with BSRBR-RA and a new Data Release Form will need to be completed.

(5) LAY SUMMARY OF DATA

See Exhibit B

Signed  **_____**

Print name___Philip Hamann_____

Dated_____07/07/2016 _____

References

13 References

1. Oxford English Dictionary (OED). Oxford University Press. Available from http://www.oed.com [accessed 24th September 2017].

2. Landre-Baeuvais, AJ. The first description of rheumatoid arthritis. Unabridged text of the doctoral dissertation presented in 1800. Joint Bone Spine. 2001 Mar 1;68(2):130–43.

3. Garrod AB. The Nature And Treatment Of Gout And Rheumatic Gout. 1860.

4. Garrod AE. A treatise on rheumatism and rheumatoid arthritis, 1890. C. Griffin and Co; 1895.

5. Short CL. The antiquity of rheumatoid arthritis. Arthritis Rheum. 1974 May;17(3):193–205.

6. Wegner N, Wait R, Sroka A, Eick S, Nguyen K-A, Lundberg K, et al. Peptidylarginine deiminase from Porphyromonas gingivalis citrullinates human fibrinogen and α-enolase: Implications for autoimmunity in rheumatoid arthritis. Arthritis Rheum. 2010 May 6;62(9):2662–72.

7. Entezami P, Fox DA, Clapham PJ, Chung KC. Historical perspective on the etiology of rheumatoid arthritis. Hand Clin. 2011 Feb;27(1):1–10.

8. Roser M. Life Expectancy. Our World in Data. Available from https://ourworldindata.org/life-expectancy/ [accessed 24th September 2017]

9. Ortner DJ, Putschar WG. Identification of Pathological Conditions in Human Skeletal Remains. Smithsonian Contributions to Anthropology. 1981; 28

10. Blondiaux J, Cotten A, Fontaine C, Hänni C, Bera A, Flipo RM. Two Roman and medieval cases of symmetrical erosive polyarthropathy from Normandy: anatomico-pathological and radiological evidence for rheumatoid arthritis. International Journal of Osteoarchaeology. 1997 Sep 1;7(5):451–66.

11. Rothschild BM, Turner KR, DeLuca MA. Symmetrical erosive peripheral polyarthritis in the Late Archaic Period of Alabama. Science. 1988 Sep 16;241(4872):1498–501.

12. Sydenham T, Latham RG. The works of Thomas Sydenham, MD. 1850.

13. Musk AW, de Klerk NH. History of tobacco and health. Respirology. 2003 Sep;8(3):286–90.

14. Goodman J. Tobacco in History. Routledge. 2005.

15. Doll R, Bradford-Hill A. smoking and carcinoma of the lung. BMJ; 1950 Sep 30;2(4682):739–48.

16. Hutchinson D. Cigarette smoking and rheumatoid arthritis. PhD thesis. University of Leicester; 2003.

17. Panush RS. Why Did rheumatoid arthritis begin in 1800? The Rheumatologist. the-rheumatologist.org/article/why-did-rheumatoid-arthritis-begin-in-1800/ (accessed 16th September 2017).

18. Källberg H, Padyukov L, Plenge RM, Rönnelid J, Gregersen PK, van der Helm-van Mil AHM, et al. Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet. 2007 May;80(5):867–75.

19. Scott DL, Wolfe F, Huizinga TWJ. Rheumatoid arthritis. Lancet. 2010 Sep 25;376(9746):1094–108.

20. Wright HL, Moots RJ, Edwards SW. The multifactorial role of neutrophils in rheumatoid arthritis. Nat. Rev. Rheumatol. 2014 Oct;10(10):593–601.

21. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. N Engl J Med. 2011 Dec 8;365(23):2205–19.

22. Goldring SR. Bone and joint destruction in rheumatoid arthritis: what is really happening? J Rheumatol Suppl. 2002 Sep;65:44–8.

23. Manara M, Sinigaglia L. Bone and TNF in rheumatoid arthritis: clinical implications. RMD Open. 2015;1(Suppl 1):e000065.

24. Kalliolias GD, Ivashkiv LB. TNF biology, pathogenic mechanisms and emerging therapeutic strategies. Nat. Rev. Rheumatol. 2016 Jan;12(1):49–62.

25. Calabrese LH, Rose-John S. IL-6 biology: implications for clinical targeting in rheumatic disease. Nat. Rev. Rheumatol. 2014 Dec;10(12):720–7.

26. Kim GW, Lee NR, Pi RH, Lim YS, Lee YM, Lee JM, et al. IL-6 inhibitors for treatment of rheumatoid arthritis: past, present, and future. Arch Pharm Res. 2015;38(5):575–84.

27. Gabay C, Lamacchia C, Palmer G. IL-1 pathways in inflammation and human diseases. Nat. Rev. Rheumatol. 2010 Apr;6(4):232–41.

28. Miossec P, Korn T, Kuchroo VK. Interleukin-17 and type 17 helper T cells. N Engl J Med. 2009 Aug 27;361(9):888–98.

29. Feldmann M. Development of anti-TNF therapy for rheumatoid arthritis. Nat. Rev. Immunol. 2002;2(5):364–71.

30. Nakken B, Munthe LA, Konttinen YT, Sandberg AK, Szekanecz Z, Alex P, et al. B-cells and their targeting in rheumatoid arthritis - current concepts and future perspectives. Autoimmun Rev. 2011 Nov;11(1):28–34.

31. Hamann P, McHugh N. Autoimmune Serology. In: Watts R, Conaghan P, Denton C, Foster H, Isaacs J, Müller-Ladner U, editors. Oxford Textbook of Rheumatology. 4 ed. 2013.

32. Taylor P, Gartemann J, Hsieh J, Creeden J. A systematic review of serum biomarkers anti-cyclic citrullinated peptide and rheumatoid factor as tests for rheumatoid arthritis. Autoimmune diseases. 2011. doi: 10.4061/2011/815038.

33. Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, et al. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology. 2002 Jul;41(7):793–800.

34. Minichiello E, Semerano L, Boissier M-C. Time trends in the incidence, prevalence, and severity of rheumatoid arthritis: A systematic literature review. Joint Bone Spine. 2016 Dec 1;83(6):625–30.

35. Silman AJ, Pearson JE. Epidemiology and genetics of rheumatoid arthritis. Arthritis Res. 2002;4(Suppl 3):S265–72.

36. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. Nat. Rev. Rheumatol. 2013 Mar;9(3):141–53.

37. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014 Feb 20;506(7488):376–81.

38. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007 Jun 7;447(7145):661–78.

39. Viatte S, Barton A. Genetics of rheumatoid arthritis susceptibility, severity, and treatment response. Semin Immunopathol. 2017 Jun;39(4):395–408.

40. Klein K, Gay S. Epigenetics in rheumatoid arthritis. Curr Opin Rheumatol. 2015 Jan;27(1):76–82.

41. Shampo MA, Kyle RA. Philip S. Hench–1950 Nobel laureate. Mayo Clinic Proceedings. 2001.

42. Qi S, Xin R, Guo W, Liu Y. Meta-analysis of oral contraceptives and rheumatoid arthritis risk in women. Ther Clin Risk Manag. 2014;10:915–23.

43. Brennan P, Silman A. Breast-feeding and the onset of rheumatoid arthritis. Arthritis Rheum. 1994 Jun;37(6):808–13.

44. Barrett JH, Brennan P, Fiddler M, Silman A. Breast-feeding and postpartum relapse in women with rheumatoid and inflammatory arthritis. Arthritis Rheum. 2000 May;43(5):1010–5.

45. Hazes JMW, Coulie PG, Geenen V, Vermeire S, Carbonnel F, Louis E, et al. Rheumatoid arthritis and pregnancy: evolution of disease activity and pathophysiological considerations for drug use. Rheumatology. 2011 Nov;50(11):1955–68.

46. Pikwer M, Bergstrom U, Nilsson J-A, Jacobsson L, Berglund G, Turesson C. Breast feeding, but not use of oral contraceptives, is associated with a reduced risk of rheumatoid arthritis. Ann Rheum Dis. 2009 Apr;68(4):526–30.

47. Karlson EW, Mandl LA, Hankinson SE, Grodstein F. Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the Nurses' Health Study. Arthritis Rheum. 2004 Nov;50(11):3458–67.

48. Tang MW, Reedquist KA, Garcia S, Fernandez BM, Codullo V, Vieira-Sousa E, et al. The prolactin receptor is expressed in rheumatoid arthritis and psoriatic arthritis synovial tissue and contributes to macrophage activation. Rheumatology. 2016 Dec;55(12):2248–59.

49. Källberg H, Ding B, Padyukov L, Bengtsson C, Rönnelid J, Klareskog L, et al. Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. Ann Rheum Dis. 2011 Mar;70(3):508–11.

50. de Rooy DPC, van Nies JAB, Kapetanovic MC, Kristjansdottir H, Andersson MLE, Forslind K, et al. Smoking as a risk factor for the radiological severity of rheumatoid arthritis: a study on six cohorts. Ann Rheum Dis. 2014 Jun 6;73(7):1384–7.

51. Hensvold AH, Magnusson PKE, Joshua V, Hansson M, Israelsson L, Ferreira R, et al. Environmental and genetic factors in the development of anticitrullinated protein antibodies (ACPAs) and ACPA-positive rheumatoid arthritis: an epidemiological investigation in twins. Ann Rheum Dis. 2015 Feb;74(2):375–80.

52. Sugiyama D, Nishimura K, Tamaki K, Tsuji G, Nakazawa T, Morinobu A, et al. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. Ann Rheum Dis. 2010 Jan;69(1):70–81.

53. Sparks JA, Karlson EW. The Roles of Cigarette Smoking and the Lung in the Transitions Between Phases of Preclinical Rheumatoid Arthritis. Curr Rheumatol Rep. 2016 Mar;18(3):15.

54. Perry E, Kelly C, Eggleton P, De Soyza A, Hutchinson D. The lung in ACPA-positive rheumatoid arthritis: an initiating site of injury? Rheumatology. 2014 Nov;53(11):1940–50.

55. Tracy A, Buckley CD, Raza K. Pre-symptomatic autoimmunity in rheumatoid arthritis: when does the disease start? Semin Immunopathol. 2017 Jun;39(4):423–35.

56. de Smit M, Westra J, Vissink A, Doornbos-van der Meer B, Brouwer E, van Winkelhoff AJ. Periodontitis in established rheumatoid arthritis patients: a cross-sectional clinical, microbiological and serological study. Arthritis Res Ther. 2012 Oct 17;14(5):R222.

57. Harvey GP, Fitzsimmons TR, Dhamarpatni AASSK, Marchant C, Haynes DR, Bartold PM. Expression of peptidylarginine deiminase-2 and -4, citrullinated proteins and anti-citrullinated protein antibodies in human gingiva. J Periodont Res. 2013 Apr;48(2):252–61.

58. Mikuls TR, Thiele GM, Deane KD, Payne JB, O'DELL JR, Yu F, et al. Porphyromonas gingivalis and disease-related autoantibodies in individuals at increased risk of rheumatoid arthritis. Arthritis Rheum. 2012 Nov;64(11):3522–30.

59. Fuggle NR, Smith TO, Kaul A, Sofat N. Hand to Mouth: A systematic review and meta-analysis of the association between rheumatoid arthritis and periodontitis. Front Immunol; 2016 Mar 2;7(23):2205–10.

60. Li S, Yu Y, Yue Y, Zhang Z, Su K. Microbial Infection and Rheumatoid Arthritis. J Clin Cell Immunol. 2013 Dec;4(6). doi:10.4172/2155-9899.1000174.

61. Carty SM, Snowden N, Silman AJ. Should infection still be considered as the most likely triggering factor for rheumatoid arthritis? Ann Rheum Dis. 2004 Nov;63 Suppl 2(suppl_2):ii46–9.

62. Rantapää-Dahlqvist S, de Jong BAW, Berglin E, Hallmans G, Wadell G, Stenlund H, et al. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum. 2003 Oct;48(10):2741–9.

63. Gómez R, Conde J, Scotece M, Gomez-Reino JJ, Lago F, Gualillo O. What's new in our understanding of the role of adipokines in rheumatic diseases? Nat Rev Rheum 2011 Aug 2;7(9):528–36.

64. Qin B, Yang M, Fu H, Ma N, Wei T, Tang Q, et al. Body mass index and the risk of rheumatoid arthritis: a systematic review and dose-response meta-analysis. Arthritis Res Ther. 2015 Mar 29;17(1):86.

65. Liu Y, Hazlewood GS, Kaplan GG, Eksteen B, Barnabe C. Impact of obesity on remission and disease activity in rheumatoid arthritis: a systematic review and meta-analysis. Arthritis Care Res. 2017 Feb;69(2):157–65.

66. Hart JE, Källberg H, Laden F, Bellander T, Costenbader KH, Holmqvist M, et al. Ambient air pollution exposures and risk of rheumatoid arthritis: results from the Swedish EIRA case-control study. Ann Rheum Dis. 2013 Jun;72(6):888–94.

67. Stolt P, Yahya A, Bengtsson C, Källberg H, Rönnelid J, Lundberg I, et al. Silica exposure among male current smokers is associated with a high risk of developing ACPA-positive rheumatoid arthritis. Ann Rheum Dis. 2010 Jun;69(6):1072–6.

68. Stolt P, Kallberg H, Lundberg I, Sjögren B, Klareskog L, Alfredsson L, et al. Silica exposure is associated with increased risk of developing rheumatoid arthritis: results from the Swedish EIRA study. Ann Rheum Dis. 2005 Apr;64(4):582–6.

69. Too CL, Muhamad NA, Ilar A, Padyukov L, Alfredsson L, Klareskog L, et al. Occupational exposure to textile dust increases the risk of rheumatoid arthritis: results from a Malaysian population-based case-control study. Ann Rheum Dis. 2016 Jun;75(6):997–1002.

70. Alamanos Y, Voulgari PV, Drosos AA. Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review. Semin Arth Rheu. 2006 Dec;36(3):182–8.

71. Silman AJ. The changing face of rheumatoid arthritis: why the decline in incidence? Arthritis Rheum. 2002 Mar;46(3):579–81.

72. 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 Mar;46(3):625–31.

73. Myasoedova E, Crowson CS, Kremers HM, Therneau TM, Gabriel SE. Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. Arthritis Rheum. 2010 Jun;62(6):1576–82.

74. Kirwan JR, Hewlett S. Patient perspective: reasons and methods for measuring fatigue in rheumatoid arthritis. J Rheumatol. 2007; 34 (5): 1171–3

75. Strand V, Wright GC, Bergman MJ, Tambiah J, Taylor PC. Patient Expectations and perceptions of goal-setting strategies for disease management in rheumatoid arthritis. J. Rheumatol. 2015 Nov;42(11):2046–54

76. Minnock P, Ringnér A, Bresnihan B, Veale D, FitzGerald O, McKee G. Perceptions of the cause, impact and management of persistent fatigue in patients with rheumatoid arthritis following tumour necrosing factor inhibition therapy. Musculoskeletal Care. 3rd ed. 2016 Feb 12;15(1):23–35.

77. Druce KL, Bhattacharya Y, Jones GT, Macfarlane GJ, Basu N. Most patients who reach disease remission following anti-TNF therapy continue to report fatigue: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. Rheumatology. 2016 Jun 21.

78. Karshikoff B, Sundelin T, Lasselin J. Role of inflammation in human fatigue: relevance of multidimensional assessments and potential neuronal mechanisms. Front Immunol. 2017;8:21.

79. Nikolaus S, Bode C, Taal E, van de Laar MAFJ. Fatigue and factors related to fatigue in rheumatoid arthritis: a systematic review. Arthritis Care Res. 2013 Jul;65(7):1128–46.

80. Almeida C, Choy EHS, Hewlett S, Kirwan JR, Cramp F, Chalder T, et al. Biologic interventions for fatigue in rheumatoid arthritis. Cochrane Database Syst Rev. 2016 Jun 6;(6):CD008334.

81. Camacho EM, Verstappen SMM, Chipping J, Symmons DPM. Learned helplessness predicts functional disability, pain and fatigue in patients with recent-onset inflammatory polyarthritis. Rheumatology. 2013 Jul;52(7):1233–8.

82. Katz P. Fatigue in rheumatoid arthritis. Curr Rheumatol Rep. 3rd ed. 2017 May;19(5):25.

83. Hewlett S, Hehir M, Kirwan JR. Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. Arthritis Rheum. 2007 Apr 15;57(3):429–39.

84. Ward MM, Guthrie LC, Dasgupta A. Direct and indirect determinants of the patient global assessment in rheumatoid arthritis: differences by level of disease activity. Arthritis Care Res. 2017 Mar;69(3):323–9.

85. Heiberg T, Kvien TK. Preferences for improved health examined in 1,024 patients with rheumatoid arthritis: pain has highest priority. Arthritis Rheum. 2002 Aug;47(4):391–7.

86. Walsh DA, McWilliams DF. Mechanisms, impact and management of pain in rheumatoid arthritis. Nat Rev Rheum. 2014 Oct;10(10):581–92.

87. Melzack R, Wall PD. Pain mechanisms: a new theory. Science. 1965 Nov 19;150(3699):971–9.

88. Meeus M, Vervisch S, De Clerck LS, Moorkens G, Hans G, Nijs J. Central sensitization in patients with rheumatoid arthritis: a systematic literature review. Semin Arth Rheu. 2012 Feb;41(4):556–67.

89. Pollard LC, Ibrahim F, Choy EH, Scott DL. Pain thresholds in rheumatoid arthritis: the effect of tender point counts and disease duration. J Rheumatol. 2012 Jan;39(1):28–31.

90. Edwards RR, Wasan AD, Bingham CO, Bathon J, Haythornthwaite JA, Smith MT, et al. Enhanced reactivity to pain in patients with rheumatoid arthritis. Arthritis Res Ther. 2009;11(3):R61.

91. Feist E, Burmester G. Rheumatoid Arthritis—Clinical Features. In: Watts R, Conaghan P, Denton C, Foster H, Isaacs J, Müller-Ladner U, editors. Oxford Textbook of Rheumatology. 4 ed. 2013.

92. Naz SM, Symmons DPM. Mortality in established rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2007 Oct;21(5):871–83.

93. Aviña-Zubieta JA, Choi HK, Sadatsafavi M, Etminan M, Esdaile JM, Lacaille D. Risk of cardiovascular mortality in patients with rheumatoid arthritis: a meta-analysis of observational studies. Arthritis Rheum. 2008 Dec 15;59(12):1690–7.

94. Avina-Zubieta JA, Thomas J, Sadatsafavi M, Lehman AJ, Lacaille D. Risk of incident cardiovascular events in patients with rheumatoid arthritis: a meta-analysis of observational studies. Ann Rheum Dis. 2012 Sep;71(9):1524–9.

95. Agca R, Heslinga SC, Rollefstad S, Heslinga M, McInnes IB, Peters MJL, et al. EULAR recommendations for cardiovascular disease risk management in patients with rheumatoid arthritis and other forms of inflammatory joint disorders: 2015/2016 update. Ann Rheum Dis. 2016; 0:1–12. doi:10.1136/annrheumdis-2016-209775

96. Shaw M, Collins BF, Ho LA, Raghu G. Rheumatoid arthritis-associated lung disease. Eur Respir Rev. 2015 Mar;24(135):1–16.

97. Rowbotham, Grainger AJ. Radiographic Imaging. In: Watts R, Conaghan P, Denton C, Foster H, Isaacs J, Müller-Ladner U, editors. Oxford Textbook of Rheumatology. 4 ed. 2013.

98. Kane D, Platt P. Ultrasound. In: Watts R, Conaghan P, Denton C, Foster H, Isaacs J, Müller-Ladner U, editors. Oxford Textbook of Rheumatology. Oxford University Press; 2013

99. Balamoody S, Marzo-Ortega H, O'Connor PJ. Magnetic Resonance Imaging. In: Watts R, Conaghan P, Denton C, Foster H, Isaacs J, Müller-Ladner U, editors. Oxford Textbook of Rheumatology. Oxford University Press; 2013.

100. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988 Mar;31(3):315–24.

101. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative

initiative. Ann Rheum Dis. 2010 Sep;69(9):1580–8.

102. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO III, et al. 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum. 2010 Aug 10;62(9):2569–81.

103. Gerlag DM, Raza K, van Baarsen LGM, Brouwer E, Buckley CD, Burmester GR, et al. EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. Ann Rheum Dis. 2012 May;71(5):638–41.

104. Mottram PL. Past, present and future drug treatment for rheumatoid arthritis and systemic lupus erythematosus. Immunol Cell Biol. 6 ed. 2003 Oct;81(5):350–3.

105. van Tubergen A, van der Linden S. A brief history of spa therapy. Ann Rheum Dis. 2002 Mar;61(3):273–5.

106. Boonen A, van de Rest J, Dequeker J, van der Linden S. How Renoir coped with rheumatoid arthritis. Vol. 315, BMJ; 1997

107. Kean WF, Kean IRL. Clinical pharmacology of gold. Inflammopharmacology. 2008 Jun;16(3):112–25.

108. Hench PS, Kendall EC. The effect of a hormone of the adrenal cortex (17-hydroxy-11-dehydrocorticosterone; compound E) and of pituitary adrenocorticotropic hormone on rheumatoid arthritis. Proc Staff Meet Mayo Clin. 1949 Apr 13;24(8):181–97

109. Lloyd M. Philip Showalter Hench, 1896–1965. Rheumatology. Oxford University Press; 2002 May 1;41(5):582–4.

110. Gubner R, August S, Ginsberg V. Therapeutic suppression of tissue reactivity.II. Effect of aminopterin in rheumatoid arthritis and psoriasis pdf. Am J Med Sci 1951 Feb;221(2):176-82

111. Black RL, O'Brien WM, Vanscott EJ, Auerbach R, Eisen AZ, Bunim JJ. Methotrexate therapy in psoriatic arthritis; double-blind study on 21 patients. JAMA. 1964 Sep 7;189:743–7.

112. Hoffmeister RT. Methotrexate therapy in rheumatoid arthritis: 15 years experience. Am J Med. 1983 Dec 30;75(6A):69–73.

113. Willkens RF, Watson MA, Paxson CS. Low dose pulse methotrexate therapy in rheumatoid arthritis. J. Rheumatol. 1980 Jul;7(4):501–5.

114. Willkens RF, Watson MA. Methotrexate: a perspective of its use in the treatment of rheumatic diseases. J Lab Clin Med. 1982 Sep 1;100(3):314–21.

115. Weinblatt ME. Methotrexate in rheumatoid arthritis: a quarter century of development. Trans Am Clin Climatol Assoc. 2013;124:16–25.

116. Weinblatt ME, Coblyn JS, Fox DA, Fraser PA, Holdsworth DE, Glass DN, et al. Efficacy of low-dose methotrexate in rheumatoid arthritis. N Engl J Med. 1985 Mar 28;312(13):818–22.

117. Williams HJ, Willkens RF, Samuelson CO, Alarcón GS, Guttadauria M, Yarboro C, et al. Comparison of low-dose oral pulse methotrexate and placebo in the treatment of rheumatoid arthritis. A Controlled Clinical Trial. Arthritis Rheum. 1985 Jul 1;28(7):721–30.

118. Thompson RN, Watts C, Edelman J, Esdaile J, Russell AS. A controlled two-centre trial of parenteral methotrexate therapy for refractory rheumatoid arthritis. J Rheumatol. 1984 Dec 1;11(6):760–3.

119. Andersen PA, West SG, O'Dell JR, Via CS, Claypool RG, Kotzin BL. Weekly pulse methotrexate in rheumatoid arthritis: clinical and immunologic effects in a randomized, double-blind study. Ann Intern Med. 1985 Oct 1;103(4):489–96.

120. Suarez-Almazor ME, Fitzgerald A, Grace M, Russell AS. A randomized controlled trial of parenteral methotrexate compared with sodium aurothiomalate (Myochrysine) in the treatment of rheumatoid arthritis. J Rheumatol. 1988;15(5):753–6.

121. Jeurissen ME, Boerbooms AM, Van de Putte LB, Doesburg WH, Mulder J, Rasker JJ, et al. Methotrexate versus azathioprine in the treatment of rheumatoid arthritis. A forty-eight-week randomized, double-blind trial. Arthritis Rheum. 1991 Aug;34(8):961–72.

122. Weinblatt ME, Polisson R, Blotner SD, Sosman JL, Aliabadi P, Baker N, 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. 1993 May;36(5):613–9.

123. Rau R, Herborn G, Menninger H, Blechschmidt J. Comparison of intramuscular methotrexate and gold sodium thiomalate in the treatment of early erosive rheumatoid arthritis: 12 month data of a double-blind parallel study of 174 patients. Br J Rheumatol. 1997 Mar;36(3):345–52.

124. Boers M, Verhoeven AC, Markusse HM. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet. 1997; 350: 309–18

125. Grigor C, Capell H, Stirling A, McMahon AD, Lock P, Vallance R, et al. Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial. Lancet. 2004 Jul;364(9430):263–9.

126. Wessels JAM, Huizinga TWJ, Guchelaar HJ. Recent insights in the pharmacological actions of methotrexate in the treatment of rheumatoid arthritis. Rheumatology. 2008 Mar;47(3):249–55.

127. Svartz N. Salazopyrin, a new sulfanilamide preparation. A. Therapeutic results in rheumatic polyarthritis. B. Therapeutic results in ulcerative colitis. C. Toxic manifestations in treatment with sulfanilamide preparations. J Intern Med. 110 (6) 1942.

128. Suarez-Almazor ME, Belseck E, Shea B, Wells G, Tugwell P. Sulfasalazine for rheumatoid arthritis. Cochrane Database Syst Rev. 2000;(2):CD000958.

129. Meier FMP, Frerix M, Hermann W, Müller-Ladner U. Current immunotherapy in rheumatoid arthritis. Immunotherapy. 2013 Sep;5(9):955–74.

130. Monaco C, Nanchahal J, Taylor P, Feldmann M. Anti-TNF therapy: past, present and future. Int Immunol. 2015 Jan;27(1):55–62.

131. Elliott MJ, Maini RN, Feldmann M, Long-Fox A, Charles P, Katsikis P, et al. Treatment of rheumatoid arthritis with chimeric monoclonal antibodies to tumor necrosis factor alpha. Arthritis Rheum. 1993 Dec;36(12):1681–90.

132. Maini R, St Clair EW, Breedveld F, Furst D, Kalden J, Weisman M, et al. Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group. Lancet. 1999 Dec 4;354(9194):1932–9.

133. Weinblatt ME, Kremer JM, Bankhurst AD. A trial of etanercept, a recombinant tumor necrosis factor receptor: Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. N Engl J Med 1999;340(4):253–9.

134. Nam JL, Winthrop KL, van Vollenhoven RF, Pavelka K, Valesini G, Hensor EMA, et al. Current evidence for the management of rheumatoid arthritis with biological disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of RA. Ann Rheum Dis. 2010 May 24;69(6):976–86.

135. Maini RN, Breedveld FC, Kalden JR, Smolen JS, Furst D, Weisman MH, et al. Sustained improvement over two years in physical function, structural damage, and signs and symptoms among patients with rheumatoid arthritis treated with infliximab and methotrexate. Arthritis Rheum. 2004;50(4):1051–65.

136. Maini RN, Breedveld FC, Kalden JR, Smolen JS, Davis D, MacFarlane JD, et al. Therapeutic efficacy of multiple intravenous infusions of anti-tumor necrosis factor α monoclonal antibody combined with low-dose weekly

methotrexate in rheumatoid arthritis. Arthritis Rheum. 1998 Sep 1;41(9):1552–63.

137. Boers M, Verhoeven AC, Markusse HM. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet. 1997 August 2; 350: 309-318

138. Goekoop-Ruiterman YPM, de Vries-Bouwstra JK, Allaart CF, van Zeben D, Kerstens PJSM, Hazes JMW, 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):3381–90.

139. Smolen JS, Landewé R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. Ann Rheum Dis 2013;0:1–18.

140. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann Rheum Dis 2016;75:3–15.

141. Silman A, Klareskog L, Breedveld F, bresnihan B, Maini R, van Riel P, et al. Proposal to establish a register for the long term surveillance of adverse events in patients with rheumatic diseases exposed to biological agents: the EULAR Surveillance Register for Biological Compounds. Ann Rheum Dis. 2000 Jun 1;59(6):419–20.

142. Silman A, Symmons D, Scott DGI, Griffiths I. British Society for Rheumatology Biologics Register. Ann Rheum Dis. 2003 Nov;62 Suppl 2:ii28–9.

143. Watson K, Hyrich K. The role of the British Society for Rheumatology Biologics Register (BSRBR) and the NICE guidelines for anti-TNF therapy. Musculoskeletal Care; 2003 Mar;1(1):58–64.

144. Ritchie DM, Boyle JA, McInnes JM, Jasani MK, Dalakos TG, Grieveson P, et al. Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. Q J Med. 1968 Jul;37(147):393–406.

145. Boers M, Tugwell P, Felson DT. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol. 1994; (suppl 41) 21:86-9.

146. Van der Heijde DM, van 't Hof MA, Van Riel PL, Theunisse LA, Lubberts EW, Van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity

score. Ann Rheum Dis. 1990 Nov;49(11):916–20.

147. Prevoo ML, van 't Hof MA, Kuper HH, Van Leeuwen MA, Van de Putte LB, Van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995 Jan;38(1):44–8.

148. Smolen JS, Breedveld FC, Schiff MH, Kalden JR, Emery P, Eberl G, et al. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology. 2003 Feb;42(2):244–57.

149. Aletaha D, Nell VPK, Stamm T, Uffmann M, Pflugbeil S, Machold K, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. Arthritis Res Ther. 2005;7(4):R796–806.

150. Hendrikx J, de Jonge MJ, Fransen J, Kievit W, Van Riel PL. Systematic review of patient-reported outcome measures (PROMs) for assessing disease activity in rheumatoid arthritis. RMD Open. 2016;2(2):e000202.

151. Pinals RS, Masi AT, Larsen RA. Preliminary criteria for clinical remission in rheumatoid arthritis. Arthritis Rheum. 1981;24(10):1308–15.

152. Prevoo ML, van 't Hof MA, Kuper HH, Van Leeuwen MA, Van de Putte LB, Van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995 Jan;38(1):44–8.

153. Fransen J, Stucki G, van Riel PLCM. Rheumatoid arthritis measures: Disease Activity Score (DAS), Disease Activity Score-28 (DAS28), Rapid Assessment of Disease Activity in Rheumatology (RADAR), and Rheumatoid Arthritis Disease Activity Index (RADAI). Arthritis Rheum. 2003 Oct 15;49(S5):S214–24.

154. Wells G, Becker J-C, Teng J, Dougados M, Schiff M, Smolen J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. Ann Rheum Dis. 2009 Jun;68(6):954–60.

155. Fleischmann R, van der Heijde D, Koenig AS, Pedersen R, Szumski A, Marshall L, et al. How much does Disease Activity Score in 28 joints ESR and CRP calculations underestimate disease activity compared with the Simplified Disease Activity Index? Ann Rheum Dis. 2014 Aug 20;:annrheumdis–2013–204920.

156. Tamhane A, Redden DT, McGwin G, Brown EE, Westfall AO, Reynolds RJ, et al. Comparison of the disease activity score using erythrocyte

sedimentation rate and C-reactive protein in African Americans with rheumatoid arthritis. J Rheumatol. 2013 Nov;40(11):1812–22.

157. Inoue E, Yamanaka H, Hara M, Tomatsu T, Kamatani N. Comparison of Disease Activity Score (DAS)28- erythrocyte sedimentation rate and DAS28- C-reactive protein threshold values. Ann Rheum Dis. 2007 Mar;66(3):407–9.

158. Rheumatoid arthritis in adults: management. NICE. 2016 Apr pp. 1–17. Available from: nice.org.uk/guidance/cg79 [Internet; Accessed 16th September 2017]

159. Prevoo ML, van 't Hof MA, Kuper HH, Van Leeuwen MA, Van de Putte LB, Van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995 Jan;38(1):44–8.

160. Van Riel PL, van Gestel AM. Clinical outcome measures in rheumatoid arthritis. Ann Rheum Dis. 2000 Nov;59 Suppl 1:i28–31.

161. Fransen J. Remission in rheumatoid arthritis: agreement of the disease activity score (DAS28) with the ARA preliminary remission criteria. Rheumatology. 2004 Jul 13;43(10):1252–5.

162. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed. NICE Technology appraisal guidance. 2016; TA 375. Available at https://nice.org.uk/guidance/ta375 [Accessed 27th September 2017].

163. Pincus T, Swearingen C, Wolfe F. Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. Arthritis Rheum. 1999 Oct;42(10):2220–30.

164. Pincus T, Furer V, Keystone E, Yazici Y, Bergman MJ, Luijtens K. RAPID3 (Routine Assessment of Patient Index Data 3) severity categories and response criteria: Similar results to DAS28 (Disease Activity Score) and CDAI (Clinical Disease Activity Index) in the RAPID 1 (Rheumatoid Arthritis Prevention of Structural Damage) clinical trial of certolizumab pegol. Arthritis Care Res. 2011 Aug;63(8):1142–9.

165. 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. 2011 Nov 7;63(S11):S14–S36.

166. Heiberg T, Austad C, Kvien TK, Uhlig T. Performance of the Rheumatoid

Arthritis Impact of Disease (RAID) score in relation to other patient-reported outcomes in a register of patients with rheumatoid arthritis. Ann Rheum Dis. 2011 Jun;70(6):1080–2.

167. Gossec L, Paternotte S, Aanerud GJ, Balanescu A, Boumpas DT, Carmona L, et al. Finalisation and validation of the rheumatoid arthritis impact of disease score, a patient-derived composite measure of impact of rheumatoid arthritis: a EULAR initiative. Ann Rheum Dis. 2011 Jun;70(6):935–42.

168. Choy EH, Khoshaba B, Cooper D, MacGregor A, Scott DL. Development and validation of a patient-based disease activity score in rheumatoid arthritis that can be used in clinical trials and routine practice. Arthritis Rheum. 2008 Feb 15;59(2):192–9.

169. Leung AMH, Farewell D, Lau CS, Choy EHS. Defining criteria for rheumatoid arthritis patient-derived disease activity score that correspond to Disease Activity Score 28 and Clinical Disease Activity Index based disease states and response criteria. Rheumatology. 2016 Oct 24;55(11):1954–8.

170. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LHD, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum. 2011 Feb 3;63(3):573–86.

171. Latremoliere A, Woolf CJ. Central sensitization: a generator of pain hypersensitivity by central neural plasticity. J Pain. 2009 Sep;10(9):895–926.

172. van Tuyl LHD, Hewlett S, Sadlonova M, Davis B, Flurey C, Hoogland W, et al. The patient perspective on remission in rheumatoid arthritis: 'You've got limits, but you're back to being you again'. Ann Rheum Dis. 2015 Jun;74(6):1004–10.

173. Kristensen LE, Bliddal H, Christensen R, Karlsson JA, Gülfe A, Saxne T, et al. Is swollen to tender joint count ratio a new and useful clinical marker for biologic drug response in rheumatoid arthritis? Results from a Swedish cohort. Arthritis Care Res. 2014 Jan 28;66(2):173–9.

174. Nguyen H, Ruyssen-Witrand A, Gandjbakhch F, Constantin A, Foltz V, Cantagrel A. Prevalence of ultrasound-detected residual synovitis and risk of relapse and structural progression in rheumatoid arthritis patients in clinical remission: a systematic review and meta-analysis. Rheumatology. 2014 Nov;53(11):2110–8.

175. Hess A, Axmann R, Rech J, Finzel S. Blockade of TNF-α rapidly inhibits pain responses in the central nervous system. PNAS. 2011 March 1; 108 (9) 3731–3736.

176. Rech J, Hess A, Finzel S, Kreitz S, Sergeeva M, Englbrecht M, et al.

Association of brain functional magnetic resonance activity with response to tumor necrosis factor inhibition in rheumatoid arthritis. Arthritis Rheum. 2013 Feb;65(2):325–33.

177. Boers M, Kirwan J, Tugwell P, Beaton D, Bingham C, Conaghan PG, et al. The OMERACT Handbook. OMERACT. 2014.

178. Singh JA, Hossain A, Tanjong Ghogomu E, Kotb A, Christensen R, Mudano AS, et al. Biologics or tofacitinib for rheumatoid arthritis in incomplete responders to methotrexate or other traditional disease-modifying anti-rheumatic drugs: a systematic review and network meta-analysis. Cochrane Database Syst Rev. 2016 May 13;(5):CD012183.

179. de Punder YMR, Fransen J, Kievit W, Houtman PM, Visser H, van de Laar MAFJ, et al. The prevalence of clinical remission in RA patients treated with anti-TNF: results from the Dutch Rheumatoid Arthritis Monitoring (DREAM) registry. Rheumatology. 2012 Sep;51(9):1610–7.

180. Smolen JS, Aletaha D, Bijlsma JWJ, Breedveld FC, Boumpas D, Burmester G, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis 2010;69:631–637.

181. Uhlig T, Kvien TK, Pincus T. Test-retest reliability of disease activity core set measures and indices in rheumatoid arthritis. Ann Rheum Dis. 2009 Jun;68(6):972–5.

182. Siemons L, Klooster ten PM, Vonkeman HE, Glas CAW, van de Laar MAFJ. Distinct trajectories of disease activity over the first year in early rheumatoid arthritis patients following a treat-to-target strategy. Arthritis Care Res. 2014 Mar 26;66(4):625–30.

183. Lopez-Olivo MA, Siddhanamatha HR, Shea B, Tugwell P, Wells GA, Suarez-Almazor ME. Methotrexate for treating rheumatoid arthritis. Cochrane Database Syst Rev. 2014 Jun 10;(6):CD000957.

184. Hyrich KL, Silman AJ, Watson KD, Symmons DPM. Anti-tumour necrosis factor alpha therapy in rheumatoid arthritis: an update on safety. Ann Rheum Dis. 2004 Dec;63(12):1538–43.

185. Goldman SA. Limitations and strengths of spontaneous reports data. Clin Ther. 1998;20 Suppl C:C40–4.

186. Pincus T. Limitations of randomized controlled clinical trials to depict accurately long-term outcomes in rheumatoid arthritis. Z Rheumatol. 1998 Feb;57(1):46–9.

187. Silman A, Klareskog L, Breedveld F. Proposal to establish a register for the long-term surveillance of adverse events in patients with rheumatic diseases exposed to biological agents: the EULAR Surveillance Register for Biological Compounds. Ann Rheum Dis. 2000;59:419–420.

188. Silman A, Symmons D, Scott DGI, Griffiths I. British Society for Rheumatology Biologics Register. Ann Rheum Dis. 2003 Nov;62 Suppl 2(Suppl 2):ii28–9.

189. Nikiphorou E, Buch MH, Hyrich KL. Biologics registers in RA: methodological aspects, current role and future applications. Nat Rev Rheum. 2017 Jun 1;23:S195

190. Little R. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc. 1988; 83 (404): 1198 - 1202.

191. Sargeant J. Concepts in missing data. [Presentation] British Society for Rheumatology Annual Conference 2016.

192. Andridge RR, Little RJA. A Review of Hot Deck Imputation for Survey Non-response. Int Stat Rev. 2010 Apr;78(1):40–64.

193. Bernoulli J. On the law of large numbers. 1713 [Translated in 2005 by Sheynin O].

194. Poisson SD. Probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilitiés. Paris; 1837.

195. Honaker J, King G, Blackwell M. Amelia II: A program for missing data. J Stat Softw. 2011 Dec; 45(7): 1 - 46.

196. Honaker J, King G. What to do about missing values in time-series cross-section data. Am J Polit Sci. 2010 Apr; 54 (2): 561 - 581.

197. Carlin JB, Li N, Greenwood P, Coffey C. Tools for analyzing multiple imputed datasets. The Stata Journal. 2003; 3 (3): 1 - 20.

198. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986 Feb 8;1(8476):307–10.

199. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20(1): 37 - 46.

200. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar; 33 (1): 159 - 174.

201. Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol. 2009 Nov 13;1(1):3–14.

202. Akaike H, Petrov BN, Csaki F. Information theory and an extension of the maximum likelihood principle. Proceeding of the Second International Symposium on Information Theory. 1973; 267 - 281.

203. Schoenberg IJ. Contributions to the problem of approximation of

equidistant data by analytic functions. Part A. On the problem of smoothing or graduation. A first class of analytic approximation formulae. Q Appl Math. 1946; 4(1): 45 - 99.

204. Wood SN. Generalized additive models: an introduction with R. Boca Rantom. 2006.

205. Bayes M, Price M. An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. Philosophical Transactions (1683-1775). 1763.

206. Galef J. A visual guide to Bayesian thinking. Available from: https://youtu.be/BrK7X_XlGB8 [Accessed 25th September 2017].

207. Kelly D, Smith C. More Complex Models for Random Durations. Bayesian Inference for Probabilistic Risk Assessment. Springer-Verlag; 2011: 39–50.

208. Schwarz G. Estimating the dimension of a model. Ann Stat. 1978; 6(2): 461 - 464.

209. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38(4): 963 - 974.

210. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: The R Package lcmm. J Stat Softw. 2017; 78(2), 1-56.

211. Cheng X, Cook D, Hofmann H. Visually exploring missing values in multivariable data using a graphical user interface. J Stat Softw. 2015;68(6):1–23.

212. Katchamart W, Johnson S, Lin H-JL, Phumethum V, Salliot C, Bombardier C. Predictors for remission in rheumatoid arthritis patients: A systematic review. Arthritis Care Res. 2010 Mar 16;62(8):1128–43.

213. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1.

214. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine. 2009; 6: p. e1000097.

215. Fransen J, Welsing P, De Keijzer R, Van Riel PCLM. Disease activity scores using C-reactive protein: CRP may replace ESR in the assessment of RA disease activity. Ann Rheum Dis. 2003; Available from: http://www.abstracts2view.com/eular/view.php?nu=EULAR03L1_2003THU0138 [Abstract. Accessed 9th August 2016]

216. Boers M, Kirwan JR, Tugwell P. The OMERACT Handbook. OMERACT. 2015.

217. Wells G, Shea B, O'Connell D, Peterson J, Welch V. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2008. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm [Accessed 19th November 2014].

218. Review Manager (RevMan)[Computer Program] Version 5.2. 3. Cochrane Collaboration. Copenhagen: The Nordic Cochrane Centre. 2014.

219. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. The Cochrane Collaboration. Available at http://handbook-5-1.cochrane.org/ [Accessed 27th September 2017].

220. Barnabe C, Homik J, Barr SG, Martin L, Maksymowych WP. The effect of different remission definitions on identification of predictors of both point and sustained remission in rheumatoid arthritis treated with anti-TNF therapy. J Rheumatol. 2014 Aug;41(8):1607–13.

221. Furst DE, Pangan AL, Harrold LR, Chang H, Reed G, Kremer JM, et al. Greater likelihood of remission in rheumatoid arthritis patients treated earlier in the disease course: Results from the Consortium of Rheumatology Researchers of North America registry. Arthritis Care Res. 2011 May 31;63(6):856–64.

222. Brocq O, Millasseau E, Albert C, Grisot C, Flory P, Roux C-H, et al. Effect of discontinuing TNF alpha antagonist therapy in patients with remission of rheumatoid arthritis. Joint Bone Spine. 2009 Jul;76(4):350–5.

223. Tanaka Y, Hirata S, Kubo S, Fukuyo S, Hanami K, Sawamukai N, et al. Discontinuation of adalimumab after achieving remission in patients with established rheumatoid arthritis: 1-year outcome of the HONOR study. Ann Rheum Dis. 2015 Feb;74(2):389–95.

224. Balogh E, Madruga Dias J, Orr C, Mullan R, Harty L, FitzGerald O, et al. Comparison of remission criteria in a tumour necrosis factor inhibitor treated rheumatoid arthritis longitudinal cohort: patient global health is a confounder. Arthritis Res Ther. 2013;15(6):R221.

225. Einarsson JT, Geborek P, Saxne T, Kapetanovic MC. Sustained remission in tumor necrosis factor inhibitor-treated patients with rheumatoid arthritis: a population-based cohort study. J Rheumatol. 2015 May;42(5):741–8.

226. Navarro-Millan I, Chen L, Greenberg JD, Pappas DA, Curtis JR. Predictors and persistence of new-onset clinical remission in rheumatoid arthritis patients. Semin Arth Rheu. 2013 Oct 1;43(2):137–43.

227. Kristensen LE, Saxne T, Geborek P. The LUNDEX, a new index of drug

efficacy in clinical practice: Results of a five-year observational study of treatment with infliximab and etanercept among rheumatoid arthritis patients in southern Sweden. Arthritis Rheum. 2006;54(2):600–6.

228. Wetteland P, Røger M, Solberg HE, Iversen OH. Population-based erythrocyte sedimentation rates in 3910 subjectively healthy Norwegian adults. A statistical study based on men and women from the Oslo area. J Intern Med. 1996 Sep;240(3):125–31.

229. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor. Health Technol Assess. 2010. Available from: https://www.nice.org.uk/guidance/ta195/resources/adalimumab-etanercept-infliximab-rituximab-and-abatacept-for-the-treatment-of-rheumatoid-arthritis-after-the-failure-of-a-tnf-inhibitor-82598558287813 [Accessed 24th September 2017]

230. Roubille C, Richer V, Starnino T, McCourt C, McFarlane A, Fleming P, et al. The effects of tumour necrosis factor inhibitors, methotrexate, non-steroidal anti-inflammatory drugs and corticosteroids on cardiovascular events in rheumatoid arthritis, psoriasis and psoriatic arthritis: a systematic review and meta-analysis. Ann Rheum Dis. 2015 Mar;74(3):480–9.

231. International Classification of Functioning, Disability and Health. 2001. Available at http://www.who.int/classifications/icf/en/ [Accessed 24th September 2017].

232. Almeida C, Choy EH, Hewlett S, Kirwan JR, Cramp F, Chalder T, et al. Biologic interventions for fatigue in rheumatoid arthritis. Cochrane Database Syst Rev. 2016 Jun 6;6:CD008334.

233. van Gestel AM, Haagsma CJ, Van Riel PL. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. Arthritis Rheum. 1998 Oct;41(10):1845–50.

234. Siemons L, Vonkeman HE, Klooster ten PM, van Riel PLCM, van de Laar MAFJ. Interchangeability of 28-joint disease activity scores using the erythrocyte sedimentation rate or the C-reactive protein as inflammatory marker. Clin Rheumatol. 2014 Jun;33(6):783–9.

235. Matsui T, Kuga Y, Kaneko A, Nishino J, Eto Y, Chiba N, et al. Disease Activity Score 28 (DAS28) using C-reactive protein underestimates disease activity and overestimates EULAR response criteria compared with DAS28 using erythrocyte sedimentation rate in a large observational cohort of rheumatoid arthritis patients in Japan. Ann Rheum Dis. 2007 Sep;66(9):1221–6.

236. DAS28 - How to calculate the DAS28 [Internet]. das-score.nl. Available from: http://www.das-score.nl/das28/en/difference-between-the-das-

and-das28/how-to-measure-the-das28/how-to-calculate-the-das28.html [Accessed 20th Sep 2016]

237. Eklund CM. Chapter 5 - Proinflammatory Cytokines in CRP Baseline Regulation. 1st ed. Vol. 48, Advances in Clinical Chemistry. Elsevier Inc; 2009.

238. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. Vol. 894, World Health Organization technical report series. 2000.

239. Fransen J, van Riel PLCM. The Disease Activity Score and the EULAR response criteria. Clin Exp Rheumatol. 2005 Sep;23(5 Suppl 39):S93–9.

240. Smolen JS. Treat-to-target: rationale and strategies. Clin Exp Rheumatol. 2012 Jul;30(4 Suppl 73):S2–6.

241. Aletaha D, Ward MM. Duration of rheumatoid arthritis influences the degree of functional improvement in clinical trials. Ann Rheum Dis. 2006 Feb;65(2):227–33.

242. Hyrich KL, Watson KD, Lunt M, Symmons DPM. Changes in disease characteristics and response rates among patients in the United Kingdom starting anti-tumour necrosis factor therapy for rheumatoid arthritis between 2001 and 2008. Rheumatology. 2011 Jan 1;50(1):117–23.

243. Smolen JS, Landewé R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Ann Rheum Dis 2010;69:964–975.

244. Manders SHM, Kievit W, Braakman-Jansen ALMA, Brus HLM, Hendriks L, Fransen J, et al. Determinants associated with work participation in patients with established rheumatoid arthritis taking tumor necrosis factor inhibitors. J Rheumatol. 2014 Jul;41(7):1263–9.

245. Verstappen SMM. Rheumatoid arthritis and work: The impact of rheumatoid arthritis on absenteeism and presenteeism. Best Pract Res Clin Rheumatol. 2015 Jun;29(3):495–511.

246. Levitsky A, Brismar K, Hafström I, Hambardzumyan K, Lourdudoss C, Van Vollenhoven RF, et al. Obesity is a strong predictor of worse clinical outcomes and treatment responses in early rheumatoid arthritis: results from the SWEFOT trial. RMD Open. 2017;3(2):e000458.

247. Gremese E, Carletto A, Padovan M, Atzeni F, Raffeiner B, Giardina AR, et al. Obesity and reduction of the response rate to anti-tumor necrosis factor α in rheumatoid arthritis: an approach to a personalized medicine. Arthritis Care Res. 2013 Jan;65(1):94–100.

248. Leffondre K, Abrahamowicz M. Modeling smoking history: a comparison

of different approaches. Am J Epidemiol 2002;156:813–823.

249. Hudson M, Lo E, Baron M, Steele R, Canadian Scleroderma Research Group. Modeling smoking in systemic sclerosis: a comparison of different statistical approaches. Arthritis Care Res. 2011 Apr;63(4):570–8.

250. Buch MH, Pavitt S, Parmar M, Emery P. Creative trial design in RA: optimizing patient outcomes. Nat Rev Rheum. 2013 Mar;9(3):183–94.

251. Detert J, Klaus P. Biologic monotherapy in the treatment of rheumatoid arthritis. Biologics. 2015;9:35–43.

252. Curtis JR, Yang S, Chen L, Park GS, Bitman B, Wang B, et al. Predicting low disease activity and remission using early treatment response to antitumour necrosis factor therapy in patients with rheumatoid arthritis: exploratory analyses from the TEMPO trial. Ann Rheum Dis. 2012 Feb;71(2):206–12.

253. Buch MH, Bingham SJ, Bryer D, Emery P. Long-term infliximab treatment in rheumatoid arthritis: subsequent outcome of initial responders. Rheumatology. 2007 Jul;46(7):1153–6.

254. Aletaha D. Nothing lasts forever-a critical look at sustained remission. Arthritis Res Ther. 2012; 14:116

255. Oncologists Partner with Watson on Genomics. Cancer Discovery 2015;5:788 [News article].

256. Bibault J-E, Giraud P, Burgun A. Big Data and machine learning in radiation oncology: State of the art and future prospects. Cancer Lett. 2016 Nov 1;382(1):110–7.

257. Asimov I. Foreword. In: Holdstock R. Encyclopedia of science fiction. Octopus Books; 1983.