AN ECONOMIC EVALUATION OF A BIOMARKER TEST TO STRATIFY TREATMENT FOR RHEUMATOID ARTHRITIS

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2017

SEAN P GAVAN

SCHOOL OF HEALTH SCIENCES

Contents

Title **Page**

Title **Page**

Title **Page**

Title **Page**

WORD COUNT

73,284 words.

List of Tables

List of Figures

List of Abbreviations

ACR	American College of Rheumatology
ADAb	Anti-drug antibody
AIC	Akaike information criteria
Anti-CCP	Anti-cyclic citrullinated peptide
bDMARD	Biologic disease-modifying anti-rheumatic drug
BIC	Bayesian information criteria
BMI	Body mass index
BRAGGSS	Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate
BSRBR	British Society for Rheumatology Biologics Register
BSRBR-RA	British Society for Rheumatology Biologics Register – Rheumatoid Arthritis
CBA	Cost-benefit analysis
CCA	Cost-consequences analysis
CCG	Clinical commissioning group
cDMARD	Conventional synthetic disease-modifying anti-rheumatic drug
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CEAF	Cost-effectiveness acceptability frontier
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CMA	Cost-minimisation analysis
CRP	C-reactive protein
CUA	Cost-utility analysis
DAP	Diagnostics Assessment Programme
DAS28	Disease Activity Score – 28 Joint Count
DES	Discrete event simulation
EED	Economic Evaluation Database
ELISA	Enzyme-linked immunosorbent assay
ENBS	Expected net benefit of sampling
ESR	Erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
EVPI	Expected value of perfect information
EVPPI	Expected value of partial perfect information
FMI	Fraction of missing information
HAQ	Health Assessment Questionnaire
HAQ-DI	Health Assessment Questionnaire – Disability Index
HTA	Health Technology Assessment
ICER	Incremental cost-effectiveness ratio

IFR	Individual funding request
INMB	Incremental net monetary benefit
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
mAb	Monoclonal antibody
MAR	Missing at random
MCAR	Missing completely at random
MCP	Metacarpophalangeal
MeSH	Medical Subject Headings
MICE	Multiple imputation by chained equations
MNAR	Missing not at random
MNL	Multinomial logistic regression
MRI	Magnetic resonance imaging
MTHFR	Methylenetetrahydrofolate reductase
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMB	Net monetary benefit
NSAID	Non-steroidal anti-inflammatory drug
ONS	Office for National Statistics
PIP	Proximal Interphalangeal
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	Probabilistic sensitivity analysis
PSS	Personal social services
QALY	Quality-adjusted life year
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies tool - 2
RA	Rheumatoid arthritis
RCT	Randomised controlled trial
ROC	Receiver operating characteristic
SRQR	Standards for Reporting Qualitative Research
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TNFi	Tumour necrosis factor-α inhibitor
TPMT	Thiopurine-methyltransferase
UK	United Kingdom
VAS	Visual analogue scale
VOI	Value of information

Glossary of Terms

2x2 table
A matrix that represents the accuracy of a medical test by presenting the number, and proportion, of patients with a true-positive, false-positive, true-negative, and false-negative test result (Macaskill et al., 2010).

Adalimumab
A specific tumour necrosis factor-α inhibitor that may be prescribed to patients with rheumatoid arthritis (European Medicines Agency, 2017).

Agency relationship
The relationship whereby one party (the agent) acts upon the objectives of another party (the principal). In the context of a routine treatment decision, the patient is the principal and the clinician is the agent (Williams, 1988; Propper, 1995).

Analytic validity (of a test)
The accuracy and reliability of a test when measuring a specific biomarker (Rogowski et al., 2009).

Anti-drug antibody
Developed during immunogenicity against a treatment. Anti-drug antibodies may (i) increase the clearance rate of a treatment and (ii) bind to a treatment and neutralise its therapeutic effect (Krieckaert et al., 2012).

Autoantibodies
Antibodies against an individual's own proteins.

Biologic disease-modifying anti-rheumatic drug
A protein-based treatment for patients with rheumatoid arthritis that acts against a specific target of inflammation.

Biomarker
A portmanteau of 'biological' and 'marker'. A biomarker is "…*a characteristic that can be objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention*" (National Institute for Health Biomarkers Definitions Working Group, 2001, p.91).

Biomarker test
A health technology that aims to reveal patient heterogeneity within routine clinical practice by detecting the presence of a biomarker.

Biosimilar
Biologic treatments with similar quality, safety, and efficacy of a reference biologic (Grabowski et al., 2014).

Care pathway
A pathway that describes the sequence of health technologies delivered to a patient over time in routine clinical practice (Brennan et al., 2000).

Clinical validity (of a test)
The ability of a biomarker to indicate a specific clinical status (Rogowski et al., 2009).

Clinical utility (of a test)
The ability to improve outcomes by making a treatment decision based on the result of a test (Rogowski et al., 2009).

Cluster-robust standard errors
Used when data are clustered (for example, patients within different hospitals) to account for the potential within-cluster correlation between observations.

Companion diagnostic
A biomarker test co-developed *ex ante* alongside a pharmaceutical and typically stated explicitly within the product label of a treatment (Milne et al., 2015).

Complementary diagnostic
A biomarker test that may be developed *ex post* as a stand-alone test after a pharmaceutical has achieved market access. Complementary diagnostics are not typically stated within the product label of a treatment (Milne et al., 2015).

Conceptual model
An abstract representation of a phenomenon of interest, often illustrated diagrammatically, to assist in determining the final structure of a *de novo* decision analytic model (Tappenden, 2014).

Cost-effectiveness acceptability curve
A graphical illustration of the probability that each alternative comparator strategy is relatively cost-effective (Y-axis) over a range of cost-effectiveness thresholds (X-axis) (Fenwick et al., 2001).

Cost-effectiveness acceptability frontier
A graphical illustration of the probability that the alternative with the highest net benefit is relatively cost-effective (Y-axis) over a range of cost-effectiveness thresholds (X-axis) (Fenwick et al., 2001).

Cost-effectiveness threshold
The additional cost that must be imposed on the budget for health care to displace one QALY elsewhere within the health care system (Claxton et al., 2015a).

Cut-off value (of a test)
The quantity of a biomarker in a sample, measured by a test, that distinguishes whether the test has a *positive* or *negative* result; the cut-off value of a test is related directly to its sensitivity and specificity (Macaskill et al., 2010).

Cytokines
Proteins secreted during an immune response that influence the interaction between, or behaviour of, specific cells (Feldmann, 2002).

Decision analytic model
A series of mathematical relationships that represent the progression of a patient's disease and the impact of a health technology on disease progression (Brennan et al., 2000). The output of a decision analytic model can be expressed in terms of the expected outcomes of interest for each alternative comparator strategy.

Decision problem
An explicit statement of the resource allocation decision under consideration (Roberts et al., 2012).

Decision uncertainty
The probability that an incorrect decision is made, in the context of resource allocation decisions for health care (O'Hagan et al., 2005).

Deterministic sensitivity analysis
To establish the sensitivity of the expected outcomes derived from an economic evaluation by performing a manual adjustment to the value(s) of a model's input parameter(s) (Briggs et al., 1999).

Diagnostic information
The information derived from the result of a medical test.

Discrete event simulation
A decision analytic modelling technique that simulates the histories of individual patients over time, characterised by the specific events that they may experience (Caro et al., 2016b).

Disease Activity Score – 28 Joint Count (DAS28)
A composite condition-specific outcome measure for rheumatoid arthritis used to determine a patient's level of disease activity (van Gestel et al., 1998).

Dominance
A health technology is *dominated* if a comparator strategy produces more health at a lower cost (Karlsson et al., 1996).

Drug interference
The inability to measure free anti-drug antibodies unless their quantity exceeds the amount of therapeutic drug within a serum sample (van Schouwenburg et al., 2013).

Early economic evaluation
An economic evaluation conducted early in the product lifecycle of a health technology. Such studies are typically characterised by limited evidence and substantial parameter uncertainty (Annemans et al., 2000).

Econometric analysis
A statistical analysis informed by economic theory (Tintner, 1953).

Economic evaluation
"…*the comparative analysis of alternative courses of action in terms of both their costs and consequences*" (Drummond et al., 2015, p.4). An *alternative* typically refers to a health technology and the *consequence* is typically expressed in terms of health benefits.

End-to-end evidence
A single study that incorporates all elements of a test-and-treatment strategy, by following a patient (i) from their observed test result, (ii) to a specific treatment decision, and then to their final (health and resource) outcomes (National Institute for Health and Care Excellence, 2011a).

Endogeneity
When an independent variable is correlated with the residual error term in a regression analysis. Endogeneity may occur due to (i) reverse-causality; (ii) measurement error; or (iii) omitted variable bias (Wooldridge, 2010).

EULAR response
Criteria of treatment response in clinical practice for patients with rheumatoid arthritis in England, measured six months after commencing any treatment. Patients may be classified as having achieved a *good, moderate, or no* EULAR response (van Gestel et al., 1998).

Extended dominance
An health technology is *extendedly dominated* if a linear combination of two alternatives produces more health at a lower cost (Karlsson et al., 1996).

Enzyme-linked immunosorbent assay
A biochemical assay that can be used to test for the presence of specific antibodies within a patient's serum sample (van Schouwenburg et al., 2013).

Health technology
An intervention used in the delivery of health care. For example, a pharmaceutical treatment, medical test, or device.

Heterogeneity (in health outcomes)
The variability in health outcomes can be explained by observed characteristics (Briggs et al., 2006).

Horizontal inequity
When patients of equal need receive unequal treatment (Wagstaff et al., 1991).

Immunogenicity
A process whereby the immune system of a patient incorrectly produces an immune response against a treatment itself, which may ultimately neutralise its therapeutic properties (Krieckaert et al., 2012).

Incidence
The rate of new cases within a specific population over a period of time.

Incremental cost
The difference in cost between two alternative interventions (Drummond et al., 2015).

Incremental cost-effectiveness ratio
The ratio of incremental costs to incremental health benefits (Drummond et al., 2015).

Incremental QALY
The difference in QALYs between two alternative interventions (Drummond et al., 2015).

Mapping algorithm
A technique to map between different instruments that measure clinical outcomes; for example, between a condition-specific and a generic outcome measure (Dakin, 2013).

Monte Carlo simulation
The process of random sampling.

Multiple imputation by chained equations
A method for handling missing data by randomly sampling values for the missing data, based on the observed data, and combining the results using Rubin's rules (White et al., 2011).

National Institute for Health and Care Excellence
The decision-making authority responsible for making recommendations regarding the allocation of population health care resources in England (National Institute for Health and Care Excellence, 2013a).

Omitted variable bias
When an independent variable is omitted from a regression that is correlated with a different independent variable and/or the dependent variable (Greene, 2012).

Opportunity cost
The benefit forgone from the next-best use of a specific resource. The opportunity cost of resource allocation decisions for health care can be expressed in the *health* benefits forgone (Claxton et al., 2015a).

Perspective
The scope of the costs that should be included in an economic evaluation. The perspective is typically defined by the budget constraint of the decision-maker. Examples include a *health care system* perspective and a *societal perspective* (Drummond et al., 2015).

Prescribing algorithm
A set of "if…then" statements that informs a subsequent prescribing decision based on relevant factors (such as a biomarker) at the time of a clinical decision (Schoenbaum et al., 1990).

Prevalence
The proportion of cases within a specific population at a specific period of time.

Probabilistic sensitivity analysis
To propagate joint parameter uncertainty through a decision analytic model by (i) characterising all input parameters as probability distributions and (ii) sampling values for all parameters by using Monte Carlo simulation (Doubilet et al., 1985).

Purposive sample
A non-random sample, used often in qualitative research, to enable an understanding of a phenomenon being researched (Silverman et al., 2008).

Quality-adjusted life year
A generic outcome measure of health benefit, calculated by multiplying each year of life by a weight that represents its health-related quality of life. Weights are calculated according to the reference points of one (full health) and zero (death); states worse than death are possible (Drummond et al., 2015).

Reference case
A pre-specified preferred criteria for conducting an economic evaluation. A reference case is typically an expression of a decision-maker's value judgements (Drummond et al., 1993).

Rheumatoid arthritis
A chronic, systemic inflammatory autoimmune disease characterised by (i) inflammation in the lining of the joints, and (ii) the progressive and irreversible destruction of joints and cartilage. Inflammation occurs predominantly within the hands and feet of a patient. Patients experience pain, a gradual decline in functional ability, and a reduction in quality of life. (Firestein, 2003; Kvien, 2004; Russell, 2008; Scott et al., 2010; McInnes et al., 2011).

Sampling frame
The individuals within a target population that were eligible for recruitment to a study (Morgan, 2008).

Secondary non-response
When a patient with rheumatoid arthritis loses response to a treatment, such as a tumour necrosis factor-α inhibitor, after having responded for at least six months previously (Jani et al., 2015b).

Semi-empirical medicine
To treat all patients equally and make adjustments by trial-and-error (Woodcock, 2007).

Semi-structured interview
A strategy for qualitative data collection that poses a series of predetermined questions, based on the research objectives, which may be asked in any order to guide the structure of an interview (Ayres, 2008).

Sensitivity (of a test)
The proportion of patients with a positive test result, of those patients who truly have the biomarker of interest (Macaskill et al., 2010).

Specificity (of a test)
The proportion of patients with a negative test result, of those patients who truly do not have the biomarker of interest (Macaskill et al., 2010).

Stratified medicine
To (i) identify subgroups of patients that share a pre-defined patient-level characteristic associated with a clinical outcome, and (ii) to make a subsequent treatment decision conditional on each patient's subgroup membership (Trusheim et al., 2007).

Thematic analysis
A qualitative method of data analysis whereby the researcher takes an active role in identifying *themes* within a set of data (Braun et al., 2006).

Therapeutic drug level
The quantity of treatment (for example, a tumour necrosis factor-α inhibitor) circulating within a patient's serum (Pouw et al., 2015).

Tumour necrosis factor-α
The cytokine responsible for the inflammatory disease process in most patients with rheumatoid arthritis (Feldmann, 2002).

Value of information
A set of methods, derived from statistical decision theory, that quantifies the potential value of producing further prospective research to reduce the parameter (and decision) uncertainty associated with making decisions based on expected outcomes (Wilson, 2015).

Variability (in health outcomes)
The chance occurrence of a clinical outcome (Briggs et al., 2006).

Abstract

Health care policy, systems, and investment in related research have increasingly advocated for the *personalisation* of routine treatment decisions in order to improve population health outcomes. Stratified medicine may facilitate such personalisation by targeting specific health technologies to subgroups of patients defined by their individual-level characteristics. The management of patients with chronic diseases, in particular, such as rheumatoid arthritis, may be revolutionised by stratified medicine if treatments are subsequently prescribed conditional on patient-level heterogeneity. Economic evidence is essential to demonstrate that stratified medicine is a cost-effective use of finite health care resources to support its expansion within the National Health Service in England.

The aim of this thesis was to provide evidence for the relative cost-effectiveness of a biomarker test to stratify treatment for patients with rheumatoid arthritis, consistent with the requirements of decision-makers for the National Health Service in England. A specific case study of stratified medicine was selected (adalimumab anti-drug antibody and drug level testing) that was early in its product lifecycle and had an emerging, but limited, supporting evidence base. The thesis addressed three specific research questions:

(i) What was the existing economic evidence for stratified medicine in rheumatoid arthritis?

(ii) How were treatment decisions with biologic therapies made for patients with rheumatoid arthritis in current practice in England?

(iii) Are treatment decisions stratified by adalimumab anti-drug antibody and drug level testing, for patients with rheumatoid arthritis in England, a relatively cost-effective use of health care resources?

The research questions were answered by using a mixed methods approach (systematic reviews; qualitative thematic framework analysis; quantitative econometric analysis; decision analytic modelling by discrete event simulation). The individual studies within the thesis built towards the overall aim sequentially and were linked together by a common theme of generating relevant evidence to inform decision-making for a new stratified medicine in England. A *de novo* early model-based economic evaluation of adalimumab anti-drug antibody and drug level testing for patients with rheumatoid arthritis in England was conducted, informed by (i) the limitations of published economic evaluations of stratified medicine for rheumatoid arthritis, (ii) a substantial characterisation of the care pathways in current practice for patients with rheumatoid arthritis in England, and (iii) an extensive process to conceptualise the structure of the decision analytic model.

Stratified medicine by adalimumab anti-drug antibody and drug level testing for patients with rheumatoid arthritis in England was found not likely to be a relatively cost-effective use of health care resources based on current evidence. There was considerable decision uncertainty associated with this result and further prospective research, in particular on test accuracy and health outcomes, were found to be of substantial value to the health care system.

This thesis made clear contributions to knowledge by advancing the economic evidence base for stratified medicine in rheumatoid arthritis and by demonstrating the value and challenges of generating economic evidence early within the product lifecycle of a new stratified medicine.

Declaration

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.

A study is reported in *Appendix 35* that was co-authored with Dr. Meghna Jani. This study was published in *Rheumatology* and was supplementary to the work submitted in this thesis.

Copyright Statement

(i) The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University Manchester certain rights to use such Copyright, including for administrative purposes.

(ii) Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

(iii) The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

(iv) Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

Acknowledgements

Foremost, I would like to thank my supervisory team, Prof. Katherine Payne and Prof. Anne Barton, for their guidance, generosity of time, and support throughout the PhD process.

Thank you to the eleven rheumatologists, who will remain anonymous, for choosing to voluntarily participate in the qualitative study reported in *Chapter Three*. Additionally, thank you to Dr. Gavin Daker-White for imparting his expert advice on the analysis and reporting of these qualitative data.

Many thanks to Dr. Nick Bansback and Dr. Mark Harrison for hosting my research visit to *The University of British Columbia* in Vancouver, Canada during October 2015, and for providing advice on decision analytic modelling in rheumatoid arthritis and the econometric analysis of patient-level data, respectively.

Thank you to Dr. Meghna Jani for sharing her knowledge regularly on the clinical research and application of testing for immunogenicity against adalimumab.

I am indebted to the following individuals for agreeing to second-screen abstracts within the systematic reviews presented in the thesis: Rose Atkins; Martin Eden; Niall Davison; Cheryl Jones; Dr. Caroline Vass; Stuart Wright.

Thank you to the *International Society for Pharmacoeconomics and Outcomes Research* and the *European League Against Rheumatism* for awarding me with competitive travel grants to present research from this thesis at their respective international conferences.

Lastly, I am very grateful to the *National Institute for Health Research* (NIHR) for supporting the research presented in this thesis through a *Manchester Musculoskeletal Biomedical Research Unit* PhD studentship. The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health.

The Author

Sean Gavan has the following previous qualifications:

- Master of Science in Health Economics (*Distinction*)
- Master of Science in Economics (*Distinction*)
- Batchelor of Arts in Economics for Business (*First-class honours*)

Dissemination

The dissemination of the research in this thesis, to date, is described below in terms of peer-reviewed publications, conference presentations, and invited academic seminars. A complete publication strategy to disseminate the research further is reported in *Appendix 1*.

Peer-reviewed Publications

The research in *Chapter Two* was published in the following peer-reviewed manuscript:

- Gavan, S. et al. (2014). "Economics of Stratified Medicine in Rheumatoid Arthritis", *Current Rheumatology Reports*, Vol. 16, 12(468), pp. 1-11.

Conference Presentations

The research in *Appendix 34* has been accepted for presentation at the following international conferences:

- Gavan, S. et al. (Forthcoming). *Measuring Adalimumab Drug Levels by ELISA to Detect Treatment Response in Rheumatoid Arthritis: A Systematic Review and Bivariate Meta-analysis*. International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 22nd Annual International Meeting. Boston, USA, 23 May 2017.

- Gavan, S. et al. (Forthcoming). *A Systematic Review and Bivariate Meta-analysis of Studies that Measured Adalimumab Drug Levels by ELISA to Detect Treatment Response in Rheumatoid Arthritis*. European League Against Rheumatism (EULAR) Annual European Congress of Rheumatology. Madrid, Spain, 16 June 2017.

The research in *Chapter Three* was presented at the following international conference:

- Gavan, S. et al. (2016). *Exploring Factors which Influence Anti-TNF Treatment Decisions for Rheumatoid Arthritis in England – A Qualitative Analysis*. European League Against Rheumatism (EULAR) Annual European Congress of Rheumatology. London, England, 10 June 2016.

The research in *Chapter Four* was presented at the following international conference:

- Gavan, S. et al. (2016). *Identifying Factors which Influence the Selection of Anti-TNF Treatment in Patients with Rheumatoid Arthritis in England*. International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 21st Annual International Meeting. Washington DC, USA, 24 May 2016.

Invited Academic Seminars

The research in *Chapter Three* and *Chapter Four* was presented at the following invited academic seminar:

- Gavan, S. (2016). *What's Influencing Treatment Decisions with Biologic Therapies in Patients with Rheumatoid Arthritis in England?* Seminar for the Academic Unit of Health Economics' Seminar Series, The University of Leeds, 5 July 2016.

Chapter 1
Introduction

Health care systems around the world have expressed, with growing interest, a common desire to improve health outcomes by tailoring treatment decisions to the characteristics of individual patients (The Academy of Medical Sciences, 2013). Different terms have been used to describe the nuances of how treatment decisions may be tailored, including, but not limited to, *personalised medicine*, *precision medicine*, and *individualised medicine* (Schleidgen et al., 2013; Pokorska-Bocci et al., 2014; Pearson, 2016). *Stratified medicine* perhaps best encapsulates the practical application of these concepts to treatment decisions made in the routine practice of health care. The aim of stratified medicine is to inform a specific treatment decision by first identifying subgroups of patients, who share a pre-defined patient-level characteristic (known as a *biomarker*), associated with a particular clinical outcome such as treatment response or an adverse drug reaction (Trusheim et al., 2007). Each patient's specific treatment decision is then subsequently made conditional on their subgroup membership (Trusheim et al., 2007). The conventional approach to making a treatment decision, by contrast, can be characterised as *semi-empirical*, whereby all patients are treated equally and subsequent adjustments are made by trial-and-error (Woodcock, 2007). Stratified medicine may improve population health outcomes, relative to the conventional semi-empirical approach, by treating only those patients likely to respond and adjusting treatment for those patients likely to experience harm (Trusheim et al., 2007). Stratified medicine may improve the management of patients with chronic diseases, in particular, such as rheumatoid arthritis (RA), some of whom, conditional on their subgroup membership, may have the potential to otherwise receive therapies for many years that are relatively costly and/or relatively less effective or safe. The focus of this thesis was to generate economic evidence that evaluated a specific example of stratified medicine for patients with RA in England.

The mantra of stratified medicine has become increasingly embedded within the discourse of national and international health policy. For example, in 2013, the *World Health Organization* produced a policy document that highlighted the importance of stratified medicine to facilitate patient access to treatments from their list of global priority medicines (Kaplan et al., 2013). The *Department of Health* in the United Kingdom (UK), also in 2013, launched a separate company called *Genomics England*, to collect whole genome data from 100,000 patients in the National Health Service (NHS) (Genomics England, 2015). The innovation agency of the UK Government, *Innovate UK*, have provided funding for the *Precision Medicine Catapult* to engender collaborations between medical test manufacturers, academia, and the NHS (Innovate UK, 2016). Finally, in 2015, the Government of the United States of America proposed an investment of $213 million to support the collection of linked biomarker and health record data from one million citizens (Collins et al., 2015; McCarthy, 2015).

The positive signal for stratified medicine as an agenda for health policy has encouraged substantial investment in related research within the UK. For example, the *Medical Research Council's Stratified Medicine Initiative* (Medical Research Council, 2016) has provided £60 million to support UK-wide research consortia across different disease areas, including RA (Barton et al., 2016). The prospective *UK Biobank* database, supported by research funding, has collected biomarker samples from 500,000 NHS patients to investigate the genetic and non-genetic determinants of disease (Sudlow et al., 2015). The *National Institute for Health Research*, funded by the *Department of Health* in the UK, has supported research into stratified medicine through investment in *Biomedical Research Centres* and *Diagnostic Evidence Cooperatives* (National Institute for Health Research, 2014). UK-wide non-profit organisations, such as the *UK Pharmacogenetics and Stratified Medicine Network*, have been launched to promote collaborative research across industrial sectors into implementing stratified medicine within the NHS (McNamme, 2016). Established national cohort studies, such as *Understanding Society* (formerly the *British Household Panel Survey*) have been increasingly investing in the collection of supplementary individual-level biomarker data for the purpose of novel empirical research (Benzeval et al., 2016).

Yet despite the high priority for health policy and the substantial investment in related research, few examples of stratified medicine have successfully translated into routine health care practice to date (Davis et al., 2009; Blair et al., 2011). Any new *health technology* (such as a pharmaceutical treatment, medical test, or device) faces two

translational gaps: (i) from basic research to market approval as a medical product, and (ii) from market approval to being recommended for routine use in the health care system (Becla et al., 2011), which ultimately relies on a favourable *economic* assessment by health care payers and decision-makers (Rogowski et al., 2008). Given that the resources available for health care in the NHS are finite, a new stratified medicine must provide evidence of its relative cost-effectiveness before being recommended for use in routine clinical practice (Drummond et al., 2015). The aim of this thesis was as follows:

Thesis aim: *to provide evidence for the relative cost-effectiveness of a biomarker test to stratify treatment for patients with RA, consistent with the requirements of decision-makers for the NHS in England.*

This introductory chapter presents a general background on economic evaluation in health care (Section 1.1), the economic evaluation of stratified medicine, in particular (Section 1.2), and the rationale and case study for stratified medicine in RA (Section 1.3). Section 1.4 reports the research questions and overall structure of the thesis.

1.1. Economic Evaluation in Health Care

There are multiple objectives facing the NHS in England but its primary objective is to improve health outcomes (Department of Health, 2016a). It is not possible, however, to recommend every effective health technology within the NHS because the annual national budget for health care is finite (Cylus et al., 2015). Decision-makers must therefore choose which health technologies to recommend in order to achieve the objective of population health maximisation, subject to the prevailing constraint on resources (National Institute for Health and Care Excellence, 2013a). The *opportunity cost* of any resource allocation decision is equivalent to the benefit forgone from the next-best use of those resources (Palmer et al., 1999a). In the context of allocating resources to recommend a specific health technology in the NHS, the opportunity cost can be expressed as the health benefit forgone if those same resources were used to provide a different (effective) health technology (Claxton et al., 2015a).

The National Institute for Health and Care Excellence (NICE) was established in 1999 (Buxton, 2006) as "*the agent of a socially legitimate higher authority [Government]*" (Claxton et al., 2010, p.16), responsible for population health care resource allocation decisions in England, whose objectives are set by the Secretary of State for Health

(Department of Health, 2014). NICE uses economic evidence, that makes the expected opportunity cost of any decision explicit (Drummond et al., 2015), to inform resource allocation recommendations during their health technology appraisal and guideline development processes (National Institute for Health and Care Excellence, 2013a). Economic evaluations are the primary method to provide such economic evidence, conventionally defined as, "…*the comparative analysis of alternative courses of action in terms of both their costs and consequences*" (Drummond et al., 2015, p.4). An *alternative* typically refers to a health technology and the *consequence* of concern to decision-makers is typically the health benefit derived from each alternative.

Evidence from an economic evaluation can provide information about how to improve the (i) *technical efficiency* of a specific health care intervention; for example, by identifying the fewest resources required to achieve a pre-specified health outcome (Shiell et al., 2002), and the (ii) *allocative efficiency* of population health care resources by enabling decision-makers to only recommend health technologies if their expected benefit exceeds their opportunity cost, consistent with the notion of population health maximisation within a finite budget for health care (Palmer et al., 1999b).

1.1.1. The Methods of Economic Evaluation

A *decision problem* is an explicit statement of the resource allocation decision under consideration (Roberts et al., 2012). Three predominant methods of economic evaluation, that may be used to inform an explicit decision problem, are summarised in Table 1.1 in terms of their defining characteristics.

Table 1.1. *Defining characteristics of three methods of economic evaluation.*

Economic Evaluation Method	Valuation of Costs	Valuation of Health Consequences
Cost-benefit analysis.	Monetary units.	Monetary units.
Cost-effectiveness analysis.	Monetary units.	Natural units.
Cost-utility analysis.	Monetary units.	Generic outcome measure.

Source: Drummond et al. (2015, p.11).

Each method of economic evaluation estimates the cost of a health technology in monetary units. Costs comprise two components: (i) an estimated quantity of resources, and (ii) a price at which those resources are valued (Drummond et al., 2015). The *incremental cost* is the difference in cost between two alternative health technologies. The *perspective* of an economic evaluation defines the scope of the costs that should be included in the analysis, which are typically those that fall on the decision-maker's budget constraint (Drummond et

al., 2015). For example, a *health care system perspective* will include all direct medical costs, whereas a *societal perspective* may include a wider set of costs, irrespective of who bears them, such as the societal cost of patients' reduced labour productivity or patients' own out-of-pocket expenditures (Weinstein, 1990).

Value judgements are necessary, given the objective of health maximisation, to define *how* health should be measured and *by whom* it should be valued (Weinstein et al., 1996). The three methods of economic evaluation differ in their underlying value judgements and, in turn, their measure of benefit and relevance to specific decision problems (Drummond et al., 2015). A cost-benefit analysis (CBA) is broadly consistent with welfarist economic theory and values the benefit of a health technology in monetary units as a function of an individual's utility (Coast et al., 2008). A cost-effectiveness analysis (CEA) and cost-utility analysis (CUA), by contrast, are consistent with an extra-welfarist evaluative framework (Brouwer et al., 2008). A CEA measures health outcomes in natural units (for example, binary endpoints) (Johannesson et al., 1996). However, as most health technologies affect survival and health status, a generic outcome measure is required that incorporates both quantity and quality of life, to facilitate comparisons across diseases (Drummond et al., 2015). A CUA, itself a form of CEA, measures benefits in terms of quality adjusted life-years (QALYs), which are a generic outcome measure of health benefit (Drummond et al., 2015). The term *cost-effectiveness analysis*, rather than *cost-utility analysis*, is used frequently within the literature as a more-common nomenclature for an economic evaluation that has measured health outcomes in terms of QALYs (National Institute for Health and Care Excellence, 2013a).

QALYs are estimated by multiplying each year of life by a weight representing its *health-related quality of life* (Brazier et al., 1999). Weights are assigned according to the reference points of one (full health) and zero (death), and health states worse than death are assumed to be possible (Drummond et al., 2015). QALY weights are estimated according to preferences elicited from the general public to ensure that desirable health states receive a higher weight (Drummond et al., 2015). The EQ-5D (EuroQol Group, 1990), SF-6D (Brazier et al., 2002), and HUI3 (Horsman et al., 2003) are three generic multi-attribute instruments that are used commonly within the literature to classify health states according to different dimensions relevant to quality of life (Drummond et al., 2015). The *incremental benefit* is the difference in QALYs between two alternative health technologies.

Costs and QALYs are conventionally discounted to their present value (Drummond et al., 2015). A positive time preference (the present is preferred to the future) implies that the present value of costs and health outcomes (derived in the future) will be less than at the time when they are realised (O'Mahony et al., 2015b). In practice, decision-makers express a preferred rate at which future outcomes should be discounted (O'Mahony et al., 2015b). The discount rate expressed by NICE has varied over time (Claxton et al., 2011; O'Mahony et al., 2014) and is currently recommended to be 3.5% per year for costs and QALYs (see *Appendix 2*) (Claxton et al., 2006b; Claxton et al., 2011; Paulden et al., 2012; National Institute for Health and Care Excellence, 2013a).

Two related methods of evaluation, generally regarded as unsuitable to inform population health care resource allocation decisions, are (i) cost-minimisation analysis (CMA) and (ii) cost-consequences analysis (CCA). A CMA assumes that the benefits derived from alternatives are equivalent, and that the choice between alternatives should be made according to the lowest cost (Briggs et al., 2001). However, the CMA method has been deemed inappropriate for decision-making because, in the presence of uncertainty, it is not possible to claim *a priori* that two alternatives produce equivalent health benefits (Briggs et al., 2001). A CCA presents decision-makers with all potential costs and benefits of each alternative in a disaggregated list (Mauskopf et al., 1998); the CCA method may be inappropriate to inform population health care resource allocation decisions because its interpretation often lacks transparency, given that decision-makers must implicitly trade-off the different benefits between competing alternatives (Owens et al., 2017).

1.1.2. The Standardisation of Economic Evidence

NICE has an explicit mandate to recommend health technologies that maximise population health subject to the NHS budget for health care (Culyer, 1997; Claxton et al., 2010), and achieves this by appraising health technologies in terms of their relative effectiveness and cost-effectiveness (National Institute for Health and Care Excellence, 2013a). Final recommendations arise from a deliberative process that balances contestable value judgements and evidence (Culyer et al., 2006; Shah et al., 2013).

The standardisation of the methods for conducting an economic evaluation can facilitate comparability between evaluations; a *reference case* is a pre-specified preferred criteria for conducting an economic evaluation that will typically represent an expression of a decision-maker's specific value judgements (Drummond et al., 1993; Sanders et al., 2016).

The NICE Reference Case (reported in *Appendix 2*), for example, includes the value judgements that (i) a QALY is an appropriate outcome measure of health benefit; (ii) the EQ-5D, in particular, can appropriately characterise all relevant aspects of health-related quality of life; and (iii) QALYs are valued equally, irrespective of who gains or loses them (Rawlins et al., 2004). The economic evaluation conducted in this thesis (reported in *Chapter Six*) conformed to the NICE Reference Case, in order to generate evidence that was relevant to decision-makers in England.

1.1.3. Decision Analytic Modelling

The decision-making process at NICE, in accordance with the NICE Reference Case, requires evidence that has simultaneously (i) accounted for all relevant comparators; (ii) included all relevant sources of evidence; (iii) appropriately characterised uncertainty; (iv) been estimated over a sufficient duration of time; and (v) has direct relevance to the decision-making context in the NHS (Buxton et al., 1997; Brennan et al., 2000; Sculpher et al., 2006a). An economic evaluation conducted alongside a single randomised controlled trial (RCT), by prospectively collecting resource use and QALY data (Glick et al., 2015), may be limited by its ability to meet the evidential requirements of the NICE Reference Case (National Institute for Health and Care Excellence, 2013a). For example, it may not be feasible to include all relevant comparators within a single RCT (Buxton et al., 1997; Brennan et al., 2000; Sculpher et al., 2006a). Model-based economic evaluations are therefore a fundamental component of the decision-making process at NICE (Akehurst, 2003; Bryan et al., 2007) because of their ability to synthesise multiple sources of evidence in a format complicit with the NICE Reference Case (National Institute for Health and Care Excellence, 2013a). The economic evaluation in this thesis (see *Chapter Six*) was therefore conducted by producing a *de novo* decision analytic model to generate evidence consistent with the requirements of NICE.

A *decision analytic model* is a series of mathematical relationships that represent the progression of a patient's disease and the impact of a health technology on disease progression (Brennan et al., 2000). The structure of a decision analytic model is typically representative of a relevant *care pathway* that describes the sequence of health technologies delivered to a patient over time in routine clinical practice (Brennan et al., 2000; Brennan et al., 2006). The *parameters* of a decision analytic model are the specific inputs whose values may be estimated from existing evidence (Briggs, 2000). For example, different potential sources of evidence for different types of input parameter are reported in

Table 1.2 (Zechmeister-Koss et al., 2014). The output of a decision analytic model can be expressed in terms of the expected outcomes (for example, mean costs and QALYs) derived from each alternative comparator strategy (Briggs et al., 2006).

Table 1.2. *Example input parameters and sources of evidence for a decision analytic model.*

Model Input Parameter	Potential Source of Evidence
Clinical effectiveness.	Randomised controlled trial; Meta-analysis.
Natural history of disease.	Observational cohort study.
Resource use.	Microcosting study; Direct observation; Clinical guidelines.
Unit costs.	National price lists.
Health-related quality of life.	Published databases; Mapping algorithms.

Source: Adapted from Zechmeister-Koss et al. (2014, p.293).

Three types of decision analytic model observed most frequently in the literature are decision trees, Markov models, and discrete event simulations of individual patients (Barton et al., 2004a; Brennan et al., 2006; Briggs et al., 2006); these three different types of decision analytic model are described further in *Appendix 3*. The choice of decision analytic model, and its structure, should be justified by an explicit and transparent model conceptualisation process (Roberts et al., 2012; Tappenden, 2014). Therefore, *Chapter Five* reports an extensive conceptualisation of the *de novo* decision analytic model in this thesis.

1.1.4. Decision Rules for Relative Cost-effectiveness

Three equivalent decision rules (reported in Table 1.3) can be applied to the output of a decision analytic model to inform whether a health technology is cost-effective (assuming the objective of health maximisation), relative to an alternative strategy (Claxton et al., 2010).

Table 1.3. *Decision rules for relative cost-effectiveness.*

Decision Rule	Definition
$\frac{\Delta C}{\Delta H} < \lambda$	The ratio of incremental costs to incremental health consequences is less than the cost-effectiveness threshold.
$\Delta H\lambda - \Delta C > 0$	The incremental net monetary benefit is greater than zero.
$\Delta H - \frac{\Delta C}{\lambda} > 0$	The incremental net health benefit is greater than zero.

Source: Adapted from Claxton et al. (2010; pp.16-17); Note: ΔC=incremental costs; ΔH=incremental QALYs; λ=cost-effectiveness threshold.

The *incremental cost-effectiveness ratio* (ICER) is the ratio of incremental costs (ΔC) to incremental health benefits (ΔH) between two alternatives ($\frac{\Delta C}{\Delta H}$). A cost-increasing health technology is said to be relatively cost-effective if its ICER is below the value of a cost-effectiveness threshold (λ) (Claxton et al., 2010). Alternative strategies are *dominated* if they produce less health at a higher cost compared with a different alternative, or *extendedly dominated* if a linear combination of two different alternatives produces more health at a lower cost (Karlsson et al., 1996). The decision rule can be expressed equivalently as an incremental net benefit by converting health and costs into the same units (Stinnett et al., 1998). A positive incremental net benefit, whether expressed in terms of money or health, is indicative of relative cost-effectiveness (Claxton et al., 2010).

1.1.5. The Cost-effectiveness Threshold

The cost-effectiveness threshold, in theory, represents the additional cost that must be imposed on the budget for health care to displace one QALY elsewhere within the health care system (Claxton et al., 2015a). The QALYs gained from an intervention health technology must be greater than its opportunity cost (the QALYs displaced elsewhere) in order to maximise population health (McCabe et al., 2008). However, the actual value of the cost-effectiveness threshold in the NHS is not known with certainty.

The most recent empirical estimate of the cost-effectiveness threshold for the NHS in England was £12,936 per QALY gained (Claxton et al., 2015a). Previous appraisal decisions by NICE have recommended health technologies with ICERs estimated in excess of £30,000 per QALY gained (Devlin et al., 2004; Dakin et al., 2015). The relative cost-effectiveness of a health technology is therefore neither a necessary nor sufficient condition for it to be recommended within the NHS because other factors (such as equity in the distribution of QALY gains) may also inform NICE's deliberative decision-making process (Shah et al., 2013).

In practice, NICE assumes that the cost-effectiveness threshold for the NHS falls within a plausible range £20,000 and £30,000 per QALY gained (National Institute for Health and Care Excellence, 2013a). The higher a health technology's estimated ICER, the lower the likelihood that it will be recommended by NICE (Rawlins et al., 2004). Exceptions may be made for health technologies used at the end of a patient's life; a cost-effectiveness threshold of £50,000 per QALY gained may be applied if the health technology (i) is suitable for patients with a life expectancy of less than twenty-four months; (ii) may extend

life by at least three months; and (iii) is indicated for a small patient population (Paulden et al., 2014b).

1.1.6. Decision-making under Uncertainty

Uncertainty is inherent in all decisions regarding the relative cost-effectiveness of health technologies and is paramount to the decision-making process at NICE (Claxton et al., 2005; National Institute for Health and Care Excellence, 2013a). *Decision uncertainty* is defined as the probability that an incorrect decision is made, by recommending a health technology that is not relatively cost-effective (O'Hagan et al., 2005). The consequence of an incorrect decision is an inefficient allocation of health care resources and, in turn, the health forgone as a consequence (O'Hagan et al., 2005; Stevenson et al., 2014).

The use of decision analytic modelling to characterise uncertainty in terms of probability statements can be regarded as an application of Bayesian reasoning (Briggs, 1999; Luce et al., 1999); for example, by estimating the probability that an alternative strategy is cost-effective *given the observed data* (O'Hagan et al., 2003; Shih, 2003; Fenwick, 2014). When a decision must be made (that includes *doing nothing*), the approach taken by NICE is to make recommendations for a new health technology according to the relative magnitude of probable expected outcomes, and not according to an inferential test designed to reject a null hypothesis of "no difference" between two alternatives (Arrow et al., 1970; Claxton, 1999b; Luce et al., 1999).

1.1.6.1. Defining Uncertainty

There are three types of uncertainty inherent in any model-based economic evaluation, defined in Table 1.4.

Table 1.4. *Types of uncertainty in decision analytic models.*

Type of Uncertainty	Definition
Methodological uncertainty.	Uncertainty in the methods of conducting an economic evaluation.
Structural uncertainty.	Uncertainty in the conceptual and mathematical representation of a decision problem.
Parameter uncertainty.	Uncertainty in the values of a model's input parameters.

Source: Stevenson et al. (2014, p.62).

Methodological uncertainty can be minimised by using a reference case that describes the appropriate methods of an economic evaluation (Briggs, 2000). Structural uncertainty can be minimised by undertaking a thorough model conceptualisation exercise *a priori* (Stevenson et al., 2014; Tappenden, 2014). Parameter uncertainty exists because the true value of any input parameter cannot be known with certainty (Briggs et al., 1999). Deterministic sensitivity analysis methods can be used to assess how the expected outcomes of a model are affected by manual adjustments to the values of one or more input parameters (Briggs et al., 1999). However, as all input parameters are jointly uncertain, a *probabilistic sensitivity analysis* (PSA) is recommended by NICE to characterise the joint uncertainty in all parameters of a model appropriately (Claxton et al., 2005; Claxton, 2008; National Institute for Health and Care Excellence, 2013a).

1.1.6.2. Probabilistic Evaluation of a Decision Analytic Model

A PSA characterises the input parameters of a decision analytic model as probability distributions, rather than point estimates, to represent the uncertainty in their true values (Doubilet et al., 1985; Baio et al., 2015). Only a subset of probability distributions, which are described further in *Appendix 4*, are appropriate to characterise the uncertainty in a model's parameters. *Monte Carlo simulation* (random sampling) then samples values for all input parameters from their respective distributions and the model estimates the expected outcomes of interest (costs, QALYs, and net benefits) for each comparator strategy. This whole process is repeated many times to produce a distribution of outcomes that is representative of the joint uncertainty of the model's parameters (Briggs et al., 1999; Claxton et al., 2005; Claxton, 2008). Based on the PSA output, uncertainty in the relative cost-effectiveness between multiple alternatives can be illustrated using cost-effectiveness acceptability curves (CEACs) and a cost-effectiveness acceptability frontier (CEAF) (Fenwick et al., 2001; Fenwick et al., 2004; Fenwick et al., 2005; Fenwick et al., 2006; Barton et al., 2008), which are described further in *Appendix 5*.

1.1.6.3. Further Research to Reduce Parameter Uncertainty

Further research, subsequent to a model-based economic evaluation, such as an RCT or observational study, has the potential benefit of reducing decision uncertainty. However, further research may also require resources that could be used elsewhere in the health care system to (i) generate health, or (ii) fund alternative research projects. Value of information (VOI) methods, based on statistical decision theory, can utilise the PSA output

to inform decisions regarding the desirability of further research by evaluating its potential benefit against its potential cost (Claxton et al., 1996; Claxton, 1999a; Claxton et al., 2001; Claxton et al., 2002; Claxton et al., 2004; Sculpher et al., 2005a; Claxton et al., 2006a; Eckermann et al., 2010; Griffin et al., 2011; Steuten et al., 2013; Drummond et al., 2015; Wilson, 2015).

VOI methods have been applied in health care decision-making during NICE technology appraisals (Claxton et al., 2006a), research prioritisation decisions by the NHS *Health Technology Assessment* programme (Claxton et al., 2004; Mohiuddin et al., 2014), and in the design of future RCTs (Thompson, 1981; Claxton et al., 1996; Chilcott et al., 2003; Kent et al., 2013). VOI comprises three methods of increasing complexity that can inform increasingly specific decisions regarding further research (Eckermann et al., 2010). The expected value of perfect information (EVPI) can be used to estimate the overall potential value of further research to reduce decision uncertainty; the expected value of perfect partial information (EVPPI) can be used to estimate the specific parameter(s) with the greatest value for further research; the expected net benefit of sampling (ENBS) can be used to estimate the expected net benefit derived from specific research designs (Claxton, 1999a). To date, all published examples of VOI have included an estimation of population EVPI (Steuten et al., 2013; Mohiuddin et al., 2014; Thorn et al., 2016); the methods for estimating population EVPI from a model-based economic evaluation are described in *Appendix 6*. The model-based cost-effectiveness analysis of stratified medicine in this thesis, reported in *Chapter Six*, estimated population EVPI.

1.1.6.4. The Early Economic Evaluation of Health Technologies

An *early economic evaluation* is the term used to describe an economic evaluation performed during an early stage of a health technology's product lifecycle (Steuten et al., 2014). In general, the model-based economic evaluations of pharmaceuticals within the NICE technology appraisal process are produced at a late-stage of the health technology's product lifecycle. For example, pharmaceuticals require a product licence before being evaluated by NICE, indicating that trial-based studies (which demonstrate evidence of safety and efficacy) have been produced. The economic evaluation of health technologies, however, may be performed iteratively and can be undertaken before the supporting clinical evidence base has reached maturity (Sculpher, 1997; Vallejo-Torres et al., 2008).

It is likely that the uncertainty inherent in any model-based economic evaluation will reduce as the quantity and quality of clinical evidence for a new health technology increases over time. As a consequence, the uncertainty in the relative cost-effectiveness of a health technology may be greatest when the supporting clinical evidence base is in its infancy (Annemans et al., 2000; Steuten et al., 2014). The results of an early economic evaluation may, therefore, be indicative of relative cost-effectiveness rather than definitive (Sculpher et al., 1997).

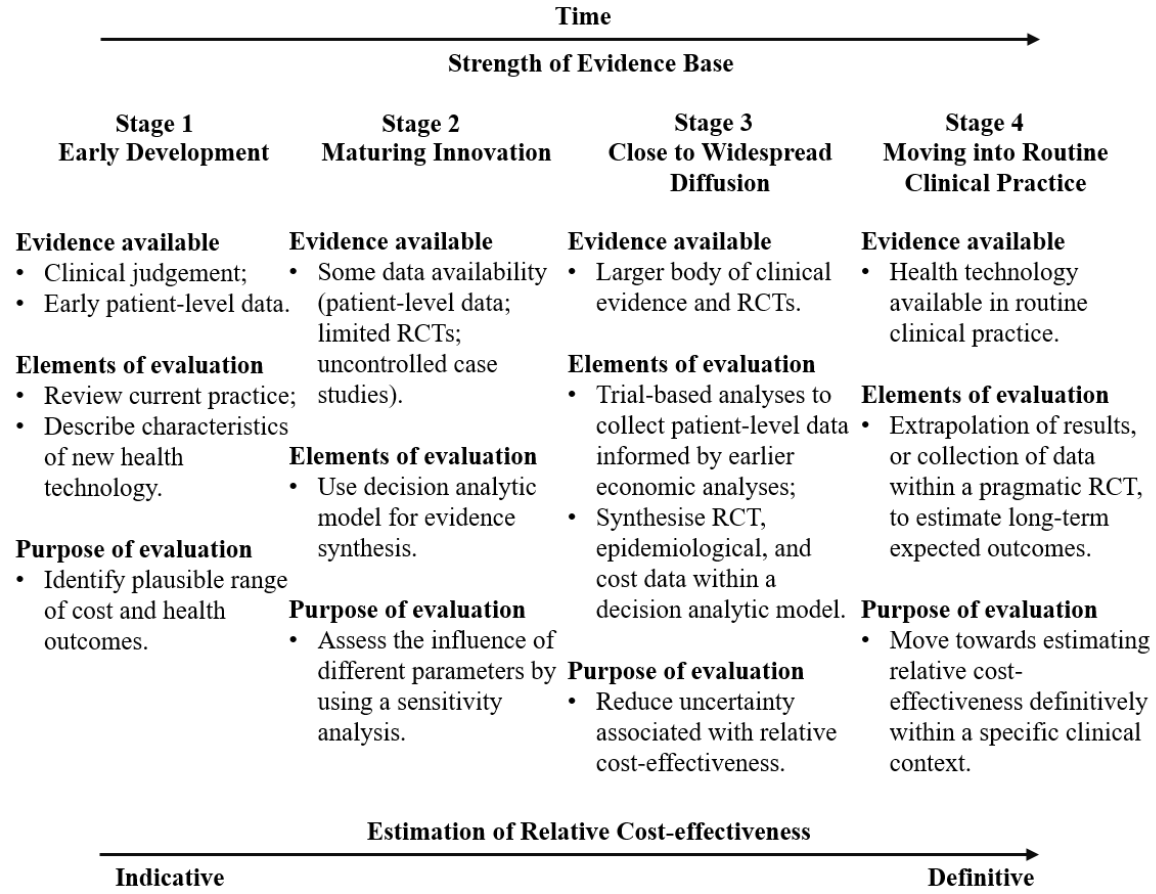
Economic evaluations, whether performed early or late, are designed to inform a decision. The specific decision, however, may vary depending on the timing of the economic evaluation. For example, a late-stage economic evaluation may be best suited to inform whether a health technology should be used widely in routine clinical practice; an early economic evaluation, by contrast, may be useful to inform the design of a subsequent clinical research study (Sculpher et al., 1997). Therefore, the use of VOI methods during an early economic evaluation, in particular, can be advantageous to identify whether such additional evidence would be of potential value to the health care system (Steuten et al., 2014).

The degree to which an economic evaluation can be interpreted as an *early* economic evaluation, conceptually, depends on the quantity and/or quality of the supporting clinical evidence base, *a priori*. The framework by Sculpher et al. (1997) posited that the economic evaluation of a new health technology could be performed iteratively over four broad time periods, defined as stages. Figure 1.1 illustrates these four stages, and describes the type of analyses that could be performed and the strength of clinical evidence that may support these analyses.

The earliest form of economic evaluation, *Stage 1*, is performed during the development phase of a health technology. The components of an economic evaluation in *Stage 1* may encompass an extensive description of current practice and of the characteristics of the new health technology, both in terms of costs and health outcomes. An economic evaluation performed when some additional clinical data has been produced (such as uncontrolled case studies or limited RCTs) is a *Stage 2* analysis within the framework. Decision analytic modelling techniques can be used during *Stage 2* to synthesise the available evidence and to assess the potential influence of specific parameters on the estimates of relative cost-effectiveness. The quantity and quality of the evidence supporting these two early stages of economic evaluation will likely be limited (compared with a later-staged evaluation), leading to an indicative, rather than definitive, estimation of relative cost-effectiveness.

Later-staged economic evaluations can make use of a more extensive clinical evidence base and pragmatic RCT study designs, potentially informed by earlier economic evaluations, to reduce the uncertainty associated with the relative cost-effectiveness of a health technology.

Figure 1.1. *Conceptual framework of an early economic evaluation according to Sculpher et al (1997).*



Source: Adapted from Sculpher et al. (1997); Abbreviations: RCT=randomised controlled trial.

The economic evaluation presented in this thesis was an example of an early economic evaluation because the clinical evidence base to support the health technology of interest (described in Section 1.3.5) was limited, characterised by a small number of test accuracy studies, and uncertainty with respect to the timing and effectiveness of testing within clinical practice. The economic evaluation, therefore, could be described as a *Stage 2* analysis; a decision analytic model was subsequently developed to provide an indicative estimate of relative cost-effectiveness, by synthesising all available evidence, assessing the sensitivity of the results to specific input parameters, and by using VOI methods to assess the value of further prospective research.

1.2. Stratified Medicine and Economic Evaluation

Stratified medicine, in the context of this thesis, required a biomarker test that informed a treatment decision, such as (i) the identification of the most appropriate treatment from a set of alternatives; or (ii) whether adjustments to a previous treatment decision (dose-adjustment, administration frequency, or change of treatment) were required (Redekop et al., 2013). A definition of a *biomarker* (a portmanteau of the words 'biological' and 'marker'), first proposed by the *US Food and Drug Administration* and assumed by this thesis, is "…*a characteristic that can be objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention*" (National Institute for Health Biomarkers Definitions Working Group, 2001, p.91).

Biomarkers that indicate a likely pharmacologic response are known as *predictive biomarkers* which, by definition, are required to implement stratified medicine (Trusheim et al., 2007). Predictive biomarkers can be measurable patient characteristics based on, but not limited to, genotypes, proteins, identified by imaging, or by physiological assessment (Trusheim et al., 2007; European Commission, 2010). Biomarker tests that only provided diagnostic information, without a subsequent treatment decision (for example, a genetic test for Huntington's disease), and the potential need for the valuation of non-health benefits (Payne et al., 2013b), were beyond the scope of this thesis.

1.2.1. Theoretical Principal of the Economics of Stratified Medicine

The theoretical principal that underpins the economics of stratified medicine is the distinction between *variability* and *heterogeneity*. Variability refers to a patient's chance occurrence of a clinical outcome (Groot et al., 2010; Groot et al., 2011); for example, a specific patient may, or may not, experience an adverse drug reaction. Heterogeneity refers to the variability that can be explained by observed characteristics (Briggs et al., 2006); for example, men may have a higher probability of an adverse drug reaction, relative to women.

The expected outcomes estimated by an economic evaluation of a conventional treatment apply to an *average* patient; population-level decisions can then be made by subsequently claiming that, on average, the treatment is (or is not) relatively cost-effective for all patients (Drummond et al., 2015). However, the costs and QALYs derived by each patient

may exhibit variability across the population (Sculpher, 2008), such that (i) some patients may receive a treatment for whom it is not individually cost-effective; or (ii) some patients may not receive a treatment for whom it would be individually cost-effective (Stevens et al., 2004; Basu et al., 2007). If such individual variability in relative cost-effectiveness could, in part, be explained by specific patient-level characteristics, an economic evaluation could potentially account for heterogeneity by estimating relative cost-effectiveness according to distinct patient subgroups (Sculpher, 2008). Decision-makers would then be able to exploit such heterogeneity to maximise population health by recommending the treatment only for those subgroups of patients whose health gains exceed the opportunity cost (Coyle et al., 2003; Sculpher, 2008).

The majority of decision-makers responsible for allocating population health care resources consider *patient* heterogeneity when making their recommendations (Ramaekers et al., 2013). For example, NICE have explicitly stated that subgroups according to geography, preferences, or non-disease related variability should not be considered within their appraisal of health technologies (National Institute for Health and Care Excellence, 2013a). Potential patient subgroups identified by an economic evaluation should therefore have biologic plausibility and be implementable in routine practice (Sculpher, 2008; Espinoza et al., 2014a). For example, subgroups could, theoretically, be defined at the level of the individual patient (Basu et al., 2007) but this may not be feasible to implement in practice (Bloss et al., 2013). Moreover, greater stratification of a patient population may be characterised by diminishing returns, such that the expected marginal opportunity cost of revealing additional heterogeneity may exceed its expected marginal health benefit (Stevens et al., 2004; Espinoza et al., 2014a).

The use of a biomarker test to inform a stratified medicine is therefore the mechanism by which patient heterogeneity can be revealed, to identify patient subgroups, and to potentially improve the relative cost-effectiveness of a treatment (van Gestel et al., 2012; Espinoza et al., 2014b). An economic evaluation of a stratified medicine, therefore, estimates the relative cost-effectiveness of treating patients by subgroup according to such a testing strategy that reveals patient heterogeneity *within* routine clinical practice.

1.2.2. Distinction between Types of Test

Biomarker tests used for stratified medicine, like all health technologies, incur costs and consequences that must be evaluated before being recommended for routine use in the

NHS (Husereau et al., 2014). However, the fundamental difference between the economic evaluation of a stratified medicine and a conventional treatment is that the health outcome attributable to a stratified medicine is principally derived from a subsequent treatment decision and not from the intervention test itself (National Institute for Health and Care Excellence, 2011a). Commercial biomarker tests, to stratify a treatment decision, can either be manufactured (i) *ex ante* during the development of a treatment; or (ii) *ex post* after a treatment had been licenced for routine practice (Garrison et al., 2007).

Tests developed *ex ante* are known as *companion diagnostics* and are typically stated within the product licence of a treatment (Milne et al., 2015). Tests developed *ex post* are known as *complementary diagnostics* (and also may be known as *stand-alone tests*), which can be used to inform treatment decisions regarding disease management, early diagnosis, and drug monitoring, without being explicitly referenced in the product licence of a treatment (Annemans et al., 2013; Milne et al., 2015). This distinction is important for the economic evaluation of stratified medicine because companion diagnostics conventionally have a stronger supporting evidence base on health outcomes compared with complementary diagnostics (Garrison et al., 2007). A model-based economic evaluation of a stratified medicine must therefore consider how the diagnostic information obtained from a biomarker test is related to a patient's final health outcome (Phillips et al., 2013), which, for complementary diagnostics in particular, has the potential to increase the uncertainty inherent in the decision-making process (Annemans et al., 2013; Fugel et al., 2014; Fugel et al., 2016). The cost-effectiveness analysis in *Chapter Six* of the thesis evaluated a stratified medicine that utilised a complementary diagnostic, with a limited evidence base on health outcomes, to inform a subsequent treatment decision.

1.2.3. Fragmented Decision-making for Stratified Medicine in England

The decision-making authorities responsible for appraising test-based treatment strategies (such as stratified medicine) for use in the NHS are relatively fragmented, compared with the case for pharmaceuticals. Five programmes that are able to evaluate a medical test in England are reported in Table 1.5; the appropriate decision-maker depends on the nature and value proposition of the test (Brockis et al., 2016).

NICE have three appraisal programmes to evaluate treatment strategies that incorporate medical testing: (i) the standard *Technology Appraisal Programme* is suitable for companion diagnostics; (ii) the *Medical Technology Evaluation Programme* is suitable for

tests that offer similar health benefits at lower costs; and (iii) the *Diagnostics Assessment Programme* (DAP) is suitable for tests that may increase costs and health outcomes (National Institute for Health and Care Excellence, 2011a). However, only the *Technology Appraisal Programme* imposes mandatory funding for a test after receiving a positive recommendation by NICE (Brockis et al., 2016). Single-gene genetic tests can be evaluated by the *UK Genetic Testing Network* by submission of a 'Gene Dossier' (Miller et al., 2011); however, the Gene Dossier does not require any evidence for the relative cost-effectiveness of testing (UK Genetic Testing Network, 2016).

Table 1.5. *Programmes that evaluate medical tests for use in the NHS.*

Evaluation Programme	Characteristic of Test
NICE Technology Appraisal Programme†.	Companion diagnostic with a new treatment.
NICE Diagnostics Assessment Programme♦.	(i) Tests with the potential to increase costs and health outcomes; (ii) Complementary diagnostic with an established treatment.
NICE Medical Technology Evaluation Programme▲.	Tests with a potential for similar health outcomes at a lower cost, or greater health outcomes at a similar cost.
UK Genetic Testing Network●.	Genetic tests for single-gene genetic disorders.
UK National Screening Committee■.	Population screening programmes.

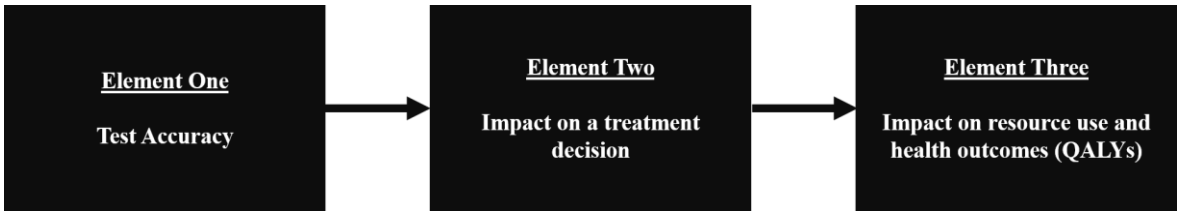
Source: †=National Institute for Health and Care Excellence (2013a); ♦=National Institute for Health and Care Excellence (2011a); ▲=National Institute for Health and Care Excellence (2011b); ●=UK Genetic Testing Network (2016); ■=UK National Screening Committee (2016).

The economic evidence generated for the stratified medicine in this thesis, which utilised a complementary diagnostic, was therefore designed to conform with the requirements of the NICE DAP (National Institute for Health and Care Excellence, 2011a).

1.2.4. Practical Challenges in the Model-based Economic Evaluation of Stratified Medicine

The use of a model-based economic evaluation, to estimate the relative cost-effectiveness of any stratified medicine, requires evidence of three related elements of a test-and-treatment strategy (Byron et al., 2014), illustrated in Figure 1.2.

Figure 1.2. *Elements of a test-and-treatment strategy for stratified medicine.*



A practical challenge when performing a model-based economic evaluation of stratified medicine, and in particular for complementary stand-alone diagnostics, is that tests frequently lack *end-to-end evidence* (National Institute for Health and Care Excellence, 2011a). End-to-end evidence is defined as a single study that incorporates all elements of a test-and-treatment strategy illustrated Figure 1.2, by observing a patient from their test result, to a treatment decision, and then to their final health and resource outcomes (National Institute for Health and Care Excellence, 2011a). The relationship between a test result and a final health outcome can be defined according to three features of any test (Rogowski et al., 2009; Phillips et al., 2013):

- *Analytic validity* – the accuracy and reliability of the test when measuring a specific biomarker;
- *Clinical validity* – the ability of the biomarker to indicate a specific clinical status;
- *Clinical utility* – the ability to improve outcomes by making a treatment decision based on the result of the test.

The lack of end-to-end evidence is driven, in part, by the current regulatory requirements for European test manufacturers, who must only produce evidence of safety and analytic validity (but not necessarily clinical utility) in order to achieve market access for a new test (Payne, 2008; 2009; The Academy of Medical Sciences, 2013; Fugel et al., 2014). As a consequence, tests that achieve rapid market access frequently lack the necessary evidence to estimate their relative cost-effectiveness (Meadows et al., 2015).

A second practical challenge is that the accuracy of testing can be characterised by four potential outcomes, and two of which indicate that a test may provide an incorrect (a false-positive or false-negative) result (Annemans et al., 2013). A *2x2 table* can be used to describe the four potential outcomes of a test (illustrated in Table 1.6) (Macaskill et al., 2010). The *sensitivity* of a test is the proportion of patients with a positive test result, of those patients who truly have the biomarker of interest; the *specificity* of a test is the proportion of patients with a negative test result, of those patients who truly do not have the biomarker of interest. An economic evaluation of a stratified medicine must therefore

be designed to include the health and resource consequences associated with making a treatment decision according to both a correct, and an incorrect, test result (Annemans et al., 2013).

Table 1.6. *2x2 table of test accuracy and the potential outcomes of a test.*

	Does Patient have Biomarker?	
Test Result	**Yes**	**No**
Test Positive	True-positive.	False-positive.
Test Negative	False-negative.	True-negative.

Source: Adapted from Macaskill et al. (2010, p.10).

1.2.5. Sources of Uncertainty in Model-based Economic Evaluations of Stratified Medicine

The following section describes five potential sources of uncertainty, that were relevant to the economic evaluation of stratified medicine presented in this thesis, by drawing on (i) nine published systematic reviews of economic evaluations of stratified medicine (Phillips et al., 2004; Vegter et al., 2008; Beaulieu et al., 2010; Vegter et al., 2010; Wong et al., 2010; Hatz et al., 2014; Berm et al., 2016; Oosterhoff et al., 2016; Plumpton et al., 2016); and (ii) a wider literature that has discussed the potential challenges to conducting an economic evaluation of stratified medicine (Becla et al., 2011; Annemans et al., 2013; Buchannan et al., 2013; Fleeman et al., 2013; Fugel et al., 2014; Rogowski et al., 2015; Shabaruddin et al., 2015; Fugel et al., 2016).

Defining the Role of Testing

An economic evaluation of a stratified medicine necessarily requires an understanding of how a test will be used in routine practice to inform a subsequent treatment decision (National Institute for Health and Care Excellence, 2011a). A lack of end-to-end evidence, however, may introduce a degree of uncertainty over the most appropriate way to use a test in practice; both the timing of a new test and the respective treatment decision can affect the delivery of care in routine practice (National Institute for Health and Care Excellence, 2011a) which, in turn, may affect the outcomes derived from treatment stratification and the relative cost-effectiveness of testing (Hatz et al., 2014; Shabaruddin et al., 2015). A model-based economic evaluation of a stratified medicine must therefore make choices regarding the timing and treatment decisions associated with testing (Annemans et al.,

2013), which can be investigated further by using model conceptualisation techniques (Tappenden, 2014). The role of testing in this thesis was subsequently investiagted in Section 5.3 of *Chapter Five*, by reviewing the clinical literature for prescribing algorithms that had incorporated the test being evaluated.

Choice of Comparator

An economic evaluation must include all relevant comparator strategies to generate relevant evidence for decision-makers (Drummond et al., 2015). The choice of comparator strategies included within an economic evaluation may be dependent on the assumptions made by the decision analyst. Different comparator strategies for a stratified medicine may be defined by using a test in different ways (for example, by changing the timing of testing) (Husereau et al., 2014; Fugel et al., 2016; Oosterhoff et al., 2016). In the context of evaluating a stratified medicine, comparing all alternative testing strategies to a common comparator (such as no testing) will underestimate the opportunity cost of the intervention and may not maximise population health (O'Mahony et al., 2015c). Moreover, given that relative cost-effectiveness is determined by the incremental comparison of a strategy to its next-best alternative, the choice of comparators can affect the estimated ICER of a particular strategy (O'Mahony et al., 2015a). Section 5.4 of *Chapter Five* in the thesis therefore explored the choice of comparator strategies by developing a novel algebraic conceptualisation technique, to inform the identification of potentially relevant comparator strategies of a stratified medicine early in its product lifecycle.

Model Complexity

The potential lack of end-to-end evidence may increase the number of structural assumptions made within a decision analytic model, which may, in turn, increase the need for a greater model complexity (Payne et al., 2010; Annemans et al., 2013). For example, it may be necessary to use a *linked-evidence approach*, commonly observed within technology appraisals for the NICE DAP (National Institute for Health and Care Excellence, 2011a), which synthesises different sources of evidence to demonstrate a link between test accuracy, treatment decisions, and patient outcomes (Merlin et al., 2013). Furthermore, it may also be necessary to design a decision analytic model that simulates patients individually, such as a discrete event simulation, if the occurrence of previous clinical events may affect future clinical events (Caro et al., 2016b). For example, future treatment response may be affected by whether the patient had developed a specific

biomarker at an earlier point in time (Rogowski et al., 2015). An explanation of the conceptualisation and development process of the *de novo* decision analytic model in this thesis, and the specific elements that had the potential to introduce complexity into the model, is provided in *Chapter Five*.

Cost of Testing

An economic evaluation of a stratified medicine must include an estimate for the unit cost of testing. However, there is no national price tariff for tests in England and the resources required to implement a new test in practice (for example, an additional clinical visit or the time taken to analyse a blood sample) may be unknown *a priori* (Buchannan et al., 2013; Fugel et al., 2016). Previous economic evaluations of stratified medicine have suggested that the relative cost-effectiveness of treatment stratification may be sensitive to the cost of testing (Phillips et al., 2004; Beaulieu et al., 2010; Vegter et al., 2010; Oosterhoff et al., 2016). *Microcosting* methods, which aim to (i) identify the specific quantity of resources required to use a health technology in practice and to (ii) value those resources at relevant prices (Frick, 2009), may be valuable to estimate the true opportunity cost of a new testing strategy. A microcosting study was therefore conducted, supplementary to this thesis, with Dr. Meghna Jani at *The University of Manchester* (Jani et al. (2016a); reported in *Appendix 35*) to estimate the unit cost of the specific test that was evaluated in this thesis. The estimated cost was subsequently incorporated as a source of evidence for the *de novo* decision analytic model in *Chapter Six*.

Accuracy of Testing

The accuracy of detecting a specific biomarker, characterised by a test's sensitivity and specificity, is affected by the test's cut-off value (Trusheim et al., 2015). The cut-off value defines the quantity of a biomarker that distinguishes a positive test result from a negative test result (Macaskill et al., 2010). However, different studies that have estimated the accuracy of a test may have used different cut-off values and the appropriate cut-off value may be unknown *a priori* (Macaskill et al., 2010). A synthesis of test accuracy evidence should therefore account for the correlation between sensitivity and specificity, if different cut-off values have been used (National Institute for Health and Care Excellence, 2011a), by performing a bivariate hierarchical meta-analysis (Dinnes et al., 2005; Reitsma et al., 2005; Harbord et al., 2008; Macaskill et al., 2010). A bivariate meta-analysis was therefore

performed to appropriately synthesise the evidence of test accuracy (reported in *Appendix 34*) for the cost-effectiveness analysis in *Chapter Six*.

1.3. Disease Application: Rheumatoid Arthritis

The economic evaluation presented in this thesis was based on a specific case study of stratified medicine in RA. This section provides a background to the epidemiology of RA (Section 1.3.1); relevant condition-specific outcome measures (Section 1.3.2); treatments for RA (Section 1.3.3); the general rationale for stratified medicine in RA (Section 1.3.4); and the specific case study addressed by this thesis (Section 1.3.5).

1.3.1. Epidemiology of Rheumatoid Arthritis

RA is a chronic, systemic inflammatory autoimmune disease characterised by inflammation within the lining of joints (the *synovium*), and the progressive destruction of joints and cartilage due to the development of invasive pannus tissue (Firestein, 2003; Scott et al., 2010; McInnes et al., 2011). Joint inflammation and destruction occurs predominantly within the hands and feet (Firestein, 2003). Patients with RA experience pain, a gradual decline in functional ability, and a reduction in quality of life (Kvien, 2004; Russell, 2008). There is no known cure for RA (National Audit Office, 2009).

The estimated *prevalence* (proportion of cases within a population at a specific time) of RA within the UK is 0.81% (Symmons et al., 2002). The annual *incidence* (rate of new cases within a population over a period of time) of RA is estimated at 48 cases per 100,000 individuals within the UK (Humphreys et al., 2013). Cases of RA are approximately three times more common in women than men (Kvien, 2004; Scott et al., 2010). The average age of disease onset occurs between 55-64 years for women and 65-75 years for men (Symmons, 2002).

The precise factors that cause an individual to develop RA are unknown; however interaction between genetic and environmental factors are known to influence disease onset (Symmons, 2002; McInnes et al., 2011; Okada et al., 2014). RA is characterised by the production of *autoantibodies* (antibodies against an individual's own proteins), namely rheumatoid factor and anti-citrullinated protein antibodies (McInnes et al., 2011). Rheumatoid factor production is associated with smoking behaviour (Albano et al., 2001), and a more aggressive form of the disease (Firestein, 2003; Scott et al., 2010). Patients

with RA also experience the overproduction of pro-inflammatory *cytokines* (proteins secreted during an immune response that influence the interaction between, or behaviour of, specific cells) (Scott et al., 2010). The cytokine *tumour necrosis factor-α*, in particular, is important in the pathogenesis of RA because of its role in promoting such an extensive inflammatory response (Feldmann, 2002; Feldmann et al., 2005; Feldmann et al., 2010).

Patients with RA experience an increased risk of mortality, with an average reduction in life expectancy of five to ten years, compared with the general population (Kvien, 2004). Early mortality is primarily related to an increase in comorbid cardiovascular disease (Meune et al., 2009). RA patients may also experience comorbid depression, fatigue, diabetes, osteoporosis and infection (Kvien, 2004; Dougados, 2016; Siebert et al., 2016), which may be underreported in practice (Baillet et al., 2016). Greater physical disability and a lower probability of improvement in disease activity are associated with a higher number of comorbidities in patients with RA (Radner et al., 2010; Raganath et al., 2013).

RA is associated with substantial costs to the health care system and to wider society. Patients with severe, active disease can escalate treatment to relatively expensive biologic therapies, which cost the NHS approximately £10,000 per patient per year (British National Formulary, 2016). There is an inverse relationship between the direct medical costs of treating a patient with RA and their functional ability (Kvien, 2004). The *National Audit Office* (2009) subsequently estimated that the direct cost (to the NHS) for treating RA was £560 million per year. Patients with RA may also incur productivity costs to the wider economy through a reduced ability to maintain employment (Zhang et al., 2011), estimated to be between £3.8 and £4.75 billion per year to the UK economy (National Institute for Health and Care Excellence, 2009).

1.3.2. Condition-specific Outcome Measures for Rheumatoid Arthritis

This section summarises the condition-specific outcome measures for RA that were used in this thesis, with respect to disease classification, and the measurement of functional ability, disease activity, and treatment response. The components of each outcome measure are reported fully in *Appendix 7*.

Disease Classification: ACR Criteria

Disease classification instruments are primarily used to ensure patient homogeneity within clinical studies (Symmons, 2002; Aggarwal et al., 2015). The *American College of Rheumatology* (ACR) *1987 Classification Criteria* remains widely used in empirical studies of RA (Arnett et al., 1988). The ACR and *European League Against Rheumatism* (EULAR) have since updated the criteria (*2010 ACR/EULAR Classification Criteria*) to classify patients with RA at an earlier stage of their disease (Aletaha et al., 2010).

Functional Ability: HAQ-DI

Functional ability can be assessed using the *Health Assessment Questionnaire – Disability Index* (HAQ-DI) (Bruce et al., 2003). The HAQ-DI has twenty questions across eight categories of functional ability (dressing, rising, eating, walking, hygiene, reach, grip, and usual activities). The total HAQ-DI score is between zero (no disability) and three (complete disability), and there are twenty-five possible numerical outcomes in increments of 0.125. Patients can have *mild-moderate* (HAQ-DI=0 to 1), *moderate-severe* (HAQ-DI=1 to 2), or *severe-very severe* (HAQ-DI=2 to 3) disability (Bruce et al., 2003).

Disease Activity: DAS28

The *Disease Activity Score-28 Joint Count* (DAS28) is a composite measure of disease activity, used in routine clinical practice in England, comprising a count of a patient's swollen joints, tender joints, and an assessment of their inflammatory markers (erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP)) (Prevoo et al., 1995; Madsen, 2013). The DAS28 is bound between zero and 9.4, and a score greater than 5.1 is indicative of high disease activity (van Gestel et al., 1998).

Treatment Response: EULAR Response

Treatment response in clinical practice for patients with RA in England is classified according to their EULAR response, defined by the absolute change in DAS28 and the prevailing level of disease activity, six months after commencing a treatment. A *good EULAR response* is equivalent to a DAS28 reduction of at least 1.2 and the presence of low disease activity (DAS28≤3.2) (van Gestel et al., 1998).

1.3.3. Treatment for Rheumatoid Arthritis

Rheumatologists in England are predominately guided by the recommendations produced by NICE (National Institute for Health and Care Excellence, 2009) when making routine treatment decisions for RA, supported by the national *British Society for Rheumatology* (Ding et al., 2010). Treatment decisions can also be informally influenced by the recommendations of the international professional rheumatology organisations, EULAR (Smolen et al., 2014) and the ACR (Singh et al., 2016b).

An international consensus has developed that regards the objective of treatment to be the achievement of clinical remission or low disease activity within six months of commencing therapy (Smolen et al., 2015). Treatment decisions for patients with RA should therefore follow an aggressive *treat-to-target* approach (Smolen et al., 2016). Regular assessments of disease activity and treatment adjustments are recommended to achieve the target of remission (Smolen et al., 2016). Health care is delivered by a multidisciplinary team and patient involvement in treatment decisions though shared decision-making is encouraged (Smolen et al., 2014; Singh et al., 2016b; Smolen et al., 2016).

Early treatment of newly-diagnosed patients with RA is recommended to prevent irreversible long-term joint damage; a period known in the literature as the *window of opportunity* (Boers, 2003; van Nies et al., 2014). Intensive treatment at an early stage of RA may increase the likelihood of achieving remission, reduce the time to achieving remission, and improve the duration of remission (Verstappen et al., 2007). Consequently, treat-to-target objectives may be relatively difficult to achieve in patients with late-staged established, active RA (van Nies et al., 2014; Smolen et al., 2015).

Patients with RA may ultimately require treatment over their lifetime due to the chronic, incurable nature of the disease (Smolen et al., 2015). There are four broad categories of treatment available for RA: (i) non-steroidal anti-inflammatory drugs (NSAIDs); (ii) glucocorticoids; (iii) conventional synthetic disease-modifying anti-rheumatic drugs (cDMARDs); and (iv) biologic disease-modifying anti-rheumatic drugs (bDMARDs). NSAIDs are used for short-term pain relief, however their long-term use is associated with adverse events due to toxicities (Scott et al., 2010; Crofford, 2013). Similarly, short-term glucocorticoid therapy can manage flares in disease activity, however their long-term use is associated with adverse events such as steroid-induced osteoporosis (National Institute for Health and Care Excellence, 2009; Scott et al., 2010).

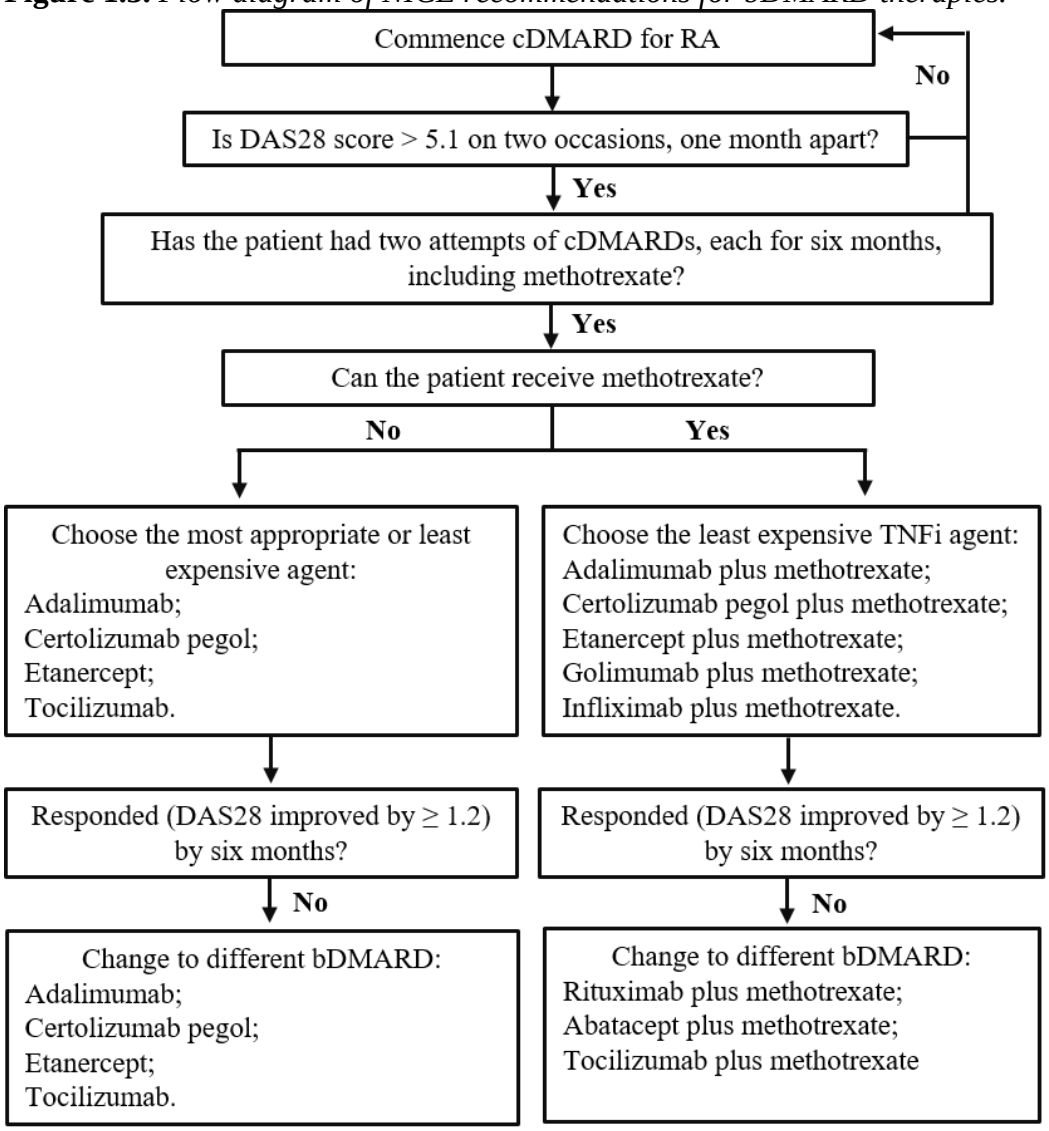
cDMARDs are a heterogeneous group of synthetic therapies that modify disease activity, reducing both inflammation and further joint damage (Scott et al., 2010). NICE recommends that patients with RA receive a combination of cDMARDs within three months of diagnosis if clinically appropriate (National Institute for Health and Care Excellence, 2009). The cDMARD methotrexate is often referred to as an *anchor drug* because of its prominent role in the management of RA (Pincus et al., 2013). Approximately two-thirds of patients will respond to methotrexate as monotherapy or in combination with other cDMARDs (Pincus et al., 2013).

bDMARD therapies are biologic protein-based treatments (commonly referred to as *biologics*) that act against different therapeutic targets of inflammation (Smolen et al., 2015). One such target is the cytokine tumour necrosis factor-α that mediates the inflammatory process in RA; bDMARDs that target tumour necrosis factor-α are known as *tumour necrosis factor-α inhibitors* or *anti-TNF therapies* (hereafter referred to as TNFi therapies) (Smolen et al., 2015). Patients typically escalate treatment to bDMARD therapies after failing to respond to conventional cDMARDs. Figure 1.3 presents a flow diagram to illustrate the conditions under which a patient with RA may receive bDMARD therapies according to clinical recommendations by NICE.

There are five TNFi therapies (reported in Table 1.7) recommended by NICE for patients with RA that meet the following criteria (National Institute for Health and Care Excellence, 2016a):

- High disease activity, measured by a DAS28 score of at least 5.1 on two occasions;
- Have previously failed two attempts of cDMARD therapy, usually including methotrexate.

Figure 1.3. *Flow diagram of NICE recommendations for bDMARD therapies.*



Source: flow diagram based on prescribing recommendations provided by NICE (National Institute for Health and Care Excellence, 2016a) and the regional care pathway for biologic therapies in Greater Manchester (Jani et al., 2015b).

All TNFi therapies for RA are licenced for use in combination with methotrexate due to greater efficacy than TNFi monotherapy (Nam et al., 2014; Choy et al., 2016). Four TNFi therapies are self-administered by subcutaneous injection with a pen device (between every one to four weeks) and infliximab is administered by intravenous infusion every two months following an initial loading dose. Four of the TNFi therapies (adalimumab, certolizumab pegol, golimumab, and infliximab) are *monoclonal antibodies* and etanercept is a *fusion protein* (Monaco et al., 2015). Up to fifty percent of patients with RA will fail to respond adequately to a TNFi therapy; bDMARDs with different therapeutic targets are subsequently recommended by NICE upon TNFi failure, including rituximab (B-cell depletion), abatacept (inhibits T-cell activation), and tocilizumab (acts against interleukin-6) (Curtis et al., 2011; Mócsai et al., 2014; National Institute for Health and Care Excellence, 2016a).

Table 1.7. *Five TNFi therapies recommended by NICE for patients with RA.*

Drug (*Trade name*)	**Concomitant methotrexate**	**Mode of administration**	**Administration Frequency**	**EU Patent expiry**	**Manufacturer**
Adalimumab† (*Humira*)	Yes	Subcutaneous injection	40mg every 2 weeks	2018	AbbVie
Certolizumab pegol♦ (*Cimzia*)	Yes	Subcutaneous injection	400mg every 4 weeks	2024	UCB Pharma
Etanercept▲ (*Enbrel*)	Yes	Subcutaneous injection	50mg every week	2015	Pfizer
Golimumab● (*Simponi*)	Yes	Subcutaneous injection	50mg every month	2024	Merck Sharp & Dohme
Infliximab■ (*Remicade*)	Yes	Intravenous infusion	3mg/kg at 0, 2, and 6 weeks initially, then every 8 weeks thereafter	2015	Merck Sharp & Dohme

Source: †=European Medicines Agency (2017); ♦=European Medicines Agency (2016a); ▲European Medicines Agency (2014); ●=European Medicines Agency (2016b); ■=European Medicines Agency (2012).

Three TNFi therapies (adalimumab, etanercept, and infliximab) are currently experiencing patent expiry within the European Union, which will facilitate the production of *biosimilar* therapies. Biosimilar therapies are treatments with similar quality, safety, and efficacy of a reference biologic (Grabowski et al., 2014). The *European Medicines Agency* has granted market approval for biosimilar infliximab and etanercept, and biosimilar adalimumab is currently in development (Dörner et al., 2016).

bDMARDs, relative to cDMARDs, impose a substantial opportunity cost on the health care system in England. Patent protection, and the high costs associated with the manufacture and development of bDMARDs, contribute to their relatively high cost to the NHS (approximately £10,000 per patient per year) (Kelly et al., 2009; British National Formulary, 2016). The cost of biosimilar TNFi therapies in the short-run may also be high due to limited competition between incumbent manufacturers (Grabowski et al., 2014). The evidence to support the relative cost-effectiveness of TNFi therapies for RA, despite their proven efficacy in clinical trials, is generally unfavourable (van der Velde et al., 2011; Joensuu et al., 2015). The most recent recommendations regarding TNFi therapies by NICE were based on an economic evaluation of bDMARD therapies (relative to methotrexate) that estimated ICERs above the conventional cost-effectiveness threshold

used in NICE decision-making (between £39,100 and £42,200 per QALY gained) (Stevenson et al., 2016). Therefore, if decision-makers continue to recommend TNFi therapies for RA within the NHS, irrespective of the unfavourable evidence regarding their relative cost-effectiveness, further investigation into alternative means of prescribing TNFi therapies may be justified to improve their relative cost-effectiveness and, in turn, maximise population health across the NHS.

1.3.4. Stratified Medicine in Rheumatoid Arthritis

Stratified medicine has been proposed within the clinical literature as a method to improve the management of patients with RA (van den Broek et al., 2013; Karsdal et al., 2014). In practice, stratified medicine in RA could be implemented at different stages of the disease, for example: (i) to predict the onset of RA and commence early treatment in asymptomatic patients, (ii) to confirm a diagnosis of RA in symptomatic patients; and (iii) to predict response, or loss of response, to therapies when managing active RA (Gibson et al., 2012). Stratified bDMARD therapy, in particular, may be desirable given their high cost per patient and the potential to reduce the side-effects associated with less-effective therapies in certain patients (Isaacs et al., 2011). Tests to predict treatment response may reduce the time that RA patients are exposed to ineffective therapies and, in turn, reduce the potential for irreversible joint damage (Lindstrom et al., 2010).

The clinical interest for stratified medicine in RA has been facilitated by a growing evidence-base of patient-level heterogeneity in the disease (Emery et al., 2011; Tak, 2012; Cuppen et al., 2016). For example, differences in disease pathogenesis and inflammatory markers have been observed between patients with RA (Tak, 2012). Such heterogeneity may be associated with different therapeutic responses within and across different classes of treatment (Emery et al., 2011). For example, different bDMARDs have different therapeutic targets (Mócsai et al., 2014), and patients may respond best to the treatment that acts against their own primary source of inflammation (Cuppen et al., 2016).

1.3.5. Case Study: Stratified Medicine in RA by Testing for Immunogenicity against Adalimumab

The economic evaluation in this thesis focused on a specific case study that involved testing for immunogenicity against a particular TNFi therapy, adalimumab, in patients with RA. This case study was chosen as a practical example of how treatment could be stratified

in the context of RA, and was a novel therapeutic approach emerging in routine practice when this thesis was conceived. This subsection introduces the case study, using the terminology presented in Section 1.2, by describing the: (i) clinical problem, (ii) source of heterogeneity, (iii) biomarkers, (iv) test to detect the biomarkers, (v) application of stratified medicine, and (vi) the potential economic rationale for stratified medicine.

Clinical Problem

Patients with RA who receive a TNFi therapy can experience two forms of treatment failure (National Institute for Health and Care Excellence, 2010; Emery, 2012; Jani et al., 2015b):

- *Primary non-response* occurs when a patient fails to respond to their TNFi within the first six months of treatment;
- *Secondary non-response* occurs when a patient loses response to their TNFi after previously experiencing a period of sustained response to the treatment.

Secondary non-response to TNFi therapies is a relatively common phenomenon; approximately 50% of RA patients will lose response to a TNFi within five years of commencing treatment (Tak, 2012). There are several factors that may influence secondary non-response to TNFi therapies; for example, a patient may fail to take concomitant methotrexate or may receive too little TNFi for their body weight (Jani et al., 2014; Jani et al., 2015b). The factor of increasing interest to the rheumatology community, due to its potential implication for stratified medicine, is *immunogenicity* against TNFi therapies (Garcês et al., 2013).

All biologic therapies may cause an immunogenic response after being administered (Schaeverbeke et al., 2016). Immunogenicity against a TNFi occurs when the body, believing that it has experienced the presence of a harmful pathogen, inappropriately produces an immune response and develops anti-drug antibodies (ADAb) against the treatment itself (Krieckaert et al., 2012). ADAbs can have a dual role in affecting treatment by (i) increasing the clearance rate of therapeutic TNFi within a patient, and (ii) by binding to the TNFi drug to neutralise its therapeutic effect (Krieckaert et al., 2012). Consequently, the amount of TNFi circulating within a patient's serum is reduced, referred to as the *therapeutic drug level* (Friedman et al., 1986), implying that there is less drug available to suppress the production of tumour necrosis factor-α, resulting in increased joint

inflammation and, ultimately, secondary non-response to the TNFi (Krieckaert et al., 2012).

The rate at which ADAbs develop is known to vary between different TNFi therapies and is a greater concern for monoclonal TNFi therapies (Schaeverbeke et al., 2016). For example, clinical evidence has indicated that immunogenic responses to etanercept are relatively uncommon (Schaeverbeke et al., 2016). Out of the monoclonal TNFi therapies available for patients with RA (described in Section 1.3.3), secondary care expenditure on adalimumab in England exceeded the expenditure on all other pharmaceutical treatments across all indications between 2014-15 (Health and Social Care Information Centre, 2015). Moreover, there existed a developing clinical evidence base for immunogenicity against adalimumab in RA, compared with the other monoclonal TNFis (Bartelds et al., 2011; Pouw et al., 2015; Jani et al., 2016b). The thesis therefore focussed specifically on the application of stratified medicine for patients with RA that were prescribed adalimumab, given the emerging clinical evidence base and the potential for its substantial impact on NHS expenditures.

Source of Heterogeneity

The development of immunogenicity against adalimumab was a (previously unobservable) source of heterogeneity in secondary non-response between patients with RA (Jamnitski et al., 2011; Jani et al., 2015a). For example, earlier clinical studies have demonstrated that patients with ADAb against adalimumab had significantly lower therapeutic drug levels (Radstake et al., 2009), and were significantly more likely to lose response to treatment over time (Garcês et al., 2013).

Biomarkers

The two biomarkers indicative of immunogenicity against adalimumab were (i) adalimumab ADAb and (ii) adalimumab drug levels. The usefulness of measuring ADAb and drug levels was of international interest within the rheumatology community, and had been raised as an important agenda for research by EULAR (Smolen et al., 2014). However, unlike predictive biomarkers that are tested before treatment initiation, ADAb and drug levels could only develop, and be measured, once a patient had commenced TNFi therapy (Krieckaert et al., 2012).

Test to Detect Biomarker

A range of health technologies exist that can detect the presence of ADAb and drug levels associated with biologic therapies (Mire-Sluis et al., 2004; Pineda et al., 2016). Drug levels are most commonly detected by a sandwich enzyme-linked immunosorbent assay (ELISA) (Chen et al., 2015b; Jani et al., 2015a). The two most prominent methods to detect TNFi ADAb in patients with RA are the bridging-ELISA and the radioimmunoassay (RIA) (van Schouwenburg et al., 2013). The bridging-ELISA is more likely to be used in routine practice due to its comparatively lower start-up and running costs, and given that there is no use of radioactive materials (Jani et al., 2014). This thesis therefore concentrated on treatment stratification by using ELISA-based technologies to test adalimumab ADAb and drug levels in patients with RA.

There are potentially three elements present within a patient's serum when immunogenicity against adalimumab therapy occurs: (i) the monoclonal antibody (adalimumab) itself, (ii) free ADAb, and (iii) drug-ADAb complexes (Schaeverbeke et al., 2016). The bridging-ELISA (and the RIA) can detect free ADAb within a patient's serum (van Schouwenburg et al., 2013). However the bridging-ELISA is limited by *drug interference* when attempting to measure free ADAb if some of the biologic therapy is also present within the serum; free ADAb are only detectable by a bridging-ELISA if their levels exceed the levels of therapeutic drug within the serum (van Schouwenburg et al., 2013).

Commercial tests are now available to measure adalimumab ADAb and drug levels, using ELISA techniques (Karsdal et al., 2014; Kiely, 2016), which are increasingly being considered for routine use within the NHS. For example, the regional care pathway for patients with RA that receive TNFi therapies in Greater Manchester has incorporated ADAb and drug level assessment (Jani et al., 2015b). The tests are an example of a complementary diagnostic because they (i) were developed after adalimumab had received market access, and (ii) are not included within the product label of adalimumab. Table 1.8 reports three commercial ELISA assays for measuring adalimumab ADAb and drug levels.

Table 1.8. *Commercial ELISA assays to detect adalimumab ADAb and drug levels.*

Name of Commercial ELISA Test		
Detect Free Anti-adalimumab Antibodies	**Detect Free Adalimumab**	**Manufacturer**
LISA-TRACKER Anti-Adalimumab	LISA-TRACKER Adalimumab	Theradiag/ Alpha Laboratories
Immunodiagnostik TNFα-Blocker Monitoring, Antibodies Against Adalimumab Drug Level (e.g. Humira®) ELISA	Immunodiagnostik TNFα-Blocker Monitoring, Adalimumab Drug Level (e.g. Humira®) ELISA	Immunodiagnostik AG BioHit Healthcare Ltd
Promonitor Anti-ADL ELISA	Promonitor ADL ELISA	Proteomika

Source: National Institute for Health and Care Excellence (2014).

Application of Stratified Medicine

Advocates of testing ADAb and/or drug levels have argued that they may be candidate biomarkers to characterise patient subgroups such that subsequent bDMARD treatment decisions could be stratified (Vincent et al., 2013). However the commercial ELISA-based ADAb and drug level tests, as with most complementary diagnostics, lacked end-to-end evidence that followed a patient from their test result to a health outcome. The (i) clinical utility, (ii) appropriate timing of testing, and (iii) the specific treatment decision following a test result were therefore uncertain *a priori*.

Potential Economic Rationale for Stratified Medicine

The potential economic rationale for testing immunogenicity against adalimumab was twofold. Firstly, a relative QALY gain may have been achieved if patients receiving adalimumab were able to be stratified to avoid the harm associated with second non-response. Secondly, a relative cost reduction may have been achieved by using the testing strategies to identify a subgroup of patients that could have sustained response after experiencing a reduction in their dose of adalimumab.

Stratified treatment decisions according to adalimumab ADAb and drug levels will, however, impose an opportunity cost because additional health care resources will be required to pay for testing. The relative cost-effectiveness of routine immunogenicity testing in England was unknown. Therefore, an economic evaluation of treatment stratification was necessary to provide health care decision-makers in England with

evidence regarding the relative cost-effectiveness of adalimumab ADAb and drug level testing.

1.4. Research Questions and Thesis Structure

The overall aim of this thesis was addressed using mixed methods (systematic reviews, qualitative thematic framework analysis, quantitative econometric analysis, and decision analytic modelling). The relative cost-effectiveness of stratifying treatment for RA in England according to a test for immunogenicity against adalimumab was estimated by conducting an early model-based cost-effectiveness analysis. However, uncertainties existed in the (i) wider evidence base regarding the relative cost-effectiveness of stratified medicine in RA, (ii) the characterisation of current treatment decisions for RA in England, and (iii) the appropriate use of adalimumab ADAb and/or drug level ELISA testing in routine practice.

Subsequently, the thesis addressed three related research questions:

Research Question 1: *What was the existing economic evidence for stratified medicine in RA?*

Research Question 2: *How were treatment decisions with biologic therapies made for patients with RA in current practice in England?*

Research Question 3: *Are treatment decisions stratified by adalimumab ADAb and drug level testing, for patients with RA in England, a relatively cost-effective use of health care resources?*

The thesis was written as a case study that developed the economic evidence base for stratified medicine in RA. The structure of the thesis is summarised in Table 1.9 by outlining the research question addressed by each chapter, and each chapter's general method and purpose within the thesis. Each chapter was written as a standalone study, with the exception of *Chapter Five*, which was written as a series of smaller sub-studies. All chapters were linked by the common theme of developing relevant evidence to inform the overall aim of the thesis.

Table 1.9. *Thesis research questions, structure, general method, and purpose of each chapter.*

Research Question	Chapter Addressed	General Method	Purpose of Chapter
1. *What was the existing economic evidence for stratified medicine in RA?*	Chapter 2	Systematic review.	To identify the exiting model-based economic evaluations of stratified medicine in RA.
2. *How were treatment decisions with biologic therapies made for patients with RA in current practice in England?*	Chapter 3	Qualitative analysis.	To explore (i) the relevant care pathways for patients with RA in England, (ii) the potential influences on treatment decisions, and (iii) the potential barriers to using ADAb and drug level testing in routine practice.
	Chapter 4	Quantitative analysis.	To estimate the patient-level factors that influenced the choice of TNFi prescribed to patients with RA in England.
3. *Are treatment decisions stratified by TNFi ADAb and drug level testing, for patients with RA in England, a relatively cost-effective use of health care resources?*	Chapter 5	Model conceptualisation.	To (i) identify how ADAb and drug level testing may be used to stratify treatment decisions in routine practice; (ii) define the decision problem of the early economic evaluation; (iii) conceptualise the care pathways and health outcomes associated with treatment stratification; (iv) select the appropriate type of model; and (v) to design the structure of the *de novo* decision analytic model.
	Chapter 6	Decision analytic modelling and evidence synthesis.	To (i) estimate relevant values for the input parameters of the decision analytic model; (ii) estimate the relative cost-effectiveness of adalimumab ADAb and drug level testing to stratify treatment for patients with RA in England; and (iii) to estimate the potential value of conducting additional prospective research to reduce the parameter uncertainty associated with testing.

Chapter Two reports a systematic review of published economic evaluations of stratified medicine in RA, which established the existing economic evidence base upon which this thesis was built. *Chapter Three* presents a qualitative study, conducted with a sample of consultant rheumatologists in England, which explored the prevailing care pathway for treating patients with RA, the factors that influenced routine treatment decisions, and the potential barriers to using ADAb and drug level testing in practice. The evidence from *Chapter Three* subsequently informed the structure of the *de novo* decision analytic model. *Chapter Four* developed the findings of the qualitative study, and reports a quantitative econometric analysis of the patient-level factors associated with a specific treatment decision (the choice of TNFi therapy) using a representative sample of patients with RA in England. *Chapter Five* presents a thorough conceptualisation of the *de novo* decision analytic model and economic evaluation of stratified medicine. The chapter begins by exploring how ADAb and drug level testing may be used to stratify treatment in routine clinical practice and presents a novel algebraic conceptualisation technique to identify potentially relevant comparator testing strategies. *Chapter Five* then defines the decision problem of the economic evaluation, conceptualises the progression of RA and the relevant care pathways, justifies the type of decision analytic model chosen to address the decision problem, and presents the final structure of the *de novo* decision analytic model. *Chapter Six* presents the full economic evaluation of the adalimumab ADAb and drug level ELISA tests to stratify treatment for patients with RA in England. *Chapter Seven* concludes the thesis by discussing the specific contributions to knowledge, the implications of the results to different stakeholders, the potential limitations of the thesis, and suggests possible topics for further research.

This thesis provided seven clear contributions to knowledge:

(i) The first synthesis and critical appraisal of existing economic evidence for stratified medicine in RA;
(ii) The first exploratory analysis of the factors that influenced rheumatologists' prescribing of biologic therapies for RA in England;
(iii) The identification of potential barriers to using ADAb and drug level testing in routine practice, perceived by rheumatologist in England;
(iv) The first quantitative analysis of the patient-level factors that influenced the choice of TNFi therapy prescribed to patients with RA in England, using data from treatment decisions observed in routine practice;

(v) The first synthesis of published clinical recommendations for using TNFi ADAb and drug level testing in routine clinical practice for patients with RA;

(vi) A novel model conceptualisation technique to facilitate the early identification of potentially relevant comparator strategies during the development of a *de novo* model-based economic evaluation of a stratified medicine;

(vii) The first estimate of the relative cost-effectiveness of adalimumab ADAb and drug level ELISA testing to stratify treatment for patients with RA in England, and the potential value of conducting further prospective research subsequent to this thesis.

Chapter 2

Existing Economic Evidence for Stratified Medicine in Rheumatoid Arthritis

Chapter Two reports an investigation into the existing economic evidence for stratified medicine in RA. A published version of *Chapter Two* (Gavan et al., 2014) is provided in *Appendix 8*; the content of this thesis chapter has been updated (in December 2016) to account for new evidence since publication of the peer-reviewed manuscript. The chapter is presented as a standalone study in terms of an introduction (Section 2.1), aim and objectives (Section 2.2), method (Section 2.3), results (Section 2.4), discussion (Section 2.5), and conclusion (Section 2.6).

2.1. Introduction

The application of stratified medicine in RA posed a timely and policy-relevant case study with the potential to improve the relative cost-effectiveness of care. Stratified medicine in RA may be valuable to the health care system in England given the high disease prevalence, the high relative cost of treatment for patients with severe disease activity, and the potential for heterogeneity in treatment response (for example, in rates of treatment failure, adverse events, and effectiveness). Moreover, bDMARD therapies remain widely available in England despite questionable evidence of their relative cost-effectiveness (see Section 1.3.3). In practice, treatment for RA could be stratified at multiple points along the care pathway, from presentation with asymptomatic early arthritis to the management of severe inflammation in later stages of the disease (see Section 1.3.4). Despite the clinical

interest, the research investment to identify biomarkers associated with treatment response, and the growing potential for stratified medicine in RA, an investigation of the existing economic evidence to support this emerging treatment paradigm had not yet been undertaken.

In the context of this thesis, the existing literature of relevance was previous economic evaluations of stratified medicine for RA that had used a decision analytic model. The purpose of reviewing existing economic evaluations differs from a review of conventional clinical evidence. Economic evaluations are primarily produced to inform a specific decision and their findings may not generalise to different decision-making contexts (Sculpher et al., 2006b; Vale, 2010). For example, decision-makers in different jurisdictions may have (i) different objectives, (ii) different budget constraints, (iii) different decision problems, (iv) different context-specific inputs such as unit costs and (iv) different reference case standards by which to perform an economic evaluation (Sculpher et al., 2006b; Anderson, 2010). Therefore, it was unlikely to be appropriate to synthesise the principle findings from a set of economic evaluations (similar to the process of a meta-analysis) to make general statements about whether a health technology was, or was not, relatively cost-effective (Anderson, 2010).

A more appropriate purpose of reviewing existing economic evaluations was to inform future decisions regarding the design of a *de novo* economic evaluation (Anderson, 2010). For example, existing economic evaluations may inform the potential type and structure of a subsequent decision analytic model, and may document the availability of evidence for particular input parameters (Anderson, 2010). The technology appraisal programme for NICE, and other health technology assessment agencies around the world, incorporates a review of the existing economic and clinical literature before generating a *de novo* decision analytic model to provide evidence for a specific resource allocation problem (Anderson, 2010; National Institute for Health and Care Excellence, 2011a). Relevant topics for this thesis regarding the existing economic evidence for stratified medicine in RA included investigating: (i) the motive to stratify treatment (for example, to improve effectiveness or avoid adverse events); (ii) how the evidence on test accuracy was identified and synthesised; (iii) the methods to characterise uncertainty; and (iv) whether (and how) any VOI analyses were performed to estimate the value of further research.

2.2. Aim and Objectives

The aim of this study was to identify and appraise the existing economic evidence for stratified medicine in RA. There were two objectives to meet this aim:

Objective 1: Identify all published model-based economic evaluations of stratified medicine for RA;

Objective 2: Critically appraise all published model-based economic evaluations of stratified medicine for RA.

2.3. Method

A systematic review and critical appraisal of published model-based economic evaluations of stratified medicine for RA was performed according to the *Preferred Reporting Items for Systematic Reviews and Meta-analyses* (PRISMA) reporting standards (Liberati et al., 2009). The completed PRISMA checklist for this study is reported in *Appendix 9*.

Study selection

The study inclusion criteria (reported in Table 2.1) was designed to incorporate as wide a set of relevant economic evaluations to the review as possible.

Table 2.1. *Systematic review inclusion criteria: economic evaluations of stratified medicine for RA.*

Study Characteristic	Inclusion Criteria
Population	Any population that included adults (> 16 years) with RA.
Intervention	Any test-based strategy of a biomarker to stratify any treatment decision with any pharmacological therapy.
Comparator	Any comparator treatment strategy.
Outcomes	Expected costs and expected patient benefits per intervention strategy.
Study Design	Full economic evaluation (cost-benefit analysis, cost-effectiveness analysis, cost-utility analysis), comparing at least two treatment strategies, using a decision analytic model.
Language	English; full-text publication.

A study was included if it was a full economic evaluation, defined by Drummond et al. (2015) as the joint comparison of at least two alternative interventions in terms of both their costs and consequences. This definition included all CUA, CEA, and CBA analyses. A stratified medicine intervention was defined broadly, consistent with the definition in Section 1.2, as any test-based strategy of a biomarker to stratify any treatment decision, within a patient population that included at least some adults with RA. Only economic evaluations based on a decision analytic model were included in the review, as they were the most relevant studies to inform the development of the subsequent *de novo* model-based economic evaluation.

Medline, Embase, Web of Science and the NHS *Economic Evaluations and Health Technology Assessment* (NHS EED and HTA) databases were searched electronically, initially between January 1990 and January 2014. These four databases were deemed appropriate because a previously published systematic review had used the same electronic databases to identify all economic evaluations of TNFi therapies for RA (Heather et al., 2014). The review searched for published studies from 1990 because (i) examples of higher-cost bDMARDs (assumed to be a motivating factor for stratified medicine) were introduced into practice in the 1990s (Mócsai et al., 2014); and (ii) the number of published economic evaluations increased during the 1990s because health care decision-makers began to demand economic evidence (Hutton, 2012). Taking into account the number of studies identified in each database, the systematic review was subsequently updated by searching the *Medline* and the NHS *EED and HTA* databases electronically from January 2014 until December 2016.

The search strategies to identify published studies from the electronic databases are reported in *Appendix 10*. The *Medline* and *Embase* search strategies combined index and free-text terms for *rheumatoid arthritis* with the published search filters to identify economic evaluations produced by *the Centre for Reviews and Dissemination* (Duffy et al., 2007). The *Web of Science* search strategy combined terms for *rheumatoid arthritis* and *economic evaluations*. The NHS *EED and HTA* databases were manually searched using the Medical Subject Headings (MeSH) term "*Arthritis, Rheumatoid*". Finally, the reference lists of all included studies were hand-searched for publications that met the inclusion criteria of the systematic review.

The title and abstract of all publications identified by the search strategies were screened by SG against the inclusion criteria in Table 2.1. Four researchers at *the Manchester*

Centre for Health Economics, The University of Manchester, were allocated an equal proportion of titles and abstracts to second-screen independently. Abstracts were not excluded at the screening stage if there were disagreements between SG and the independent reviewers. Studies that remained after the screening stage were read in full by SG to determine whether a full economic evaluation of a stratified medicine for RA had been performed.

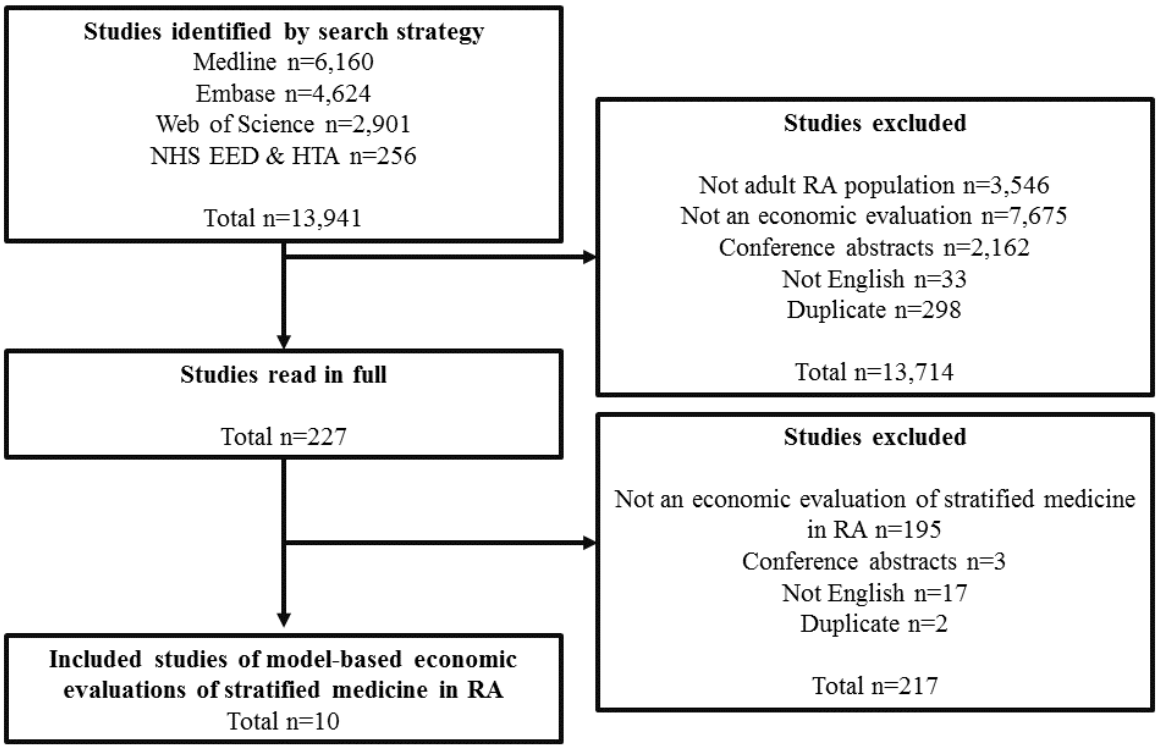
Data Extraction and Analysis

A data extraction form was designed according to the recommendations of the *Centre for Reviews and Dissemination* for appraising published economic evaluations (Craig et al., 2007). The form comprised (i) key study features (target population, intervention and comparator, evaluation vehicle); (ii) sources of evidence; (iii) components of the economic evaluation (costs included, measure of effectiveness, uncertainty analysis); and (iv) the results of the economic evaluation (Craig et al., 2007). Particular attention for this systematic review was given to the test characteristics in each economic evaluation, including the source of evidence for test performance, the assumed timing of testing, and the reported health and resource consequences of testing. All data extraction was performed by SG. Key features of each study were summarised in tabular form and a critical appraisal of the included economic evaluations was performed by a narrative synthesis of the evidence base.

2.4. Results

A flow diagram of the study inclusion procedure is illustrated in Figure 2.1. The search strategy identified 13,941 study titles and abstracts, and 227 manuscripts were read in full. Ten studies met the inclusion criteria by reporting a model-based economic evaluation of a stratified medicine for RA, published between January 1990 and December 2016.

Figure 2.1. *Systematic review flow diagram of included studies: economic evaluations of stratified medicine for RA.*



The completed data extraction forms for the ten economic evaluations included in the systematic review are reported in *Appendix 11*. An overview of the stratified medicine evaluated by each study is provided in Table 2.2, detailing the source of heterogeneity within the patient population, the test to reveal this heterogeneity, and the stratified treatment decision informed by the test result.

Table 2.2. *Systematic review: ten economic evaluations of stratified medicine for RA.*

Author (Year)	Source of Heterogeneity	Test to Reveal Heterogeneity	Stratified Treatment Decision
Purpose of Stratified Medicine: Reduce Adverse Drug Reactions			
Bergquist et al. (1995).	Patients may develop toxicity to methotrexate whilst receiving treatment.	Liver biopsy during treatment.	Continue or discontinue methotrexate.
Kim et al. (2006).	Patients with a MTHFR mutation may develop toxicity to methotrexate whilst receiving treatment.	Genotype test (MTHFR) before commencing treatment.	Different initial dose of methotrexate according to MTHFR mutation.
Kowada. (2010).	Patients may have latent tuberculosis infection, increasing the risk of adverse events during TNFi therapy.	Interferon-gamma release assay to test for tuberculosis infection before TNFi therapy.	Commence treatment protocol for tuberculosis infection.
Marra et al. (2002).	Patients with a TPMT mutation may experience severe adverse drug reactions to full dose azathioprine.	Genetic test (TPMT) activity before starting treatment.	Different initial dose of azathioprine according to TPMT mutation.
Oh et al. (2004).	Patients with a TPMT mutation may experience severe adverse drug reactions to full dose azathioprine.	Genetic test (TPMT) activity before starting treatment	Different initial dose of azathioprine according to TPMT mutation.
Solomon et al. (2000).	Patients with low bone mineral density may develop osteoporosis after receiving corticosteroids.	Test bone mineral density by x-ray before treatment.	Treat with corticosteroids according to bone mineral density test.
.			
Purpose of Stratified Medicine: Improve Treatment Effectiveness			
Konnopka et al. (2008).	Patients with early undifferentiated arthritis may subsequently develop RA.	Anti-CCP test to diagnose cases of RA earlier at clinical presentation.	Earlier effective treatment of RA with cDMARDs.
Nair et al. (2015)	Patients with differential inflammation may benefit from intense, tight-control, methotrexate.	Monitoring of inflammation using *handscan* imaging.	Adjustment of methotrexate dose according to inflammation.
Suter et al. (2011).	Patients may have differential risks of radiographic disease progression.	Include MRI scan in the standard risk-stratification protocol.	Early treatment with combination cDMARD in patients at-risk of severe progression.
Purpose of Stratified Medicine: Reduce Unnecessary Health Care Resources			
Krieckaert et al. (2015).	Patients receiving adalimumab may have drug levels greater than, or less than, a target value to achieve an effective response.	ELISA test of adalimumab drug levels 28 weeks after commencing treatment.	Continue treatment, reduce-dose of adalimumab, or change treatment to a different bDMARD.

Note: Anti-CCP=antibodies against cyclic citrullinated peptides; bDMARD=biologic disease-modifying antirheumatic drug; cDMARD=conventional disease-modifying antirheumatic drug; ELISA=enzyme-linked immunosorbent assay; MRI=magnetic resonance imaging; MTHFR=methylenetetrahydrofolate reductase; TNFi=tumour necrosis factor-α inhibitor; TPMT=thiopurine-methyltransferase.

The results reported by each economic evaluation were mostly favourable towards stratified medicine within their respective jurisdiction (Konnopka et al., 2008), and the approach to stratified medicine was identified as the dominant alternative in six studies (Marra et al., 2002; Oh et al., 2004; Kim et al., 2006; Kowada, 2010; Krieckaert et al., 2015; Nair et al., 2015). However, the validity of these principal results should be viewed as conditional on the decisions made for each decision analytic model in terms of the identification of supporting evidence, the structure of the model, and the characterisation of parameter uncertainty.

Four studies performed an economic evaluation by using a simple decision tree (Bergquist et al., 1995; Marra et al., 2002; Oh et al., 2004; Kim et al., 2006), four studies used a Markov model (Solomon et al., 2000; Suter et al., 2011; Krieckaert et al., 2015; Nair et al., 2015), and two studies combined a decision tree and Markov model (Konnopka et al., 2008; Kowada, 2010). Three economic evaluations were conducted for a lifetime time horizon (Solomon et al., 2000; Kowada, 2010; Suter et al., 2011) whereas the seven remaining economic evaluations had much shorter time horizons, ranging between six months to ten years. No study documented the conceptualisation procedure of model development and only one economic evaluation varied the timing of testing (Bergquist et al., 1995; tested patients at five years and ten years).

The majority of economic evaluations provided limited evidence regarding the estimation of test accuracy. Four studies assumed that testing was perfectly accurate with a sensitivity and specificity of 100% (Bergquist et al., 1995; Solomon et al., 2000; Kim et al., 2006; Nair et al., 2015). A single published source of evidence was used to identify test accuracy in three studies (Marra et al., 2002; Oh et al., 2004; Kowada, 2010). Krieckaert et al. (2015) did not report the accuracy of adalimumab drug level testing. Suter et al. (2011) identified multiple sources of evidence to derive values for test sensitivity and specificity, however no method of evidence synthesis was performed with these test accuracy data. By contrast, Konnopka et al. (2008) derived the sensitivity and specificity of anti-CCP testing from a published systematic review and meta-analysis.

The resource use (and subsequent cost) of testing in the majority of studies (n=9) was assumed to be represented by the unit cost of testing. Only Kowada (2010) accounted for the additional resources required to operationalise testing in practice, including the time to collect a sample of blood, the need for an additional patient visit to the health care professional, and the time for a laboratory technician to evaluate the samples.

All studies characterised parameter uncertainty with a deterministic *one-way sensitivity analysis* by making manual adjustments to the values of individual input parameters. The prior prevalence of patients with the target condition addressed by the stratified medicine (for example, the prevalence of patients with a particular genetic mutation who subsequently experience an adverse drug reaction) was a key driver of relative cost-effectiveness in five studies (Bergquist et al., 1995; Solomon et al., 2000; Kim et al., 2006; Kowada, 2010; Suter et al., 2011). Other common drivers of relative cost-effectiveness were the cost of testing (Marra et al., 2002), and the severity and likelihood of treatment according to an incorrect test result (Konnopka et al., 2008; Suter et al., 2011). A PSA was performed in five economic evaluations (Konnopka et al., 2008; Kowada, 2010; Suter et al., 2011; Krieckaert et al., 2015; Nair et al., 2015). No study undertook a VOI analysis to estimate the value of further research to reduce parameter uncertainty.

2.5. Discussion

This study identified and critically appraised ten model-based economic evaluations of different examples of stratified medicine for RA. In all included studies, a test was used to reveal unobserved heterogeneity in a population of patients with RA, which was then subsequently used to stratify a specific treatment decision. Consistent with the clinical literature that addressed the potential for stratified medicine in RA (see Section 1.3.4), the economic evaluations incorporated testing strategies that revealed heterogeneity between patients at all points in the care pathway, including: (i) pre-diagnosis of RA (Konnopka et al., 2008; Suter et al., 2011), (ii) first-line cDMARD therapy (Bergquist et al., 1995; Marra et al., 2002; Oh et al., 2004; Kim et al., 2006; Nair et al., 2015), (iii) prior to commencing bDMARD therapy (Kowada, 2010) or corticosteroid therapy (Solomon et al., 2000), and finally, (iv) during treatment with a bDMARD (Krieckaert et al., 2015).

The studies generally reported favourable evidence for the relative cost-effectiveness of stratified medicine. However, the reliability of the studies' findings, and their usefulness for informing decision-making, may be affected by the potential limitations of their analytic approaches (Anderson, 2010; Tappenden et al., 2014). Ultimately, these potential limitations had several implications for the subsequent design of the model-based economic evaluation of adalimumab ADAb and drug level testing presented in this thesis (in *Chapter Six*).

Current best-practice for developing a *de novo* decision analytic model for RA recommended to represent the chronic nature and gradual worsening of the disease through changes to a patient's HAQ score over time (Madan et al., 2015). For example, all academic models that informed previous NICE appraisals of TNFi therapies in England estimated HAQ progression by an individual patient simulation (Jobanputra et al., 2002; Barton et al., 2004b; Malottki et al., 2011; Stevenson et al., 2016). The use of decision tree models in this review did not incorporate the chronic worsening of RA due to their lack of an explicit time component (Sonnenberg et al., 1993; Briggs et al., 1998). No study performed an individual patient simulation, yet three studies represented disease progression over time in a Markov model (Konnopka et al., 2008; Krieckaert et al., 2015; Nair et al., 2015). However, Markov models may also be limited for evaluating a stratified medicine if the status of the unobserved biomarker differs between patients over time, and that biomarker affects the probability of a subsequent clinical event occurring (Caro et al., 2016b).

The economic evaluations in this review predominately compared a treatment strategy informed by a test to a strategy that represented current clinical practice without testing. The definition of current practice for RA was often unclear in the reported studies. Treatment decisions in England are made with reference to national recommendations by NICE (National Institute for Health and Care Excellence, 2009; 2016a); however, there remained evidence of substantial regional variation in the treatments prescribed to patients with RA (The British Society for Rheumatology, 2015). The studies were poor at explaining the conceptualisation of current clinical practice, which may bring into question its suitability as a relevant comparator strategy for stratified medicine.

Current clinical practice may not have been the only relevant comparator if the test featured in each economic evaluation could have been used in different ways (for example, by testing patients earlier or having a different treatment decision) (O'Mahony et al., 2015c). Only one study investigated variation in the timing of testing (Bergquist et al., 1995). Different ways of using a test to stratify treatment decisions can affect its relative cost-effectiveness, as patients may accrue costs and QALYs at a different rate between different testing strategies (Hatz et al., 2014; Shabaruddin et al., 2015) (see Section 1.2.5). Therefore, the credibility of the incremental analyses reported by each economic evaluation may have been improved if multiple relevant testing strategies were included within the respective decision problems (O'Mahony et al., 2015a).

A stratified medicine may require resources in addition to the test itself, such as clinician time to make a treatment decision and a patient appointment to collect a biomarker sample (Buchannan et al., 2013). Only one economic evaluation in this systematic review included all the resources necessary for testing (Kowada, 2010). Previous systematic reviews have identified that the relative cost-effectiveness of different stratified medicines were sensitive to the cost of testing (Phillips et al., 2004; Beaulieu et al., 2010; Oosterhoff et al., 2016). The majority of economic evaluations identified in this review may have underestimated the unit cost (and opportunity cost) of testing by omitting these additional resources (Luce et al., 1990) which, as a consequence, may have inaccurately improved the relative cost-effectiveness of the stratified medicine (Drummond et al., 2015).

While screening the abstracts for model-based economic evaluations, one RCT-based economic evaluation of a stratified medicine for RA was identified by Thompson et al. (2014). This prospective study investigated the relative cost-effectiveness of TPMT testing to stratify azathioprine therapy using a *pragmatic trial* design. Pragmatic trials are performed in routine clinical practice to investigate the effectiveness of an intervention health technology (Roland et al., 1998); in the case of Thompson et al. (2014), clinicians could make their own treatment decisions but were guided by pre-defined recommendations on how to interpret the TPMT test result. Clinicians must make the appropriate treatment decision according to a test result to realise the benefit of treatment stratification in practice (Garrison et al., 2006; Annemans et al., 2013; Buchannan et al., 2013). However, Thompson et al. (2014) found that clinicians may not have followed the test result, as described in the trial protocol, when making their treatment decisions. Consequently, the estimated cost-effectiveness of a stratified medicine derived from a model-based economic evaluation may not be achieved if clinicians perceive that there are barriers to testing in routine practice.

Limited information was provided by the economic evaluations in this systematic review regarding the identification of evidence for test accuracy. The NICE DAP programme, when appraising a stratified medicine, requires that the evidence of a test's accuracy must be identified by a systematic review (National Institute for Health and Care Excellence, 2011a); however only one economic evaluation in this study identified a test's sensitivity and specificity from a systematic review (Kowada, 2010). Three studies assumed perfect test accuracy, meaning that the consequences of an incorrect (false-positive and false-negative) test result were not incorporated in the respective evaluations (Annemans et al., 2013). Suter et al. (2011) identified multiple sources of evidence for test accuracy,

however no formal method of evidence synthesis, such as a bivariate meta-analysis, was used to combine these data (Harbord et al., 2008; Macaskill et al., 2010). The relative cost-effectiveness of a stratified medicine may ultimately be sensitive to the proportion and/or consequence of incorrect treatment decisions due to test inaccuracies (Phillips et al., 2004; Annemans et al., 2013; Oosterhoff et al., 2016).

Decision-makers customarily require model-based economic evaluations to propagate parameter uncertainty with a PSA (Claxton et al., 2005; National Institute for Health and Care Excellence, 2013a); however, only half of the economic evaluations included in this review performed a PSA. No study investigated the value of additional research by performing a VOI analysis, which has been increasingly utilised in early model-based economic evaluations characterised by limited supporting evidence (Mohiuddin et al., 2014; Buisman et al., 2016). The studies could have used VOI methods, given the limited evidence reported for some parameter inputs related to stratified medicine (for example, test accuracy or the impact of treatment on patient outcomes), to investigate whether the cost of additional prospective research may exceed its potential value in reducing uncertainty in the estimate of relative cost-effectiveness (Claxton et al., 2001; Wilson, 2015).

Limitations

One potential limitation of this systematic review was that the target populations of two economic evaluations included patients with different inflammatory diseases (not only RA) (Marra et al., 2002; Oh et al., 2004). However the two studies met the inclusion criteria in Table 2.1, ensuring that the widest-possible set of relevant model-based economic evaluations were included in the systematic review.

A second potential limitation was that only *Medline* and the NHS *EED & HTA* databases were searched electronically to update the systematic review in December 2016. This decision was unlikely to have affected the results because the original database search (in January 2014) did not identify any unique economic evaluations from *Embase* or *Web of Science* that were subsequently included in the systematic review.

A third potential limitation was the exclusion of trial-based analyses from the systematic review. Decision analytic models conventionally make simplifying assumptions that may overlook important contextual factors related to the use of a health technology in routine

clinical practice. For example, the pragmatic trial by Thompson et al. (2014) reported that clinicians may not prescribe the correct treatment according to the result of a test; the decision analytic models within this systematic review, however, assumed that treatments were prescribed correctly according to each testing strategy. The absence of such contextual factors in model-based economic evaluations, which may be revealed within pragmatic trial-based analyses, may limit the generalisability of the results generated by a decision analytic model. However, the search strategy identified only one trial-based economic evaluation of a stratified medicine for RA and it was unlikely that additional studies of a similar design were also present within the literature.

Implications for Future Research

The findings of this systematic review provided seven implications for the conceptualisation of the *de novo* decision analytic model presented in this thesis. These seven implications are now summarised:

(i) The decision analytic model may benefit from being implemented as an individual patient simulation, consistent with current best-practice in RA, given that the status of adalimumab ADAb and drug levels may change over time, which may also affect the likelihood of future clinical events (Caro et al., 2016b);

(ii) Given the uncertainty in defining current practice for RA, an exploration of prescribing decisions for RA in England may help to inform the characterisation of the care pathway within the economic model (Tappenden, 2014);

(iii) An extensive review of the potential ways to incorporate adalimumab ADAb and drug level testing into the care pathway for RA was required, to ensure that all relevant comparator testing strategies were included in the model (Buisman et al., 2016);

(iv) An investigation into the potential barriers to routine adalimumab ADAb and drug level testing perceived by rheumatologists in England may be beneficial, in light of evidence that clinicians may not choose to follow the result of a test to stratify a treatment in clinical practice (Thompson et al., 2014);

(v) All resources associated with adalimumab ADAb and drug level testing must be quantified and valued to account for their opportunity cost (Oosterhoff et al., 2016);

(vi) Test accuracy data must be identified by a systematic review and combined using appropriate methods of evidence synthesis (National Institute for Health and Care Excellence, 2011a);

(vii) A VOI analysis may be beneficial to investigate the value of further prospective research regarding adalimumab ADAb and drug level testing (Steuten et al., 2014).

2.6. Conclusion

The aim of this study was to identify appraise the existing economic evidence supporting the use of stratified medicine in RA. A systematic review identified ten published model-based economic evaluations in RA that had estimated the relative cost-effectiveness of a stratified medicine. The examples of stratified medicine used different means of testing to reveal unobserved heterogeneity in the patient population at different points in the care pathway for RA. The results suggested that the existing evidence for the cost-effectiveness of stratified medicine in RA, albeit generally favourable within their respective jurisdiction, may be limited by the analytic choices and assumptions made within each decision analytic model.

The results of the systematic review provided seven implications for conducting the *de novo* economic evaluation of adalimumab ADAb and drug level ELISA testing in this thesis. These seven implications are subsequently addressed in the forthcoming chapters, including the characterisation of current practice (*Chapter Three* and *Chapter Four*), the conceptualisation of the decision problem and decision analytic model (*Chapter Five*), and the estimation of relative cost-effectiveness (*Chapter Six*).

Chapter 3

Understanding Treatment Decisions for Rheumatoid Arthritis in Current Practice

Chapter Three presents a qualitative investigation of treatment decisions for RA made in current practice with a sample of senior consultant rheumatologists in England. The study provided a source of evidence for: (i) conceptualising the structure of the subsequent decision analytic model (described in *Chapter Five*), and (ii) the potential barriers to introducing ADAb and drug level testing into routine practice. An earlier version of this study was presented at the *EULAR 2016 Annual European Congress of Rheumatology* (Gavan et al., 2016a). The chapter comprises an introduction (Section 3.1), aim and objectives (Section 3.2), method (Section 3.3), results (Section 3.4), discussion (Section 3.5), and conclusion (Section 3.6).

3.1. Introduction

The structure of any decision analytic model, used to conduct an economic evaluation, is conventionally designed according to an existing care pathway, so that its outcomes are estimated with relevance to a specific decision-making context (National Institute for Health and Care Excellence, 2011a; Tappenden, 2014). The model-based economic evaluation of adalimumab ADAb and drug level testing to stratify treatment, presented in this thesis, therefore, required an understanding of current practice for RA in England. The results of the systematic review in *Chapter Two* indicated that previous model-based economic evaluations of stratified medicine in RA had not reported their characterisations

of current practice with clarity. *Chapter Three* was subsequently designed to understand the treatment decisions made in current practice for patients with RA in England.

Prevailing clinical recommendations and guidelines may be used as one potential source of evidence to understand the treatment decisions within an existing care pathway (Tappenden, 2014). For example, it may be possible to deduce the appropriate care pathway for RA from NICE recommendations, which are used by rheumatologists in England to inform routine treatment decisions for patients in the NHS (National Institute for Health and Care Excellence, 2009; 2016a). NICE recommendations, in theory, were established to improve the standard of care in the NHS, promote a cost-effective use of population health care resources, and minimise inappropriate regional variation in clinical practice (Walker et al., 2007). In the context of a treatment decision for a specific patient, however, NICE recommendations are fundamentally advisory (Cookson et al., 2001) and do not preclude the expression of clinical judgement (National Institute for Health and Care Excellence, 2016b). As a potential consequence, recent studies have identified considerable regional variation in the treatment decisions for patients with RA in England, despite the existence of national evidence-based recommendations by NICE (Tugnet et al., 2013; Blake et al., 2014; The British Society for Rheumatology, 2015). Such regional variation indicated that clinical recommendations and guidelines may have been necessary, but not sufficient, sources of evidence for this thesis to understand the care pathway for RA in England.

The NICE clinical guideline for RA, for example, includes seven recommendations for patient management (National Institute for Health and Care Excellence, 2009). In 2014-15, *The British Society for Rheumatology* performed a high-profile national audit of rheumatology practice against these seven recommendations by analysing patient data (n=6,354) from the majority of NHS rheumatology providers (n=135) in England and Wales (The British Society for Rheumatology, 2015). The audit reported evidence of non-trivial regional variation in the implementation of the seven recommendations; for example, (i) access to rheumatology services within three weeks of referral was observed in 55% of trusts in London and 32% of trusts in the Midlands and East England; and (ii) education and self-management activities were provided in 63% of cases in the North of England and 38% of cases in London (The British Society for Rheumatology, 2015).

Tugnet et al. (2013) surveyed 311 patients with RA from nineteen rheumatology units located in the Midlands of England and found that, even by restricting the analysis to this

specific geographic region, between-hospital variability was also present in the uptake of the same recommendations by NICE. For example, only 25% of the rheumatology units within the sample conducted monthly assessments of disease activity (DAS28) and 40% of units treated their patients with combination cDMARDs. Statistically significant differences in implementing recommendations were also observed between rheumatology units with, and without, an early inflammatory arthritis clinic (Tugnet et al., 2013).

Blake et al. (2014) reported a related study that investigated the extent to which patients with RA, who changed treatment between two bDMARD therapies, were compliant with NICE recommendations. Data were collected from all patients with RA who changed bDMARD therapies across eighteen rheumatology units in the East and West Midlands, England. Decisions to change between bDMARDs were compliant with NICE recommendations in 65% of cases. Additionally, individual hospitals varied in their rates of compliance with NICE recommendations (between 50% of patients to 100% of patients) which may suggest that unobserved hospital-specific factors had an influence on treatment decision-making (Blake et al., 2014).

An alternative source of evidence, that has been recommended to understand the treatment decisions made within an existing care pathway, is to explore the views of experts within the health care system for their input (Chilcott et al., 2010; National Institute for Health and Care Excellence, 2011a; Tappenden, 2014). Qualitative methods of research, in general, can be used to formalise the interpretation and understanding of the views provided by different individuals (Snape et al., 2003). Such methods, in the context of this thesis, had the potential to be valuable, by enabling an understanding of specific observations within a relevant clinical scenario (Coast, 1999; Obermann et al., 2013), such as regional variation in treatment decisions. Qualitative methods have also been explicitly recommended as one approach to inform the structure of a *de novo* decision analytic model during the conceptualisation and development stages of an economic evaluation (Husbands et al., 2017). In particular, the use of clinical experts to inform the design of a decision analytic model can be valuable during an early-stage economic evaluation in order to describe the characteristics of current practice, as described within the conceptual framework presented in Figure 1.1 (Sculpher et al., 1997).

No previous research had used qualitative methods to explore routine bDMARD prescribing decisions for patients with RA with a sample of consultant rheumatologists in England. Evidence provided by consultant rheumatologists had the potential to inform the

understanding of three relevant topics of interest for this thesis: (i) the care pathway for patients with RA in England, (ii) the factors that may influence treatment decisions in current practice, and (iii) the potential barriers to introducing ADAb and drug level testing into routine clinical practice for RA.

3.2. Aim and Objectives

The aim of this study was to understand the current prescribing and treatment practices for the management of patients with RA with TNFi therapies in England. Three objectives were addressed to meet this aim:

Objective 1: To explore the differences between rheumatologists' approaches to managing patients with RA in England;

Objective 2: To understand the reasoning for why rheumatologists in England choose to treat their patients with RA in particular ways;

Objective 3: To explore the potential barriers perceived by rheumatologists regarding the use of ADAb and drug level testing in routine practice.

3.3. Method

This study used a qualitative thematic framework analysis of semi-structured one-to-one in-depth interviews with senior consultant rheumatologists across England. The study was performed according to the best-practice *Standards for Reporting Qualitative Research* (SRQR) checklist (O'Brien et al., 2014); the completed *SRQR* checklist for this study is reported in *Appendix 12*. The method is reported according to the study population and sample (Section 3.3.1); data collection (Section 3.3.2); data analysis (Section 3.3.3); and ethical approval (Section 3.3.4).

3.3.1. Study Population and Sample

The target population of this study was senior consultant rheumatologists that had experience of treating patients with RA in England. A purposive sample of individuals were recruited, consistent with this target population, according to their professional role of

employment. A *purposive sample* is a non-random sample collected with a purpose of enabling an understanding of the phenomenon being researched (Silverman et al., 2008).

A *sampling frame* defines the individuals, within a target population, that were eligible for recruitment to a particular study (Morgan, 2008). The sampling frame for this study, from which participants were recruited, was the list of principle investigators (n=45) of the *Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate* (BRAGGSS) cohort study. The quantitative study, reported in *Chapter Four*, analysed the patient-level data from BRAGGSS; the use of the BRAGGSS principal investigators in the sampling frame of this study therefore introduced a degree of consistency between the research reported in *Chapter Three* and *Chapter Four*. Each principle investigator was a consultant rheumatologist, experienced in treating RA patients with TNFi therapies, from a different hospital across England. The sample inclusion criteria were: (i) consultant rheumatologists in England; (ii) who were principle investigators of BRAGGSS; (iii) with experience of treating patients with RA.

An identical participant recruitment email and participant information sheet was sent individually to all rheumatologists in the sampling frame in December 2014. Rheumatologists who did not respond were sent a follow-up recruitment email in March 2015. No further contact was made with non-responding rheumatologists after March 2015. The participant recruitment email, and the participant information sheet that was attached to each email, are reported in *Appendix 13*.

3.3.2. Data Collection

In the context of qualitative research, interviews can facilitate a conversation with a purpose (Miller, 1995) to explore the views and experiences of individuals regarding a specific phenomenon (Gill et al., 2008). In-depth semi-structured interviews were conducted with open-ended questioning that enabled the rheumatologists to provide expansive responses. A *semi-structured interview* is a strategy for data collection in which a series of predetermined questions are posed, based on the research objectives, which may be asked in any order according to the responses of each participant (Ayres, 2008). The interview schedule addressed the three objectives of this study in seven topics of questioning, reported in Table 3.1. A pilot interview was conducted with a clinical research fellow (Dr. Meghna Jani) at *The University of Manchester*, who had experience of treating

patients with RA in England, to evaluate the suitability of the interview questions and its structure. The complete semi-structured interview schedule is reported in *Appendix 14*.

Table 3.1. *Seven topics addressed by the semi-structured interview schedule.*

Research Objective	Topic Addressed by Interview Schedule
1 & 2	The interpretation of NICE recommendations.
1 & 2	Procedures to ensure compliance with NICE recommendations.
2 & 3	Assessing the suitability of TNFi therapy.
2 & 3	The decision of choosing the first TNFi.
2 & 3	Treatment decisions after TNFi failure.
2 & 3	Beliefs about the five TNFis recommended by NICE.
3	The use of ADAb and drug level testing in routine practice.

Additional interview questions were posed according to the responses of preceding participants, consistent with grounded theory method of qualitative research (Glaser et al., 1967). For example, if one participant described a unique experience, the questions posed to subsequent participants were updated to assess whether they had shared a similar experience.

Participants were interviewed by SG over the telephone to increase the feasibility of conducting one-to-one interviews with a sample of practicing clinicians distributed across England, compared with scheduling a face-to-face meeting, during working-hours (Miller, 1995). To minimise the burden of participating, the rheumatologists that provided consent were able to schedule the interview at a date and time that was most convenient for them. SG conducted all telephone interviews from the *Manchester Centre for Health Economics, The University of Manchester*.

The telephone interviews were recorded by a digital audio recorder and the content of all interviews was transcribed verbatim by SG using the method reported by Poland (1995). The audio of each interview was replayed in full after it had been transcribed to ensure data integrity and congruence between the audio and the transcript (Poland, 2008). The transcripts were made anonymous by SG by removing references to the names of individuals, geographic locations, and labelling each transcript alphabetically in ascending order of the interview date. The rheumatologists that participated were therefore unable to remove their data from the study after transcription.

3.3.3. Data Analysis

The interview transcripts were analysed by an inductive thematic framework analysis (Ritchie et al., 2001; Braun et al., 2006; Gale et al., 2013). A *thematic analysis* is a qualitative method of data analysis in which the researcher takes an active role in identifying themes within the data (Braun et al., 2006). An *inductive analysis* indicated that such patterns were identified within the data without specifying of how those patterns should be defined *a priori* (Patton, 2003). The analysis comprised six stages (familiarisation, coding, developing the framework, applying the framework, generating the framework matrix, and interpretation); these six stages are now described.

Stage One: *Familiarisation* – The initial stage of the framework analysis, common to most forms of qualitative research, was to become *immersed* in the data (Ritchie et al., 2001; Braun et al., 2006). Therefore, all telephone interviews were (i) conducted by SG and (ii) transcribed by SG, which are conventionally regarded to be the time period at which data familiarisation begins (Braun et al., 2006). The transcripts were then read repeatedly by SG, in an active way (for example, by making notes in the margins), to identify initial patterns of responses within and between individual transcripts (Braun et al., 2006).

Stage Two: *Coding* – A *code* was defined as a descriptive label that was applied to elements of a transcript that conveyed a meaning in relation to the research objectives (Braun et al., 2006). Codes may have referred to matters of fact (for example, an explicit statement made within a transcript) or to the emotions conveyed by the rheumatologists (for example, an expression of frustration towards a particular situation) (Gale et al., 2013). A specific code was applied by SG to each line of the transcripts to label excerpts from the interviews. Coding was systematic and thorough which enabled different sections of the transcripts to be compared between the sample of rheumatologists (Braun et al., 2006; Gale et al., 2013). Supplementary coding was performed by KP and Dr. Gavin Daker-White, who had extensive experience in the analysis of qualitative data. In addition, to enhance the trustworthiness of the analysis, a researcher at the *Manchester Centre for Health Economics, The University of Manchester,* who was independent from the study and blind to the study's design, provided supplementary coding in twenty percent of transcripts that were selected at random.

Stage Three: *Developing the Framework* – The codes *within* and *between* individual transcripts were grouped together according to their similarity. A *theme* was the term used

to describe a distinct set of codes that shared a common element, and reflected a pattern of responses across the rheumatologists, that were important in relation to the research objectives (Braun et al., 2006). All themes were constructed by SG.

Stage Four: *Applying the Framework* – The themes and codes that were identified within earlier transcripts were then applied to subsequent transcripts. The analytic framework was refined if excerpts of subsequent transcripts did not relate to a theme or code that was identified earlier (Ritchie et al., 2001). The interview transcripts were therefore analysed continually during data collection, rather than after all data had been collected (Glaser et al., 1967).

Stage Five: *Generating the Framework Matrix* – The transcribed data were then input into a matrix by a process known as *charting* (Gale et al., 2013). A separate matrix was created for each theme. Each column of the matrix represented a specific code and each row represented a specific participant (Gale et al., 2013). A summary of each participant's data was written for each code, within each cell along their specific row within the matrix (Ritchie et al., 2001). The matrix maintained a link to the original transcribed data by referring to specific page numbers of the transcripts and quotations (Ritchie et al., 2001; Gale et al., 2013). The data were charted from the transcripts to the framework matrix by SG.

Stage Six: *Interpretation* – The responses of all rheumatologists were compared across each code and theme after all data were charted to the framework matrix (Ritchie et al., 2001). Table 3.2 provides an example of a framework analysis matrix for one specific theme.

Table 3.2. *Example of a framework analysis matrix.*

	Theme 1		
Rheumatologist No.	**Code a**	**Code b**	**Code c**
Rheumatologist 1			
Rheumatologist 2			
Rheumatologist 3			

The matrix in Table 3.2 is based on the interview transcripts of three rheumatologists. The coding and framework development stages in this example identified one overall theme within the data (*"Theme 1"*) and three codes were identified that related to this theme

(*"Code a", "Code b", "Code c"*). A summary of the rheumatologists' data for each code (referring back to the original transcripts) would then be written within each empty cell of the matrix. These summarised data would then be analysed and interpreted by reading down a column to compare each rheumatologists' response for a specific code. The results were presented for each research objective by using a narrative synthesis with supporting quotations.

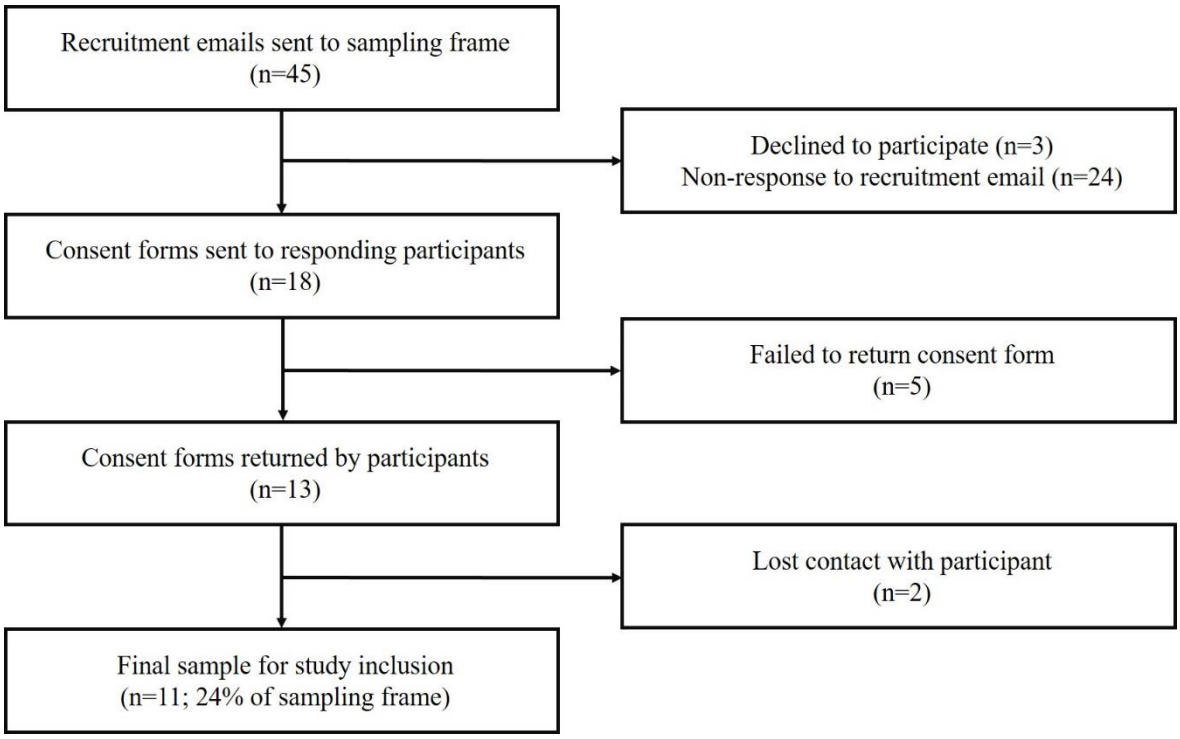
3.3.4. Ethical Approval

This research received ethical approval by *The University of Manchester Research Ethics Committee 2* (reference number: 14147). All rheumatologists contributed voluntarily and received no financial compensation. All participants provided informed consent and agreed to: (i) the recording of their telephone interview, and (ii) the publication of anonymous quotations within the final PhD thesis and peer-reviewed journal articles.

3.4. Results

Figure 3.1 illustrates a flow diagram of participant recruitment. Recruitment emails were sent to the sampling frame of consultant rheumatologists (n=45) in December 2014. Seventeen individuals responded, of whom three declined to participate. Follow-up recruitment emails were sent to the twenty-eight non-responding individuals in March 2015 and four additional rheumatologists agreed to participate. Thirteen completed consent forms were returned by rheumatologists who agreed to be interviewed. Contact was lost with two individuals after returning completed consent forms. Telephone interviews were conducted with the remaining participants (n=11; 24% of the sampling frame) between January and September 2015. All participants met the study inclusion criteria. The mean duration of the telephone interviews was 30 minutes (range: 16 minutes to 56 minutes).

Figure 3.1. *Flow diagram of participant recruitment to the qualitative study.*



The summary demographics of the final sample are reported in Table 3.3. The sample was distributed evenly by geographic region across England. The mean self-reported population size of patients with RA that were treated at each hospital was 2,450 patients.

Table 3.3. *Summary demographics of the final interviewed sample.*

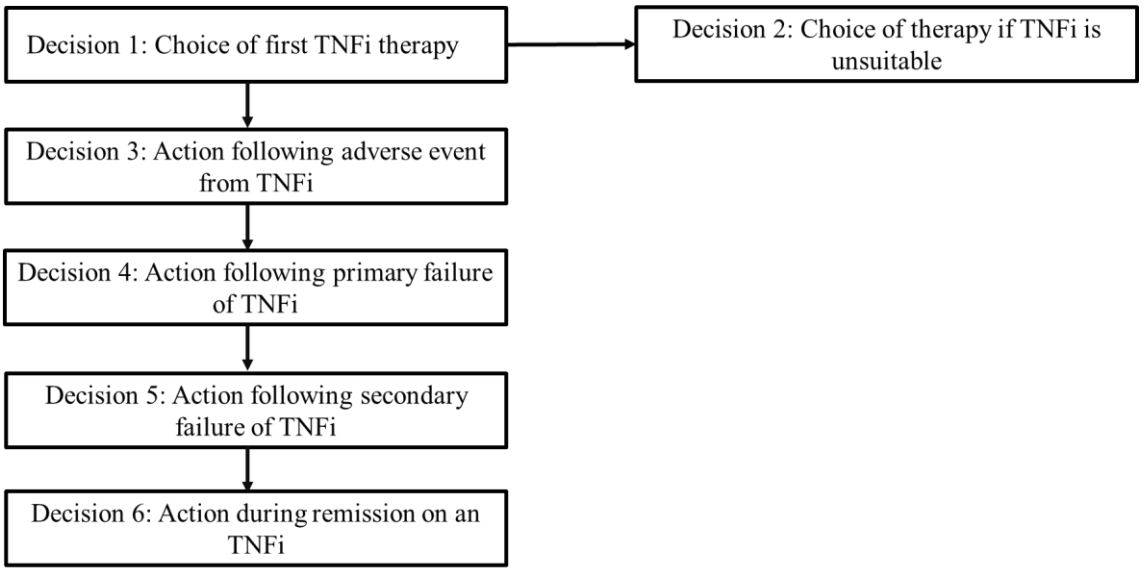
Sample Characteristic	**n (%)**
Sex	
Male	9 (82%)
Female	2 (18%)
Region in England	
North	4 (36%)
Midlands	4 (36%)
South	3 (27%)

The results are now presented in three sections for each research objective: the differences in treatment decisions (Section 3.4.1), the reasoning for treatment decisions (Section 3.4.2), and the potential barriers to ADAb and drug level assessment in routine practice (Section 3.4.3).

3.4.1. Objective 1: Differences in Routine Treatment Decisions

The interview schedule addressed six treatment decisions that could be made along the care pathway for patients with RA (illustrated in Figure 3.2) who became eligible for TNFi therapy. A table that reports the eleven rheumatologists' approaches to treatment at each decision point is provided in *Appendix 15*. This section explores the similarities and differences between the rheumatologists' responses.

Figure 3.2. *Six treatment decisions along the care pathway for RA.*



Decision 1: Choice of First TNFi

The participants reported three conceptually different approaches to choosing the first TNFi: (i) facilitating free patient choice; (ii) choosing the TNFi according to patient characteristics; and (iii) choosing the TNFi according to the recommendations of hospital-level guidelines or regional health care commissioners.

The set of TNFi therapies that were offered to patients varied (a restricted set versus all five TNFi therapies recommended by NICE), if the rheumatologists facilitated a free patient choice. Participants who made treatment decisions according to the requirements of regional health care commissioners varied in the choice of first TNFi (however, three participants reported that their health care commissioners required the first TNFi to be certolizumab pegol). Etanercept and infliximab were repeatedly cited as candidate first-line TNFi therapies in patients with infection or compliance issues, respectively.

Decision 2: when TNFi was Unsuitable

The rheumatologists explained their preferred treatment approaches for patients with RA when TNFi therapy was unsuitable, for example: (i) tocilizumab monotherapy for patients unable to receive concomitant methotrexate; (ii) abatacept for patients with an infection risk or multiple sclerosis; and (iii) rituximab for patients with evidence of malignancy, infection, or lung disease.

Decision 3: following an Adverse Event from TNFi

The treatment decision after a TNFi adverse event was reported to depend on its severity. Minor adverse events (such as an injection site reaction) were commonly followed with a second TNFi, in some cases switching between monoclonal and non-monoclonal agents. Alternatively, one rheumatologist favoured to continue the first TNFi and treat any pain with anaesthetic.

The rheumatologists argued that more severe adverse events (such as an infection) would prompt a decision to change treatment to a different class of bDMARD, such as rituximab. A minority of rheumatologists argued that they may use a second TNFi, if their patient had responded well to the first TNFi, following a severe adverse event.

Decision 4: following Primary Failure of TNFi

Primary failure of the first TNFi was followed by changing treatment to a different bDMARD. Only one rheumatologist claimed that they may attempt to control disease activity with a second TNFi, conditional on the patient's approval.

Decision 5: following Secondary Failure of TNFi

One rheumatologist claimed that they may attempt dose-escalation specifically after secondary failure of infliximab. However, a number of participants did not consider TNFi dose-escalation to be an appropriate treatment strategy. Most participants explained that treatment would be changed to a different bDMARD (such as rituximab) after secondary failure of a TNFi. Two rheumatologists explained that they may attempt a second TNFi if their patient responded well to an earlier TNFi.

Decision 6: during TNFi-induced Remission

Six rheumatologists explained that they would increase the interval between TNFi injections (to reduce the effective dose) for patients that had entered remission while receiving a TNFi. One rheumatologist preferred to maintain the regular TNFi injection intervals for patients in remission. Two rheumatologists argued that testing TNFi drug levels and ADAb may be helpful to inform treatment decisions in remission. In contrast, three rheumatologists argued to reduce a patient's concomitant methotrexate in remission (two of which explained this was only suitable for patients that had previously experienced an adverse drug reaction to methotrexate).

3.4.2. Objective 2: Reasons for Treatment Decisions

Given that differences between the rheumatologists' routine treatment decisions were reported along the care pathway for RA (Section 3.4.1), an exploratory analysis of the factors that influenced these treatment decisions may inform why such differences were reported. The rheumatologists' responses suggested that TNFi treatment decisions were made within a system characterised by three nested themes of influential factors, the relative strength of which appeared to inform treatment variation across the sample. The three themes comprised influences on treatment decisions: (i) from the wider context in which a hospital operated (*External Environment Influences*); (ii) from within the rheumatologists' hospitals (*Internal Hospital Influences*); and (iii) from the day-to-day factors closest to each rheumatologist (*Individual-level Influences*). Influences in the external environment affected those at the hospital level, and both had an impact on the rheumatologists' individual-level influences (illustrated in Figure 3.3).

The specific influences, identified across the rheumatologists' transcripts, were classified as subthemes within each theme (reported in Table 3.4). This section discusses each theme and subtheme, using direct quotations from the participants to support the results.

Figure 3.3. *The interaction of influences that affected rheumatologists' treatment decision-making.*

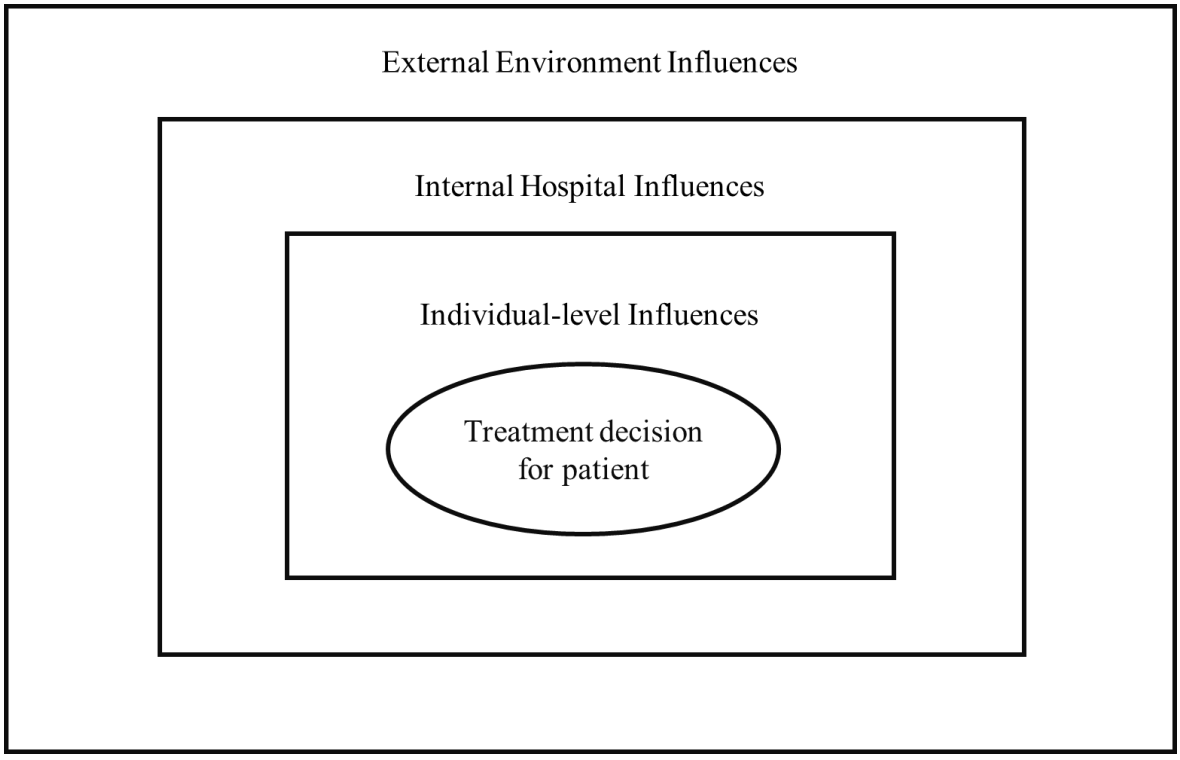


Table 3.4. *Specific factors that influenced treatment decisions categorised by broad theme.*

Theme	Subtheme (Specific Factor of Influence)
External Environment Influences	NICE recommendations; Clinical commissioning groups; Cost pressures; Published clinical evidence; Colleagues in Different Hospitals; Pharmaceutical companies;
Internal Hospital Influences	Systems to promote compliance with NICE recommendations; Internal treatment pathways; Hospital Culture;
Individual-level Influences	Patient influence; Consultant autonomy; Consultant experience; Perceptions of DAS28;

3.4.2.1. External Environment Influences

Six factors in the external environment that influenced treatment decisions (Table 3.4) were identified within the rheumatologists' transcripts; these are now described.

NICE Recommendations

The rheumatologists framed the discussions of TNFi treatment decisions around NICE recommendations, which were claimed to be used without question or forethought:

> Rheumatologist E: *"Obviously DAS28, obviously patients have to have had rheumatoid for over six months".*

NICE recommendations were perceived as suitable for most patients. However, recommendations were occasionally difficult to interpret, in particular, when a patient presented with RA that did not conform to the conventional NICE eligibility criteria:

> Rheumatologist J: *"…NICE guidance tends to provide a linear algorithm of the way to go and…if you don't end up on that linear algorithm, then…it's just not clear what's allowed".*

> Rheumatologist C: *"[NICE recommendations are] contradicting and confusing. They're not always good".*

The rheumatologists' interpretations of NICE recommendations were characterised by two extreme positions. One rheumatologist, for example, interpreted NICE recommendations as too restrictive:

> Rheumatologist I : *"[NICE guidance] is not open enough, in my view. It should be more open".*

A different rheumatologist, in contrast, thought that the same recommendations were flexible and open to interpretation:

> Rheumatologist H: *"…almost all NICE guidance is open to interpretation…Guidance is guidance. It's not…the law that has to be followed, otherwise you go to jail or something".*

Clinical Commissioning Groups

Regional health care services are commissioned by clinical commissioning groups (CCGs) in England (Cylus et al., 2015) and the rheumatologists perceived their CCG as an enforcer

of NICE recommendations. The relationship that each rheumatologist had with their CCG varied across the sample. Some explained how a good relationship, facilitated by following NICE recommendations closely, improved their ability to persuade their CCG to approve a treatment outside of NICE recommendations. In contrast, other rheumatologists argued that their CCG (typically those in a worse financial position) imposed the choice of first TNFi according to NICE recommendations:

> Rheumatologist K: *"Our CCG is a bit strapped for cash so we are not allowed to deviate one iota [from NICE recommendations]".*

The rheumatologists within the sample expressed displeasure towards treatment decisions imposed by their CCG, principally because of the necessary restriction on clinical autonomy:

> Rheumatologist G: *"I should say, it's not a particularly popular decision with the clinicians…because we want to have free choice of biologics".*

The conflict between clinical autonomy and CCG-imposed treatment decisions may have arisen due to differences in the interpretations of NICE recommendations (literal compared with advisory):

> Rheumatologist J: *"…the payers often use guidance as…somewhat less flexible than it's intended to be".*

Cost Pressures

The influence of cost on treatment decisions was a contentious issue across the sample, best summarised by one rheumatologist, who claimed:

> Rheumatologist K: *"We've debated [the influence of cost] fairly aggressively within our department".*

The pressure to prescribe TNFi therapies according to their cost, in alignment with NICE recommendations, appeared greatest in the hospitals whose CCGs were experiencing financial difficulties. Most rheumatologists, however, suggested that cost had a limited influence on their treatment decisions, despite NICE recommending to use of the lowest-cost TNFi in routine practice:

Rheumatologist A: *"I think there's lip service paid to total acquisition cost"*.

Rheumatologist I: *"Personally speaking, I try and ignore [cost in treatment decisions]"*.

However, a sense of duty was expressed by other rheumatologists when considering the sustainability of high-cost treatments in the NHS:

Rheumatologist B: *"It's a hellish expensive total when you start adding up what we're spending on biologics as a department, and are we getting our value for money out of it…or is it just…an ever-expanding expense mushrooming"*.

Rheumatologist F: *"I think we should be obliged, as clinicians, to consider costs with every treatment decision we make…Every anti-TNF drug we start takes money out of the health service that could be used for other purposes or other patients"*.

The rheumatologists expressed that it was difficult to make cost-savings and certain treatment approaches (such as the escalation of TNFi doses following loss of response) were rarely considered due to their relative expense. The implications of failing to save costs also appeared to influence the treatment decisions of some rheumatologists in the sample. For example, one participant preferred to use a cheaper biosimilar therapy than reduce the number of nurses in the hospital:

Rheumatologist K: *"We prefer to use a…biosimilar than sack nurses, quite honestly"*.

Two strategies to mitigate high treatment costs included regional price negotiations of TNFi therapies and engagement in research studies to receive an experimental treatment free of charge.

Published Clinical Evidence

There was some consensus between the rheumatologists that developments in the clinical literature had superseded the treatment recommendations of NICE. A developing clinical evidence base appeared to encourage prescribing decisions outside of NICE recommendations, facilitated by individual funding requests on a per-patient basis or CCG-approved changes to the local care pathway:

Rheumatologist E: *"[The aim of the local treatment algorithm is to]…keep up to date with what's going on, which is sometimes a little bit ahead of what NICE is actually saying"*.

A *lack* of clinical evidence was also described by some rheumatologists as a factor that influenced treatment decisions, for example, by having national drives to generate evidence on less-utilised therapies (such as the newer TNFi therapies). There was no consensus, however, on treatment decisions made outside of NICE recommendations; such decisions were often guided by clinical intuition instead when the published clinical evidence was uncertain:

Rheumatologist I: *"None of this stuff [treatment decisions outside of NICE recommendations] is really very well decided or agreed, and it comes down to your clinical…feeling, really"*.

Colleagues in Different Hospitals

The rheumatologists expressed that differences in treatment practices were likely between hospitals across England and most stated an interest in understanding how treatment decisions were made in other rheumatology units. However, the extent to which their own practice was influenced by that of others varied. Some rheumatologists valued the treatment experiences shared at national professional meetings, whereas others assigned a greater value to their own personal experiences of patient management:

Rheumatologist I: *"…what I find most helpful is to go to sessions to hear people talk about their clinical experience. Particularly people who are using drugs in different ways"*.

Rheumatologist D: *"…all that really matters to me is…knowing that what I do works for my patients"*.

An awareness of the approaches to treatment decisions in other rheumatology units also appeared to facilitate informal comparisons in best-practice between the rheumatologists. For example, some rheumatologists argued that their use of early, intensive combination cDMARD therapy or subcutaneous methotrexate led to relatively fewer patients receiving (and subsequently reduced total expenditure on) bDMARD therapies:

> Rheumatologist A: *"…we could actually bring drug spend up significantly by doing what many people do, which is…don't use combinations, use small-dose methotrexate, don't use high-dose subcutaneous methotrexate…then put loads of people on biologics"*.

Pharmaceutical Companies

The rheumatologists perceived the claims of pharmaceutical companies with scepticism and argued that they did not influence their treatment decisions. However, some concern was expressed that the pharmaceutical industry, in general, may have exerted an influence to promote treatment towards bDMARD therapies. The rheumatologists speculated that *rheumatology nurse specialists* involved in patient care, including the provision of information regarding bDMARD therapies (Palmer et al., 2010), may be vulnerable to the marketing messages of pharmaceutical companies, particularly when meeting with promotional representatives:

> Rheumatologist E: *"…one of the things I'm always a little bit concerned about is…drug reps speaking to us, speaking to nurses…Because so often…you ask [the patient] to see the specialist nurse who can speak about anti-TNF treatment with them…That decision making can be influenced"*.

It was suggested that one benefit of prescriptive treatment recommendations was to minimise the influence of pharmaceutical manufacturers on routine treatment decisions.

3.4.2.2. Internal Hospital Influences

There were three factors at the hospital-level (Table 3.4) that were reported to influence treatment decisions; these are now described.

Systems to Promote Compliance with NICE Recommendations

The rheumatologists reported the existence of internal systems to promote compliance with NICE recommendations, often implemented by pressure from their CCG:

> Rheumatologist F: *"Our CCG have imposed that on us…they actually set an ambitious target of ninety-five percent adherence to NICE"*.

Rheumatologist I: *"We have to [comply with NICE recommendations] we're very governed here"*.

Internal audits of practice were cited as a means to ensure that treatment decisions were made within the boundaries of NICE recommendations. However, the frequency of auditing was variable across the sample:

Rheumatologist I: *"[The CCG] want certain bits of our data every three months"*.

Rheumatologist H: *"…every year, there is a full report that comes out"*.

Rheumatologist C: *"…we audit each drug…every now and again"*.

The approaches to auditing were characterised by two extreme positions. One position demonstrated a proactive investment in a computerised system to monitor indicators of treatment response:

Rheumatologist E: *"We've got a [computerised monitoring system] which…we've paid for, which is a database of all our patients with, not just rheumatoid, but all inflammatory arthritis"*.

The second position contrasted by reporting that there was no formal approach to auditing:

Rheumatologist D: *"No [we don't have a system to ensure adherence to NICE]. I think we deliberately blur the margins"*.

Computerised prescribing systems were perceived by some rheumatologists as a means to enforce NICE recommendations, whereas others considered such systems to be fallible:

Rheumatologist K: *"…we've got the discipline of the computerised prescribing system"*.

Rheumatologist A: *"[The computerised prescribing system] logs if they've not responded to things, and you can…duck and dive a bit there"*.

Internal Treatment Pathways

Most rheumatologists explained how their treatment decisions were, or were soon to be, guided by an internal treatment pathway based on NICE recommendations. Internal recommendations often incorporated deviations from national recommendations by NICE:

> Rheumatologist E: *"…what we do now, in consideration, is use our [local] guidelines rather than NICE guidance"*.

> Rheumatologist I: *"…we have our own pathway which…is basically the NICE pathway but there's one or two minor exceptions"*.

Hospital Culture

The general approach to treatment was reported to have become more aggressive in recent years. However, some participants positively referred to treatment aggression in terms of rapid escalation to bDMARDs, whereas others referred to treatment aggression in terms of early arthritis clinics to delay bDMARD therapy:

> Rheumatologist B: *"I think with the kind of evidence that comes though about early treatment…we're probably more rapidly putting more patients on [bDMARDs]"*.

> Rheumatologist K*: "…we've got nearly 180 people on [combination cDMARDs], so probably more than the average unit, and we think that's partly responsible for us having a relatively low biologics use"*.

Divergent views were presented on the influence that rheumatology nurse specialists had on treatment decisions, from being an additional enforcer of NICE recommendations to taking a passive role in treatment decisions:

> Rheumatologist F: *"…if our biologics nurse has received referrals with as DAS below 5.1, for example…they'd just bat that straight back to…the lead consultant for that patient"*.

> Rheumatologist B: *"Patients become very laissez-faire about being on their biologics…and I think the nurses [are the same]. As we have more and more*

experience, [the nurses say] 'ah, they've very safe' and they can just…put everyone on them".

The rheumatologists identified that their ability to treat patients aggressively may have been restricted by capacity limitations:

> Rheumatologist K: *"We've got four rheumatologists, we could probably do with a fifth. There's been a capacity issue which has meant that…the behaviour of the unit has been a bit suboptimal"*.

3.4.2.3. Individual-level Treatment Influences

There were four factors that were reported to influence routine treatment decisions at the individual-level of the rheumatologist (Table 3.4); these are now described.

Patient Influence

The extent to which patients influenced treatment decisions varied between rheumatologists' responses, reflected by the set of TNFi therapies offered to patients:

> Rheumatologist C: *"If everything else is fine…we're happy to go with…whatever [TNFi] the patient wants"*.

> Rheumatologist F: *"…we don't give them options of five agents…you don't want to bewilder patients"*.

The rheumatologists expressed some concern that patients may have surreptitiously modified their treatment regime without later making that known during a consultation:

> Rheumatologist A: *"…the other problem is that when they're doing well on the biologics, they [the patients] wind down their other treatments…the methotrexate et cetera"*.

> Rheumatologist B: *"Often they [the patients]…self-regulate [their TNFi injections] with minor infections. So if they get a cold, they will stop them…It's rare they'll take, you know, fifty-two etanercept injections in a year"*.

Some rheumatologists were sceptical over patients' ability to make an informed treatment decision:

> Rheumatologist D: *"…you know what it's like with patients. Even if somebody changes the colour of their paracetamol, they're convinced it isn't working as well"*.

> Rheumatologist I: *"…what we've found is when we sort of said…'do you want this one or that [TNFi]?', [the patient]…said, 'well, what do you recommend?'…I kinda' think you've got to make a clinical decision, really"*.

The ability for patients to directly influence treatment decisions was considered to be sacrificed when treatment decisions were imposed by a CCG:

> Rheumatologist K: *"…we've compromised patient choice in the interests of the health economy"*.

It was believed that patients acquired information about treatments primarily from rheumatology nurse specialists. In-house, charity, and pharmaceutical manufacturer information leaflets also provided patients with information. The rheumatologists suggested that their patients had typically expressed a preference over the TNFi injection frequency and mode of administration; however, they also recognised that preferences over specific treatments were variable between different patients.

> Rheumatologist H: *"…it's pretty obvious that they [patients] think [in] their own way…we cannot really figure out how they think"*.

Consultant Autonomy

The ability to exert clinical autonomy over treatment decisions was valued by the rheumatologists in the sample. Some rheumatologists indicated that they made treatment decisions strategically to maintain a wider set of recommended bDMARD therapies later in the care pathway, potentially highlighting a concern that clinical autonomy appeared to be compromised:

> Rheumatologist I*: "…my view would be…to have every agent available first-line and then second-line…and third-line and, being really greedy, then have a fourth-*

line option which currently we don't have…I think that's partly why we don't move away from anti-TNF as our first-line, because it gives us more options [later] in the pathway".

The desire for clinical autonomy was expressed in the measures taken when attempting to approve treatments outside of NICE recommendations:

> Rheumatologist A: *"…we've been around various loops [to get a treatment outside of NICE approved], we've written business cases and [been] told that we must submit to different bodies…There's the national overarching body – turned out they were too busy, then there's a regional overarching body – turned out they were too busy, and now we're back to doing individual business cases, and that can take a long time, and that's foolish"*.

There was variability across the sample to facilitate clinical autonomy by using individual funding requests (IFRs) as a mechanism to obtain approval from the CCG for prescribing treatments outside of NICE recommendations:

> Rheumatologist H: *"Quite a lot [of IFRs have been undertaken]…in our area, we have actually never had a problem…I cannot recollect a single occasion that we were refused funding"*.

> Rheumatologist D: *"…it takes an absolute year of paperwork to get them [IFRs], because you've got to go through hundreds of different committees…so I'm very glad I haven't needed to"*.

Further strategic behaviour was demonstrated by using previously successful IFRs as templates for future IFRs. Alternatively, some rheumatologists argued that changing the hospital's internal treatment recommendations, rather than using IFRs, was a more effective strategy to achieve clinical autonomy:

> Rheumatologist F: *"…these things come up over and over again, so you've got a kind of 'Situation X IFR' that you can use, cut and paste"*.

> Rheumatologist I: *"…we never get anywhere with IFRs…In the end, after a lot of wrangling, got [a treatment outside of NICE recommendations] though as*

a…change to the pathway…IFRs…for rheumatoid just don't wash 'cos it's not an individual fight".

Consultant Experience

The rheumatologists' previous experience of TNFi therapies had an influence on their decision making. In particular, the older TNFi therapies were viewed positively due to the rheumatologists having more experience with using them:

> Rheumatologist J: *"…we tend to just use what we're familiar with, and the ones [TNFi therapies] that have been around the longest"*.

Treatment decisions were occasionally referred to as habitual, based on how previous patients had been treated:

> Rheumatologist K: *"…some of the nurses have been [treating patients] for…the best part of twenty years…so they're set in their ways"*.

Negative treatment experiences, such as an increase in infections, were found to dissuade the rheumatologists from subsequently using that treatment in the future:

> Rheumatologist A: *"We don't use a lot of leflunomide in combination with biologics because…we've seen more infections. But that's only in small numbers, but it does influence you"*.

Perception of DAS28

The use of the DAS28 assessment of disease activity to determine eligibility for TNFi therapy, in alignment with NICE recommendations, was generally perceived negatively across the sample:

> Rheumatologist I: *"…[the DAS28] is probably as good as we've got [to measure disease activity]…albeit it's not fantastic"*.

The rheumatologists expressed that the DAS28 may be unsuitable for patients with RA predominately in their ankles or feet, and that it may overestimate or underestimate disease activity:

Rheumatologist E: *"…we all recognise that some patients actually score highly on DAS28 because they've got a lot of tender joints, for example, and yet we know that there's…actually not…all active disease"*.

In particular, rheumatologists reported experiencing cases where low inflammatory markers or low patient self-reporting on the visual analogue scale (VAS) led the DAS28 to underestimate disease activity:

Rheumatologist F: *"…there are a few patients who we will put for anti-TNF even if their DAS is below 5.1 if…they're 'copers' or 'habituators'…where we…from our…expert judgement feel that their disease is far more active than perhaps their VAS would indicate"*.

Rheumatologist A: *"…then there's some people who don't put up their inflammation tests, their ESR/CRP, and that is quite a big part of the composite [DAS28] score. So there are a lot of people who you think, 'if only their inflammation test went up', and we could [giver them TNFi therapy]"*.

Uncertainty in the validity of a DAS28 score prompted the rheumatologists to exercise their clinical judgement over disease activity:

Rheumatologist F: *"…I think that if you just go on the DAS…you lose a lot of that information [about a patient's disease], and it's kind of a…drone's approach to medicine…I think you've got to rationalise and justify…every DAS score really…to say what it actually means…in the context of that patient"*.

A number of rheumatologists argued that the NICE-recommended DAS28 threshold for TNFi eligibility was too high, and consequently patients with less active disease suffered from the inability to access TNFi therapies:

Rheumatologist H: *"The problem is with those people who have DAS28s of 4, 4.5, 4.2…continuously…We know they've got a considerable amount of disease activity, but we cannot actually give them [a TNFi]"*.

Rheumatologist J: *"…we have got…quite a lot of patients [with a DAS28<5.1] …They smoulder in modest, moderate disease activity…probably slowly damaging their joints"*.

Some rheumatologists advocated a range of methods to strategically game the DAS28 assessment, to enable more patients to receive a TNFi therapy (reported in Table 3.5). One rheumatologist expressed concern that inaccurate DAS28 assessments may bias subsequent empirical analysis that used the data collected by large national patient registers:

Rheumatologist A: *"…I think most people lie actually [about DAS28 scores]…most people make it up…the problem for…the [biologics] registry is that people make up the numbers…to keep the CCG happy…but then give those spurious numbers to the registry"*.

Table 3.5. *Methods to game the DAS28 assessment reported by the sample of rheumatologist.*

Rheumatologist	Method of Gaming DAS28 Assessment
A	Measure disease activity using a different instrument (such as RAPID3) and map to DAS28.
B	Claim that the patient has psoriatic arthritis because fewer active joints are required to receive TNFi therapy, relative to RA.
D & K	Only perform one DAS28 assessment.
H	Stop a patient's steroids to increase their DAS28.
I	Perform DAS28 when patient is having a flare.
I	Increase frequency of DAS28 assessment to increase the likelihood of obtaining two DAS28 scores below 5.1.

3.4.3. Objective 3: Potential Barriers to ADAb and Drug Level Testing

The rheumatologists were invited to discuss ADAb and drug level testing of TNFi therapies in routine practice for RA. All participants demonstrated an awareness that health technologies were available for measuring TNFi ADAb and drug levels and some participants reported being in contact with commercial test manufacturers:

Rheumatologist A: *"People are trying to sell us a little kit to check drug levels and antibody levels"*.

The majority of the sample, however, apart from one rheumatologist, did not consider testing for TNFi ADAb or drug levels in their current routine management of patients with RA.

> Rheumatologist D: *"I know the tools exist and we have toyed with using them, but we haven't"*.

> Rheumatologist I: *"To be honest, we don't…routinely measure any antibodies here. So, I mean, I know it's a kind of interesting area…but it doesn't really alter our clinical practice here as of yet"*.

> Rheumatologist C: *"…we have already worked with our immunologist who have got some antibody assays to use…our plan is to change our guidelines at some point in the near future. So, we'll include screening for antibodies…"*.

These responses provided insight into the potential barriers perceived by rheumatologists regarding the introduction of TNFi ADAb and drug level testing into routine practice. Four potential barriers were identified across the sample.

Barrier 1: Recognition of a Clinical Problem

Despite a growing academic literature documenting immunogenicity against TNFi therapies (Radstake et al., 2009; Krieckaert et al., 2012), the participants discussed that it was generally not a recognisable problem in their own clinical practice.

> Rheumatologist E: *"Put it this way, I know it's described [immunogenicity against TNFi therapies], but we don't see it particularly"*.

> Rheumatologist B: *"…we've not really seen a lot of problems with immunogenicity"*.

Barrier 2: Understanding of the Purpose of Testing in Routine Practice

An understanding of the potential purpose of TNFi ADAb and drug level testing, and its role to inform subsequent treatment decisions, was variable between the rheumatologists in the sample. Some participants were optimistic:

> Rheumatologist J: *"I think if we were having this conversation in five years' time, we'd probably be looking routinely at that [TNFi immunogenicity assessment]…One algorithm would be all patients have drug levels checked automatically at, maybe two or three times a year…whether or not they're doing well…I think drug levels are gonna' allow us to reduce doses of drugs as well, which is the other reason to use them"*.

Some more sceptical:

> Rheumatologist H: *"…we've known about immunogenicity for twenty years. So why hasn't it, you know, taken off?"*.

Some were uncertain:

> Rheumatologist A: *"It [TNFi immunogenicity assessment] appeals to us. We don't know why it appeals to us. We think it's just, you know, a shiny little gizmo…it's interesting"*.

Barrier 3: Evidence Supporting TNFi Immunogenicity Testing

The rheumatologists explained how a lack of clinical evidence supporting TNFi ADAb and drug level measurement may dissuade them from testing in routine practice.

> Rheumatologist A: *"I think those…things like drug antibody kits and drug levels should be evaluated in proper controlled studies, really…The danger is that the market will be flooded with kits and everybody will think, 'that's great, I'll have a go', and…no one will know, in the end, what's actually happening"*.

> Rheumatologist H: *"I think that if…we have robust, you know, reliable methods…to measure antibodies and relate them to clinical response, then possibly they could go in the therapeutic algorithm. But I really think that we are a long way away"*.

Barrier 4: Capacity and Resource Constraints

The introduction of ADAb and drug level testing into routine practice will likely require additional resources (see Section 1.3.5). The rheumatologists, however, explained that they

may not have the budget to introduce testing or that the availability of laboratories to analyse samples may be inadequate.

> Rheumatologist D: *"…It's [measuring ADAb] a bit of a faff, and again it's more time, more money, more thought around it…I'm still not convinced that for the majority of patients it really changes management".*

> Rheumatologist B: *"I don't think we've got the capability [to measure ADAb]. There's only a couple of places in the country that'll do it, so it's not something that we are getting too concerned about".*

3.5. Discussion

This study explored the decisions for treating patients with RA in current practice by performing in-depth semi-structured telephone interviews with eleven senior consultant rheumatologists in England. Regional variation, consistent with published national audits of practice, was observed at key decision points along the care pathway for RA. The factors that appeared to influence treatment decisions were categorised according to three themes (external environment influences, internal hospital influences, and individual-level influences). The rheumatologists also reported four potential barriers to introducing ADAb and drug level testing in routine practice.

NICE recommendations state that when choosing a TNFi therapy to prescribe, for most patients with RA who meet the eligibility criteria, the lowest cost therapy should be selected (National Institute for Health and Care Excellence, 2016a) (see Section 1.3.3). Given the growing evidence that the approach to using bDMARDs, as recommended by NICE, may impose a substantial opportunity cost on population health (van der Velde et al., 2011; Joensuu et al., 2015; National Institute for Health and Care Excellence, 2016a), any systematic deviations from these recommendations are unlikely to be a cost-effective use of NHS resources. This study, however, identified that factors other than cost influenced the choice of TNFi in routine practice and some rheumatologists ignored costs entirely. By failing to acknowledge cost in treatment decisions, which ultimately reduces the health care resources available for patients elsewhere in the NHS, an implicit judgement was made that the health of *identifiable* patients with RA was valued more than

the health of *unidentifiable* patients elsewhere in the NHS (Claxton et al., 2015b; Cookson, 2015).

Contention was observed between the rheumatologists' interpretations of clinical recommendations (from advisory to mandatory) and the methods to ensure compliance with NICE recommendations were inconsistent and perceived as fallible. Sheldon et al. (2004) evaluated the implementation of eleven examples of NICE guidance using routinely collected data, patient case notes, and qualitative interviews with the leads of clinical specialities, governance, and chief executives. Consistent with the findings of this study, the participants in Sheldon et al. (2004) also exhibited divergent beliefs regarding the role of guidance and in the extent to which regular audits of practice occurred.

The rheumatologists that agreed to deviate from NICE recommendations in principle did not agree on how to do so in practice. Moreover, the rheumatologists that agreed to follow NICE recommendations interpreted those recommendations differently; for example, whether an *aggressive* treatment regime meant rapid escalation to bDMARD therapy or an intensive use of combination cDMARDs. The previous quantitative observations of regional variation in current practice for RA (Tugnet et al., 2013; Blake et al., 2014; The British Society for Rheumatology, 2015) may therefore be due to: (i) divergent interpretations of clinical recommendations, (ii) successful attempts to consciously treat outside of NICE recommendations, or (iii) a combination of the two.

The specific decisions that were made by the rheumatologists along the care pathway for RA (Section 3.4.1, *Appendix 15*) were used to inform the design of the *de novo* decision analytic model in *Chapter Five*. For example, the rheumatologists described that following secondary non-response of a TNFi therapy, patients were likely to be prescribed rituximab and not a second TNFi. In addition, the rheumatologists described how patients that remained in remission whilst receiving a TNFi therapy may experience a reduction in the intensity of their treatment. The use of such expert clinical input, in turn, facilitated a characterisation of current practice, which was highlighted as a relevant objective during the early economic evaluation of a health technology (see Section 1.1.6.4, Figure 1.1).

The participants reported that making cost savings was challenging in practice as hospitals were already operating at capacity. The immediate financial pressure was therefore expected to fall on the hospitals' labour force (the employment of nurses and consultant rheumatologists). Since conducting data collection for this study, most NHS Trusts in

England have entered a financial deficit (Lafond et al., 2016). Therefore, more rheumatology units in England may subsequently experience CCG-imposed treatment regimes, as reported in this study, in the future. As a consequence, both clinical autonomy and patient involvement in decision making (however defined) may be further sacrificed across England.

Examples of previous studies that had used qualitative methods to explore the factors that influenced the prescription of TNFi therapies for patients with RA were sparse, and none of those studies had been performed in England. Kee et al. (2005) investigated consultants' beliefs of whether patients with RA should continue infliximab therapy in the Republic of Ireland. Some evidence was found to indicate that consultants were gaming the DAS28 assessment, in particular when patient-reported symptoms were more severe than suggested by the DAS28 score (Kee et al., 2005). Similarly to Kee et al. (2005), the sample of English rheumatologists in the present study also reported strategies to game the DAS28 assessment to improve patient eligibility for TNFi therapy and concern was expressed over whether these artificial values were entered into national patient registers. The presence of artificial DAS28 scores in patient registers may, in part, explain why previous studies have been unsuccessful at identifying robust predictive biomarkers of treatment response for patients with RA (a change in DAS28) (Emery et al., 2011; Gibson et al., 2012).

Kalkan et al. (2014) performed a qualitative analysis by exploring the influences of senior consultant rheumatologists' bDMARD prescribing decisions in Sweden. Despite differences between health care systems, comparable influential factors were also reported by the sample in the present study, such as the influence of a developing clinical evidence base, colleagues, departmental culture, and budget constraints (Kalkan et al., 2014).

Four potential barriers to using ADAb and drug level testing were discussed by the rheumatologists in this study. The exploration of potential barriers to using a new health technology can provide useful information to support an early economic evaluation because the views of clinicians may inform how such a health technology may, or may not, be used in routine clinical practice (Figure 1.1; Section 1.1.6.4). Raghavan et al. (2014) conducted a review of studies that investigated the potential barriers, perceived by physicians, to incorporating genetic testing in routine practice; the barriers identified by Raghavan et al. (2014) (for example, a lack of knowledge of genetic testing; a desire for more trial evidence; unclear benefit of testing; a challenge to include additional health technologies in a time and resource-constrained environment) overlapped with those

reported by the rheumatologists within this study. Similarly, Jones et al. (2013) conducted a systematic review of seven qualitative studies that explored clinicians' attitudes towards using point-of-care testing in primary care (such as a urine test that provided an immediate result); the potential barriers to testing identified by Jones et al. (2013) included the additional cost associated with testing and an uncertainty regarding the usefulness of testing. These findings, when considered collectively, may therefore suggest that the introduction of any new test (to stratify treatment) into clinical practice may experience a set of potential barriers that are common to all types of medical test.

Reflexive Statement

Qualitative research necessarily requires an interaction between the researcher and the phenomenon being researched (Patton, 2003). For example, the data generated during an interview must derive from a conversation between the participant and the researcher (Miller, 1995). Therefore, best-practice recommends the inclusion of a reflexive statement, that demonstrates self-awareness on behalf of the researcher regarding how their behaviour (interactions; prior beliefs; perceived status with the participants) may have influenced the results (Coast, 1999; Patton, 2003; Coast et al., 2004).

The primary researcher (SG) had never met any of the rheumatologists prior to conducting the telephone interviews and existed in a distinct professional discipline (health economics) compared with the participants (rheumatology). In the terminology of the qualitative literature, the researcher was therefore an *outsider* because they did not share the same prior experiences or professional role of the participants (Dwyer et al., 2009). Therefore, it was possible that the sample of rheumatologists did not address technical issues in detail because of a belief that the primary researcher would not understand. Alternatively, the participants were aware that the research was being conducted under conditions of anonymity at an academic institution for a doctoral thesis, and may have responded less-candidly if the primary researcher had been a member of the general public or a manufacturer of a commercial health technology instead. The responses of all rheumatologists were detailed and insightful, which may suggest that a good rapport was established during the interviews between SG and the participants. Ultimately, it is not necessary for a qualitative researcher to be an *insider* of the group being researched in order to have an understanding of the participants' perspectives and experiences (Dwyer et al., 2009).

Limitations

The relatively small sample size of this study could be claimed as a potential limitation. However, twenty-four percent of the sampling frame were recruited as participants, which was equivalent to the proportion of rheumatologists recruited in the similar qualitative study by Kalkan et al. (2014) (twenty-five percent of their sampling frame). Data were not collected until saturation due to the purposive nature of the sample. However, the sample was sufficient to reveal variation in previously undocumented factors that influenced rheumatologists' treatment decisions in England. Issues of sample size are most relevant when the purpose of research is to infer generalisations to a wider population, based on a single *reality* that is assumed to be measurable (consistent with a positive, empirical ontology). The ontology of qualitative research, by contrast, presupposes that *reality* is, to some degree, a construct of the participants' perceptions (Norum, 2008). In this study, for example, participants provided different interpretations of the appropriate way to treat patients with RA; yet, all believed that their interpretation was correct. Multiple *realities* may have existed in the sample and the purpose of exploratory data analysis, therefore, was to understand these divergent perceptions further (Silverman et al., 2008). A larger sample size, consequently, is neither a necessary, nor sufficient, to obtain an *understanding* of a phenomenon being researched (Sandelowski, 1995).

Secondly, the use of telephone interviews as a means of data collection may have had its limitations, compared with face-to-face interviews conducted in person. For example, meaningful non-verbal communication (such as head nods and hand gestures) were not observable over the telephone and pauses in the conversation (to provide the participant with time to think) may have been misinterpreted as the conclusion of a sentence (Miller, 1995). The rheumatologists' responses were therefore interpreted at face-value. However, the data collected for this study were comprehensive; it was therefore unlikely that the interpretation of the results would have changed substantially if these data had been collected by face-to-face interviews instead.

A third potential limitation of this study was that a rheumatologist's account of the factors that influenced treatment decisions was just one perspective that could have been explored. The perspective of other stakeholders in the decision-making process (for example, the CCG, nurses, and patients) may have revealed different influences on routine treatment decisions for RA.

Implications for Future Research

Given the competing influences on treatment decisions identified by this study, a future qualitative study may benefit to explore the decision-making process at the CCG and, in particular, the evidence required at a local level to recommend a new health technology in routine practice. Health care commissioners, for example, may require an explicit business case to be made (Lourenco et al., 2011) for a new stratified medicine. In addition, a future qualitative study could interview a sample of rheumatology nurse specialists, given the divergent perceptions of their influence reported in this study, to explore their own perceived influence on treatment decisions within the care pathway for RA.

The availability of patient-level data may facilitate a quantitative analysis of the factors that influenced treatment decisions observed in routine practice. A quantitative analysis of TNFi choice, in particular, may provide evidence of an implicit treatment stratification mechanism, given that cost was not the only factor reported to influence the rheumatologists' prescribing decisions in this study. The results of this qualitative study may suggest that such a patient-level analysis should attempt to control for influential factors at the hospital-level and in the external environment (which may be unobserved). *Chapter Four* therefore develops the results of this study to estimate the patient-level factors associated with the choice of TNFi prescribed to patients with RA in England, by using the data collected by the BRAGGSS cohort.

Finally, the need for an economic evaluation of adalimumab ADAb and drug level testing was justified by the rheumatologists in this study on three grounds:

(i) Inconsistent treatment decisions were recommended across the sample for patients who experienced TNFi-induced remission or secondary non-response of a TNFi, both of which may be informed by testing (Vincent et al., 2013). The participants also reported that TNFi dose-reduction strategies already occurred in routine practice; consequently, these strategies should be included as relevant comparators in the conceptualisation and development of the *de novo* decision analytic model in this thesis (reported in *Chapter Five*);

(ii) Commercial ADAb and drug level test manufacturers had reportedly attempted market access with the rheumatologists in the sample; however, no evidence for

the relative cost-effectiveness of testing in England was available;

(iii) The rheumatologists' requested additional primary research regarding the use of ADAb and drug level testing in routine clinical practice. Therefore a VOI analysis would provide useful evidence of whether such further research would be valuable to the NHS (Claxton et al., 2001; Wilson, 2015).

3.6. Conclusion

This was the first study to explore the factors that influenced routine treatment decisions of TNFi therapies for patients with RA in England. This research was timely given the documented regional variation in RA treatment decisions and the questionable evidence supporting the relative cost-effectiveness of TNFi therapies for RA in England.

This study provided relevant evidence for the subsequent cost-effectiveness analysis of adalimumab ADAb and drug level testing in RA (reported in *Chapter Six*). The rheumatologists that were interviewed demonstrated a clinical need for the testing strategies by describing differences in treatment decisions following secondary failure of a TNFi and during TNFi-induced remission. The treatment decisions, reported by the rheumatologists, along the care pathway for RA were subsequently used to conceptualise the structure of a *de novo* decision analytic model in *Chapter Five*. The specific questions raised by the rheumatologists regarding how to use ADAb and drug level testing in practice, and whether to conduct further research on the tests, are addressed in *Chapter Five* and *Chapter Six* of the thesis, respectively.

The eleven rheumatologists that participated in this study revealed that factors other than those recommended by NICE may influence routine treatment decisions. A quantitative analysis of observed treatment decisions using a nationally representative sample patient-level data could be valuable to estimate whether implicit treatment stratification had occurred in current practice. Building on the results of this chapter, *Chapter Four* investigates current practice further by quantifying the patient-level factors that influenced the choice of TNFi therapy prescribed to patients with RA across England.

Chapter 4

Estimating the Patient-level Factors that Influence the Choice of TNFi Prescribed for Rheumatoid Arthritis in Current Practice

Chapter Four presents a quantitative econometric analysis of prescribing decisions for patients with RA in England. This study built on the results of *Chapter Three*, to develop an understanding of current practice further, by estimating the patient-level factors that may influence TNFi prescribing decisions and, subsequently, may indicate the presence of implicit treatment stratification for patients with RA. An earlier version of this study was presented at the *International Society for Pharmacoeconomics and Outcomes Research 21st Annual International Meeting* in May 2016 (Gavan et al., 2016b). The chapter is structured according to the following subsections: an introduction (Section 4.1), aim and objectives (Section 4.2), method, (Section 4.3), results (Section 4.4), discussion (Section 4.5), and conclusion (Section 4.6).

4.1. Introduction

The systematic review in *Chapter Two* identified that previous model-based economic evaluations of stratified medicine in RA had not described "current practice" with sufficient clarity. The qualitative analysis in *Chapter Three*, therefore, explored treatment decisions along the care pathway for RA and identified that factors may influence a

rheumatologist's routine treatment decisions within the external environment, at the hospital-level, and at the individual-level. Three potential limitations of this qualitative analysis, however, were that: (i) the self-reported beliefs of individual rheumatologists, in the context of a telephone interview, may not align with their actual prescribing decisions in practice; (ii) it was not possible to identify each factor's relative magnitude of influence; and (iii) the results may not generalise to other rheumatologists outside of the sample. The use of observational patient-level data, derived from routine treatment decisions, has been recommended as an additional source of evidence to understand care pathways in current practice (Tappenden, 2014) and may counteract the potential limitations of the qualitative study. Therefore, this chapter presents a quantitative analysis of current practice for RA by using patient-level data from actual treatment decisions observed in England.

The economic rationale for stratified medicine is that population health outcomes may be maximised by allocating health care resources to an *explicit* biomarker testing strategy, that can identify subgroups of patients for whom the relative cost-effectiveness of a treatment may be improved (see Section 1.2.1) (Coyle et al., 2003; Sculpher, 2008; Espinoza et al., 2014a; Espinoza et al., 2014b). Treatment decisions in current practice may be *implicitly* stratified if they are made according to systematic differences in observable patient characteristics (FitzGerald et al., 2017). The choice of TNFi prescribed to patients with RA in England, according to NICE recommendations, should, on average, be determined by its cost to the NHS (National Institute for Health and Care Excellence, 2016a). However, the rheumatologists in *Chapter Three* explained that cost had a limited influence on their routine decision-making (see Section 3.4.2.1). A relevant topic for further research was to investigate whether patient-level factors were being used to implicitly stratify the TNFi therapy prescribed to patients with RA within current practice in England.

Three previous studies from North America have estimated the patient-level factors that influenced TNFi prescribing decisions by using data from patients with RA enrolled to various health care plans (DeWitt et al., 2006; Carter et al., 2012; Zhang et al., 2013). *Appendix 16* describes the features of these three studies and the potential limitations of their analytic design. No previous study had estimated the patient-level factors associated with the choice of TNFi, by using data from actual treatment decisions, for patients with RA in England.

4.2. Aim and Objectives

The aim of this study was to identify the patient-level characteristics that influenced the choice of TNFi prescribed to patients with RA in England. There were two objectives to meet this aim:

Objective 1: Identify a nationally representative sample of patients with RA in England that included observations of actual TNFi prescribing decisions;

Objective 2: Test the hypothesis that no patient-level characteristics systematically influenced the choice of TNFi, after controlling for unobservable influences on prescribing decisions in the external environment and at the hospital-level.

4.3. Method

This study was a pooled cross-sectional analysis of prescribing decisions observed in routine clinical practice in England. The study was reported according to the standards outlined by the *Strengthening the Reporting of Observational Studies in Epidemiology* (STROBE) checklist; the completed *STROBE* checklist is reported in *Appendix 17*. The method of this study is reported in three subsections: Section 4.3.1 describes the elements of economic theory that informed the analysis; Section 4.3.2 describes the dataset; and Section 4.3.3 reports the method of analysis.

4.3.1. Supporting Economic Theory

Statistical analyses informed by economic theory are referred to as *econometric* analyses (Tintner, 1953). There were three relevant elements of economic theory that informed this study: (i) *the determinants of demand for health care*; (ii) *the agency relationship*; and (iii) *inequities in health care utilisation*.

4.3.1.1. The Determinants of Demand for Health Care

The estimation of factors that influence the choice of a specific good or service is an example of a study that estimates the *determinants of demand* (Morris et al., 2012). Health care, in general, has a set of unique characteristics when compared with other types of

tradable commodity (Arrow, 1963). For example, the demand for health care is derived from the demand for health itself (Grossman, 1972). In addition, non-health factors such as a patient's socioeconomic characteristics and the cost of treatment are known to influence health care demand (Zuvekas, 2014). The cost of TNFi therapies in England, however, are determined at the regional-level via unobservable price negotiations with treatment manufacturers, and are not borne by the patients themselves (Stokoe et al., 2011). Therefore, an econometric analysis of the factors that influence the choice of TNFi should account for patient-level health characteristics, the cost of treatment at the hospital-level, and patient-level non-health socioeconomic characteristics.

Endogeneity may occur in a regression-based analysis when an independent variable is correlated with the residual error term, principally due to (i) reverse-causality, (ii) measurement error, or (iii) omitted variable bias (Wooldridge, 2010). *Omitted variable bias*, in particular, occurs when an independent variable is excluded from a regression that is correlated with a different independent variable and/or the dependent variable (van der Gaag et al., 1991; Greene, 2012). A *confounding* variable shares a similar definition within the epidemiology literature; however, an omitted variable within the econometric literature often refers to an unmeasured confounding variable (for example, unobserved heterogeneity at the geographic-level) (Zohoori et al., 1997; Gunasekara et al., 2008). Based on the results of *Chapter Three*, failing to control for the influences on treatment decisions in the external environment and at the hospital-level in this study may have led to endogenous patient-level influences due to omitted variable bias.

4.3.1.2. The Agency Relationship in Health Care

The market for health care in England is characterised by a hierarchy of agency relationships, whereby one party (the agent) acts upon the objectives of another party (the principal) (Propper, 1995; Baxter et al., 2008). The agency relationship that characterises a treatment decision can be exemplified by the information asymmetry between an informed clinician (the agent) and a less-informed patient (the principal). An individual patient, given this agency relationship, was therefore unable to directly influence the demand for specific treatments (Williams, 1988; Mooney et al., 1993).

Perfect agency occurs when the principal acts as the agent would in the case of no information asymmetry (Morris et al., 2012). Strategies to overcome information asymmetry may occur through a process of shared decision-making by consulting the

patient in a treatment decision, as advocated by the treatment guidelines for RA produced by the international professional organisations for rheumatology, EULAR and the ACR (Smolen et al., 2014; Singh et al., 2016b). However, there may be barriers to perfect agency; for example, patients and clinicians may have competing objectives (Ryan, 1994) and clinicians (as agents) may influence both the supply and demand for care (Maynard, 1979; Ferguson, 1985). Wider organisational constraints are also likely to affect treatment decisions because the clinician may be regarded as a double agent, responsible for the patient and payer's competing objectives of quality health care delivery in a cost-effective manner (Blomqvist, 1991; Shortell et al., 1998). The implication for this study was that the choice of TNFi therapy may, to some extent, have been influenced by the preferences of patients; however, the influence of patients may be diminished by the wider objective of ensuring cost-effective treatment decisions. For example, the rheumatologists in *Chapter Three* described how patient choice was sacrificed if their CCG imposed the use of the cheapest TNFi.

4.3.1.3. Horizontal Inequity in Health Care Utilisation

The econometric literature regarding the estimation of inequities in health care utilisation is related closely to the *demand for health care* literature. *Inequalities* in health care occur when different patients receive different health care resources; *inequities* in health care utilisation occur when different patients do not receive the health care they need (Gravelle et al., 2006). *Horizontal inequity* arises when patients of equal need receive unequal treatment (Wagstaff et al., 1991).

The analysis of horizontal inequity necessarily requires a value judgement regarding the patient characteristics that define *need* and those that, by implication, define *non-need* (Morris et al., 2005; Gravelle et al., 2006; Fleurbaey et al., 2009; Kjellsson et al., 2015; Wagstaff, 2015). Studies that have estimated the demand for health care have predominantly represented *need* with patient-level variables for health status and *non-need* with variables representing socioeconomic characteristics (Vallejo-Torres et al., 2014). A regression analysis of the determinants of health care demand may therefore detect horizontal inequity if the coefficient of a *non-need* independent variable is estimated to be non-zero and statistically significant (Morris et al., 2005; Gravelle et al., 2006; Vallejo-Torres et al., 2014). In this study, it was assumed that evidence of horizontal inequity was present if TNFi treatment decisions were implicitly stratified by non-need patient characteristics.

4.3.2. Data

This study used patient-level data from the *Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate* (BRAGGSS), which was a nationally representative prospective cohort of patients with RA in England whose genetic and genomic data were collected alongside information on response to bDMARD therapies (Maxwell et al., 2008; Jani et al., 2015a). Patients within BRAGGSS were followed for twelve months and data were collected at four time points (baseline, three months, six months, and twelve months). Recruitment for BRAGGS occurred between 2008 and 2014 from fifty-seven NHS hospitals in England (hereafter referred to as *hospitals*). The full list of contributing hospitals is provided in *Appendix 18*.

The sample was restricted to the patient-level baseline observations from BRAGGSS which provided information on the treatment prescribed, and the *need* and *non-need* characteristics of each patient. All patients were observed only once (at baseline). The sample was also restricted to 2009 to 2014, to cover a period of time in which no changes were made to the recommendations for managing RA by NICE (National Institute for Health and Care Excellence, 2009).

The structure of the BRAGGSS cohort was such that, patient-level observations were clustered within a hospital, which were additionally clustered by year of inclusion into the cohort (illustrated in Figure 4.1). Using the terminology from the qualitative interviews in *Chapter Three*, the clustered design of the cohort could be exploited to control for heterogeneity in the unobservable environmental and hospital-level factors that influenced treatment decisions (Rice et al., 1997). For example, controlling for variation over time was assumed to account for the factors that influenced TNFi choice in the external environment (such as a developing clinical evidence base). Additionally, controlling for variation between hospitals was assumed to account for unobserved heterogeneity in hospital-level influences on TNFi choice (such as a local care pathway, systems to promote compliance with NICE, or the cost of each TNFi therapy).

4.3.2.1. Patient Inclusion Criteria

The criteria for inclusion to the study is reported in Table 4.1. Individuals were included in the sample if they were adults with RA, as classified by the ACR 1987 criteria (Arnett et al., 1988) (see *Appendix 7* for a description of the criteria). Patients must have been

prescribed any TNFi at baseline and must not have been exposed to a biologic agent previously. Non-Caucasian patients were excluded from the sample given their infrequency in the BRAGGSS cohort.

Figure 4.1. *Structure of BRAGGSS cohort.*

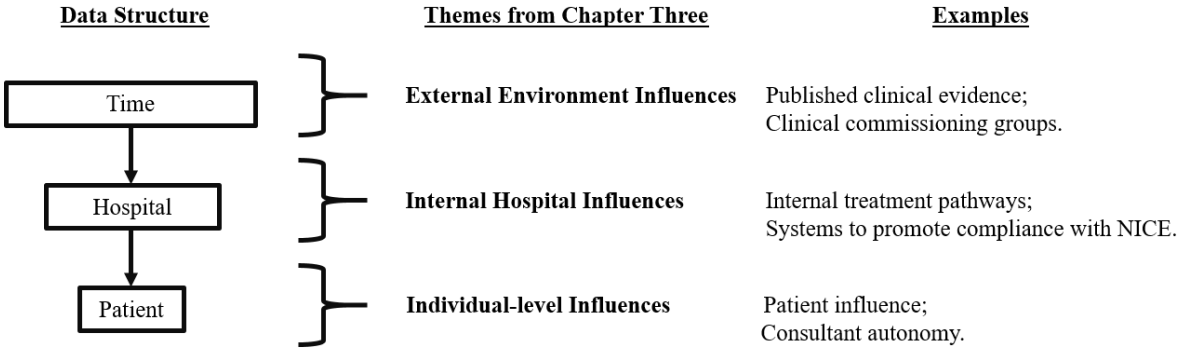


Table 4.1. *Patient inclusion criteria.*

Patient Characteristic	Inclusion Criteria
Disease.	RA (ACR 1987 Classification Criteria†).
Age.	Adult (>18 years).
Sex.	Any.
Ethnicity.	Caucasian.
Treatment.	Any TNFi, first-line.
Year of prescribing decision.	2009 to 2014.

Source: † Arnett et al. (1988).

4.3.2.2. Dependent Variable

The dependent variable (defined as $TNFi_{Prescribed}$) used in the regression analysis was an unordered categorical variable that reported the TNFi prescribed to each patient at baseline. The variable had three mutually exclusive categories, based on clinical plausibility, to indicate whether the patient had been prescribed (i) a non-monoclonal antibody (etanercept), (ii) an older monoclonal antibody (adalimumab or infliximab), or (iii) a newer monoclonal antibody (certolizumab pegol or golimumab). The ability to collapse the five TNFi therapies into three categories was verified by a likelihood ratio test (reported fully in *Appendix 19*).

4.3.2.3. Independent Variables

Table 4.2 describes the independent variables that were included in the analysis. The selection of independent variables was informed by the wider econometric literature regarding the determinants of demand for health care (Zuvekas, 2014), described in Section 4.3.1, conditional on the data availability within BRAGGSS. A distinction was

made between *health* and *non-health* variables, consistent with previous studies that had estimated the influence of patient-level factors on treatment decisions (Vallejo-Torres et al., 2014).

Table 4.2. *Description of independent variables.*

Variable	Description
Health Variables	
DAS28	DAS28 score. Calculated using C-reactive protein level.
Woman	Dummy variable, =1 if patient was a woman.
Age/10	Patient's age divided by 10.
HAQ	Patient's Health Assessment Questionnaire-Disability Index score.
Totaldrug	Total number of previous cDMARDs.
Totalcomorb	Total number of comorbidities.
YearswithRA	Years with RA.
BMIover	Dummy variable, =1 if patient had BMI $\geq$ 25.
MTX	Dummy variable, =1 if patient was receiving methotrexate.
Non-health Variables	
Alcohol	Dummy variable, =1 if patient self-reported alcohol consumption.
Smoke	Dummy variable, =1 if patient self-reported smoking daily.
Employment Status	
Work1	Dummy variable, =1 if patient was employed (full-time/part-time) or a student.
Work2	Dummy variable, =1 if patient was on sick/disability leave or retired early due to arthritis.
Work3	Dummy variable, =1 if patient was unemployed or retired for reasons unrelated to poor health.
Marital Status	
Marital1	Dummy variable, =1 if patient was married/living with partner.
Marital2	Dummy variable, =1 if patient was single.
Marital3	Dummy variable, =1 if patient was divorced/widowed.
Homecare	Dummy variable, =1 if patient had a carer at home.
Non-patient Variables	
Year	Time trend for each year of analysis (1=2009, 6=2014).
$Hospital_{Dummy}$	Dummy variable for each hospital, =1 if patient enrolled there.

Abbreviations: BMI=Body mass index.

Nine independent variables were included in the analysis that sought to encompass each patient's health status. Three clinical health variables, used to inform decision making in routine clinical practice for RA (*DAS28*, *HAQ*, *BMIover*), were included in BRAGGS and were measured at baseline by a health professional. For this study, *DAS28* was calculated using C-reactive protein (CRP), rather than using the erythrocyte sedimentation rate (ESR),

because more CRP samples were observed at baseline. *BMIover* was calculated by dividing the patient's weight (in kilograms) by their height (in metres, squared), and dichotomised to indicate an overweight body mass index (BMI) (BMI≥25) according to the *World Health Organization's* BMI classification score (World Health Organization, 2000). *Totalcomorb* was the sum of each patient's self-reported comorbidities, recorded at baseline, from the following list (high blood pressure, angina, heart attack, heart failure, stroke, epilepsy, asthma, chronic bronchitis, peptic ulcer, liver disease, renal disease, tuberculosis, demyelination, diabetes, hyperthyroidism, depression, cancer). Similarly, *Totaldrug* was the sum of unique cDMARDs used by each patient at baseline, from the following list (methotrexate, cyclophosphamide, auranofin, intramuscular gold, leflunomide, azathioprine, hydroxychloroquine, cyclosporine, penicillamine, sulphasalazine). Both *Totalcomorb*, *Totaldrug*, and the years that a patient reported to have had RA (*YearswithRA*) were included in the analysis as proxy variables for the severity of disease (the greater the quantity of comorbidities, previous cDMARDs used, or years since diagnosis, the more severe the patient's disease was likely to have been). NICE recommended that a subset of TNFi therapies may be prescribed as monotherapy for patients with contraindications to methotrexate (National Institute for Health and Care Excellence, 2016a); therefore, a dummy variable was included to indicate concomitant methotrexate use (*MTX*).

Five variables were available in BRAGGSS, self-reported by each patient, that encapsulated the potential *non-health* influences on the demand for TNFi choice. Two dummy variables were included that were assumed to reflect a patient's health behaviours (*Smoke*, *Alcohol*); both variables may have indicated a lack of patient involvement in treatment decision-making because (i) clinicians were assumed to have informed patients that smoking may increase the severity of RA and reduce treatment effectiveness (Westhoff et al., 2008), and (ii) the product label for methotrexate described that alcohol consumption was contraindicated and should have been avoided (British National Formulary, 2016). The presence of a carer at home (*Homecare*) was included in the analysis as a potential non-health influence on treatment selection because it may proxy for resources available to the patient at home, for example, to assist with, or record the frequency of, treatment administration.

A patient's employment status was classified according to their activity in the labour market and whether their self-reported exit from the labour market was due to their health condition; a patient may have been (i) employed and active in the labour market, full-time,

part-time, or as a student (*Work1*); (ii) inactive in the labour market due to health reasons, for example, by reporting sick leave, disability leave, or early retirement due to arthritis; or (iii) inactive in the labour market due to reasons unrelated to their health status, for example, by reporting unemployment or retirement due to non-health reasons (*Work3*). Marital status was categorised according to whether the patient was in a relationship (*Marital1*), single (*Marital2*), or had formerly been in a relationship (*Marital3*). The employment and marital status of patients are examples of socioeconomic characteristics, available in BRAGGSS, that have been found to influence the demand for health care in previous econometric studies (Zuvekas, 2014).

Two sets of non-health independent variables were included, by exploiting the clustered design of BRAGGSS (illustrated in Figure 4.1), to control for factors that may have influenced routine prescribing decisions above the level of a patient's characteristics. An annual time trend (*Year*) was included to control for unobservable temporal changes in external environmental influences, and a dummy variable for each hospital was included to control for unobservable heterogeneity in hospital-level influences ($Hospital_{Dummy}$).

4.3.2.4. Missing Data

The occurrence of variables with missing observations is a common problem in quantitative analyses of large datasets (Little et al., 1987; Rubin, 1987). There were some individuals with missing data within the BRAGGS dataset (described in *Appendix 20*). Simple solutions to handle missing data, such as (i) excluding observations with missing data (known as *a complete-case analysis*) or (ii) replacing the missing value with the sample mean, may have led to imprecise parameter estimates and biased inference (Sterne et al., 2009; Janssen et al., 2010; Rezvan et al., 2015). Multiple imputation was a more appropriate method to account for uncertainty in the true values of the missing data, by estimating a series of plausible values based on the observed data instead (White et al., 2011). Multiple imputation was therefore used to address the missing data in this study by using chained equations (Royston, 2009; Royston et al., 2011; White et al., 2011; Romaniuk et al., 2014). *Appendix 20* describes the multiple imputation method that was used in this study.

4.3.3. Analysis

The data analysis used a multinomial logistic (MNL) regression to estimate the factors that influenced the choice of TNFi prescribed for each patient. MNL regression was appropriate because the dependent variable was an unordered, categorical variable (Greene, 2012). The analysis pooled the cross-sections of BRAGGSS over time because the unit of observation (the prescribing decision of an individual patient) was observed only once (and not replicated over time, as would be required to analyse BRAGGSS as panel data) (Wooldridge, 2010). This subsection describes the following analysis methods: MNL regression (Section 4.3.3.1), addressing potential endogeneity (Section 4.3.3.2), model specification (Section 4.3.3.3), cluster-robust standard errors (Section 4.3.3.4) and measures of model fit (Section 4.3.3.5).

4.3.3.1. Multinomial Logistic Regression

MNL regression can be used to model a random variable (Y) that may take the value $\{0, \dots, J\}, J \geq 2$. The dependent variable in this study had three categories $(J = 3)$. A vector (V) of (k) independent variables can be defined (see Table 4.2) with a unit intercept term (Wooldridge, 2010).

The analysis of factors that influenced TNFi choice focused on how unit changes in the variables within (V) affected the probability of choosing an alternative in (Y). The probability that alternative (j) was chosen could be expressed as shown in Equation 4.1:

$$Prob(Y = j|V) = \frac{\exp(\beta_j V)}{1+\sum_{h=1}^{J} \exp(\beta_h V)} =, j = 1, \dots, J \qquad \textbf{(Equation 4.1)}$$

Where β_j was a $(kx1)$ vector of unknown parameters to be estimated for alternative (j), and (h) was indexed over all alternatives. As the probability of choosing an alternative TNFi category must sum to one, the probability of choosing $(j = 0)$ could be expressed as in Equation 4.2 (Wooldridge, 2010):

$$Prob(Y = 0|V) = \frac{1}{1+\sum_{h=1}^{J} \exp(\beta_h V)} \qquad \textbf{(Equation 4.2)}$$

Changes in the probability of choosing an alternative were determined by the partial effects of the independent variables in (V). For any independent variable $x_k \epsilon V$, the partial effect

was not provided by its estimated coefficient alone (β_{jk}), but instead by Equation 4.3 (Wooldridge, 2010):

$$\frac{\partial Prob(Y=j|V)}{\partial x_k} = Prob(Y=j|V) * (\beta_{jk} - \frac{[\sum_{h=1}^{J} \beta_{hk} * \exp(\beta_h V)]}{1+\sum_{h=1}^{J} \exp(\beta_h V)}) \quad \textbf{(Equation 4.3)}$$

The sign and magnitude of the partial effect for a unit change in (x_k), for any alternative (j), therefore depended on the estimated coefficients across all other alternatives β_h (Wooldridge, 2010). Each partial effect was interpreted as the average change in probability of being prescribed a category of TNFi due to a unit increase in the independent variable of interest, *ceteris paribus*. The MNL regression was estimated by maximum likelihood in *STATA Version 13* (StataCorp, 2013).

4.3.3.2. Endogeneity

The vector (V) contained patient-level health and non-health characteristics. The demand for health care literature (see Section 4.3.1.1) described that patient-level characteristics may be endogenous if correlated with the residual error term, principally due to omitted variable bias. The results of *Chapter Three* suggested that two candidate sources of omitted variable bias were the influences on TNFi prescribing decisions within the external environment and within individual hospitals. Examples of such influences, that may be correlated with the independent and/or dependent variables in the MNL regression, are reported in Table 4.3.

Table 4.3. *Potential unobservable influences of TNFi choice.*

External Environment Influences	Hospital-level Influences
Change in evidence base or clinical guideline.	Local care pathways.
National drives to use specific TNFi therapies.	Productive capacity of hospital.
National price of a TNFi therapy.	Propensity of a hospital to negotiate the price of a TNFi therapy.

Therefore, the inclusion of the time-trend in vector (V) was anticipated to minimise endogeneity by controlling for unobservable changes in the external environment that may have influenced TNFi choice and/or patient-level variables over time. The inclusion of hospital-level dummy variables in vector (V) was assumed to control for the endogenous and unobservable supply-side fixed effects of hospitals over time (Gravelle et al., 2003).

4.3.3.3. Model Specification and Sensitivity Analysis

The base-case results estimated four different multivariable model specifications (reported in Table 4.4). The model specifications analysed the influence of health variables alone, and in conjunction with non-health variables, both with, and without, hospital-level fixed effects.

Table 4.4. *Regression specifications.*

Regression	Independent Variables	Variable List†
Regression A	Health variables, Time.	*DAS28, Woman, Age/10, HAQ, Totaldrug, Totalcomorb, YearswithRA, BMIover, MTX, Year.*
Regression B	Health variables, Time, Hospital fixed effects.	*DAS28, Woman, Age/10, HAQ, Totaldrug, Totalcomorb, YearswithRA, BMIover, MTX, Year, Hospital dummy variables.*
Regression C	Health variables, Non-health variables, Time.	*DAS28, Woman, Age/10, HAQ, Totaldrug, Totalcomorb, YearswithRA, BMIover, MTX, Smoke, Work1, Work3, Marital1, Marital3, Homecare, Year.*
Regression D	Health variables, Non-health variables, Time, Hospital fixed effects.	*DAS28, Woman, Age/10, HAQ, Totaldrug, Totalcomorb, YearswithRA, BMIover, MTX, Smoke, Work1, Work3, Marital1, Marital3, Homecare, Year, Hospital dummy variables.*

Note: †Variables are defined in Table 4.2.

Two sensitivity analyses of the base-case results were performed. One sensitivity analysis investigated whether the results were sensitive to changes in the measure of a patient's health behaviour, by replacing the *Smoke* variable with the *Alcohol* variable and re-estimating the base-case results. The second sensitivity analysis re-estimated the base-case results by omitting hospitals with fewer than ten patients from the sample, to determine whether the results were sensitive to the number of patients clustered within each hospital (Cameron et al., 2015).

4.3.3.4. Cluster-robust Standard Errors

Statistical inference requires precision in both the parameter estimate and the associated standard error (Cameron et al., 2015). Regression-based analyses of individual-level data typically assume that residual error terms are uncorrelated between observations (Greene, 2012). However, if the observations are clustered, as was the case in this study (patients were clustered by hospital), residual error terms may be uncorrelated across clusters but

correlated *within* clusters (Morris et al., 2005). Within-cluster residual correlation, if left unaddressed, may have overestimated the magnitude of the standard errors and artificially reduced the associated p-value (Williams, 2000). Therefore, cluster-robust standard errors by hospital were used for all MNL regressions (Wooldridge, 2010). The partial effect of an independent variable was deemed to be significantly different from zero if its associated p-value was less than 0.01, 0.05, or 0.1, demonstrating decreasing levels of statistical significance.

4.3.3.5. Measures of Model Fit

Three statistical measures of model fit (*McFadden's Pseudo-R^2*, the *Akaike Information Criteria* (AIC), and the *Bayesian Information Criteria* (BIC)) were used to estimate the relative ability of the four model specifications to fit the patient-level data (Akaike, 1974; McFadden, 1974; Schwarz, 1978). Each measure used the log-likelihood from the full model (LL^{Full}), and the pseudo-R^2 calculation required the log-likelihood from a restricted model including only the intercept term ($LL^{Restrictred}$). These measures were estimated from the complete-case data, given the difficulty in estimating relative model performance from multiply imputed datasets (Wood et al., 2008)

4.3.3.5.1. McFadden's Pseudo-R^2

The pseudo-R^2 (McFadden, 1974) was calculated according to Equation 4.4:

$$\text{Pseudo-}R^2 = 1 - \left(\frac{LL^{Full}}{LL^{Restricted}}\right) \qquad \textbf{(Equation 4.4)}$$

The pseudo-R^2 would have equalled zero if the log-likelihood from the two models were identical. The greater the log-likelihood of the full model, the greater the pseudo-R^2 and the better the relative fit of the model. However, the value of a pseudo-R^2 would have always increased as the number of parameters in the regression model increased (Long et al., 2012).

4.3.3.5.2. Akaike Information Criteria

The AIC (Akaike, 1974) was calculated according to Equation 4.5:

$$\text{AIC} = -2LL^{Full} + 2K \qquad \textbf{(Equation 4.5)}$$

where (K) was equal to the number of parameters in the model. The lower the AIC statistic, the better the model fit. Unlike the pseudo-R^2, the AIC imposed a penalty as the number of parameters in the regression model increased (Akaike, 1974).

4.3.3.5.3. Bayesian Information Criteria

The BIC (Schwarz, 1978) was calculated according to Equation 4.6:

$$\text{BIC} = -2LL^{Full} + K(LnN) \qquad \textbf{(Equation.4.6)}$$

where (N) was equal to the sample size. The lower the BIC statistic, the better the relative fit of the model. The penalty imposed by including additional parameters was greater than for the AIC, which indicated that the BIC gave preference to more parsimonious models (Schwarz, 1978).

4.4. Results

The results indicated a rejection of the hypothesis, stated in *Objective 2* (see Section 4.2), that no patient-level characteristics systematically influenced the choice of TNFi, after controlling for unobservable influences on prescribing decisions in the external environment and at the hospital-level. The results of this study are reported according to the summary statistics (Section 4.4.1), base-case results (Section 4.4.2), and the results of the sensitivity analyses (Section 4.4.3).

4.4.1. Summary Statistics

Figure 4.2 illustrates a flow diagram of the patients with RA that were included in the study. There were 1,788 baseline observations within BRAGGSS that reported a prescribing decision which involved any bDMARD. A total of 894 patients with RA met the study inclusion criteria (Table 4.1) and had specifically been prescribed any TNFi for the first time between 2009 and 2014 across forty-nine NHS hospitals. Table 4.5 presents the distribution of patients to each category of the dependent variable.

Figure 4.2. *Flow diagram of patients that met the study inclusion criteria.*

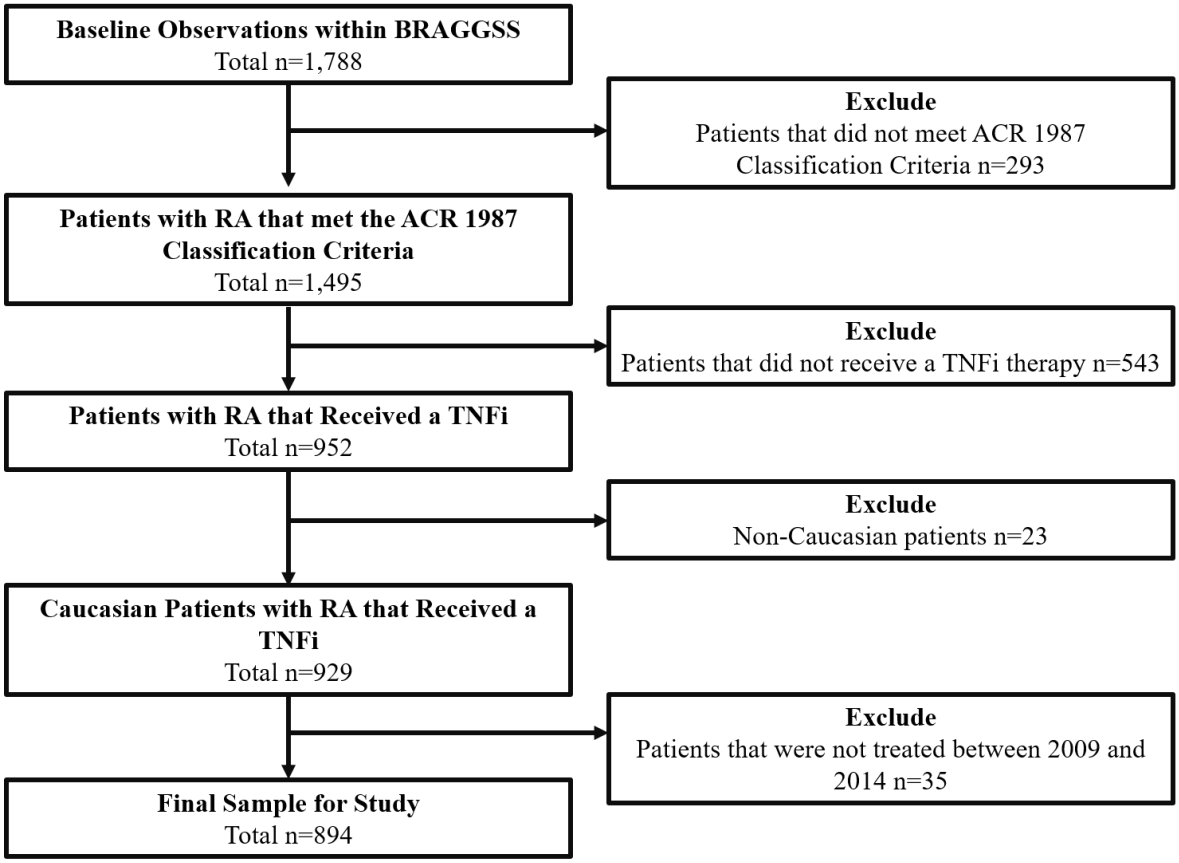


Table 4.5. *Distribution of patients in the sample to each category of the dependent variable.*

Category Name	TNFi	Number of Patients (n)
Non-monoclonal TNFi.	Etanercept.	357
Older monoclonal TNFi.	Adalimumab and infliximab.	373
Newer monoclonal TNFi.	Certolizumab pegol and golimumab.	164

The proportion of patients in the sample that were prescribed each TNFi varied over time (illustrated in Figure 4.3). The newer monoclonal TNFi therapies were prescribed less frequently than the other treatments between 2009 and 2014. The non-monoclonal TNFi, etanercept, became the most frequently prescribed treatment in the sample from 2012. Figure 4.4 depicts the proportion of patients in each of the forty-nine hospitals that were prescribed each category of TNFi treatment, to illustrate the presence of hospital-level variation in prescribing decisions across the sample.

Figure 4.3. *The proportion of patients in the sample that were prescribed each category of TNFi per year.*

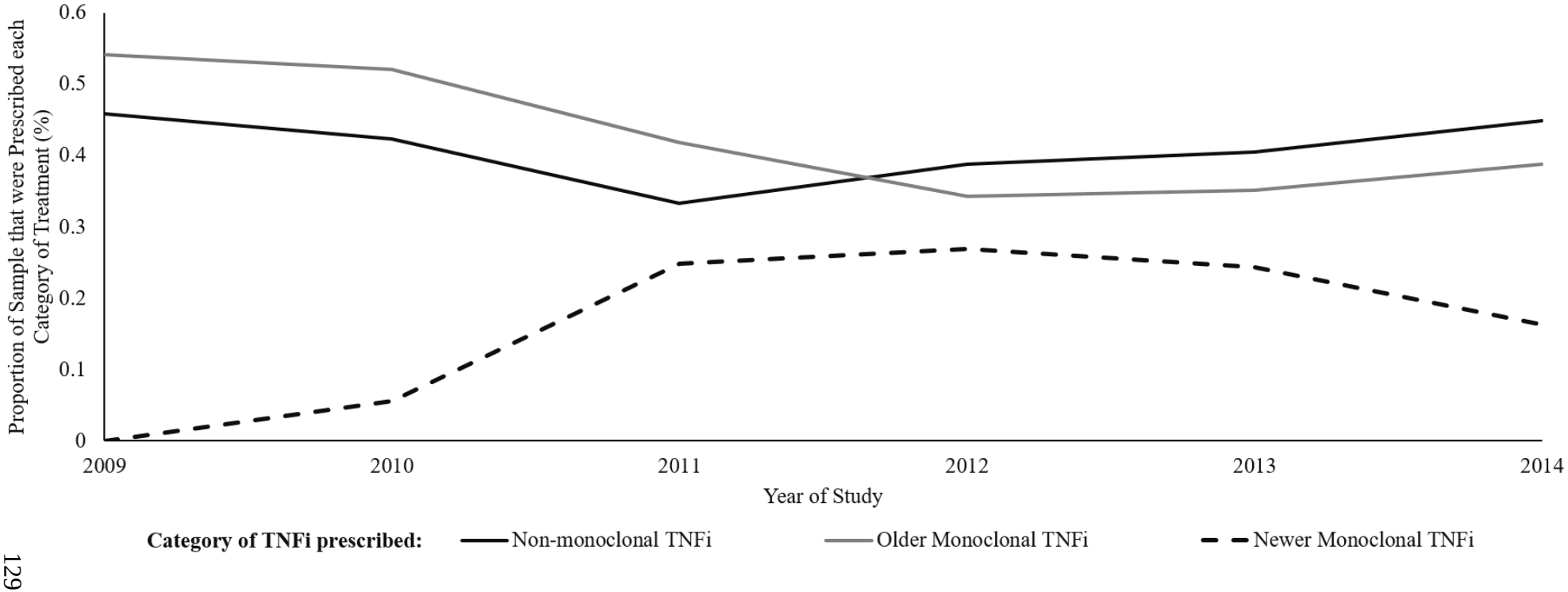


Figure 4.4. *The proportion of patients in each hospital that were prescribed each category of TNFi.*

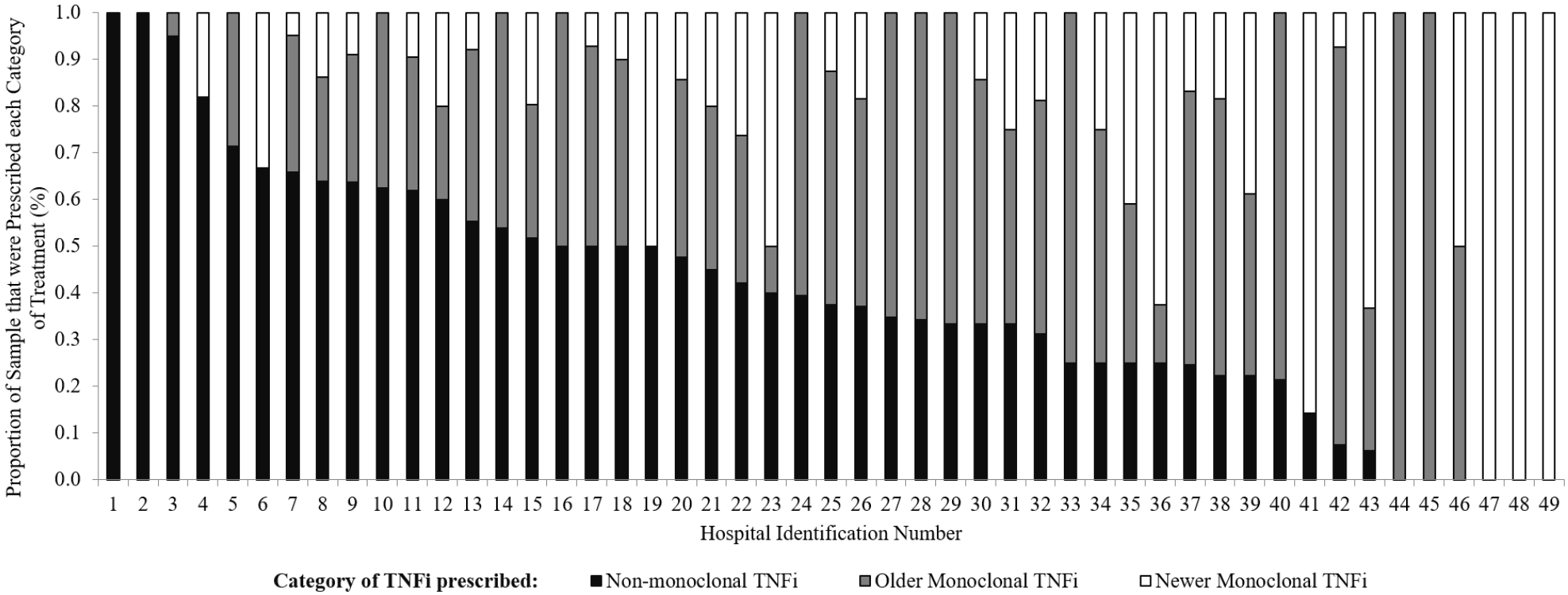


Table 4.6 reports the summary statistics for each independent variable in the analysis for (i) the full sample and (ii) for each category of the dependent variable. The mean values of the summary statistics for the full sample were consistent across each category of TNFi therapy.

The mean DAS28 score for the full sample was 5.41, which indicated a population with active and severe RA (Prevoo et al., 1995). Three quarters of the sample were women and the average patient age was 57 years old, both of which were consistent with the population-level epidemiological characteristics of RA (see Section 1.3.1). Patients had, on average, one comorbid condition alongside their RA diagnosis, and had previously received four different cDMARDs. Patients in the sample had a diagnosis of RA for an average of ten years prior to enrolment in the study. The majority of patients (79%) received concomitant methotrexate with their TNFi and 68% had a BMI indicative of being overweight.

Over half of the sample reported regularly drinking alcohol (59%), whereas 20% were a self-reported daily smoker. Over a third of the sample (38%) reported being employed; 44% were unemployed due to health reasons and 18% had left the labour market for non-health reasons. The majority of patients (74%) considered themselves to be in a relationship and almost all (90%) had a carer available to assist them at home.

Table 4.6. *Summary statistics of independent variables in the full sample and for each category of the dependent variable.*

	Total Sample (n=894)				Non-monoclonal TNFi (n=357)		Older monoclonal TNFi (n=373)		Newer monoclonal TNFi (n=164)	
Variable	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Health Variables										
DAS28	5.41	0.81	1.86	8.59	5.45	0.79	5.40	0.84	5.38	0.75
Woman	0.75	0.43	0	1	0.77	0.42	0.75	0.43	0.73	0.44
Age/10	5.71	1.23	1.9	8.5	5.76	1.22	5.62	1.26	5.80	1.18
HAQ	1.70	0.64	0	3	1.69	0.64	1.70	0.63	1.72	0.65
Totaldrug	4.14	1.63	0	12	4.25	1.74	4.07	1.58	4.04	1.50
Totalcomorb	1.09	1.18	0	7	1.15	1.18	1.04	1.22	1.08	1.09
YearswithRA	10.48	10.28	0	60	10.57	10.64	11.03	10.33	9.03	9.14
BMIover	0.68	0.47	0	1	0.69	0.46	0.67	0.47	0.67	0.47
MTX	0.79	0.41	0	1	0.74	0.44	0.81	0.39	0.86	0.35
Non-health Variables										
Alcohol	0.59	0.49	0	1	0.61	0.49	0.58	0.49	0.60	0.49
Smoke	0.20	0.40	0	1	0.18	0.38	0.21	0.41	0.23	0.42
Employment Status										
Work1	0.38	0.48	0	1	0.40	0.49	0.39	0.49	0.31	0.46
Work2†	0.44	0.50	0	1	0.43	0.50	0.45	0.50	0.46	0.50
Work3	0.18	0.38	0	1	0.17	0.37	0.16	0.37	0.23	0.42
Marital Status										
Marital1	0.74	0.44	0	1	0.73	0.44	0.76	0.43	0.73	0.45
Marital2†	0.09	0.29	0	1	0.11	0.32	0.08	0.27	0.07	0.25
Marital3	0.17	0.37	0	1	0.15	0.36	0.16	0.37	0.21	0.41
Homecare	0.90	0.30	0	1	0.90	0.30	0.89	0.31	0.91	0.29
Year	3.39	1.45	1	6	3.36	1.51	3.15	1.49	4.01	0.97

Note: SD = standard deviation; † denotes omitted categorical variable. Non-monoclonal TNFi=etanercept; Older monoclonal TNFi=infliximab & adalimumab; Newer monoclonal TNFi= certolizumab pegol & golimumab.

4.4.2. Base-case Results

The base-case results for each regression specification are presented in Table 4.7. The inclusion of hospital-level fixed effects improved the model fit according to all criteria (Regression B Vs. Regression A; Regression D Vs. Regression C). Regression D fitted the data best, compared with other model specifications, according its lower AIC and BIC statistics and higher pseudo-R^2 (Pseudo-R^2=0.40; AIC=447; BIC=593).

The partial effects of all clinical assessment variables (*DAS28*, *HAQ*, *BMIover*) were not significantly different from zero (for each variable, p>0.1 for all TNFi categories across all model specifications). This result may indicate that implicit stratification of TNFi therapies, according to clinical disease status, did not occur in routine practice.

The positive and significant partial effect of the *Year* variable (which was a proxy for the influence of factors in the external environment) may suggest that the probability of being prescribed a newer monoclonal TNFi increased over time (In *Regression C*: $Year_{\text{partial effect}|\text{newer monoclonal TNFi}}$=0.0705; p=0.00). However, the influence of factors in the external environment appeared to be absorbed by including hospital-level fixed effects to control for influences on treatment decisions at the hospital-level (In *Regression D*: $Year_{\text{partial effect}|\text{newer monoclonal TNFi}}$=0.0045; p=0.00). Moreover, the influence of patient-level factors on being prescribed a newer monoclonal TNFi were also reduced after controlling for unobservable heterogeneity between hospitals. These findings may provide evidence that hospital supply-side factors were more influential than patient-level characteristics, in determining whether a newer monoclonal TNFi was prescribed.

Regression D indicated that, after controlling for unobservable influences at the hospital-level and in the external environment, a ten-year increase in a patient's age was associated with a 4.5% greater probability of receiving the non-monoclonal TNFi etanercept (p=0.01), *ceteris paribus*. In contrast, patients that were prescribed concomitant methotrexate had an 11% lower probability of also being prescribed etanercept (p=0.02), *ceteris paribus*.

The socioeconomic characteristics of patients were found to be associated with the prescription of etanercept, adalimumab, and infliximab; patients reported to be in a relationship were were 16% more likely to have received an older monoclonal TNFi (p=0.02), *ceteris paribus*. The influence of socioeconomic characteristics may provide evidence for horizontal inequity in TNFi prescribing decisions (see Section 4.3.1.3.) if a

value judgement was made that non-health factors (such as a patient's relationship status) should not have influenced a patient's health care utilisation (Morris et al., 2005; Gravelle et al., 2006).

4.4.3. Sensitivity Analyses

Table 4.8 reports the mean partial effects from the multinomial logistic regression when the *Smoke* variable was replaced with the *Alcohol* variable. These two variables were a proxy for each patient's individual health behaviours, which may have been correlated with their involvement in routine treatment decision-making (see Section 4.3.3.3). The partial effects of the base-case analysis were robust to changes in the variable that measured each patient's individual health behaviours. *Regression D* remained the preferred model specification, due to its relatively higher pseudo-R^2 and relatively lower AIC and BIC test statistics. The same independent variables were statistically significant in *Regression D* (*Age, MTX, Marital1* in particular for etanercept, adalimumab, and infliximab) and demonstrated the same magnitude of partial effect, when compared with the base-case results in Table 4.7. There was therefore no evidence that a patient's self-reported health behaviours influenced the choice of TNFi therapy prescribed in routine clinical practice.

Appendix 21 reports the results of the sensitivity analysis when the regression was estimated on a restricted sample of hospitals that contributed at least ten patients to BRAGGSS. The sample size was reduced to 816 patients and the relative magnitude and direction of the base-case results were robust to performing the analysis on the restricted sample (*Appendix 21*; Table A21.1).

Table 4.7. *Base-case results: mean partial effects from multinomial logistic regression.*

	Regression A			Regression B			Regression C			Regression D		
	Non-mAb	**Older mAbs**	**Newer mAbs**	**Non-mAb**	**Older mAbs**	**Newer mAbs**	**Non-mAb**	**Older mAbs**	**Newer mAbs**	**Non-mAb**	**Older mAbs**	**Newer mAbs**
DAS28	0.0216	-0.0212	-0.0004	0.0481	-0.0472	-0.0009	0.0219	-0.0224	0.0004	0.0480	-0.0473	-0.0007
	(0.0287)	(0.0299)	(0.0201)	(0.0363)	(0.0363)	(0.0012)	(0.0292)	(0.0304)	(0.0197)	(0.0369)	(0.0369)	(0.0011)
Woman	0.0455	-0.0144	-0.0311	0.0536	-0.0510	-0.0025	0.0481	-0.0105	-0.0376	0.0529	-0.0505	-0.0024
	(0.0371)	(0.0409)	(0.0346)	(0.0445)	(0.0443)	(0.0022)	(0.0399)	(0.0428)	(0.0361)	(0.0476)	(0.0475)	(0.0019)
Age/10	0.0129	**-0.0280****	0.0150	**0.0277****	**-0.0279****	0.0002	**0.0335****	**-0.0352****	0.0016	**0.0452*****	**-0.0450*****	-0.0002
	(0.0123)	(0.0135)	(0.0113)	(0.0130)	(0.0130)	(0.0005)	(0.0147)	(0.0163)	(0.0111)	(0.0175)	(0.0175)	(0.0005)
HAQ	-0.0395	0.0161	0.0233	-0.0134	0.0117	0.0016	-0.0272	0.0200	0.0072	-0.0034	0.0027	0.0007
	(0.0295)	(0.0316)	(0.0196)	(0.0368)	(0.0369)	(0.0012)	(0.0285)	(0.0312)	(0.0212)	(0.0352)	(0.0353)	(0.0011)
Totaldrug	0.0165	0.0091	-0.0256	0.0006	-0.0004	-0.0002	0.0148	0.0095	-0.0243	-0.0017	0.0018	-0.0001
	(0.0202)	(0.0163)	(0.0187)	(0.0180)	(0.0177)	(0.0006)	(0.0202)	(0.0160)	(0.0185)	(0.0180)	(0.0178)	(0.0005)
Totalcomorb	0.0139	-0.0114	-0.0025	0.0216	-0.0217	0.0001	0.0142	-0.0096	-0.0046	0.0192	-0.0192	0.0000
	(0.0150)	(0.0153)	(0.0094)	(0.0183)	(0.0182)	(0.0005)	(0.0158)	(0.0157)	(0.0093)	(0.0193)	(0.0192)	(0.0005)
YearswithRA	-0.0008	0.0025	-0.0018	-0.0015	0.0017	**-0.0002****	-0.0009	**0.0029***	-0.0019	-0.0018	0.0020	**-0.0002*****
	(0.0020)	(0.0017)	(0.0014)	(0.0021)	(0.0021)	(0.0001)	(0.0021)	(0.0017)	(0.0015)	(0.0022)	(0.0021)	(0.0001)
BMIover	0.0245	0.0032	-0.0277	0.0060	-0.0053	-0.0007	0.0205	0.0032	-0.0237	0.0013	-0.0010	-0.0003
	(0.0470)	(0.0467)	(0.0294)	(0.0575)	(0.0572)	(0.0017)	(0.0471)	(0.0478)	(0.0293)	(0.0576)	(0.0575)	(0.0015)
MTX	**-0.1208*****	0.0577	**0.0631****	**-0.1049****	**0.1013****	**0.0036****	**-0.1237*****	0.0616	**0.0621****	**-0.1141****	**0.1109****	**0.0032*****
	(0.0393)	(0.0399)	(0.0287)	(0.0460)	(0.0459)	(0.0015)	(0.0410)	(0.0406)	(0.0288)	(0.0467)	(0.0465)	(0.0012)
Smoke							-0.0393	0.0255	0.0138	-0.0547	0.0522	**0.0025***
							(0.0442)	(0.0459)	(0.0261)	(0.0549)	(0.0549)	(0.0015)
Work1							0.0548	0.0193	**-0.0742****	0.0349	-0.0318	**-0.0031****
							(0.0578)	(0.0530)	(0.0330)	(0.0673)	(0.0670)	(0.0015)
Work3							-0.0279	0.0417	-0.0138	-0.0357	0.0371	-0.0014
							(0.0623)	(0.0636)	(0.0406)	(0.0808)	(0.0806)	(0.0017)
Marital1							**-0.1587****	**0.1052****	0.0535	**-0.1629****	**0.1613****	0.0016
							(0.0629)	(0.0529)	(0.0374)	(0.0671)	(0.0666)	(0.0016)
Marital3							**-0.1614****	0.0488	0.1127	-0.1319	0.1285	0.0034
							(0.0760)	(0.0786)	(0.0695)	(0.1084)	(0.1082)	(0.0033)
Homecare							0.0073	-0.0330	0.0258	0.0406	-0.0403	-0.0003
							(0.0610)	(0.0560)	(0.0346)	(0.0805)	(0.0803)	(0.0019)
Year	-0.0152	**-0.0515****	**0.0667*****	-0.0011	-0.0037	**0.0048*****	-0.0164	**-0.0541*****	**0.0705*****	-0.0011	-0.0034	**0.0045*****
	(0.0162)	(0.0206)	(0.0173)	(0.0186)	(0.0186)	(0.0010)	(0.0173)	(0.0205)	(0.0171)	(0.0202)	(0.0201)	(0.0008)
Hospital Dummy	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Pseudo R2		0.072			0.3634			0.1138			0.4022	
AIC		613.7943			474.8829			612.1243			447.0357	
BIC		694.2261			628.4346			736.428			593.2754	

Note: Standard errors are reported in parentheses. Non-mAbs = Non-monoclonal antibody (etanercept); Older mAbs = Older monoclonal antibodies (infliximab & adalimumab); Newer mAbs = Newer monoclonal antibodies (certolizumab pegol & golimumab). *, **, *** indicates statistical significance at 10%, 5% and 1%, respectively. Partial effects of hospital dummy variables are unreported

Table 4.8. *Sensitivity analysis: mean partial effects from multinomial logistic regression after replacing smoking behaviour with self-reported alcohol intake.*

	Regression C			Regression D		
	Non-mAb	**Older mAbs**	**Newer mAbs**	**Non-mAb**	**Older mAbs**	**Newer mAbs**
DAS28	0.0221	-0.0218	-0.0002	0.0482	-0.0474	-0.0009
	(0.0290)	(0.0306)	(0.0199)	(0.0368)	(0.0367)	(0.0012)
Woman	0.0525	-0.0154	-0.0371	0.0576	-0.0548	-0.0028
	(0.0388)	(0.0418)	(0.0345)	(0.0471)	(0.0470)	(0.0021)
Age/10	**0.0347****	**-0.0355****	0.0008	**0.0461*****	**-0.0458*****	-0.0003
	(0.0143)	(0.0156)	(0.0106)	(0.0172)	(0.0172)	(0.0005)
HAQ	-0.0259	0.0154	0.0105	-0.0029	0.0020	0.0010
	(0.0274)	(0.0331)	(0.0231)	(0.0341)	(0.0342)	(0.0013)
Totaldrug	0.0157	0.0080	-0.0238	-0.0002	0.0003	-0.0001
	(0.0199)	(0.0164)	(0.0175)	(0.0181)	(0.0179)	(0.0006)
Totalcomorb	0.0155	-0.0116	-0.0040	0.0212	-0.0212	0.0000
	(0.0156)	(0.0153)	(0.0094)	(0.0192)	(0.0191)	(0.0005)
YearswithRA	-0.0009	**0.0029***	-0.0020	-0.0016	0.0018	**-0.0002*****
	(0.0020)	(0.0017)	(0.0014)	(0.0021)	(0.0021)	(0.0001)
BMIover	0.0260	-0.0027	-0.0233	0.0085	-0.0081	-0.0004
	(0.0477)	(0.0486)	(0.0292)	(0.0590)	(0.0588)	(0.0016)
MTX	**-0.1210*****	0.0569	**0.0641***	**-0.1110****	**0.1075****	**0.0035*****
	(0.0407)	(0.0402)	(0.0280)	(0.0466)	(0.0465)	(0.0013)
Alcohol	0.0250	-0.0451	0.0201	0.0226	-0.0235	0.0009
	(0.0393)	(0.0438)	(0.0348)	(0.0460)	(0.0464)	(0.0016)
Work1	0.0587	0.0162	**-0.0749****	0.0418	-0.0383	**-0.0035****
	(0.0574)	(0.0521)	(0.0332)	(0.0670)	(0.0667)	(0.0017)
Work3	-0.0238	0.0365	-0.0127	-0.0276	0.0292	-0.0016
	(0.0625)	(0.0633)	(0.0415)	(0.0799)	(0.0796)	(0.0018)
Marital1	**-0.1591****	**0.1065****	0.0526	**-0.1624****	**0.1606****	0.0018
	(0.0627)	(0.0524)	(0.0377)	(0.0669)	(0.0663)	(0.0018)
Marital3	**-0.1633****	0.0481	**0.1152***	-0.1353	0.1309	0.0045
	(0.0747)	(0.0781)	(0.0700)	(0.1074)	(0.1071)	(0.0038)
Homecare	0.0121	-0.0393	0.0272	0.0467	-0.0463	-0.0004
	(0.0619)	(0.0570)	(0.0355)	(0.0824)	(0.0823)	(0.0022)
Year	-0.0160	**-0.0545*****	**0.0705*****	-0.0009	-0.0040	**0.0049*****
	(0.0173)	(0.0205)	(0.0170)	(0.0199)	(0.0198)	(0.0009)
Hospital Dummy	No	No	No	Yes	Yes	Yes
Pseudo R2		0.1138			0.4022	
AIC		612.1243			447.0357	
BIC		736.428			593.2754	

Note: Standard errors are reported in parentheses. Non-mAbs = Non-monoclonal antibody (etanercept); Older mAbs = Older monoclonal antibodies (infliximab & adalimumab); Newer mAbs = Newer monoclonal antibodies (certolizumab pegol & golimumab). *, **, *** indicates statistical significance at 10%, 5% and 1%, respectively. Partial effects of hospital dummy variables are unreported.

4.5. Discussion

This study aimed to identify the patient-level characteristics that influenced the choice of TNFi prescribed to patients with RA in England. The analysis built on the results of *Chapter Three* which suggested that treatment decisions for RA may be influenced by factors in the external environment and at the hospital-level. MNL regression was used to model the patient-level factors that may influence first-line TNFi prescribing decisions observed in a representative sample of 894 patients with RA across England between 2009 and 2014, identified within the BRAGGSS cohort. There was evidence to reject the hypothesis that patient-level factors did not influence the prescription of etanercept, adalimumab, and infliximab which may indicate that these therapies were subject to

implicit stratification in current practice. There was also evidence that patient-level factors had a limited influence on the prescription of newer monoclonal TNFi therapies (certolizumab pegol and golimumab), which may suggest that, on average, these treatments were prescribed according to their lower cost, in alignment with NICE recommendations.

Older patients were found to have a greater probability of being prescribed etanercept, relative to other TNFi therapies, *ceteris paribus*. The risk of infection from bDMARD therapies may be greater in older patients (Lahiri et al., 2015) and the rheumatologists in *Chapter Three* perceived that etanercept was the most suitable TNFi for patients at-risk of infection. Therefore, the potential implicit stratification of etanercept therapy by age may have been mediated by an unmeasured patient-level risk of infection. The three North American studies of the patient-level factors that influenced TNFi choice (described in *Appendix 16*), by contrast, estimated that older patients were more likely to have been prescribed infliximab (DeWitt et al., 2006; Carter et al., 2012; Zhang et al., 2013). However, direct cross-country comparisons of health care use may be inappropriate, due to structural differences between health care systems in the delivery of care, payment mechanisms, and the patient case-mix (McPherson, 1989).

Patients that were prescribed concomitant methotrexate were more likely to have been co-prescribed an older monoclonal TNFi (infliximab or adalimumab), *ceteris paribus*. This result also had clinical plausibility because etanercept monotherapy was recommended by NICE for patients that were intolerant to methotrexate (National Institute for Health and Care Excellence, 2016a). The importance of including concomitant methotrexate use as an independent variable was illustrated by the magnitude of its partial effect; omission of concomitant therapies, and methotrexate in particular, has previously been demonstrated to confound empirical studies in RA that used patient-level data (Hudson et al., 2010)

The results also indicated that a patient's socioeconomic characteristics, measured by proxy according to their marital status, may have influenced the choice of TNFi therapy prescribed in current practice. The potential influence of a patient's socioeconomic characteristics on their health care use has been documented widely in the econometric literature regarding the demand for health care (van Doorslaer et al., 2004; Gravelle et al., 2006; Zuvekas, 2014), including in publically-funded health care systems such as the NHS in England (Laudicella et al., 2012). However, a patient's marital status, as a multinomial variable, had nominal properties which made it an imperfect proxy of socioeconomic *status* because the direction of its influence was unknown *a priori*, unlike variables with

ordinal properties (Erreygers et al., 2011) such as income or level of education. Recommendations for managing patients with RA by EULAR and the ACR suggested that treatment decisions should be made in a shared decision-making context with the patient (Smolen et al., 2014; Singh et al., 2016b). It may be possible that a patient's socioeconomic characteristics correlated with their own preferences of certain treatments, and these preferences were expressed during a shared decision-making process (Vick et al., 1998). However, if a value judgement was made, *a priori*, that non-health patient characteristics should not affect routine treatment decisions, then the influence of a patient's marital status in this study may have suggested the presence of horizontal inequity in TNFi prescribing (Section 4.3.1.3).

This study was the first quantitative analysis of the patient-level factors that influenced routine TNFi prescribing decisions for patients with RA in England and had three distinct advantages. Firstly, the analysis utilised data from real prescribing decisions that were observed in routine clinical practice, such that the observations represented a clinician's revealed preference for a certain treatment (Mark et al., 2004), rather than hypothetical choices made in a stated preference analysis (Wardman, 1988; Kievit et al., 2010). Secondly, to mitigate the potential for endogeneity through omitted variable bias, hospital-level fixed effects and a time trend were included in the MNL regression to control for unobservable non-patient-level influences on prescribing decisions. Thirdly, multiple imputation was used to handle missing data and retain the largest possible sample; the results of related studies that were published previously, such as DeWitt et al. (2006), excluded patients with missing data, which may have introduced a selection bias into their analysis.

Limitations

A limitation experienced by most studies that have estimated the determinants of demand for health care is that data are often characterised by a trade-off between the precision of measurement and availability (Propper et al., 2005). Data are typically characterised as either (i) smaller samples, specific to a disease or treatment, that have measured variables for health and health care use with high precision, but measured non-health variables imperfectly, or (ii) larger-scale analyses of national household surveys that have measured non-health variables with greater precision, but measured health status imperfectly (Propper et al., 2005). This study was an example of the former; high-quality data were available on the treatment received by patients and on their health status (clinical

assessments were performed by a health professional) but non-heath variables were self-reported by patients and were imperfect proxies for socioeconomic status. If these socioeconomic characteristics were excluded, however, the analysis may have been subject to an additional source of omitted variable bias which, in turn, may have reduced the robustness of the results.

A second limitation of this study was that the cost of each TNFi therapy was not directly observable in the dataset. The study therefore assumed that the cost of each TNFi may have differed between hospitals, and the hospital-level fixed effects were sufficient to control for this unobservable heterogeneity. This assumption was plausible because the cost of treatments could be negotiated at the regional-level in the NHS (Stokoe et al., 2011). The hospital-level fixed effects, however, may have additionally accounted for other unobservable hospital-level influences that were described by the rheumatologists in *Chapter Three* (such as a local treatment algorithm or a system to promote compliance with NICE recommendations). Therefore, despite the inability to observe the cost of treatment directly, by controlling for unobservable hospital-level heterogeneity, the risk of endogeneity due to omitted variable bias was reduced, which may have confounded the results of similar previous studies.

A third potential limitation was that this study used a smaller sample relative to the three North American-based quantitative analyses of TNFi prescribing decisions (DeWitt et al., 2006; Carter et al., 2012; Zhang et al., 2013). However, the sample of patients in this study may have been more homogenous, compared to those other studies, given the existence of NICE recommendations that explicitly stated the characteristics of patients who were eligible to receive TNFi therapies in England (National Institute for Health and Care Excellence, 2016a). Therefore, despite the smaller sample size, the greater homogeneity of the sample may have improved the study's sensitivity to detect the factors that influenced routine prescribing decisions.

A fourth limitation of the study was that, given the sample size, the choice of TNFi was determined according to a three-category multinomial dependent variable rather than a five-category dependent variable (one category for each TNFi therapy). This study, however, was the first to consider the patient-level factors that influenced the choice of all five TNFi therapies licenced for patients with RA. In addition, the decision to collapse the dependent variable into three categories had clinical plausibility and statistical plausibility according to the results of a likelihood ratio test.

Lastly, the specification of the regression included a dummy variable for each individual hospital, which resulted in a substantial number of independent variables to be estimated within the analysis. Although the inclusion of the hospital-level dummy variables improved the fit of the regression, a potential limitation of this approach was that the frequency of patients that received each category of TNFi within each hospital may be been small. Alternatively, a dummy variable could have been included for each hospital's CCG, rather than for each hospital, and would have resulted in fewer independent variables within the regression analysis. However, the results of *Chapter Three* indicated that hospital-level influences on prescribing decisions for RA were important, compared with external influences (such as those from a CCG) and, therefore, should have been included within the empirical analysis of prescribing decisions to avoid omitted variable bias.

Implications for Future Research

The BRAGGSS cohort was a subset of patients enrolled to the larger *British Society for Rheumatology Biologics Register* (BSRBR) study (Hyrich et al., 2011). A follow-up study could therefore be undertaken, to replicate the findings of this research, by using the MNL regression method with the larger BSRBR sample. A potential advantage of performing the MNL regression on a larger sample may be that more patients would be distributed to the categories of the dependent variable and individual hospitals, which may subsequently reduce the practical problems associated with non-converging estimations by maximum likelihood (Rabe-Hesketh et al., 2005).

The data for this study were collected prior to the introduction of biosimilar TNFi therapies in current practice. Health care systems are expected to express a preference towards biosimilar therapies due to their potentially lower cost, compared with branded biologic drugs (Grabowski et al., 2014). It is unclear, however, whether clinicians and patients are willing to use biosimilar therapies in practice (Casey, 2016). Patient registers for RA in England have since started to collect data on patients treated with biosimilar TNFi therapies in routine practice (The British Society for Rheumatology, 2016). The results of this study could therefore be developed further, by using these new observational data, to investigate the patient-level factors that may influence the prescribing of biosimilar TNFi therapies, compared with their branded alternatives.

Finally, hospital-level factors were found to have an influence on the prescribing of TNFi therapies in general, and on the newer monoclonal TNFi therapies in particular. Further

exploration of these hospital-level influences was beyond the scope of this study and the inclusion of additional variables in the analysis may not have been appropriate, given the sample size. However, to develop a clearer understanding of the hospital-level influences on treatment decisions for RA, a future quantitative study could condition on additional hospital-level variables to potentially explain some of the unobservable between-hospital heterogeneity in this study. For example, treatment decisions may be influenced by hospital quality (which has previously been measured by proxy using evidence of thirty-day mortality from emergency admissions) (Dranove, 2012), the size of the hospital, and whether the hospital was research-oriented and aligned with the medical school of a university (Beckert et al., 2012).

4.6. Conclusion

This quantitative study built on the results of *Chapter Three* and utilised economic theory regarding the demand for health care, agency relationships, and inequity in health care use, to estimate the patient-level factors that influenced the choice of TNFi for patients with RA in England. Representative data were obtained from a national cohort study of 894 treatment decisions observed in routine practice between 2009 and 2014. By controlling for annual and hospital variation, to account for the environmental and hospital-level influences on treatment decisions identified by the rheumatologists in *Chapter Three*, the results identified that etanercept, adalimumab, and infliximab prescribing decisions may be implicitly stratified according to a patient's age, concomitant methotrexate use, and socioeconomic characteristics. The results also suggested that newer monoclonal TNFi therapies (certolizumab pegol and golimumab), by contrast, may have been prescribed in alignment with the recommendations by NICE.

Having explored current practice for RA in England to a greater extent than the previous model-based economic evaluations of stratified medicine (identified by the systematic review in *Chapter Two*), the remaining chapters of this thesis are focused on the early cost-effectiveness analysis of adalimumab ADAb and drug level testing to stratify treatment. *Chapter Five* begins to develop the economic evaluation by conceptualising and producing the structure of the *de novo* decision analytic model. *Chapter Six* subsequently reports the methods and final results of the economic evaluation.

Chapter 5

Early Model-based Economic Evaluation of Stratified Medicine: Decision Problem, Model Conceptualisation, and Structure

Chapter Five describes the development of the early model-based economic evaluation of adalimumab ADAb and drug level ELISA testing to stratify treatment for patients with RA in England. This chapter is presented as a series of six sequential studies that developed the decision problem of the economic evaluation and the final structure of the *de novo* decision analytic model. A diagram to illustrate the structure of the chapter is provided in Figure 5.1.

5.1. Introduction

The primary purpose of any economic evaluation is to inform a health care resource allocation decision by providing sufficient evidence of the incremental (costs and health) outcomes derived from a particular health technology (Drummond et al., 2015). Two preliminary stages of any model-based economic evaluation, that must be addressed before estimating the relative cost effectiveness of a health technology, are to (i) define the decision problem, and (ii) to conceptualise the decision analytic model (Roberts et al.,

2012). A *decision problem* provides a clear statement of the resource allocation decision under consideration (Roberts et al., 2012) and, consequently, facilitates an economic evaluation to be designed in order to generate relevant evidence for the decision-making context (Chilcott et al., 2010). A *conceptual model* is an abstract representation of a phenomenon of interest (such as the use of a new test to stratify treatment in a care pathway), often illustrated diagrammatically, that can assist in determining the final structure of a *de novo* decision analytic model (Tappenden, 2014). *Chapter Five* focussed on these two preliminary elements of the early economic evaluation of adalimumab ADAb and drug level testing by ELISA to stratify treatment for RA (see Section 1.3.5 for a description of the case study).

Decision problems, in general, are characterised by an explicit objective that details: the policy context of the economic evaluation, the target patient population, the perspective of the analysis, the time horizon, the relevant costs and health outcomes, and the intervention health technologies under consideration (Roberts et al., 2012). A specific challenge with the early economic evaluation of new a test to stratify a treatment decision, however, is that there may be multiple ways to potentially use that test in practice (Buisman et al., 2016; Smith et al., 2016). For example, commercial ELISA tests were available to measure adalimumab ADAb and drug levels (see Table 1.8), but the specific ways in which these tests could be used to inform a treatment decision was uncertain. Therefore, the first half of *Chapter Five* addressed the decision problem of the early economic evaluation by clarifying the role of ADAb and drug level testing to stratify treatment.

The second half of *Chapter Five* addressed the conceptualisation and development of the *de novo* decision analytic model. A transparent conceptualisation procedure can enhance the credibility of a decision analytic model by making explicit, and providing justification for, its key simplifications and assumptions (Robinson, 2008). For example, a conceptual model may aid the identification of key input parameters to include in a final model and the relationships between those parameters (Chilcott et al., 2010; Tappenden, 2014). A distinction was made between conceptual models that were: (i) problem-oriented, and (ii) design-oriented (Tappenden, 2014).

A *problem-oriented conceptual model* described the system in which the decision problem existed (Tappenden, 2014). There were two types of problem-oriented conceptual model: (i) a *disease logic model* described the true underlying disease status of a patient over time, in terms of the clinical events that may have occurred and disease states that may have

been experienced; (ii) a *service pathway model* described the sequence of health care interventions along a care pathway that a patient may have received (Tappenden, 2014). In the context of the early economic evaluation of adalimumab ADAb and drug level testing, problem-oriented conceptual models may have been beneficial because they could be used to rationalise the potential treatment decisions following a test result, and the potential health consequences of making a treatment decision according to a correct or incorrect test result.

A *design-oriented conceptual model* combined the two problem-oriented conceptual models to identify a feasible structure for the final decision analytic model (Tappenden, 2014). A design-oriented conceptual model made the structural modelling decisions explicit and provided early identification of the potential evidence requirements for the economic evaluation (Tappenden, 2014). Problem-orientated and design-orientated conceptual models were recommended to be used in sequence, to move from the care pathway observed in clinical practice to the final structure of the decision analytic model (Tappenden, 2014).

5.2. Aim and Objectives

The aim of *Chapter Five* was to report the design of the economic evaluation of adalimumab ADAb and drug level ELISA testing to stratify treatment for patients with RA in England. There were six sequential objectives to meet this aim:

Objective 1: Identify how ADAb and drug level testing may be used to stratify treatment for patients with RA that received a TNFi therapy in routine practice;

Objective 2: Identify the subset of relevant comparator testing strategies to include in the decision problem of the economic evaluation;

Objective 3: Define the decision problem of the economic evaluation;

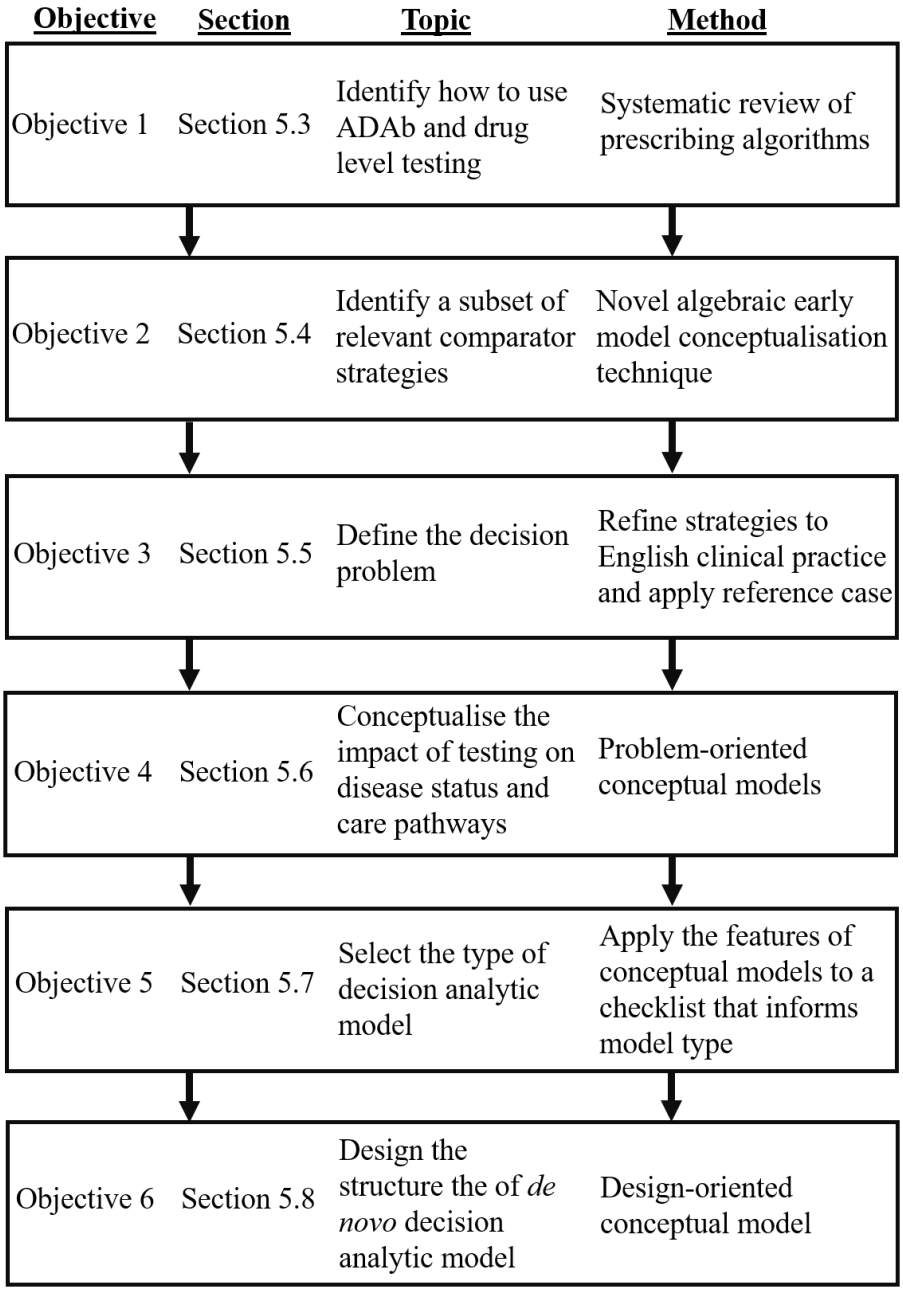
Objective 4: Conceptualise the potential impact of adalimumab ADAb and drug level testing to stratify treatment decisions on a patient's subsequent (i) disease status, and (ii) care pathways;

Objective 5: Select the type of decision analytic model to be used for the economic evaluation;

Objective 6: Determine the structure of the *de novo* decision analytic model to address the decision problem.

Each chapter objective became the aim of a specific sub-study reported within *Chapter Five*. Figure 5.1 presents a flow diagram, for the purpose of orientation, that provides a visual overview of the chapter's structure, the topic addressed by each sub-study, and the method that was used for each sub-study.

Figure 5.1. *Flow diagram to illustrate the structure of Chapter Five.*



5.3. Chapter Objective 1: Identifying Prescribing Algorithms for TNFi ADAb and Drug Level Testing in Rheumatoid Arthritis

This study addressed the first challenge of conducting an economic evaluation of ADAb and drug level testing, to stratify treatment for patients with RA, by investigating how testing may be performed within routine practice. The results of this study were subsequently used to inform the comparator strategies within the final decision problem.

5.3.1. Introduction

An economic evaluation of any stratified medicine requires an understanding of how a testing strategy can be used in clinical practice to inform a subsequent treatment decision (National Institute for Health and Care Excellence, 2011a; Annemans et al., 2013; Shabaruddin et al., 2015). The use of adalimumab ADAb and drug level testing necessarily required a change to the existing care pathway for patients with RA because testing was not currently performed in routine practice. A challenge reported in the literature for the early economic evaluation of any new testing strategy, such as the ELISA-based ADAb and drug level tests (see Section 1.3.5), was that there may have been numerous ways in which testing could potentially be used within an existing care pathway (Buisman et al., 2016). One potential solution to this challenge was to review the clinical literature to help identify how such new health technologies may be incorporated into routine practice (Tappenden, 2014; Buisman et al., 2016).

Treatment algorithms, of which prescribing algorithms are a subset, comprise "if…then" statements that describe the appropriate actions to take according to relevant factors at the time of a clinical decision (Schoenbaum et al., 1990; Woolf et al., 1999). Independent treatment algorithms for a new health technology may be published as its supporting clinical evidence base begins to develop over time (Tak, 2012). Prescribing algorithms may help to understand how a biomarker test could be used to stratify a subsequent prescribing decision which, in the absence of end-to-end evidence, was essential information in order to conduct a model-based economic evaluation of a new stratified medicine (National Institute for Health and Care Excellence, 2011a).

A number of independent prescribing algorithms have been published that state how and when to use TNFi ADAb and drug level testing explicitly, such as that of Vincent et al. (2013), which were, in theory, applicable to all diseases that were treated with TNFi therapies (hereafter termed "generic prescribing algorithms"). Some of the recommendations within these generic prescribing algorithms, however, may not have been applicable to patients with RA (den Broeder et al., 2013). For example, some generic prescribing algorithms may have recommended TNFi dose-escalation strategies (Vincent et al., 2013), but the rheumatologists in *Chapter Three* explained that it would not be appropriate to escalate the dose of bDMARD therapies for their patients with RA in England (see Section 4.3.1).

Disease-specific prescribing algorithms (hereafter termed "RA-specific prescribing algorithms") may have been more likely to recommend prescribing decisions that had clinical relevance to patients with RA. For example, an RA-specific prescribing algorithm may be based on a care pathway, or assume treatment availability, that was more representative of routine practice for RA. There had been no previous systematic investigation of the RA-specific prescribing algorithms that incorporated TNFi ADAb and drug level testing in routine practice.

5.3.2. Aim and Objectives

The aim of this study was to identify how ADAb and drug level testing may be used to stratify treatment for patients with RA that received a TNFi therapy in routine practice. The study had two objectives:

Objective 1: Identify all RA-specific prescribing algorithms that included ADAb and drug level testing for any TNFi therapy;

Objective 2: Synthesise and appraise the recommendations within the RA-specific prescribing algorithms regarding the use of TNFi ADAb and drug level testing.

5.3.3. Method

A systematic review of published RA-specific prescribing algorithms was performed according to the *PRISMA* recommendations (Liberati et al., 2009). The completed *PRISMA*

checklist for this study is reported in *Appendix 22*. The systematic review included all algorithms that reported an explicit TNFi ADAb and/or drug level test-and-treat strategy within an explicit algorithm for an RA population. The study inclusion criteria is reported in Table 5.1. Studies were excluded if they only provided recommendations for testing within the text of the manuscript.

Table 5.1. *Systematic review inclusion criteria: TNFi ADAb and drug level prescribing algorithms for RA.*

Study Feature	Inclusion Criteria
Population.	Adults with rheumatoid arthritis.
Intervention.	Any assay to measure TNFi ADAb and/or drug levels.
Outcome.	Prescribing algorithm for using the intervention in the defined population.
Study Design.	Any design in a peer-reviewed publication (exclude conference abstracts).
Language.	English.

Study Selection

This systematic review built upon a published study conducted as part of the NICE DAP for the appraisal of TNFi ADAb and drug level testing in patients with Crohn's disease in England (National Institute for Health and Care Excellence, 2016c). The search strategy from an independent systematic review conducted during the appraisal process was amended to identify RA-specific prescribing algorithms by replacing disease-specific terms for Crohn's disease with terms relating to RA (reported in *Appendix 23*).

Medline and *Embase* were then searched electronically using this search strategy for studies published from the date of inception until August 2016. *Medline* and *Embase* were deemed to be appropriate because they were the databases used most frequently within previous systematic reviews of health care interventions (Centre for Reviews and Dissemination, 2009).

All titles and abstracts identified by the search strategy were screened by SG against the inclusion criteria. Six researchers at the *Manchester Centre for Health Economics*, *The University of Manchester* were allocated an equal proportion of abstracts to second-screen. Studies were not excluded if there were disagreements at the screening stage. All studies included after the initial screening were read in full by SG to identify whether an explicit prescribing algorithm, that included TNFi ADAb and/or drug level testing, was reported.

The reference lists of all included studies were hand-searched to identify additional studies that met the inclusion criteria.

Data Extraction and Analysis

The following information was extracted from all prescribing algorithms that met the inclusion criteria: (i) the time and frequency of testing; (ii) whether recommendations were reported for a specific TNFi; (iii) the treatment recommendations for each test result; (iv) the country of the study; and (v) the evidence used to inform the algorithm. Study features were summarised in tabular form and treatment recommendations were appraised by narrative synthesis according to the population eligible for testing.

5.3.4. Results

The search strategy identified 4,006 database records and 119 manuscripts were read in full, after title and abstract screening, against the inclusion criteria (flow diagram reported in Figure 5.2). Nine publications reported an explicit RA-specific prescribing algorithm with TNFi ADAb and/or drug level testing and were included in the systematic review. The features of each prescribing algorithm are reported in Table 5.2.

Figure 5.2. *Systematic review flow diagram of included studies: TNFi ADAb and drug level prescribing algorithms for RA.*

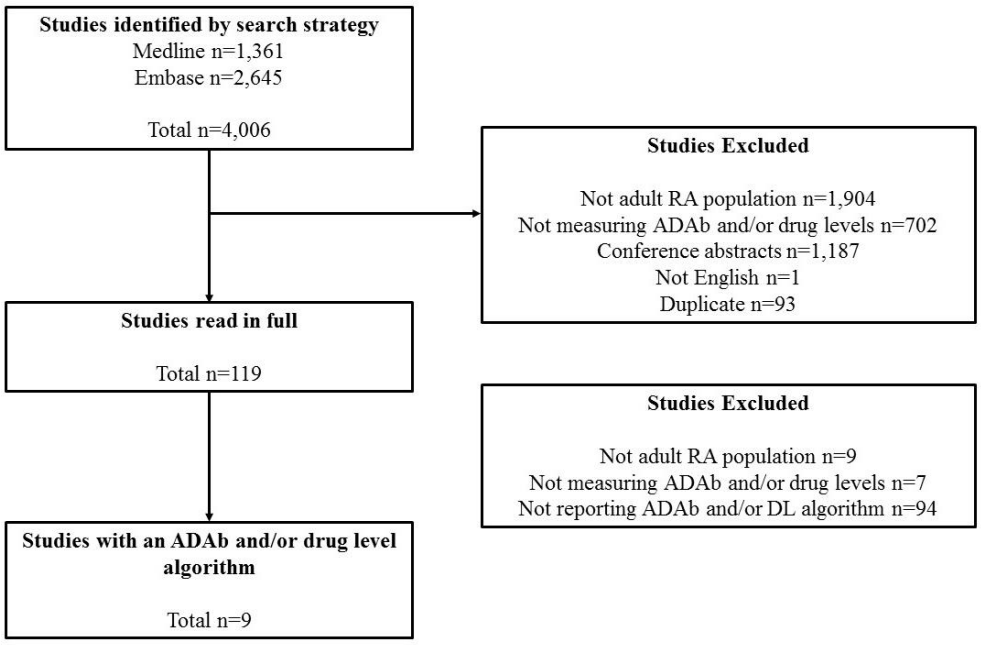


Table 5.2. *Features of nine RA-specific prescribing algorithms with TNFi ADAb and drug level testing.*

Author (Year) Country	Population Eligible for Testing: After loss of response to TNFi	Population Eligible for Testing: While responding to TNFi	Timing of Testing	Frequency of Testing (while responding)	Tests in Algorithm	Source of Evidence for Algorithm	Specific TNFi
Bendtzen. (2011). Country: Not reported.	Yes.	Yes.	(i) Early (2-3 months); (ii) Late (6 months or more).	Not reported.	(i) Drug level in early testing; (ii) ADAb and drug level in late testing.	Non-systematic review.	Not reported.
Chen et al. (2015). Country: Taiwan.	Yes.	Yes.	3-6 months after starting TNFi.	Not reported.	(i) Drug level in responders; (ii) ADAb and drug level in non-responders.	Non-systematic review & accompanying cohort study.	Not reported.
Daïen et al. (2012). Country: France.	Yes.	No.	3 months after starting TNFi.	Not applicable.	(i) Drug level.	Non-systematic review & accompanying cohort study.	Etanercept.
Garcês et al. (2014). Country: Portugal.	Yes.	Yes.	3 months after starting TNFi.	Every 3 months.	(i) Drug level only (if detectable); (ii) ADAb if drug levels are not detected.	Non-systematic review & accompanying cohort study.	Not reported.
Krieckaert et al. (2015). Country: Netherlands.	Yes.	Yes.	7 months after starting TNFi.	Not applicable.	(i) Drug level.	Non-systematic review & economic evaluation.	Adalimumab.
Mok et al. (2016). Country: Asian countries.	Yes.	Yes.	6 months after starting TNFi.	Every 6 months or when efficacy reduces.	(i) ADAb and drug level.	Authors' opinions.	Monoclonal TNFi.
Mulleman et al. (2009). Country: France.	No.	Yes.	Before next infusion.	Before next infusion.	(i) Drug level.	Non-systematic review.	Infliximab.
Mulleman et al. (2012). Country: Not reported.	Yes.	Yes.	(i) Following adverse drug reaction; (ii) Not reported for responders or loss of response.	Not reported.	(i) Drug level in responders and patients that lost response; (ii) ADAb after adverse drug reaction.	Non-systematic review.	Not reported.
Rosas et al. (2014). Country: Spain.	No.	Yes.	6 months after starting TNFi.	Every 6 months.	(i) Drug level; (ii) ADAb if drug level is low and disease activity is high.	Non-systematic review & accompanying cross-section study.	Adalimumab.

Note: TNFi = tumour necrosis factor-α inhibitor, ADAb = anti-drug antibody.

Four of the prescribing algorithms were designed for specific TNFi therapies (Mulleman et al., 2009; Daïen et al., 2012; Rosas et al., 2014; Krieckaert et al., 2015), and one algorithm was designed for patients that received any monoclonal TNFi (Mok et al., 2016). The majority of the prescribing algorithms (n=6) included both ADAb and drug level testing. Three of the algorithms for specific TNFi therapies considered only drug level testing. Testing within six months of commencing treatment was recommended by over half of the prescribing algorithms (n=5).

The evidence to support eight of the algorithms was based on a non-systematic review of the literature, five of which were supplemented by an accompanying study (Daïen et al., 2012; Garcês et al., 2014; Rosas et al., 2014; Chen et al., 2015b; Krieckaert et al., 2015). Garcês et al. (2014) evaluated the effectiveness of their algorithm, and Krieckaert et al. (2015) evaluated the relative cost-effectiveness of testing. The source of evidence for one algorithm was reported to be based on the authors' opinions (Mok et al., 2016).

The prescribing algorithms were categorised by two groups according to the population eligible for testing: (i) patients that had responded to TNFi therapy (n=8), and (ii) patients that had already lost response to TNFi therapy (n=6). Most algorithms (n=6) provided recommendations for testing in both eligible populations. Table 5.3 reports the recommended prescribing decisions for each test result by eligible population, as reported explicitly by each published algorithm.

5.3.4.1. Testing RA Patients that had Responded to TNFi Therapy

Testing patients that had respond to TNFi therapy could be used to monitor treatment and inform a subsequent prescribing decision. The strategies recommended by the algorithms when patients respond to their TNFi are reported in Table 5.3i. For this population, five algorithms tested drug levels only (Mulleman et al., 2009; Mulleman et al., 2012; Rosas et al., 2014; Chen et al., 2015b; Krieckaert et al., 2015) and three algorithms tested both ADAb and drug levels (Bendtzen, 2011; Garcês et al., 2014; Mok et al., 2016).

Table 5.3. *Strategies recommended by TNFi ADAb and drug level prescribing algorithms for patients who had (i) responded and (ii) lost response.*

	(i) Testing when Patients had Responded to a TNFi.		(ii) Testing when Patients had Lost Response to a TNFi.	
Test Result	**Author**	**Reported Recommendation**	**Author**	**Reported Recommendation**
	Drug Level Testing Only		***Drug Level Testing Only***	
High Drug Levels	Bendtzen. (2011).	Reduce intensity.	Bendtzen. (2011);	Switch to non-TNFi bDMARD.
	Chen et al. (2015).	Lower dose/frequency.	Daïen et al. (2012);	
	Garcês et al. (2014).	Eventually decrease dose.	Mulleman et al. (2012).	
	Krieckaert et al. (2015).	Prolong interval.		
	Mulleman et al. (2012).	Reduce dose.	Chen et al. (2015).	Switch to bDMARD with a different mechanism of action.
	Rosas et al. (2014).	Increase Interval between doses.	Krieckaert et al. (2015).	Switch to rituximab.
			Mulleman et al. (2009).	Switch to a different bDMARD.
Normal Drug Levels.	Mulleman et al. (2009); Mulleman et al. (2012); Krieckaert et al. (2015).	Continue treatment.	Mulleman et al. (2009); Mulleman et al. (2012).	Consider dose increase.
	Rosas et al. (2014).	Continue treatment and testing.		
Low Drug Levels.	Krieckaert et al. (2015).	Stop treatment.	Daïen et al. (2012).	Increase dose or switch to a different TNFi or class of bDMARD.
			Krieckaert et al. (2015).	Switch to etanercept.
	Mulleman et al. (2009); Mulleman et al. (2012).	Continue treatment.		
			Mulleman et al. (2009); Mullemen et al. (2012).	Increase dose.
	Drug Level and ADAb Testing		***Drug Level and ADAb Testing***	
Normal Drug Levels; Positive ADAb.	Mok et al. (2016).	Continue treatment & repeat test.	Mok et al. (2016).	Switch to non-TNFi bDMARD.
Low Drug Levels; Positive ADAb.	Bendtzen. (2011).	Pause treatment.	Garcês et al. (2014); Chen et al. (2015).	Switch to bDMARD with less immunogenicity.
	Garcês et al. (2014).	Consider stopping treatment.		
			Mok et al. (2016).	Switch to a different TNFi or non-TNFi bDMARD.
			Rosas et al. (2014).	Switch to etanercept.
Low Drug Levels; Negative ADAb.	Bendtzen. (2011).	Continue treatment.	Bendtzen. (2011).	Increase treatment intensity.
	Garcês et al. (2014).	Consider stopping treatment.	Chen et al. (2015).	Consider dose adjustment.
			Garcês et al. (2014).	Repeat immunogenicity test.
			Mok et al. (2016).	Switch to different TNFi or increase dose.
			Rosas et al. (2014).	Switch therapeutic target.

Note: Recommendations are reported as-written within each manuscript. *Switch* refers to a change of treatment; bDMARD = biologic disease-modifying antirheumatic drug; TNFi = tumour necrosis factor-α inhibitor.

There was a consensus between the algorithms to reduce the intensity of treatment when high TNFi drug levels were detected. However, the algorithms did not specify precisely when the intensity of treatment should be reduced following testing. Furthermore, the specific phrases to describe *reduced intensity* were not standardised between the algorithms; for example, Chen et al. (2015b) referred to lowering the dose whereas Rosas et al. (2014) referred to increasing the interval between doses.

The algorithms were also inconsistent with respect to prescribing decisions when low drug levels were detected; the addition of ADAb testing did not resolve these inconsistencies. Two algorithms recommended that treatment should be stopped when low drug levels were detected (Garcês et al., 2014; Krieckaert et al., 2015) whereas three algorithms recommended that treatment should be continued (Mulleman et al., 2009; Bendtzen, 2011; Mulleman et al., 2012). None of the identified algorithms recommended that testing should be used to pre-emptively change the treatment of patients with detectable ADAb (who were likely to experience secondary non-response).

5.3.4.2. Testing RA Patients that had Lost Response to TNFi Therapy

Testing patients that had lost response to a TNFi therapy may be used to inform a rheumatologist's next prescribing decision. Table 5.3ii reports the strategies recommended by the identified prescribing algorithms when patients had lost response to their TNFi.

The algorithms were consistent in recommending that treatment should be changed when high drug levels were detected after the failure of a TNFi, but imprecise with respect to what this treatment should be. For example, three algorithms recommended that a non-TNFi bDMARD should be prescribed (Bendtzen, 2011; Daïen et al., 2012; Mulleman et al., 2012), which could have referred to either rituximab, abatacept, or tocilizumab; Krieckaert et al. (2015) was the only algorithm that had recommended a specific non-TNFi treatment (rituximab) in this scenario.

The recommendations for patients that had low drug levels after losing response to their TNFi were conflicting. Daïen et al. (2012) recommended increasing the TNFi dose or changing treatment to any different bDMARD (TNFi or otherwise), Mulleman et al. (2009) recommended increasing the TNFi dose, and Krieckaert et al. (2015) recommended changing treatment to etanercept (a second TNFi).

Recommendations were also inconsistent after TNFi failure with detectable ADAb. In this scenario, Rosas et al. (2014) specifically recommended prescribing etanercept. The other algorithms recommended prescribing (i) a bDMARD with less immunogenicity which, as written, may or may not have been a second TNFi (Garcês et al., 2014; Chen et al., 2015b), (ii) a different non-specific TNFi (Bendtzen, 2011), or (iii) any other bDMARD (Mok et al., 2016).

There were also inconsistent recommendations following TNFi failure with low drug levels and undetectable ADAb. Repeat testing was recommended by one algorithm (Garcês et al., 2014) and three algorithms provided recommendations for dose-adjustment (Bendtzen, 2011; Chen et al., 2015b; Mok et al., 2016). Of the two algorithms that recommended changing treatment with low drug-levels and undetectable ADAb, one recommended using a different TNFi (Mok et al., 2016) whereas another recommended using a bDMARD with a different therapeutic target (Rosas et al., 2014).

5.3.5. Discussion

The aim of this study was to identify how ADAb and drug level testing may be used to stratify treatment for patients with RA that received a TNFi therapy in routine practice. A systematic review identified nine distinct RA-specific prescribing algorithms that provided recommendations for the timing of TNFi ADAb and drug level testing and the appropriate prescribing decision to take.

The algorithms lacked clarity with regards to certain recommendations which, at face value, may impede their ability inform prescribing decisions in routine practice consistently. For example, (i) the appropriate timing and frequency of testing was inconsistent between algorithms and (ii) the algorithms that recommended strategies to reduce treatment intensity failed to report when exactly such actions should occur. The appropriate timing of testing and dose reduction strategies, however, could be investigated further in a model-based economic evaluation, by estimating their impact on the relative cost-effectiveness of treatment stratification (Shabaruddin et al., 2015).

Another potential deficiency of the algorithms was that all omitted recommendations for when a prescribing decision may cause an adverse event. For example, a proportion of patients who undergo a reduction in TNFi intensity may subsequently experience a flare in disease activity (Fautrel et al., 2015). Disease flares are adverse events that may result in

an increased cost of care and a reduction in health benefit (National Institute for Health and Care Excellence, 2009; Bykerk et al., 2016). However no prescribing algorithm provided recommendations for managing patients that flared following a decision to reduce or stop TNFi therapy.

Two algorithms explicitly recommended prescribing a second TNFi if ADAb were detected after failure of an earlier TNFi (Rosas et al., 2014; Mok et al., 2016). Clinical justification for this recommendation was based on previous studies that concluded patients may respond better to a second TNFi if they had developed ADAb against an earlier TNFi, compared with those patients who were ADAb-negative (Jamnitski et al., 2011). However, NICE have determined that sequential TNFi therapy for patients with RA was unlikely to be a relatively cost-effective use of health care resources in England (National Institute for Health and Care Excellence, 2010). A comparison with more relevance to English clinical practice was whether patients with ADAb against an earlier TNFi obtained a greater response to a second TNFi compared with rituximab (the next appropriate treatment in the care pathway for RA in England (National Institute for Health and Care Excellence, 2016a)); otherwise, sequential TNFi therapy in patients with ADAb against an earlier TNFi was also unlikely to be relatively cost-effective, *ceteris paribus*.

Previous studies have demonstrated that patients with RA who developed ADAb against a TNFi were more likely to experience secondary non-response (Garcês et al., 2013). ADAb status could therefore be used as a predictive biomarker of treatment response and be potentially valuable to implement a stratified approach to treatment. Patients at-risk of TNFi failure may be more likely to respond to a bDMARD with a different mechanism of action (Tak, 2012) and may therefore benefit from an earlier change in treatment. None of the algorithms, however, recommended to change the treatment of patients that were responding, irrespective of ADAb status.

Limitations

One potential limitation of this systematic review was that the inclusion criteria did not consider the prescribing recommendations embedded within the text of individual studies. However, it was assumed that such recommendations were supplementary and unlikely to be derived from a primary study objective. Therefore, focussing on the recommendations within published prescribing algorithms was unlikely to bias the findings of this review.

A general limitation of the algorithms in this review and, more broadly, of reviewing prescribing algorithms to identify the role of testing within a care pathway was that the clinical evidence to support the appropriate prescribing recommendations was limited, consistent with the conceptual framework for an early-stage economic evaluation of a new health technology (Figure 1.1). Given this limited clinical evidence base, prescribing algorithms published at a later date may have been informed by the recommendations of algorithms published earlier which, in turn, may be problematic if the recommendations that persisted over time were founded on relatively poor clinical evidence. Moreover, the authors that publish recommendations for a new health technology early within its product lifecycle may, themselves, have a vested interest in advancing that health technology into routine clinical practice. Therefore, an appropriate caveat for this study, and for subsequent reviews of prescribing algorithms, was that the set of strategies identified may not be sufficient to inform how a health technology could be used within an existing care pathway.

Implications for Future Research

The results of this systematic review had four implications for further research that were addressed within this thesis. These implications are now summarised:

(i) The decision problem of an economic evaluation must include all relevant comparator strategies (Drummond et al., 2015). This systematic review highlighted one of the challenges associated with estimating the relative cost-effectiveness of a stratified medicine early in the product lifecycle of a new biomarker test; without an established evidence base *a priori*, there may potentially be many ways to stratify treatment according to the result of a new test (Buisman et al., 2016). The feasibility of including every possible comparator strategy in an economic evaluation may reduce as the number of ways to stratify a treatment decision increases (Owens et al., 2017). Therefore, an implication for future research, addressed by this thesis, was to develop an early economic evaluation conceptualisation technique for stratified medicine that identified a subset of comparator strategies most relevant for inclusion in a decision problem (see Section 5.4);

(ii) A problem-oriented conceptual model could be developed to adapt the prescribing decisions recommended by the algorithms to the clinical context for

RA in England (see Section 5.6) (Tappenden, 2014);

(iii) An early model-based economic evaluation of TNFi ADAb and drug level testing may be able to estimate the appropriate timing of events (such as (i) testing or (ii) dose-reduction strategies) that were omitted from the prescribing algorithms in the systematic review (see *Chapter Six*) (Shabaruddin et al., 2015);

(iv) This systematic review provided details of how treatment decisions could be made according to the results of a biomarker test (for TNFi ADAb and drug levels). The study constitutes a single element in the chain of evidence to estimate the relative cost-effectiveness of a stratified medicine in the absence of end-to-end evidence (see Section 1.2.4; Figure 1.2). It was also necessary to investigate the accuracy of testing and the impact of a stratified treatment decision on (health and resource) outcomes (see *Chapter Six*).

Summary of Key Findings

The early model-based economic evaluation of ADAb and drug level testing, to stratify treatment for patients with RA, required an understanding of how the biomarkers could be tested in routine practice to inform a subsequent prescribing decision. This systematic review identified nine RA-specific prescribing algorithms that included ADAb and drug level testing for patients that were treated with any TNFi therapy. The algorithms recommended inconsistent prescribing decisions to stratify treatment, which suggested that, as with most new medical tests (Buisman et al., 2016; Smith et al., 2016), there was uncertainty over how best to incorporate the testing strategies into existing care pathways for RA. Three general ways in which testing could be used were to: (i) test patients after they had lost response to a TNFi to inform their next choice of treatment; (ii) test patients during response to a TNFi to pre-emptively change their treatment (to avoid loss of response); and (iii) test patients during remission whilst receiving a TNFi to inform dose-reduction strategies.

An immediate challenge for this thesis was to identify the subset of relevant comparator strategies to be included in the decision problem of the economic evaluation. Building on the results of this systematic review, the following study presents a novel early economic evaluation conceptualisation technique that was used to select the relevant comparator strategies for the economic evaluation of TNFi ADAb and drug level testing.

5.4. Chapter Objective 2: Early Identification of Relevant Comparator Strategies to Stratify Treatment when Multiple Candidate Strategies Exist

This study presents a novel algebraic technique to facilitate the early identification of potentially relevant comparator strategies during the conceptualisation phase of a *de novo* model-based economic evaluation of a stratified medicine. The technique was subsequently applied to identify the candidate ADAb and drug level testing strategies that were relevant comparators for the decision problem in this thesis.

5.4.1. Introduction

An economic evaluation must include all relevant comparator strategies to ensure that incremental outcomes are estimated appropriately, by comparing each intervention strategy to its next-best alternative (Neyt et al., 2011; O'Mahony et al., 2015a; O'Mahony et al., 2015c). A relevant comparator can be defined, in the broadest of terms, as a strategy that has the possibility of being worthwhile (Drummond et al., 2015). New health technologies to stratify treatment (such as a test), however, may be characterised by structural uncertainty with respect to their positioning and purpose within an existing care pathway, (Shabaruddin et al., 2015; Buisman et al., 2016) (see Section 1.2.5). This structural uncertainty may be greatest during an early-stage economic evaluation when limited clinical evidence is available to inform the role of testing in routine clinical practice (see Figure 1.1, *Stage 1*). For example, the rheumatologists interviewed in *Chapter Three* expressed uncertainty regarding how TNFi ADAb and drug level testing could be used to stratify treatment decisions in current practice for RA. Moreover, the systematic review of RA-specific prescribing algorithms in Section 5.3 concluded that there were three general ways to stratify treatment by testing TNFi ADAb and drug levels (summarised in Table 5.4). By contrast, a later-stage economic evaluation may be supported by a mature evidence base, and legitimate care pathways may be observable from individual patient-level data if the health technology had diffused in routine clinical practice (see Figure 1.1, *Stage 4*).

Given the constraints that decision analysts may face (such as time, physical equipment, or technical competency), it may not be practically feasible to include every possible comparator strategy in an economic evaluation (Claxton et al., 2005; Owens et al., 2017).

It may therefore be necessary to make a decision, during the conceptualisation phase of an economic evaluation, to identify a subset of potentially relevant comparator strategies to include in a decision problem (Buisman et al., 2016). One technique that may help to identify a subset of potentially relevant comparator strategies was to conceptualise each general testing strategy, *a priori*, in terms of the incremental net benefit framework.

The incremental net benefit framework can be used in an economic evaluation to inform whether a strategy is cost-effective compared with a relevant alternative (Stinnett et al., 1998) (described in Table 1.3; Section 1.1.3). The incremental net benefit of a stratified medicine, in general, comprises the impact of a test on the incremental costs and/or health outcomes between comparator strategies (Phelps et al., 1988). Overall incremental outcomes arise from differences in the elements of resource use and health experienced between alternatives (Weinstein et al., 1977); in the specific case of testing strategies to stratify treatment, differences in incremental outcomes typically occur over time (after testing) along a care pathway (Mushlin, 1999). For example, one element that may have affected the overall incremental health outcomes of testing adalimumab drug levels to inform dose-reduction strategies was the downstream proportion of patients that subsequently flared in disease activity.

It may be possible, based on existing evidence, to describe a hypothetical profile of costs and health outcomes over time associated with each potential comparator test-and-treatment strategy. By formalising these profiles using algebraic notation, the incremental net benefit framework, and comparative statics analysis, it may be possible to identify (i) potentially *relevant* comparator strategies to include in the decision problem of an economic evaluation, and (ii) the anticipated effect of changing the value of an input parameter on the incremental net benefit of each strategy, before implementing a final quantitative decision analytic model.

5.4.2. Aim and Objectives

The aim of this study was to identify the subset of relevant comparator testing strategies to include in the decision problem of the economic evaluation presented in this thesis. The study had three objectives:

Objective 1: Describe the candidate strategies to stratify treatment decisions using ADAb and drug level testing for patients with RA;

Objective 2: Determine whether each candidate strategy was a plausible relevant comparator;

Objective 3: Investigate how the relative cost-effectiveness of each plausible relevant comparator may be affected by exogenous changes in the values of input parameters.

5.4.3. Method

The method that underpinned this novel conceptualisation technique may be regarded as an approach to describe the introduction of a new health technology within an existing care pathway during an early economic evaluation (for example, Stage 1 in Figure 1.1). The technique utilised algebraic notation to conduct a theoretical incremental cost-effectiveness analysis of alternative strategies by using the incremental net benefit framework. Justifiable assumptions were made regarding each strategy's anticipated profile of cost and QALYs over time, and the incremental net benefit framework provided a transparent rationale for excluding specific strategies from the final decision problem. The following explains the method in five steps by applying the technique to the adalimumab ADAb and drug level testing case study.

Step 1: Study Perspective

The perspective of an economic evaluation defines the scope of relevant costs that should be included within the analysis (Drummond et al., 2015). The perspective assumed by the conceptualisation technique should be consistent with the perspective of the final decision problem. The decision problem for the economic evaluation in this thesis followed the NICE Reference Case (Section 1.1.2); therefore, the appropriate perspective for this study was the NHS and personal social services budget (National Institute for Health and Care Excellence, 2013a).

Step 2: Identification of Candidate Strategies

Candidate strategies for inclusion in a decision problem should be justified according to existing *a priori* evidence. Testing strategies for stratified medicine were described in general terms (for example, 'test patients that responded'); the specific features of each candidate strategy (for example, the frequency of testing) could be determined at a later

stage during a full model-based economic evaluation (Shabaruddin et al., 2015). This study used the systematic review of RA-specific prescribing algorithms in Section 5.3 as the source of evidence to identify candidate strategies for ADAb and drug level testing.

Step 3: Characterisation of Candidate Strategies

Existing clinical evidence was used to characterise the candidate strategies by their key elements. Key elements were features that were anticipated to affect the cost and/or QALYs derived from each strategy; for example, the cost of testing, the cost of treatments, and the proportion of patients with ADAb. Each element was defined using its own notation to facilitate algebraic manipulation. The ordinal relationships between key elements were specified, where necessary (for example, the cost of full-dose therapy was assumed to be greater than the cost of reduced-dose therapy).

The anticipated cost and QALY profiles of each candidate strategy were illustrated on separate graphs that plot time (X-axis) against cost or QALYs (Y-axis) gained. Time was divided into four distinct periods as reference points for the sequence of clinical events that were assumed to occur over time. A comparator strategy (current practice) was plot on the same graph to illustrate the cost and QALY profiles that were assumed without treatment stratification.

Step 4: Identification of Plausible Relevant Comparator Strategies

The total cost (TC_i) and total QALYs (TQ_i) associated with each candidate strategy (i), and with current practice (TC_{CP}, TQ_{CP}), were estimated algebraically by calculating the area under the time profiles. The incremental cost and QALY of each candidate strategy (i), relative to current practice, was calculated using the formulas in Equation 5.1 and Equation 5.2, respectively, by subtracting the area for current practice from the area for the candidate strategy.

$$\text{Incremental Cost}_i = TC_i - TC_{CP} \quad \textbf{(Equation 5.1)}$$

$$\text{Incremental QALY}_i = TQ_i - TQ_{CP} \quad \textbf{(Equation 5.2)}$$

The incremental net monetary benefit (INMB) of each strategy was calculated algebraically using the formula in Equation 5.3, based on the decision rules reported in Table 1.3

$$\text{INMB}_i = (\lambda \times \text{Incremental QALY}_i) - \text{Incremental Cost}_i \quad \textbf{(Equation 5.3)}$$

where (λ) was the notation for the (non-zero) cost-effectiveness threshold. A non-negative INMB indicated that the strategy was potentially cost-effective; such strategies were considered to be plausible relevant comparators for the final decision problem.

Step 5: Comparative Statics Analysis

Comparative statics analysis is a technique used in mathematical economics to formalise how the value of an objective function may change (increase or decrease) with respect to exogenous changes in its constituent elements (Silberberg, 1974; Currier, 2000; Gravelle et al., 2004). The algebraic INMB for each plausible relevant comparator was differentiated with respect to each key element of the function. The sign of each first-derivative indicated the anticipated effect that an exogenous change in a key element would have on the relative cost-effectiveness of a plausible relevant comparator, *ceteris paribus*. For example, a negative first-derivative indicated that an increase in the value of the key element (for example, the cost of testing) would lead to a reduction in the INMB. This comparative statics analysis was conceptually equivalent to performing a one-way sensitivity analysis in a full model-based cost-effectiveness analysis (Briggs et al., 1999).

5.4.4. Results

The systematic review in Section 5.3 concluded that ADAb and drug level testing could be used to stratify treatment for patients with RA in three general ways. These candidate strategies are described in Table 5.4 alongside their key elements that were anticipated to affect costs and QALYs. The algebraic notation that was used to define the key elements of each strategy is reported in Table 5.5.

The cost and QALY profiles that characterised *Strategies A*, *B*, and *C* are illustrated in Figures 5.3, 5.4, and 5.5, respectively. The solid lines represented the anticipated outcomes derived from current practice and the dashed lines represented the anticipated outcomes derived from the candidate comparator strategy. Lines were drawn in parallel, for the purpose of exposition, if they followed the same cost or QALY profile over time. The

incremental outcomes for each candidate strategy are described in the following three sections. A full derivation of the algebraic solutions is reported in *Appendix 24*.

Table 5.4. *Three candidate strategies for TNFi ADAb and drug level testing.*

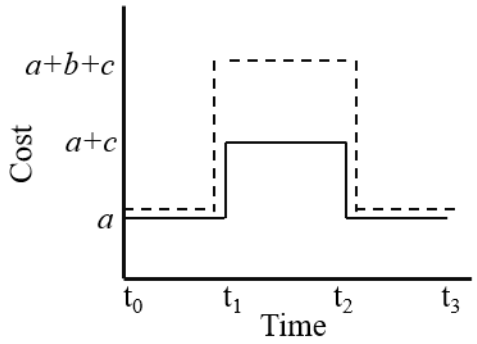
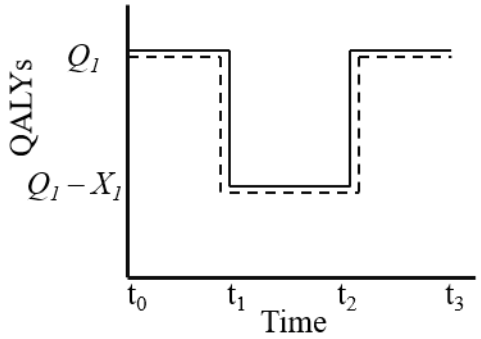
Strategy	Action	Description	Key Elements
Strategy A.	Test ADAb and drug levels after loss of response to TNFi.	Test TNFi ADAb and drug levels in patients that had lost response to inform a subsequent bDMARD prescribing decision.	• Cost of test; • Cost of treatments; • QALYs lost from treatment failure.
Strategy B.	Test ADAb and drug levels while responding to TNFi.	Test TNFi ADAb and drug levels to monitor treatment in patients that were responding. Treatment may be changed for patients that developed ADAb against their TNFi to avoid the harm associated with earlier secondary non-response.	• Cost of test; • Cost of treatments; • QALYs lost from treatment failure; • Proportion of ADAb-positive patients.
Strategy C.	Test drug levels in remission.	Test the drug levels of patients in remission to inform whether TNFi dose-reduction strategies were possible.	• Cost of test; • Cost of treatments; • QALYs lost from disease flare; • Proportion of patients with high drug levels; • Proportion of patients that flare.

Table 5.5. *Algebraic notation for the key elements of each strategy.*

Notation	Description
λ	Cost-effectiveness threshold.
$t_i, i = \{0, \dots, 3\}^{\dagger}$	Four time periods, where $t_0 < t_1 < t_2 < t_3$.
a	Unit cost of biologic therapy.
b	Unit cost of testing.
c	Unit cost of treatment for a patient that experienced secondary non-response.
y	Reduction in unit cost of TNFi therapy following dose reduction.
P	Proportion of patients $[0,1]^{\dagger}$ that developed ADAb.
d	Proportion of patients $[0,1]^{\dagger}$ that had high TNFi drug-levels in remission.
q	Proportion of patients $[0,1]^{\dagger}$ that flared after TNFi dose reduction.
Q_1	QALY gained for successful response to TNFi therapy.
Q_2	QALY gained for successful TNFi therapy in remission.
X_1	QALY loss associated with secondary non-response/adverse events.
X_2	QALY loss associated with a flare.

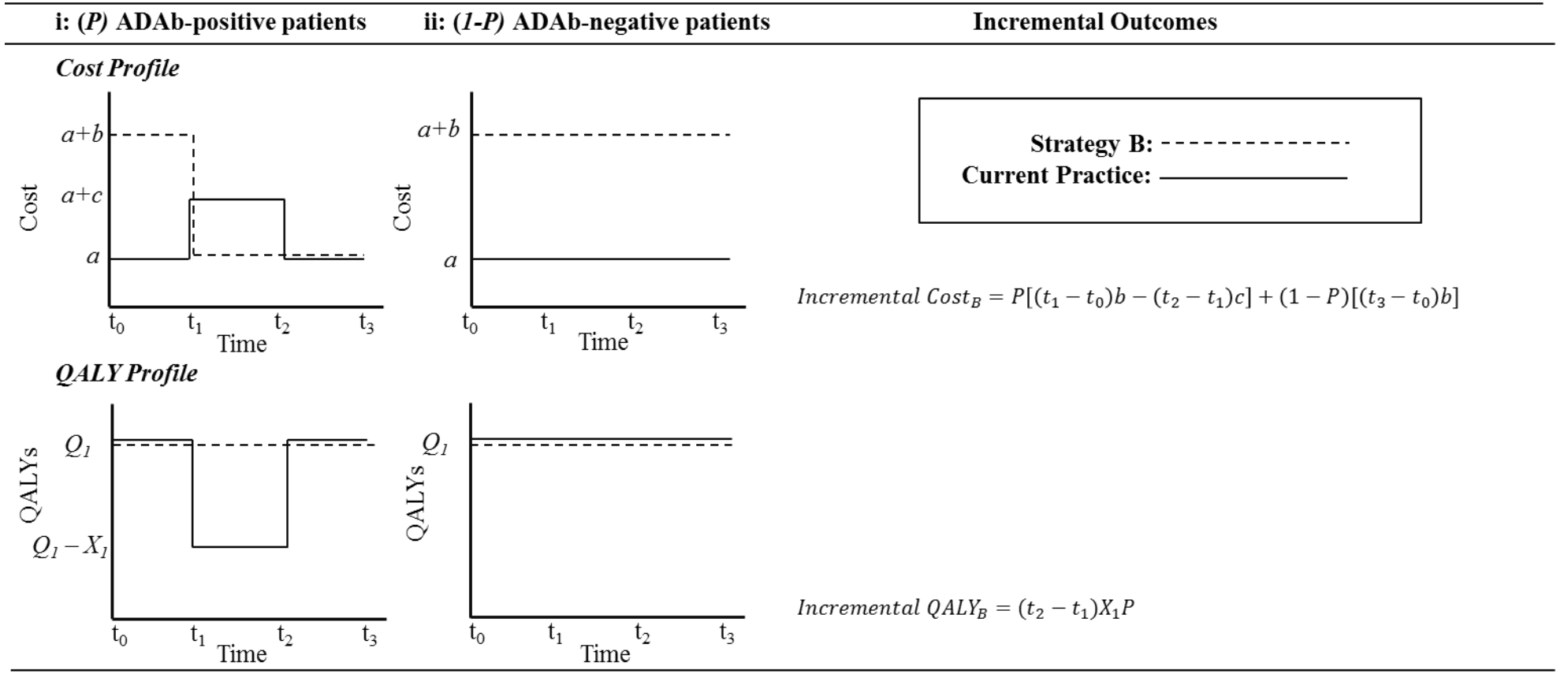
Note: †=in set theory notation, [0,1] indicates that a variable, x, can take any value within the interval ($0 \leq x \leq 1$); the notation {a,b} indicates that a variable, i, can take the value of an element specified within the set (either i=a or i=b).

Figure 5.3. *Conceptual cost and QALY profile for Strategy A.*

i: (*P*) ADAb-positive patients	ii: (*1-P*) ADAb-negative patients	Incremental Outcomes
Cost Profile 	Not applicable	**Strategy A**: - - - - - - - - - - - - - **Current Practice:** ——————— $Incremental\ Cost_A = (t_2 - t_1)bP$
QALY Profile 	Not applicable	$Incremental\ QALY_A = 0$
$Incremental\ Net\ Monetary\ Benefit_A = -(t_2 - t_1)bP$		

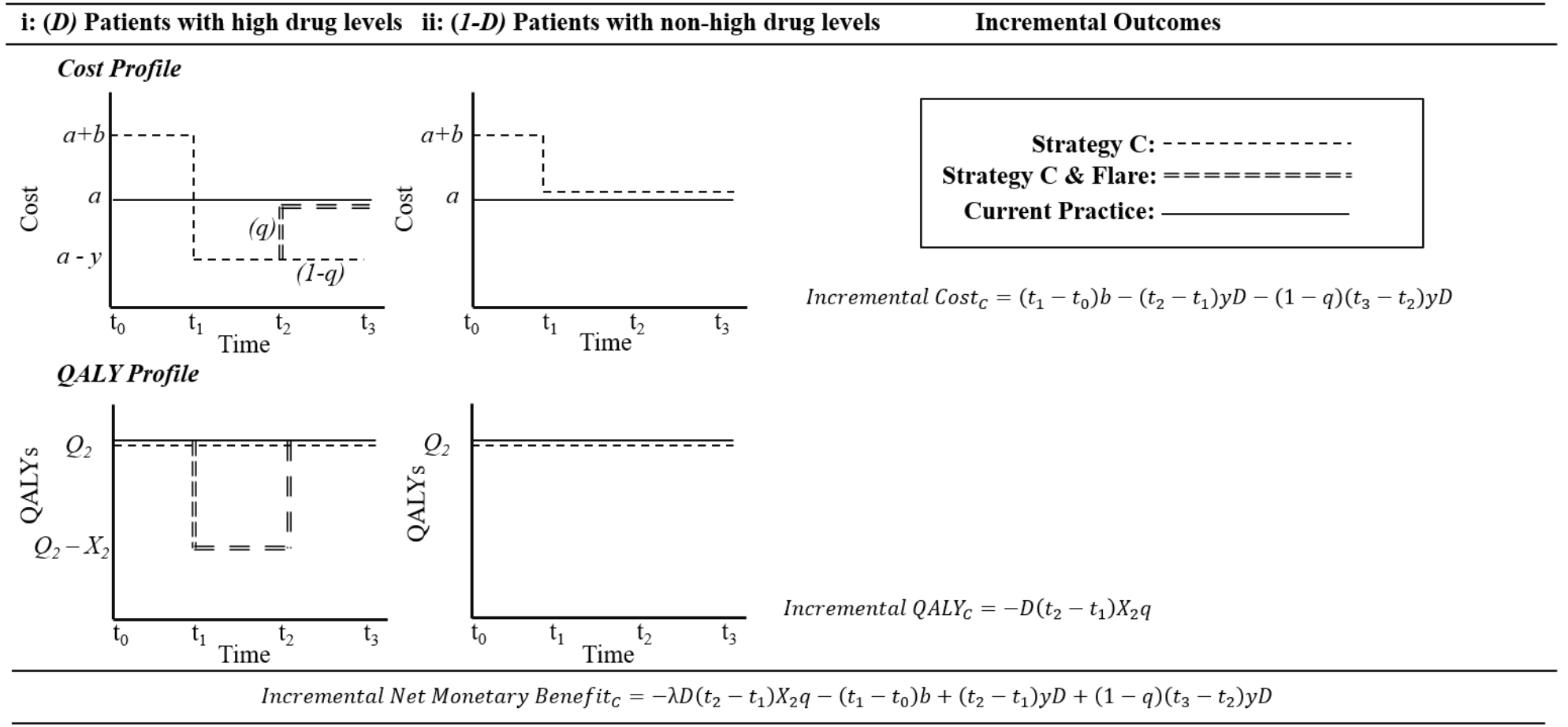
Note: λ=cost-effectiveness threshold; ti=time period; a=cost of biologic treatment; b=cost of test; c=treatment cost for secondary non-response; y=cost reduction following TNFi dose de-escalation; P=proportion of ADAb-positive patients; D=proportion of patients with high TNFi drug levels in remission; Q1=QALY gain from response to TNFi; Q2=QALY gain from response to TNFi in remission; X1=QALY loss due to secondary non-response; X2=QALY loss due to flare.

Figure 5.4. *Conceptual cost and QALY profile for Strategy B.*



Note: λ=cost-effectiveness threshold; ti=time period; a=cost of biologic treatment; b=cost of test; c=treatment cost for secondary non-response; y=cost reduction following TNFi dose de-escalation; P=proportion of ADAb-positive patients; D=proportion of patients with high TNFi drug levels in remission; Q1=QALY gain from response to TNFi; Q2=QALY gain from response to TNFi in remission; X1=QALY loss due to secondary non-response; X2=QALY loss due to flare.

Figure 5.5. *Conceptual cost and QALY profile for Strategy C.*



Note: λ=cost-effectiveness threshold; ti=time period; a=cost of biologic treatment; b=cost of test; c=treatment cost for secondary non-response; y=cost reduction following TNFi dose de-escalation; P=proportion of ADAb-positive patients; D=proportion of patients with high TNFi drug levels in remission; Q1=QALY gain from response to TNFi; Q2=QALY gain from response to TNFi in remission; X1=QALY loss due to secondary non-response; X2=QALY loss due to flare.

5.4.4.1. Strategy A

- **Clinical scenario**: Patients were assumed to have commenced a TNFi therapy at t_0. A proportion of those patients (P) were assumed to have developed ADAb at a time between t_1 and t_0. The ADAb-positive patients lost response to treatment at t_1 which, in turn, was associated with a reduction in QALYs and an increased cost of care.
- **Current Practice:** Patients that lost response at t_1 had their treatment changed to a different bDMARD at t_2.
- **Strategy A**: Patients that lost response at t_1 had their drug levels tested at t_1 to inform the change of treatment at t_2.

Incremental Costs

The top-half of Figure 5.3 is a graphical illustration of the incremental cost of *Strategy A* compared with current practice. All bDMARD therapies were assumed to cost (a).

- **Current Practice**: The cost of treatment was assumed to equal (a) units and was escalated by (c) units for the period of time that the patient had lost response $(t_2 - t_1)$.
- **Strategy A**: The cost of treatment was assumed to equal (a) units and was escalated by $(b + c)$ units for the period of time that the patient had lost response $(t_2 - t_1)$ due to the additional cost of testing (b units).

The incremental cost of *Strategy A* (subtracting the area under the solid line from the area under the dashed line) was equal to the value in Equation 5.4 (full derivation in *Appendix 24*):

$$\text{Incremental Cost}_{\text{A}} = (t_2 - t_1)bP \qquad \textbf{(Equation 5.4)}$$

The incremental cost of *Strategy A* was always positive, and depended on the cost of testing (b), the proportion of patients that developed ADAb (P), and the time between TNFi failure and commencing a different bDMARD $(t_2 - t_1)$.

Incremental QALYs

The bottom-half of Figure 5.3 is a graphical illustration of the incremental QALYs of *Strategy A* compared with current practice. All bDMARD therapies were assumed to be equally effective.

- **Current Practice**: The (P) ADAb-positive patients gained (Q_1) QALYs while responding and lost (X_1) QALYs for the duration of treatment failure $(t_2 - t_1)$.
- **Strategy A**: Given that bDMARD therapies were assumed to be equally effective, the (P) ADAb-positive patients also gained (Q_1) QALYs while responding and lost (X_1) QALYs for the duration of treatment failure $(t_2 - t_1)$.

Under these assumptions, patients derived no incremental benefit from *Strategy A* compared with current practice (Equation 5.5).

$$\text{Incremental QALY}_{\text{A}} = 0 \qquad \textbf{(Equation 5.5)}$$

Incremental Net Monetary Benefit

The INMB for *Strategy A* (Equation 5.6) was calculated by substituting (Equation 5.4) and (Equation 5.5) into the formula in Equation 5.3 (full derivation in *Appendix 24*).

$$\text{INMB}_{\text{A}} = -Pb(t_2 - t_1) \qquad \textbf{(Equation 5.6)}$$

The INMB for *Strategy A* was negative. Given that testing imposed positive incremental costs $Pb(t_2 - t_1)$, *Strategy A* must have subsequently provided a QALY-gain in order to have potentially demonstrated a positive INMB. Therefore, *Strategy A* was not considered to be a potentially relevant comparator strategy for the full cost-effectiveness analysis in this thesis.

5.4.4.2. Strategy B

- **Clinical scenario**: Patients were assumed to have commenced a TNFi therapy at t_0. A proportion of those patients (P) were assumed to have developed ADAb at a time between t_1 and t_0.

- **Current Practice**: The (P) patients that developed ADAb experienced treatment failure at (t_1) and had their treatment changed to a different bDMARD at (t_2). The $(1-P)$ ADAb-negative patients maintained response to their TNFi therapy.
- **Strategy B**: All patients experienced routine testing from (t_0) to monitor for the development of ADAb. Testing was assumed to be perfectly accurate. The (P) patients that developed ADAb had their treatment changed pre-emptively to a second bDMARD at (t_1) because testing detected the presence of ADAb.

Incremental Costs

The top-half of Figure 5.4 is a graphical illustration of the costs profile for current practice and *Strategy B* over time. The outcomes of the (P) ADAb-positive and $(1-P)$ ADAb-negative patients are on the left and right of Figure 5.4, respectively.

- **Current Practice**: The cost of treatment was assumed to equal (a) units and was escalated by (c) units for the period of time that the (P) ADAb-positive patients had lost response $(t_2 - t_1)$.
- **Strategy B**: The cost of care for the $(1-P)$ ADAb-negative patients was assumed to encompasses the cost of testing (b) and treatment (a) for all time periods (Figure 5.4ii). The cost of care for the (P) ADAb-positive patients encompassed the cost of testing (b) and treatment (a) for the initial time period $(t_1 - t_0)$, and was reduced to (a) units (no further testing) after changing treatment to a different bDMARD at (t_1).

The incremental cost of *Strategy B* (subtracting the area under the solid line from the area under the dashed line) was equal to the value in Equation 5.7 (full derivation in *Appendix 24*):

$$\text{Incremental Cost}_{\text{B}} = P[\big((t_1 - t_0)b - (t_2 - t_1)c\big] + (1-P)[(t_3 - t_0)b] \quad \textbf{(Equation 5.7)}$$

The incremental cost of *Strategy B* could be positive or negative depending on the unit cost of testing (b), the cost of treatment following loss of response (c), and the proportion of ADAb-positive patients (P). The cost of *Strategy B* was (b) units higher than current practice for the $(1-P)$ ADAb-negative patients for all time periods $(t_3 - t_0)$ (see Figure 5.4ii). For the (P) ADAb-positive patients (Figure 5.4i), the initial cost of *Strategy B* was (b) units higher than current practice, and (c) units lower than current practice after pre-emptively changing treatment.

Incremental QALYs

The bottom-half of Figure 5.4 is a graphical illustration of the QALY profile for current practice and *Strategy B* over time. All bDMARD therapies were assumed to be equally effective.

- **Current Practice**: Patients that responded to treatment were assumed to gain (Q_1) QALYs. The (P) ADAb-positive patients experienced a loss of response at (t_1), which reduced QALYs by (X_1) until the second-line bDMARD treatment $(t_2 - t_1)$, when QALYs were assumed to revert to their original level (Q_1) (see Figure 5.4i).
- **Strategy B**: The (P) patients that developed ADAb maintained the (Q_1) QALY gain by pre-emptively changing treatment at (t_1) to avoid the loss of response.

The incremental QALY of *Strategy B* (subtracting the area under the solid line from the area under the dashed line) was equal to the value in Equation 5.8 (full derivation in *Appendix 24*):

$$\text{Incremental QALY}_{\text{B}} = (t_2 - t_1)X_1P \quad \textbf{(Equation 5.8)}$$

The incremental QALY of *Strategy B* was always positive (due to avoiding the QALY-loss in current practice), and depended on the proportion of patients that developed ADAb (P), the magnitude of the QALY-loss associated with treatment failure (X_1), and the duration of treatment failure $(t_2 - t_1)$.

Incremental Net Monetary Benefit

The INMB for *Strategy B* (Equation 5.9) was calculated by substituting (Equation 5.7) and (Equation 5.8) into the formula in Equation 5.3 (full derivation in *Appendix 24*).

$$\text{INMB}_{\text{B}} = \lambda(t_2 - t_1)X_1P - (t_1 - t_0)bP + (t_2 - t_1)cP - (t_3 - t_0)b + (t_3 - t_0)bP \quad \textbf{(Equation 5.9)}$$

It was not possible to determine whether the incremental net monetary benefit of *Strategy B*, compared with current practice, was positive or negative based on the algebraic analysis alone. The sign of the incremental net monetary benefit depended on whether (i) the monetary value of the incremental QALY gains ($\lambda(t_2 - t_1)X_1P$) *and* (ii) the cost-reduction

associated with avoiding the need to treat patients that had lost response ($(t_2 - t_1)cP$) were greater than the additional costs imposed by testing ($-(t_1 - t_0)bP - (t_3 - t_0)b + (t_3 - t_0)bP$). Therefore, *Strategy B* was a potentially relevant comparator strategy to be included in the final decision problem.

Table 5.6 reports the sign of the first-derivative of Equation 5.9 with respect to five parameters (full derivation in *Appendix 24*). The positive first-derivative for three parameters (cost of treating loss of response; proportion of ADAb-positive patients; the QALY loss associated with treatment failure) indicated that the relative cost-effectiveness of *Strategy B* may increase if the magnitude of these parameters were to increase, *ceteris paribus*. The two parameters with a negative first-derivative (cost of testing; time taken to develop ADAb) indicated that the relative cost-effectiveness of *Strategy B* may reduce if the magnitude of these parameters were to increase, *ceteris paribus*.

Table 5.6. *First-derivative of INMB with respect to individual parameters: Strategy B*

Parameter		Sign of First-derivative†
Notation	**Definition**	
c	Cost of treating loss of response.	$\frac{\partial \text{INMB}}{\partial \text{C}} > 0$
b	Cost of testing.	$\frac{\partial \text{INMB}}{\partial \text{b}} < 0$
$(t_1 - t_0)$	Time taken to develop ADAb.	$\frac{\partial \text{INMB}}{\partial (\text{t}_1\text{-t}_0)} < 0$
P	Proportion of patients with ADAb.	$\frac{\partial \text{INMB}}{\partial \text{P}} > 0$
X_1	QALY loss associated with treatment failure.	$\frac{\partial \text{INMB}}{\partial \text{X}_1} > 0$

Note: †=full algebraic derivation of first-derivative is provided in *Appendix 24*.

5.4.4.3. Strategy C

- **Clinical scenario**: All patients were assumed to be in remission at (t_0) and were receiving full-dose TNFi therapy. A proportion of those patients (D) were assumed to have high drug levels by (t_1).
- **Current Practice**: Patients maintained full-dose TNFi therapy for the full duration of the analysis.

- **Strategy C**: All patients had their drug levels tested between (t_0) and (t_1). The intensity of treatment was reduced if high drug levels were detected at (t_1). A proportion (q) of patients that received reduced-dose TNFi were assumed to flare in disease activity at (t_2), which prompted treatment to revert to its original dose.

Incremental Costs

The top-half of Figure 5.5 is a graphical illustration of the costs profile for current practice and *Strategy C* over time.

- **Current Practice**: The cost of care encompassed the cost of treatment (a) for all time periods.
- **Strategy C**: The initial $(t_1 - t_0)$ cost of care for all patients encompassed the cost of treatment (a) and testing (b). At (t_1), the cost of treatment was reduced by (y) units for the (D) patients with high drug levels. Patients could follow one of two mutually exclusive pathways at (t_2) depending on whether they flared after their TNFi dose had been reduced (Figure 5.5i). Flaring patients (q) had their TNFi dose-reduction reverted to its original value (a) (bold dashed pathway). Non-flaring patients $(1 - q)$ continued to receive reduced-dose TNFi treatment.

The incremental cost of *Strategy C* (subtracting the area under the solid line from the area under the dashed line) was equal to the value in Equation 5.10 (full derivation in *Appendix 24*):

$$\text{Incremental Cost}_C = (t_1 - t_0)b - (t_2 - t_1)yD - (1 - q)(t_3 - t_2)yD \quad \textbf{(Equation 5.10)}$$

The incremental cost of *Strategy C* may be positive or negative depending on the prevalence of high drug levels (D), the cost-reduction associated with reduced-dose TNFi treatment (y), and the proportion of flares (q).

Incremental QALYs

The bottom-half of Figure 5.5 is a graphical illustration of the QALY profile for current practice and *Strategy C* over time.

- **Current Practice**: The patients with (D) and without $(1 - D)$ high drug levels gained same amount of QALYs (Q_2) (because they didn't flare).

- **Strategy C**: Of the (D) patients that experienced reduced-dose TNFi therapy under *Strategy C* (Figure 5.5i), (i) those that flared (q) experienced a QALY-loss (X_2) for the duration of the flare $(t_2 - t_1)$; (ii) those that did not flare obtained the same QALYs (Q_2) as those that received full-dose TNFi. QALYs were assumed to return to their initial value (Q_2) in patients that flared, when their dose of TNFi was reverted to its initial value.

The incremental QALYs of *Strategy C* (subtracting the area under the solid line from the area under the dashed line) was equal to the value in Equation 5.11.

$$\text{Incremental QALY}_{\text{C}} = -D(t_2 - t_1)X_2 q \quad \textbf{(Equation 5.11)}$$

The incremental QALY of *Strategy C* was negative assuming that dose-reduction strategies, based on TNFi drug-level testing in remission, only led to potential QALY-losses (from patients that flared). The magnitude of the incremental QALY-loss depended on the proportion of patients with high drug levels (D), the proportion of patients that flared (q), the duration of the flare $(t_2 - t_1)$, and the magnitude of the QALY-loss associated with a flare (X_2).

Incremental Net Monetary Benefit

The INMB for *Strategy C* (Equation 5.12) was calculated by substituting (Equation 5.10) and (Equation 5.11) into the formula in Equation 5.3 (full derivation in *Appendix 24*).

$$\text{INMB}_{\text{C}} = -\lambda D(t_2 - t_1)X_2 q - (t_1 - t_0)b + (t_2 - t_1)yD + (1 - q)(t_3 - t_2)yD \quad \textbf{(Equation 5.12)}$$

It was not possible to determine whether the INMB for *Strategy C* (Equation 5.12) was positive or negative from the algebraic analysis alone. *Strategy C* had the potential to be cost-effective, relative to current practice, if the reduction in QALYs (expressed in monetary units: $-\lambda D(t_2 - t_1)X_2 q$) and the additional cost of testing ($(t_1 - t_0)b$) were offset by the lower cost of reduced-dose TNFi therapy in patients that did not flare ($(t_2 - t_1)yD + (1 - q)(t_3 - t_2)yD$). Therefore, *Strategy C* was a potentially relevant comparator strategy to be included in the final decision problem.

Table 5.7 reports the sign of the first-derivative of Equation 5.11 with respect to five parameters (full derivation in *Appendix 24*). The positive first-derivative of the cost-

reduction associated with lower dose TNFi therapy indicated that the relative cost-effectiveness of *Strategy C* may increase if the magnitude of this cost-reduction increased, *ceteris paribus*. The negative first-derivative of four parameters (cost of testing; cost-effectiveness threshold; proportion of patients that flared; QALY-loss associated with a flare) indicated that the relative cost-effectiveness of *Strategy C* may reduce if the magnitude of these parameters were to increase, *ceteris paribus*. It was not possible to make *a priori* comparative predictions regarding one parameter (proportion of patients with high drug levels) because the direction of its first-derivative could not be determined from the algebraic analysis alone.

Table 5.7. *First-derivative of INMB with respect to individual parameters: Strategy C.*

Parameter		Sign of First-derivative†
Notation	**Definition**	
b	Cost of testing.	$\frac{\partial \text{INMB}}{\partial \text{b}} < 0$
y	Cost-reduction associated with lower dose TNFi.	$\frac{\partial \text{INMB}}{\partial \text{y}} > 0$
λ	Cost-effectiveness threshold.	$\frac{\partial \text{INMB}}{\partial \lambda} < 0$
D	Proportion of patients with high drug levels.	$\frac{\partial \text{INMB}}{\partial \text{P}} = \text{undefined}^{\blacklozenge}$
q	Proportion of patients that flared from reduced-dose TNFi.	$\frac{\partial \text{INMB}}{\partial \text{q}} < 0$
X_2	QALY-reduction associated with flare.	$\frac{\partial \text{INMB}}{\partial \text{X}_2} < 0$

Note: †=full algebraic derivation of first-derivative is provided in *Appendix 24*; ♦=not possible to determine whether the first-derivative was positive or negative.

5.4.5. Discussion

This study demonstrated a novel conceptualisation technique (using algebraic notation, the incremental net benefit framework, and comparative static analysis) to inform the choice of potentially relevant comparator strategies for the *de novo* economic evaluation of stratified medicine presented in this thesis. Two potentially relevant comparator strategies were identified from a set of three candidate strategies that described how treatment for RA could be stratified by TNFi ADAb and drug level testing. This conceptualisation technique could, in principal, help to identify the relevant comparators within any early model-based

economic evaluation of a stratified medicine, if multiple candidate strategies exist because the role of testing in a care pathway is unclear.

A fundamental component of an early economic evaluation, as posited by the conceptual framework in Section 1.1.6.4 (Figure 1; Stage 1), is to describe how a new health technology may enter an existing care pathway. This study presented an approach to formalise such a description by using algebraic manipulation to identify potentially relevant comparator strategies, in the absence of robust data to inform care pathways, resource use and health outcomes. By contrast, relevant comparator strategies are likely to be known during a later-stage economic evaluation when the clinical evidence base has matured (Figure 1.1; Stage 4). Therefore, the purpose of such an early-stage conceptualisation technique should be to produce indicative, rather than definitive, arguments of relative cost-effectiveness (Sculpher et al., 1997).

The results suggested that TNFi ADAb and drug level testing in RA had the potential to be cost-effective when used as a predictive biomarker of treatment response (*Strategy B*) and to guide dose-reduction in remission (*Strategy C*). It was notable that testing after loss of response to a TNFi (*Strategy A*) was unlikely to be cost-effective (imposing additional cost with no additional QALY gain) given that nine prescribing algorithms in Section 5.3. recommended variants of this strategy. The results for *Strategy A* were contingent on the assumption that different bDMARDS were equally effective, which can be supported by previously published meta-analyses of RCT evidence (Aaltonen et al., 2012; Stevenson et al., 2016). Additionally, no prescribing algorithm in Section 5.3 recommended a pre-emptive change of treatment, based on TNFi ADAb or drug level status, to potentially avoid a QALY loss associated with treatment failure; yet *Strategy B* in this study was potentially cost-effective when compared with current practice.

Limitations

The candidate strategies were characterised in general terms as a pragmatic decision to enable algebraic manipulation. One potential limitation of this study was that, in practice, there may have been different specific strategies that conformed to the characterisation of each general strategy. For example, the general strategy to routinely monitor patients that responded to a TNFi (*Strategy B*) did not specify the frequency of testing. It was more appropriate, however, to estimate the relative cost-effectiveness of such specific strategies

in a full model-based economic evaluation and not during a conceptualisation exercise *a priori* (Shabaruddin et al., 2015).

A potential limitation of the comparative statics analysis, after differentiating the INMB function with respect to individual parameters, was the assumption that the predictions were only valid if the values of all other parameters remained constant. A one-way sensitivity analysis within a full model-based economic evaluation would also be limited by the same assumption (Briggs et al., 1999). Practically, the internal validity of a full model-based economic evaluation could be appraised by comparing the actual results of a one-way sensitivity analysis with the direction of change predicted by the first-derivative in the conceptual algebraic analysis.

Lee et al. (2013) have used a similar algebraic technique to produce a *generalised case* model-based cost-effectiveness analysis of a companion diagnostic test. However, the approach by Lee et al. (2013) requires the input of empirical data (for example, on test accuracy or QALYs gained) and evaluates testing over a short time horizon. The advantage of the present study, therefore, was that (i) longer time horizons could be accounted for by considering the potential profile of costs and QALYs over time, and (ii) the algebraic manipulation enabled an informative analysis to be performed during an early stage of a test's product lifecycle without the need for empirical data.

Implications for Future Research

There were two possible ways that this early economic evaluation conceptualisation technique could be extended. Firstly, a fully incremental analysis could be performed using the algebraic notation, by comparing all strategies with each other, without loss of generality. This study used a common comparator (current practice) to illustrate the analytic concept; a fully-incremental analysis was not necessary to inform the relevant comparator strategies for the decision problem reported this thesis. Secondly, future research could extend the conceptualisation technique, contingent on the decision-maker's required perspective, to describe the time profile of other potentially relevant outcomes that were beyond the scope of this thesis (such as patient out-of-pocket expenditures or the QALY-loss borne by carers).

A practical implication for this thesis was that the relative cost-effectiveness of the two general strategies with the potential to be cost-effective, identified by this study (*Strategy B*

and *Strategy C*), should be investigated further in a full model-based economic evaluation of adalimumab ADAb and drug level testing to stratify treatment.

Summary of Key Findings

This study presented an early economic evaluation conceptualisation technique after recognising that decision-analysts may have had difficulty in identifying relevant comparator strategies for a stratified medicine that incorporates a new test when: (i) there was limited evidence to support the use of testing, (ii) testing could have been performed in multiple ways, and (iii) constraints existed on the time available to conduct a full economic evaluation. The results were presented as a general case analysis by using algebraic manipulation and the incremental net benefit framework. Treatment stratification by testing adalimumab ADAb and drug levels during response, and testing drug levels during remission, both had the potential to be cost-effective, relative to current practice. In contrast, informing treatment decisions by testing after a patient had lost response was less likely to be cost-effective, relative to current practice. The following section of *Chapter Five* uses the results derived from this early conceptualisation technique to define the relevant comparator strategies in the decision problem for the full economic evaluation of treatment stratification according to adalimumab ADAb and drug level testing in patients with RA.

5.5. Chapter Objective 3: The Decision Problem

This study defines the decision problem for the early economic evaluation of adalimumab ADAb and drug level ELISA testing to stratify treatment for patients with RA in England.

5.5.1. Introduction

The primary purpose of any economic evaluation is to inform a decision; namely, a decision to allocate population health care resources by providing sufficient evidence of the incremental costs and benefits of a health technology (Drummond et al., 2015). A decision problem must therefore be specified, before conducting any economic evaluation, that describes the specific resource allocation decision under consideration (Roberts et al., 2012; Drummond et al., 2015). For example, the decision problem within a NICE technology appraisal process is specified at the preliminary *scoping* phase (National Institute for Health and Care Excellence, 2011a; 2013a).

The proposed early model-based economic evaluation of adalimumab ADAb and drug level testing to stratify treatment for RA was designed to provide evidence for decision-makers responsible for resource allocation decisions in NHS England. It was therefore necessary for the decision problem to conform to the evidence requirements of a NICE technology appraisal. The decision problem, whilst being relevant to decision-makers, must have also been relevant to the clinical context of managing patients with RA in England. It was also necessary, therefore, to refine the comparator testing strategies identified within the literature for relevance to clinical practice in England.

5.5.2. Aim and Objective

The aim of this study was to define the decision problem for the economic evaluation of adalimumab ADAb and drug level testing to stratify treatment for patients with RA in England. There were two objectives:

Objective 1: Refine the ADAb and drug level testing strategies for relevance to clinical practice in England;

Objective 2: Define the decision problem in terms that conformed to the requirements of decision-makers responsible for resource allocation in NHS England.

5.5.3. Method

The two potentially relevant comparator strategies, identified by the early conceptualisation technique in Section 5.4, were made appropriate for clinical practice in England by defining the specific treatment decision associated with each test result. This decision was informed by published clinical evidence regarding testing for TNFi immunogenicity (see Section 5.3), clinical recommendations for managing patients with RA (Ding et al., 2010; Smolen et al., 2014; National Institute for Health and Care Excellence, 2016a; Singh et al., 2016b), and the interviews with rheumatologists in *Chapter Three*.

The NICE *Guide to the Methods of Technology Appraisal* outlined the appropriate evidence requirements of an economic evaluation used to inform the appraisal of health technologies in NHS England (National Institute for Health and Care Excellence, 2013a). The decision problem for the economic evaluation in this thesis, therefore, conformed to the requirements specified in the NICE Reference Case (described in *Appendix 2*) to generate results applicable to decision-makers in NHS England.

5.5.4. Results

A description of the prescribing decisions associated with each test result, and how those decisions were deemed to be relevant for routine clinical practice in England, is reported in *Appendix 25*. The final decision problem addressed by the economic evaluation in this thesis, consistent with a NICE Reference Case analysis, is summarised in Table 5.8.

The decision problem considered a population of adult patients with RA that were were (i) already receiving adalimumab and (ii) met the NICE eligibility for TNFi therapy (patients must have had a DAS28 score of at least 5.1 and must have failed two previous attempts of cDMARD therapy) (National Institute for Health and Care Excellence, 2016a). The intervention strategies to stratify treatment followed the test-and-treatment decisions that were relevant to clinical practice in England (*Appendix 25*; Table A25.1). In addition to testing, dose-reduction of adalimumab in all patients (irrespective of drug level status) was included as a relevant comparator strategy, based on the reported practice of the rheumatologists that were interviewed in *Chapter Three*. The expected outcomes from the intervention strategies were compared to a common comparator (current practice) and to each other in a fully incremental analysis.

Table 5.8. *The decision problem.*

Element of Decision Problem	Description
Population	Adults with RA in England, receiving adalimumab, who had: • (i) a DAS28 score of at least 5.1, and; • (ii) failed at least two attempts of synthetic cDMARD therapy, including methotrexate.
Intervention Technology	• (i) Test ADAb and drug levels by ELISA while responding to first-line adalimumab to inform an early change to rituximab therapy†; • (ii) Test ADAb only by ELISA while responding to first-line adalimumab to inform an early change to rituximab therapy†; • (iii) Test first-line adalimumab drug levels only by ELISA in remission to inform a dose-reduction strategy†. • (iv) Dose-reduction strategies in all patients irrespective of drug-level status.
Comparator Technology	• (i) Current practice adalimumab therapy; • (ii) All intervention strategies.
Perspective	NHS England and personal social services.
Measure of health outcome	EQ-5D quality-adjusted life years.
Costs considered	Include direct medical costs, that comprised: • (i) treatment costs; • (ii) test costs; • (ii) cost of hospitalisations. Exclude indirect and productivity costs.
Outcomes	• Expected incremental costs; • Expected incremental QALYs; • Incremental cost-effectiveness ratio; • Net monetary benefits, • Incremental net monetary benefits, • Expected value of perfect information (EVPI).
Time horizon	Lifetime.
Discount rate	Costs = 3.5%, QALYs = 3.5%.
Cost-effectiveness Threshold	£20,000 to £30,000 per QALY gained.
Sensitivity Analysis	• One-way sensitivity analysis; • Two-way sensitivity analysis; • Probabilistic sensitivity analysis.

Note: †=The rationale for the appropriateness of these treatment decisions are reported in *Appendix 25*.

5.5.5. Discussion

This study defined the decision problem for testing adalimumab ADAb and drug levels in order to stratify treatment for patients with RA in England. The testing strategies that were identified in the clinical literature (Section 5.3 and Section 5.4) were refined in accordance with current clinical practice in England. Given that rituximab therapy was assumed to be the next appropriate therapy (after adalimumab) along the care pathway for RA (see *Appendix 25*), monitoring strategies that detected the presence of low adalimumab drug levels and/or ADAb were assumed to recommend a pre-emptive change of treatment to rituximab. The decision problem required that incremental costs and QALYs derived from each intervention strategy were estimated over a lifetime time horizon.

The remaining sections of *Chapter Five*, having defined the decision problem, report how the *de novo* decision analytic model was conceptualised and developed. Section 5.6 builds on the decision problem by presenting problem-oriented conceptual models of the clinical events and care pathways that patients may have experienced over time; Section 5.7 selects the appropriate type of decision analytic model for the economic evaluation; and Section 5.8 describes the final structure of the quantitative *de novo* decision analytic model.

5.6. Chapter Objective 4: Problem-Oriented Conceptual Models

This study describes how the use of adalimumab ADAb and drug ELISA level testing to stratify treatment in current practice was conceptualised. Two problem-oriented conceptual models (disease-logic and service-pathway) are presented that described the potential health and resource consequences associated with testing in current practice. The problem-oriented conceptual models were subsequently used to select the most suitable type of decision analytic model (in Section 5.7) for the early economic evaluation in this thesis.

5.6.1. Introduction

A *de novo* decision analytic model should be structured to include the relevant events that may be experienced by a patient over time, to address an explicit decision problem (Tappenden, 2014). Structural uncertainty is inherent in any model-based economic evaluation (Briggs, 2000; Bojke et al., 2009) and may arise, for example, when the relevant events that should be included in a *de novo* decision analytic model not clear (Bojke et al., 2009). In the context of this thesis, two sources of structural uncertainty regarding adalimumab ADAb and drug level testing were the characterisation of (i) disease progression and (ii) the care pathways available to patients before, after, and without treatment stratification. Explicit model conceptualisation techniques may reduce such structural uncertainties by making transparent, and providing justification for, the choice of relevant events to include in a *de novo* decision analytic model (Tappenden, 2014).

5.6.2. Aim and Objectives

The aim of this study was to conceptualise the potential impact of treatment stratification by adalimumab ADAb and drug level testing on a patient's subsequent disease status and care pathways. There were two objectives:

Objective 1: Conceptualise the progression of RA with and without the use of adalimumab ADAb and drug level testing to stratify treatment;

Objective 2: Conceptualise the care pathway for RA with and without the use of adalimumab ADAb and drug level testing to stratify treatment.

5.6.3. Methods

Two problem-oriented conceptual models were developed to address each objective. A disease-logic conceptual model was developed first that described the true clinical events that were assumed to have been experienced by a patient with RA over time (Tappenden, 2014). A service-pathway conceptual model was developed second that described the health technologies that were assumed to be received by a patient with RA over time, based on their known characteristics (Tappenden, 2014).

Both problem-oriented conceptual models used flow diagrams to illustrate the anticipated sequence of events and were supported by descriptive text (Tappenden, 2014). The flow diagrams were used as a communication tool to obtain clinical input from Prof. Anne Barton and Dr. Meghna Jani (both of whom had experienced treating patients with RA) to ensure that the conceptualisation was relevant to clinical practice in England. The conceptualisation process followed the best-practice recommendations made by the *International Society for Pharmacoeconomics and Outcomes Research* (ISPOR) (Roberts et al., 2012) and the *NICE Decision Support Unit* (Kaltenthaler et al., 2011).

5.6.4. Results

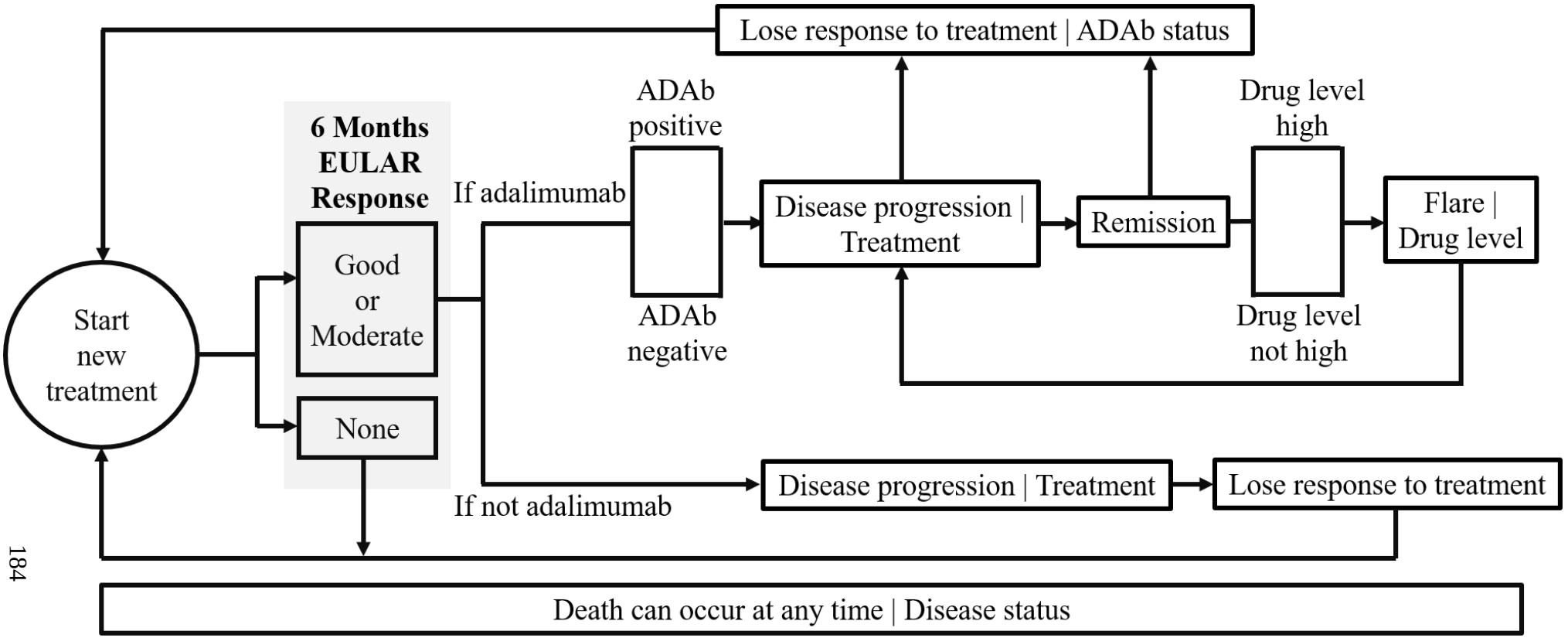
The results section reports each descriptive problem-oriented conceptual model separately.

5.6.4.1. Disease-logic Model

Figure 5.6 illustrates the disease-logic conceptual model for RA that was developed for this early model-based economic evaluation of stratified medicine.

The management of RA was characterised by the control of disease activity and the minimisation of disease progression over time (Upchurch et al., 2012). Patients were assumed to be eligible for a sequence of treatments, that began with adalimumab, which they could potentially receive over their lifetime (Tosh et al., 2014). A representative sample of patients with RA living in England were assumed to have heterogeneous characteristics at baseline (for example, age, sex, and disease activity), which may have affected the likelihood of future clinical events (Hyrich et al., 2011). A patient with RA could have died at any time over the lifetime course of treatment.

Figure 5.6. *Disease-logic conceptual model.*



Note: Arrows denote the logic of clinical events over time; *ADAb positive*=the patient had developed adalimumab anti-drug antibodies and was assumed to have low drug levels.

The disease-logic model was designed to be consistent with clinical practice in England; a patient's response to treatment was assessed using the EULAR response criteria (van Gestel et al., 1996) (described in *Appendix 7*) which was determined six months after receiving any treatment (National Institute for Health and Care Excellence, 2016a). Primary non-response occurred when no EULAR response was attained at six months, and treatment was subsequently changed to the next therapy in the sequence. A good EULAR response, by definition, was associated with a greater reduction in disease activity than a moderate EULAR response (van Gestel et al., 1996).

Patients that received a treatment in the sequence, other than adalimumab, were assumed to maintain therapy until they experienced a loss of response. Disease was assumed to progress over time at a rate conditional on the type of treatment received; the clinical evidence indicated that patients treated with cDMARD therapies had a faster rate of disease progression that those treated with bDMARD therapies (Michaud et al., 2011).

If the patient was treated with adalimumab, there was a probability that they developed ADAb (termed *ADAb-positive*). Adalimumab drug levels were assumed to correlate perfectly with ADAb status, such that ADAb-positive patients had lower drug levels (Bartelds et al., 2011; Jani et al., 2015a).

Secondary non-response to adalimumab may have occurred at any time during the course of treatment, which was assumed to be conditional on the patient's underlying ADAb status (positive or negative) and drug level status (high or low) (Bartelds et al., 2011). A patient's adalimumab ADAb status and drug levels were unobservable unless a rheumatologist used an ELISA assay to measure them (for example, see the prescribing algorithms in Section 5.3). Patients with adalimumab ADAb and low drug levels were assumed to be more likely to lose response to their treatment earlier (Garcês et al., 2013). Disease was assumed to progress over time as with the other treatments.

A distinction between treatments was made by assuming that patients treated with adalimumab could enter an explicit period of remission. In reality, patients may enter remission when receiving non-adalimumab therapies as well; however modelling this phenomenon was not relevant to the decision problem as defined (see Section 5.5). Testing non-adalimumab therapies in remission was a different resource allocation problem which likely required a different testing strategy and prescribing algorithm. Underlying adalimumab drug levels in remission could have been characterised as *high* or *not high*.

The likelihood of a flare in disease activity during remission was assumed to be a function of a patient's true underlying adalimumab drug level (Bykerk et al., 2016). For example, based on the clinical literature, patients were more likely to flare if their adalimumab dose was reduced and their drug levels were not high *a priori* (Chen et al., 2016).

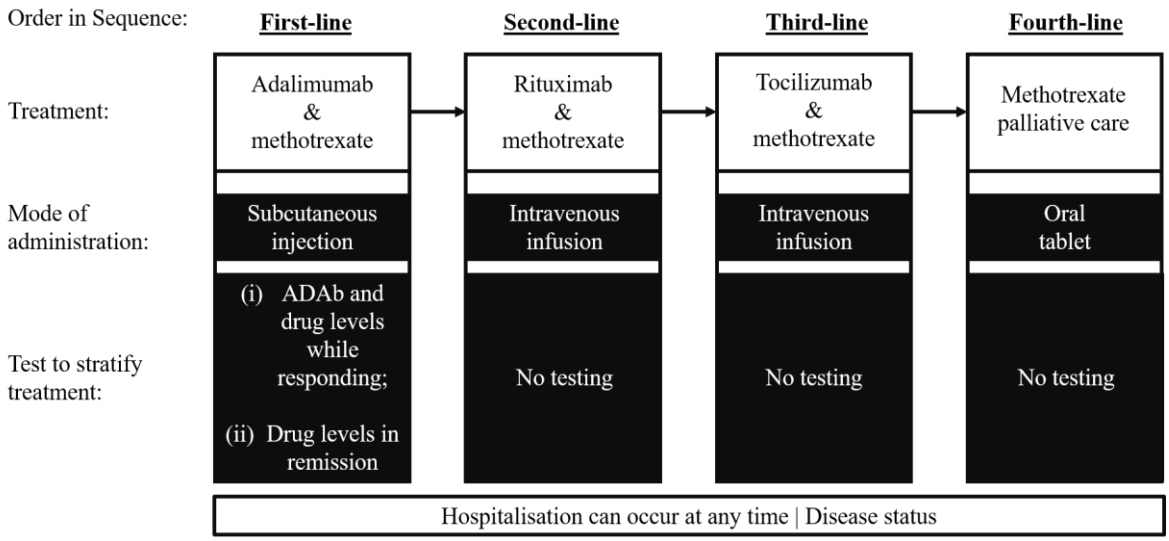
Patients were assumed to gain QALYs over their lifetime depending on the clinical events experienced in relation to their underlying disease activity. A EULAR response to treatment reduced disease activity and, in turn, increased QALYs. A loss of response to treatment increased a patient's disease activity (Prevoo et al., 1995) and, consequently, was assumed to reduced QALYs. ADAb-positivity wasn't assumed to affect QALYs directly, but instead made the QALY-loss from treatment failure occur earlier. Gradual disease progression resulted in a diminishing rate of QALY gain over time while responding to treatment. A disease flare in remission was associated with a short period (approximately one week) (Bykerk et al., 2014) of QALY-loss due to the increase in disease activity.

Given that patients who developed adalimumab ADAb may respond better to a treatment with a different therapeutic target (Tak, 2012), an early change of treatment to second-line rituximab therapy (upon ADAb detection) was assumed to provide a health benefit, reducing disease activity and increasing QALYs. However, an inappropriately early change to rituximab therapy for patients without adalimumab ADAb was assumed to cause harm by increasing disease activity and reducing QALYs (because previous inhibition of tumour necrosis factor-α production was effective).

5.6.4.2. Service-pathway Model

Figure 5.7 illustrates the service-pathway conceptual model for this early model-based economic evaluation of stratified medicine. The service-pathway conceptual model described the care pathway that was assumed for patients with RA in England, and how adalimumab ADAb and drug level testing was integrated into this care pathway.

Figure 5.7. *Service-pathway conceptual model.*



Current practice for RA was characterised by a sequence of therapies prescribed over a patient's lifetime (Tosh et al., 2014). The appropriate sequence was identified according to the qualitative interviews in *Chapter Three* and the recent recommendations for managing RA by NICE (National Institute for Health and Care Excellence, 2016a). The sequence was also consistent with recommendations by the *British Society for Rheumatology*, which advocated the prescription of rituximab after the failure of a TNFi therapy (Bukhari et al., 2011) and tocilizumab with methotrexate after inadequate response to cDMARDs (Malaviya et al., 2014). Furthermore, the sequence was consistent with the general recommendations of the international bodies for rheumatology; the *EULAR* treatment guidelines for RA recommended (i) using a bDMARD with methotrexate if the patient had an insufficient response to cDMARD therapy; and (ii) using a second bDMARD if the patient had failed their first bDMARD (Smolen et al., 2014). The *ACR* treatment guidelines for RA recommended using a sequence of up to two non-TNFi bDMARDs (with or without methotrexate) if a patient's disease activity remained high following first-line TNFi therapy (Singh et al., 2016b).

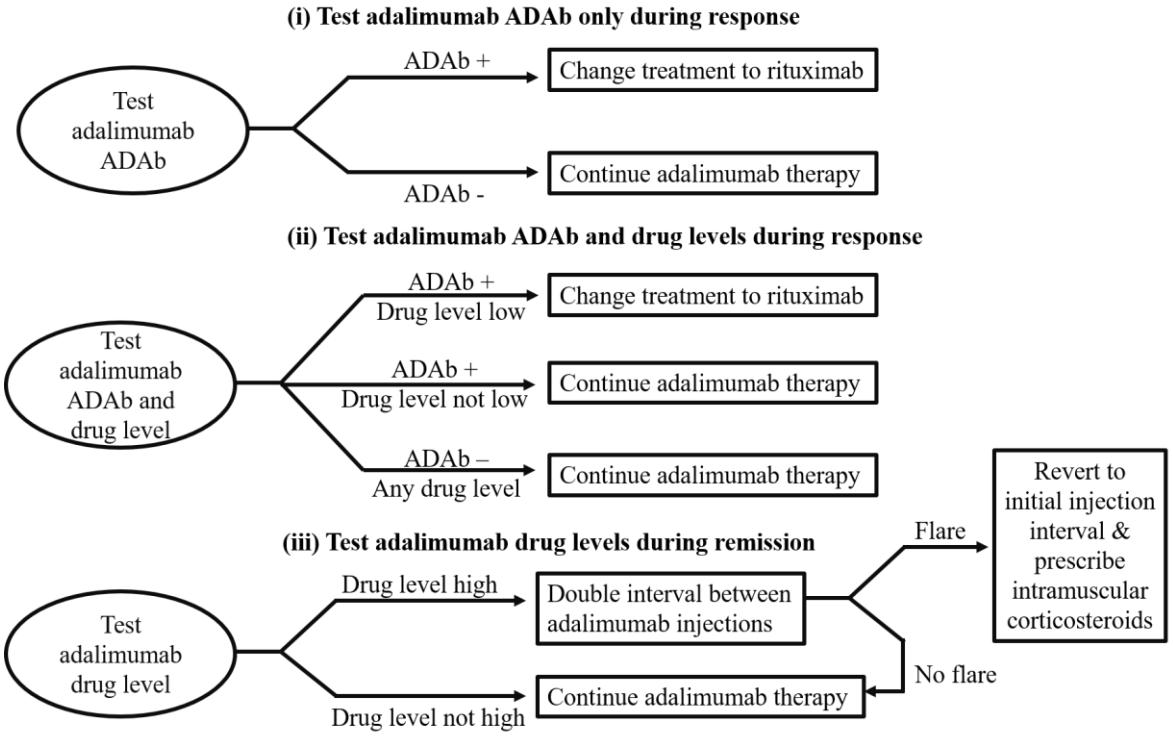
All patients were assumed to commence the sequence with adalimumab therapy. All bDMARD treatments (adalimumab, rituximab, tocilizumab) were assumed to be prescribed with concomitant methotrexate (National Institute for Health and Care Excellence, 2016a). Patients with RA were assumed to receive each treatment until they experienced a loss of response. Patients that received adalimumab could also change treatment if specified by a testing strategy. Patients were assumed to receive palliative

cDMARD therapy for the rest of their lives if they lost response to three bDMARD therapies.

Rituximab and tocilizumab required additional direct health care resources to account for their administration by intravenous infusion (British National Formulary, 2016). No direct health care resources for infusion were required for adalimumab therapy, however, a proportion of subcutaneous injections may have been administered by a nurse in routine practice (Stevenson et al., 2016). Patients were assumed to potentially require hospitalisation, at any point in their lifetime, which increased in likelihood as the patients' disease status worsened (Bansback et al., 2008)

ADAb and drug level testing was assumed to occur during response to adalimumab only. The ELISA-based testing strategies were assumed to require additional health care resources (for example, an additional outpatient appointment). The appropriate frequency of testing was unknown *a priori* and this was later informed by the incremental net benefit of testing in the full economic evaluation. Figure 5.8 illustrates the relevant prescribing decisions for the outcome of each testing strategy.

Figure 5.8. *Service pathway conceptual model of testing strategies for (i) adalimumab ADAb testing only during response; (ii) adalimumab ADAb and drug level testing during response; and (iii) adalimumab drug level testing during remission.*



Abbreviations: ADAb=Anti-drug antibody.

If only ADAb were tested and detected during response to adalimumab (Figure 5.8i), treatment was pre-emptively changed to the next bDMARD in sequence (rituximab). If patients were tested for both adalimumab ADAb and drug levels (Figure 5.8ii), treatment was changed to rituximab if ADAb and low drug levels were detected. If high adalimumab drug levels were detected in patients during remission (Figure 5.8iii), the dose was halved by doubling the time between subcutaneous injections (Smolen et al., 2014). Patients that flared were assumed to receive intramuscular steroids and had their adalimumab dose reverted to the original injection schedule (National Institute for Health and Care Excellence, 2009).

5.6.5. Discussion

This study aimed to conceptualise the impact of treatment stratification by adalimumab ADAb and drug level testing on patients' subsequent disease status and care pathways. Two problem-oriented conceptual models were developed (disease-logic and service-pathway models) that described the clinical events and health technologies that patients were assumed to experience over time. The results of this study represented the initial stage of developing a *de novo* decision analytic model to address the decision problem reported in Section 5.5.

Conceptual models are descriptive tools that have the advantage of enabling decision-analysts to explore the structural uncertainty inherent in any model-based economic evaluation, by considering alternative ways to structure a final quantitative decision analytic model (Tappenden, 2014). The process of model conceptualisation is an essential element of any model-based economic evaluation (Roberts et al., 2012) that is rarely documented in practice (Chilcott et al., 2010). The two problem-oriented conceptual models were based on relevant evidence derived from (i) published clinical studies, (ii) clinical recommendations for RA, and (iii) previous research conducted in this thesis. An advantage of these evidence-based conceptual models, therefore, was that they provided a transparent justification for how to structure the final decision analytic model of adalimumab ADAb and drug level testing.

Implications for Future Research

The problem-oriented conceptual modelling exercise had two implications for future research that were addressed by this thesis:

- A decision was required to select the type of model (decision tree, Markov model, discrete event simulation) that should be used for the final decision analytic model. The elements of the decision problem highlighted by this conceptual modelling study may be able to inform the appropriate model type for the economic evaluation of adalimumab ADAb and drug level testing (Tappenden, 2014);

- The disease-logic and service-pathway conceptual models must be combined to finalise the design of the *de novo* decision analytic model. A design-oriented conceptual model may therefore be used to justify the simplifications and assumptions of the final quantitative model.

Summary of Key Findings

This study conceptualised the health states and care pathways associated with stratifying treatment for patients with RA according to adalimumab ADAb and drug level testing. Two problem-oriented conceptual models were constructed according to evidence from a range of sources, including relevant clinical recommendations, published clinical evidence and research presented earlier in this thesis (qualitative interviews in *Chapter Three* and the systematic review of prescribing algorithms in Section 5.3). The following study reported in *Chapter Five* built on these problem-oriented conceptual models to select the type of decision analytic model that was used to address the decision problem in Section 5.5.

5.7. Chapter Objective 5: Selection of Model Type

This study reports how the type of decision analytic model was chosen for the *de novo* model-based economic evaluation of stratified medicine presented in this thesis.

5.7.1. Introduction

The next stage of developing the economic evaluation, having defined the decision problem (Section 5.5) and conceptualised the clinical states and care pathways over time (Section 5.6), was to select the most appropriate type of decision analytic model. Three candidate types of model (decision tree, Markov model, discrete event simulation) were described previously in *Appendix 3* in terms of their key features, advantages, and disadvantages. The choice of model type, however, was not arbitrary; the most appropriate type of model depended on the specific characteristics of the decision problem (Roberts et al., 2012).

5.7.2. Aim and Objectives

The aim of this study was to determine the appropriate type of decision analytic model for the economic evaluation presented in this thesis. There were two objectives:

Objective 1: Identify characteristics of the decision problem that were relevant to inform the appropriate type of decision analytic model;

Objective 2: Select the type of decision analytic model that was appropriate to provide evidence for the decision problem in Section 5.5.

5.7.3. Method

The two problem-oriented conceptual models reported in Section 5.6 were used to identify the characteristics of the decision problem that were relevant to inform the appropriate type of decision analytic model. These specific characteristics of stratified medicine for RA, according to adalimumab ADAb and drug level testing, were then applied to the published fifteen-item checklist, reported by Brennan et al. (2006), to inform the choice of decision analytic model.

5.7.4. Results

There were four characteristics of the decision problem, identified by the problem-oriented conceptual models, which may have been relevant to inform the choice of model type:

(i) The baseline characteristics of patients with RA in England were assumed to be heterogeneous (for example, by age and disease severity) (Hyrich et al., 2011);

(ii) Patients were assumed to receive a sequence of treatments over their lifetime, characterised by the recurring events of treatment response and failure (Tosh et al., 2014);

(iii) Patient-level characteristics were assumed to affect clinical outcomes (for example, patients with greater severity of disease were assumed to have a higher risk of mortality) (Stevenson et al., 2016);

(iv) The occurrence of future clinical events were assumed to depend on previous clinical events (for example, the development of adalimumab ADAb was assumed to affect the loss of response to treatment in the future) (Garcês et al., 2013).

The completed fifteen-item checklist to inform the choice of model type, developed by Brennan et al. (2006), is reported in *Appendix 26*. The key elements of the decision problem that informed the choice of decision analytic model are reported below. Each *Issue* referrers to a specific item within the Brennan et al. (2006) checklist.

- The use of a decision tree was deemed to be inappropriate (*Issue 8*) because treatments were prescribed in a sequence which, conceptually, led to each patient experiencing a lifetime of recurring clinical events (treatment response and failure);

- The dimensionality of the decision problem was likely to be too great for a model that simulated a homogeneous cohort (*Issue 7*). For example, (i) patient characteristics were assumed to affect mortality and the time to treatment failure, and (ii) the type of treatment prescribed was assumed to affect the rate of disease progression;

- The decision problem was also characterised by a range of competing clinical events that could have occurred within any unit of time (such as treatment failure, disease

progression, death, development of adalimumab ADAb, or a test for ADAb and drug levels). The checklist therefore suggested that the appropriate modelling approach should have advanced time continuously or in small time cycles (*Issue 12*);

- The development of ADAb and low drug levels during response to treatment was assumed to affect the time to secondary non-response of adalimumab. This assumption, which was the source of exploitable heterogeneity to stratify subsequent treatment decisions, implied that: (i) the timing of events was important (*Issue 9*), (ii) covariates (at the patient-level) caused an interaction effect (*Issue 5*), and (iii) the timing of events was inherently non-Markovian (*Issue 6*) because the previous clinical events experienced by a patient may have affected their likelihood of future clinical events. The checklist therefore recommended a non-Markovian individual-level model that utilised simulation methods.

The appropriate type of model to address the decision problem in Section 5.5 was therefore selected to be an individual-level discrete event simulation (DES) (Caro et al., 2016b).

5.7.5. Summary of Key Findings

This study selected the most appropriate type of decision analytic model for the economic evaluation of stratified medicine in this thesis by (i) identifying the key characteristics of the decision problem, that may have informed the type of model, from the problem-oriented conceptual models in Section 5.6, and (ii) by applying these characteristics to a published checklist that informed model selection by Brennan et al. (2006). An individual-level discrete event simulation was selected to be the most appropriate type of model to address the decision problem in Section 5.5. The key element of the decision problem that informed this choice was the need for a non-Markovian model; the model required a memory of patients' histories because the development of adalimumab ADAb and low drug levels were assumed to affect the time to subsequent treatment failure.

The final stage of model development, having (i) defined the decision problem (Section 5.5); (ii) conceptualised the system in which the decision problem existed (Section 5.6); and (iii) selected the appropriate type of model (Section 5.7), was to design the structure of the final quantitative decision analytic model. This final stage, reported in the next section of *Chapter Five*, was achieved by using a design-oriented conceptual model that brought together the two problem-oriented conceptual models within a DES framework.

5.8. Chapter Objective 6: Design-oriented Conceptual Model

This study reports how the structure of the final decision analytic model was designed, within the framework of a DES, to address the decision problem in Section 5.5.

5.8.1. Introduction

The structure of the final decision analytic model of adalimumab ADAb and drug level testing, to stratify treatment for patients with RA in England, was informed by combining the two problem-oriented conceptual models that were described in Section 5.6. There were, however, a number of different ways in which to structure the final decision analytic model (Tappenden, 2014). There is an extensive literature on the differences in structural assumptions between different published decision analytic models for RA (Bansback et al., 2005; Drummond et al., 2005; Bansback et al., 2008; Barton, 2011; Madan et al., 2011; Tosh et al., 2011; Tsao et al., 2012; Scholz et al., 2014; Tosh et al., 2014; Ganz et al., 2015; Madan et al., 2015). Therefore, previous model-based economic evaluations for RA, and in particular those that performed an individual-patient simulation, also had the potential to inform the design of the economic evaluation reported in this thesis (Tappenden, 2014)

5.8.2. Aim and Objectives

The aim of this study was to determine the structure of the *de novo* decision analytic model to address the decision problem in Section 5.5. There were two objectives to meet this aim:

Objective 1: Identify the modelling assumptions that were made by similar individual level model-based economic evaluations for RA;

Objective 2: Develop the structure of the *de novo* DES model for the early economic evaluation of stratified medicine in this thesis.

5.8.3. Method

A systematic review of published decision analytic models for RA that had performed an individual-level patient simulation is reported in *Appendix 27*. The purpose of this systematic review was to identify the distinct structural assumptions that were made within

different published decision analytic models for RA that had a similar design as the DES in this thesis.

A design-oriented conceptual model was then developed within a DES structure, which combined the two descriptive problem-oriented conceptual models from Section 5.6, informed by the results of the systematic review in *Appendix 27*.

5.8.4. Results

The results of the systematic review of published individual-level models for RA is reported, in full, in *Appendix 27*. The systematic review identified twenty-nine studies that reported an individual-level model for RA. Each model was categorised into eight 'families' that had similar structural assumptions. The models were, broadly, designed to simulate a patient individually through a sequence of treatments over time according to four phases: (i) estimate treatment response; (ii) estimate the immediate and temporal change in disease activity following treatment response; (iii) estimate loss of response to treatment and commence the next treatment in the sequence; (iv) estimate whether the patient has died, at which point, commence the simulation of the next patient within the cohort. The models made assumptions about the simulated population, the progression of disease, and the estimation of direct health care resource use and QALYs gained over time. These assumptions are described in detail within *Appendix 27* and are used throughout *Chapter Six* to justify the specific assumptions within the *de novo* decision analytic model in this thesis. The results reported here provide an overview of the final decision analytic model's structure in the thesis (Section 5.8.4.1) and describe the specific care pathways that a patient may have experienced within the model (Section 5.8.4.2).

5.8.4.1. Overview of the Model

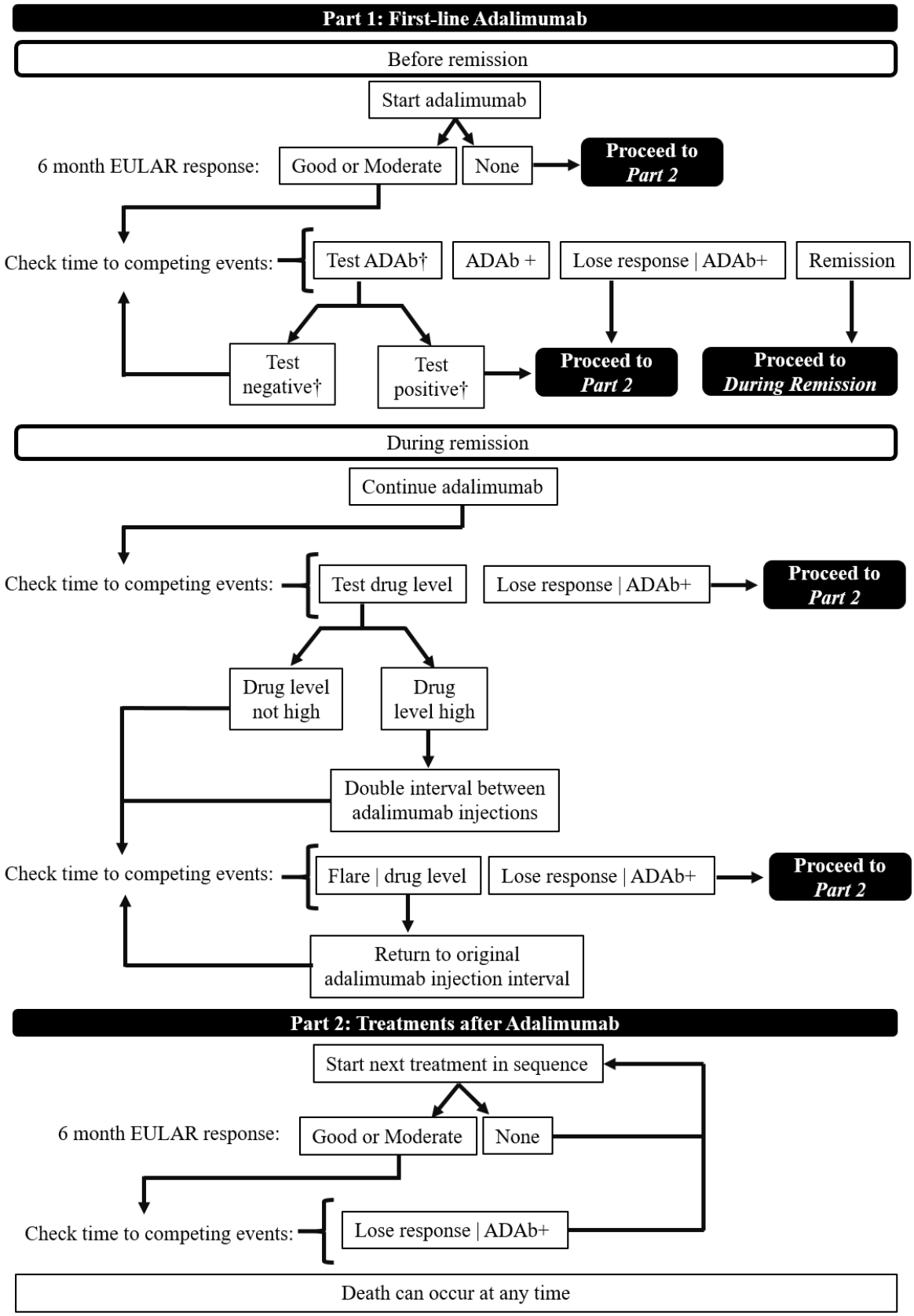
The structure of the *de novo* decision analytic model characterised the sequence of treatments that could be prescribed to a patient with RA over their lifetime. The model assumed a structure that emulated current clinical practice for RA in England to ensure relevance to the decision problem in Section 5.5. For example, (i) the appropriate sequence of treatments was informed by clinical recommendations (produced by NICE, the *British Society for Rheumatology*, and EULAR) and qualitative interviews with consultant rheumatologists in England (*Chapter Three*); and (ii) response to any treatment was characterised in terms of a EULAR response (van Gestel et al., 1998). ADAb and drug

level ELISA testing were included the model, for patients that were prescribed adalimumab, to facilitate the stratification of a subsequent treatment decision.

The decision analytic model was implemented as a DES (described in *Appendix 3*) (Caro, 2005; Caro et al., 2016a; Caro et al., 2016b). Conceptually, the DES simulated patients individually through the structure of the model by exposing each patient to a set of relevant competing events that could have occurred over their lifetime. Twelve published model-based economic evaluations have also simulated patients by estimating time-to-event values individually for each patient (*see Appendix 27;* Section A27.4). Figure 5.9 illustrates the design-oriented conceptual model of the DES simulation.

The model was divided into two parts; Part One corresponded to the period when a patient received adalimumab and Part Two corresponded to the period when they received any other subsequent treatment in the sequence. Part One was divided further into the time before and after adalimumab-induced remission, to enable different strategies of stratified medicine (based on the time of testing) to be evaluated.

Figure 5.9. *Structure of the decision analytic model to stratify treatment for RA using adalimumab ADAb and drug level testing.*



Note: †: Different testing strategies were used for routine monitoring of adalimumab: (i) Test ADAb only – *Test positive* = ADAb detected; (ii) Test ADAb and drug levels – *Test Positive* = ADAb and low drug levels detected. Abbreviations: ADAb+=Anti-drug antibody positive.

5.8.4.2. Description of Care Pathways

This section describes the different care pathways that a hypothetical individual patient could have experienced through the model.

Part One: Adalimumab before Remission

Patients that entered the model were assumed to commence adalimumab therapy. There were six competing events that a patient may have experienced, before they entered remission, whilst receiving adalimumab: (i) a EULAR response; (ii) loss of response; (iii) enter remission; (iv) develop adalimumab ADAb; (v) test ADAb and drug levels; or (vi) death. The first event was usually a EULAR response to treatment, unless death or loss of response occurred within six months of model entry. The patient moved to Part Two of the model if no EULAR response was attained; the remaining events were experienced in ascending order of time if a good or moderate EULAR response to adalimumab was attained at six months.

If a patient developed adalimumab ADAb, their time to experiencing a loss of response to adalimumab was reduced (to occur earlier). The frequency of ADAb and drug level testing, and the time to enter remission, were determined in advance to characterise different strategies to stratify treatment. If testing detected adalimumab ADAb (and low drug levels, if both tests were used), the patient proceeded to Part Two of the model and changed their treatment to the next bDMARD in the sequence (rituximab).

Part One: Adalimumab during Remission

There were three competing events that a patient could have experienced during adalimumab-induced remission: (i) loss of response; (ii) test adalimumab drug levels; or (iii) death. If the drug level test detected high drug levels, the interval between the patient's adalimumab injections was doubled (to reduce the dose). The patient may have subsequently flared based on the true status of their underlying drug level. The interval between injections was reverted to its original schedule if a patient flared. Patients remained on adalimumab until loss of response or death.

Part Two: Treatments after Adalimumab

A patient's treatment was changed to the next in sequence upon entering Part Two of the model. The three competing events that could occur after commencing any treatment in Part Two were: (i) a EULAR response; (ii) loss of response; or (iii) death. A patient that lost response to a treatment was prescribed the next therapy in the sequence. Patients that entered Part Two remained there until death; at which point, the next individual patient entered the model to simulate their specific lifetime of events.

Specific Testing Strategies

The different ways to use a testing strategy in order to stratify treatment were defined as different comparator strategies in the economic evaluation (identified in *Chapter Six*). For example, permutations of testing, consistent with the decision problem, included (i) only testing adalimumab drug levels in remission; (ii) only testing adalimumab ADAb (and not drug levels) whilst responding; and (iii) testing at different frequencies (for example, every three or six months). Additionally, the decision problem included consideration of adalimumab dose-reduction strategies in all patients without testing (Section 5.5). The care pathways in the decision analytic model were therefore adjustable to implement different comparator strategies.

5.8.5. Summary of Key Findings

This study presented the design of the final *de novo* decision analytic model of adalimumab ADAb and drug level ELISA testing to stratify treatment for patients with RA in England. The structure was designed with explicit relevance to current clinical practice in England to satisfy the requirements of the decision-makers responsive for allocating population health care resources in the NHS. A design-oriented conceptual model was used to combine the two descriptive problem-oriented conceptual models from Section 5.6, informed by a systematic review of published individual-level economic evaluations.

Having developed the structure of the *de novo* decision analytic model, the next stage of the early economic evaluation was to populate the model with relevant evidence to inform (i) the relative cost-effectiveness of using the ADAb and drug level tests to stratify treatment, and (ii) the potential value of conducting further research to reduce parameter uncertainty in the model.

5.9 Conclusion

This chapter provided a transparent explanation of how the early economic evaluation, reported in this thesis, was developed. An extensive model conceptualisation process was undertaken, which was informed by six sub-studies within the chapter: (i) a systematic review of RA-specific prescribing algorithms that included ADAb and drug level testing for any TNFi; (ii) a novel algebraic conceptualisation technique to identify relevant relevant comparator strategies from a wider set of candidate strategies; (iii) a clear definition of the decision problem; (iv) two problem-oriented conceptual models that described disease progression and the relevant care pathways; (v) a clear justification for selecting the type of decision analytic model; and (vi) a design-oriented conceptual model that described the final structure of the *de novo* decision analytic model.

In order to implement the final DES model, evidence was required to estimate the values of the input parameters for clinical outcomes, resource use, unit costs, and QALYs. *Chapter Six* presents the methods that were used to perform the economic evaluation, which addressed the decision problem in Section 5.5, by estimating: (i) the relative cost-effectiveness of adalimumab ADAb and drug level testing to stratify treatment for patients with RA in England, and (ii) the potential value of conducting further research to reduce uncertainty in the relative cost-effectiveness of treatment stratification.

Chapter 6

Early Model-based Economic Evaluation of Adalimumab Anti-drug Antibody and Drug Level Testing to Stratify Treatment for Rheumatoid Arthritis in England

Chapter Six presents an early model-based economic evaluation of adalimumab ADAb and drug level ELISA testing to stratify treatment for patients with RA in England. The study describes how the *de novo* decision analytic model was implemented as a DES and how the specific input parameters, and expected outputs, of the model were estimated.

A published microcosting study, co-authored with Dr. Meghna Jani to supplement the research within this thesis, is provided in *Appendix 35* (Jani et al., 2016a). The systematic review and bivariate meta-analysis of test accuracy studies, reported in *Appendix 34*, has been accepted for presentation at the *International Society for Pharmacoeconomics and Outcomes Research 22nd Annual International Meeting* and the *EULAR 2017 Annual European Congress of Rheumatology*. The thesis chapter is structured by the following sections: the introduction (Section 6.1), aim and objectives (Section 6.2), method (Section 6.3), results (Section 6.4), discussion (Section 6.5), and conclusion (Section 6.6).

6.1. Introduction

Chapter Five documented the development process of the early cost-effectiveness analysis presented in this thesis by (i) defining the decision problem, (ii) selecting the most appropriate type of model, (iii) conceptualising disease progression and care pathways, and (iv) finalising the structure of the decision analytic model. This chapter reports how the final decision analytic model was implemented, in its quantitative form, to estimate the relative cost-effectiveness of adalimumab ADAb and drug level testing to stratify treatment for RA in NHS England.

6.2. Aim and Objectives

The aim of this study was to determine whether adalimumab ADAb and drug level ELISA testing, to stratify treatment for patients with RA in England, was a relatively cost-effective use of health care resources, according to the decision problem in Table 5.8 of Section 5.5. There were three objectives to meet this aim:

Objective 1: Construct a quantitative *de novo* decision analytic model that was appropriate for the decision problem in Table 5.8;

Objective 2: Estimate relevant values and distributions for the input parameters of the decision analytic model;

Objective 3: Estimate the (i) relative cost-effectiveness and (ii) decision uncertainty associated with adalimumab ADAb and drug level testing to stratify treatment for patients with RA in the NHS in England, and (iii) the potential value of further research to reduce parameter uncertainty within the model.

6.3. Method

The methods of the early economic evaluation are reported over five subsections, which describe: the study's design (Section 6.3.1), the target population (Section 6.3.2), the estimation of the model's input parameters (Section 6.3.3), validation of the model (Section 6.3.4) and the analysis (Section 6.3.5). The economic evaluation was reported according to the requirements of the *Consolidated Health Economic Evaluation Reporting Standards* (CHEERS) statement (Husereau et al., 2013). The completed *CHEERS* statement is reported in *Appendix 28*.

6.3.1. Study Design

This study was designed as a model-based economic evaluation, to inform health care resource allocation decision-making in England by following the requirements specified by the NICE Reference Case (National Institute for Health and Care Excellence, 2011a; 2013a).

6.3.1.1. Perspective

The study was conducted from the perspective of the NHS and personal social services in England, defined by the decision problem in *Chapter Five* (see Table 5.8; Section 5.5). This perspective was deemed to be relevant for decision-makers that had responsibility for allocating resources from the NHS England's budget for health care (National Institute for Health and Care Excellence, 2011a; 2013a). The relevant opportunity cost of a comparator strategy was therefore estimated according to the incremental direct health care costs that were imposed on this budget constraint. Indirect costs (such as the productivity loss of a patient) were not included in the analysis because they were beyond the scope of the study's perspective (National Institute for Health and Care Excellence, 2013a).

6.3.1.2. Time Horizon

Relevant cost and QALY differences between intervention strategies were expected to occur for the duration of each patient's life due to the chronic nature of RA, and the consequential sequence of therapies that each patient would likely have experienced over their lifetime. The study was therefore conducted over a lifetime time horizon.

6.3.1.3. Form of Economic Evaluation

The economic evaluation was a cost-effectiveness analysis, with health outcomes expressed in terms of QALYs. In the NICE Reference Case, the term *cost-effectiveness analysis* was used interchangeability with *cost-utility analysis*, which was a more-common nomenclature to define an economic evaluation that used QALYs to quantify health consequences (National Institute for Health and Care Excellence, 2013a).

By expressing the health consequences of each intervention strategy in QALYs, statements could be made regarding the relative cost-effectiveness of treatment stratification in England with reference to a range of cost-effectiveness thresholds used by NICE (National Institute for Health and Care Excellence, 2013a).

6.3.1.4. Test and Treatment Strategies

All patients received a sequence of therapies, identified during model conceptualisation (Figure 5.7; Section 5.6), representative of current practice in England, which comprised (i) adalimumab plus methotrexate, (ii) rituximab plus methotrexate, (iii) tocilizumab plus methotrexate, and (iv) methotrexate for their remaining lifetime. The health technologies (treatments and tests) were assumed to be delivered in a secondary-care setting within the NHS.

ADAb and drug level testing was performed when the patient received adalimumab. The decision problem (Table 5.8; Section 5.5) included three relevant comparator strategies: (i) use the ELISA tests to monitor treatment in patients that responded to adalimumab, (ii) use the drug level ELISA test in remission to inform adalimumab dose-reduction strategies, and (iii) reduce the dose of adalimumab in all patients without testing. Choices were required over the frequency of monitoring, and the timing of testing/dose reduction, after responding to adalimumab. Additionally, the ADAb test could be used with, or without, the drug level test when used to monitor treatment. The economic evaluation was therefore structured around thirteen potential comparator strategies (reported in Table 6.1).

Eight of these comparator strategies used testing to monitor treatment, to inform a pre-emptive change of treatment to the next therapy in the sequence (rituximab plus methotrexate), while patients were responding to adalimumab (*Strategies 1-4*, *7-10*). The accuracy of monitoring ADAb alone (*Strategies 7-10*) was imperfect and may have

provided false-positive test results (because the patient truly didn't have adalimumab ADAb and low drug levels). Drug level testing could be used in combination with ADAb testing (*Strategies 1-4*), at a greater cost to the NHS, to verify whether the ADAb test was correct. It was assumed that immunogenicity monitoring could occur every three or six months, to conform with published RA-specific prescribing algorithms identified in Section 5.3 (Garcês et al., 2014; Mok et al., 2016).

Table 6.1. *Thirteen comparator strategies included in the cost-effectiveness analysis of adalimumab ADAb and drug level testing.*

Strategy	Type of Testing Strategy	Description
Current Practice	Not applicable†.	Usual care for RA patients with no testing of ADAb or drug levels.
Strategy 1	Monitoring.	ADAb and drug level testing every **3 months**.
Strategy 2	Monitoring.	ADAb and drug level testing every **6 months**.
Strategy 3	Monitoring & dose reduction.	ADAb and drug level testing every **3 months**, drug level test in remission after **2 years**.
Strategy 4	Monitoring & dose reduction.	ADAb and drug level testing every **3 months**, drug level test in remission after **3 years**.
Strategy 5	Dose reduction.	Drug level test in remission after **2 years**.
Strategy 6	Dose reduction.	Drug level test in remission after **3 years**.
Strategy 7	Monitoring.	ADAb testing only every **3 months**.
Strategy 8	Monitoring.	ADAb testing only every **6 months**.
Strategy 9	Monitoring & dose reduction.	ADAb testing only every **3 months**, drug level test in remission after **2 years**.
Strategy 10	Monitoring & dose reduction.	ADAb testing only every **3 months**, drug level test in remission after **3 years**.
Strategy 11	Not applicable†.	No testing, just half dose in remission after **2 years**.
Strategy 12	Not applicable†.	No testing, just half dose in remission after **3 years**.

Note: †=Strategy did not include testing; Abbreviations: ADAb=Anti-drug antibody.

Six strategies tested adalimumab drug levels after patients responded to treatment for two or three years (*Strategies 3-6, 9-10*). It was assumed that patients who maintained response at two years were characterised as being in remission. The standard course of full-dose adalimumab was halved upon detecting high drug levels by doubling the interval between

adalimumab injections. Four strategies combined both routine monitoring of adalimumab and remission drug level testing (*Strategies 3, 4, 9, 10*).

Three strategies did not include any testing to stratify treatment. Patients that followed the '*Current Practice*' strategy were treated according to the usual care pathway for RA. Two strategies halved adalimumab doses in all patients after responding for two or three years, irrespective of their true drug level (*Strategies 11, 12*).

6.3.1.5. Model Structure

A *de novo* decision analytic model was developed as a DES (Caro et al., 2016b), the logic for which was written in the programming language *R* (R Core Team, 2015). *R* has been used, and recommended, increasingly within the literature to develop decision analytic models that simulate the histories of individual patients over time (Tosh et al., 2008; Holland et al., 2016; Jalal et al., 2017). The DES was built by following the practical guidance on programming individual patient-level simulation models issued by the NICE *Decision Support Unit* (Davis et al., 2014). The DES simulated patients individually through the model structure illustrated in (Figure 5.9; Section 5.8.4.1).

Each individual patient received all thirteen strategies reported in Table 6.1 to facilitate an incremental analysis. Patients remained in the model until their death and received the therapies in the pre-defined sequence. Treatment decisions based on testing adalimumab ADAb and drug levels may have caused a benefit or harm, depending on whether testing provided a true-positive, false-positive, true-negative, or false-negative result. Reduced-dose adalimumab may have caused a patient to flare if they did not have high drug levels *a priori*.

The model was implemented by simulating the progression of each patient's HAQ-DI score over time, referred to hereafter as the *HAQ score*, which quantified the functional ability of each patient (Bruce et al., 2003). The simulation of each patient's individual HAQ score over time has been recommended by a multidisciplinary working party for modelling the relative cost-effectiveness of bDMARD therapies for RA (Madan et al., 2015), and has been performed by other individual-level model-based economic evaluations in RA (these published model-based economic evaluations are described extensively in *Appendix 27*).

Each patient's EULAR response was assumed to be assessed six months after starting any treatment, consistent with current practice in England (National Institute for Health and Care Excellence, 2016a). A EULAR response (*good* or *moderate*) to treatment was assumed to cause a reduction a patient's HAQ score. Upon treatment failure, the HAQ score increased by a magnitude equal to the initial reduction, referred to in the literature as a *perfect rebound* (see *Appendix 27; Treatment Withdrawal*). A patient's HAQ was assumed to gradually worsen over time to reflect the chronic nature of RA and the natural progression of the disease (Madan et al., 2015). QALYs and annual days of hospitalisation were estimated from a patient's HAQ profile over time.

The DES handled the simulation of patients through the model by using the *time-to-event method* (Caro et al., 2016b). The time-to-event method involved advancing the DES simulation clock according to the pre-defined times at which certain events were scheduled to occur for each patient. Each patient had a unique list of events that comprised nine types of event (Table 6.2) that could have occurred over time. The times to five events were estimated by parametric survival analysis; a description of the parametric survival analysis methods that were used in this thesis are reported in *Appendix 29*. The times to events that incorporated testing were determined according to the comparator strategy in Table 6.1. The time to each EULAR response and HAQ progression were fixed by assumption (six months after starting any treatment and annually, respectively, by definition). Each patient's list of events were scheduled in ascending order of time. If a patient developed adalimumab ADAb, the time to adalimumab failure was updated to occur earlier.

Table 6.2. *Nine time-to-event input parameters and their estimation method.*

Event	Estimation Method
Time to death.	Parametric survival analysis.
Time to adalimumab failure.	Parametric survival analysis.
Time to rituximab failure.	Parametric survival analysis.
Time to certolizumab failure.	Parametric survival analysis.
Time to developing adalimumab ADAb.	Parametric survival analysis.
Time to routine ADAb and drug level testing.	Varied by comparator strategy.
Time to remission testing.	Varied by comparator strategy.
Time to EULAR response.	Fixed.
Time to HAQ progression.	Fixed.

6.3.2. Model Population

The relevant population simulated by the model, consistent with the decision problem in Table 5.8; Section 5.5, was representative of all patients with RA in England that were eligible for TNFi therapy according to NICE recommendations. Patients were assumed to be bDMARD-naïve, had failed two previous attempts of cDMARD therapy (at least one being methotrexate), and had a DAS28 score of at least 5.1 indicating high disease activity (National Institute for Health and Care Excellence, 2016a).

Patients were defined at the start of the simulation by three attributes: their (i) age, (ii) sex, and (iii) HAQ score. The majority of published individual-level model-based economic evaluations in RA, identified by the systematic review in *Appendix 27*, have also described their population according to these three attributes (see *Appendix 27; Model Population*). The mean and standard deviation of each attribute (reported in Table 6.3) were obtained from the baseline summary statistics of RA patients that were recruited to the *British Society for Rheumatology Biologics Register – Rheumatoid Arthritis* (BSRBR-RA) cohort in 2004 (Hyrich et al., 2011). The BSRBR-RA cohort was a representative sample of patients with RA in the UK who were prescribed a TNFi therapy according to the eligibility criteria of NICE. The year 2004 was chosen because the BSRBR-RA recruited the largest number of eligible patients (n=3,138) to the cohort in that year.

Table 6.3. *Mean values of patient attributes derived from the BSRBR-RA cohort.*

Attribute	BSRBR-RA Value for Population
Mean Age (Standard deviation).	56.7 years (12.1).
Percentage of Women in cohort.	76%.
Mean HAQ score (Standard deviation).	2.04 (0.56).

Source: Hyrich et al. (2010, pp.119-120).

The values in Table 6.3 were used to simulate each individual patient's (i) age from a normal distribution, (ii) sex from a uniform distribution, and (iii) baseline HAQ from a normal distribution (bounded between zero and three) and rounded to the nearest 0.125 to represent a legitimate HAQ score (Bruce et al., 2003). Attributes were not simulated from correlated distributions because the variance-covariance matrix of the patient-level characteristics was not available. In general, published DES models, in the absence of individual patient data, have also not induced correlation between attributes (Caro et al., 2016b). This modelling decision was unlikely to have impacted the expected cost and QALY outcomes estimated by the model, providing that a sufficient number of patients

were simulated. However, uncorrelated patient-level attributes may have increased the uncertainty associated with the estimate of relative cost-effectiveness.

6.3.3. Model Parameters

An advantage of using a decision analytic model, to inform decision-making, was its ability to synthesise all relevant existing evidence (Shemilt et al., 2010). Every input parameter of any decision analytic model must be estimated (Kaltenthaler et al., 2013). The most appropriate source of evidence may vary depending on nature of the parameter being estimated (Kaltenthaler et al., 2011). The NICE Reference Case required that the sources of evidence to inform a model's parameters were identified in a systematic way (National Institute for Health and Care Excellence, 2013a). Coyle et al. (2010) produced a *hierarchy of data sources for health economic analyses* that described the quality of a source for different types of model parameters (clinical effect size; baseline clinical data; resource use; unit costs; utilities) in terms of how the source was estimated and its relevance to the specific decision problem; an application of the hierarchy to the data sources used to populate this decision analytic model is reported in *Appendix 30*. Section 6.2.3 reports all values of the model's parameters, and how these values were estimated, in the order of: clinical effectiveness (Section 6.3.3.1), QALYs (Section 6.3.3.2), resource use (Section 6.3.3.3), and unit costs (Section 6.3.3.4).

6.3.3.1. Clinical Effectiveness Parameters

Table 6.4 provides a summary of all clinical input parameter values used in the model.

Table 6.4. *Clinical input parameter values for the decision analytic model*

Parameter	Deterministic Analysis	Probabilistic Analysis		Source
		Distribution	Parameters	
Time to Event Parameters				
All-cause mortality: men.	Gompertz Shape: 0.103924; Rate: 0.0000154	Multivariate normal†.	Shape(0.021284) Rate(0.000040) Cov(-0.000263)	Survival analysis (*Section 6.3.3.1.1*.) using ONS (2015) data.
All-cause mortality: women.	Gompertz Shape: 0.1162922; Rate: 0.000004307	Multivariate normal†.	Shape(0.000627) Rate(0.000000) Cov(-0.000007)	Survival analysis (*Section 6.3.3.1.1*.) using ONS (2015) data.
Time to biologic failure.	Weibull Shape: 1.351142; Scale: 4.708286305	Multivariate normal†.	Shape(0.000012) Rate(0.000015) Cov(-0.000005)	Survival analysis (*Section 6.3.3.1.5*.) using Souto et al. (2016) data.
Time to develop ADAb.	Log-normal Mean: 1.146684; SD: 0.7284289	Multivariate normal†.	Shape(0.216342) SD(0.069402) Cov(0.056025)	Survival analysis (*Section 6.3.3.1.7*.) using Bartelds et al. (2011) data.
Time to testing.	Fixed.		Fixed.	Determined by comparator strategy.
RA-specific mortality hazard ratio.	**Baseline HAQ** / **Hazard (95% CI)**			
	0 / 1 (Reference)			
	0.125 - 0.375 / 1.4 (1.1-1.8)	LogNormal.	(0.34, 0.13)	Michaud et al. (2012)
	0.5 - 0.875 / 1.5 (1.2-1.9)		(0.41, 0.12)	
	1 - 1.37 / 1.8 (1.4-2.2)		(0.59, 0.12)	
	1.5 - 1.875 / 2.7 (2.2-3.5)		(0.99, 0.12)	
	2 - 2.375 / 4 (3.1-5.2)		(1.39, 0.13)	
	2.5 - 3 / 5.5 (3.9-7.7)		(1.70, 0.17)	
Clinical Parameters				
Consequence of developing adalimumab ADAb.	**Relative Risk of Loss of Response (95% CI)** 0.47 (0.33-0.65)	LogNormal.	(-0.76, 0.17)	Garcês et al. (2013)
Annual HAQ progression.	**Treatment** / **HAQ Increase**			Previous NICE Appraisals♦.
	bDMARD / 0		Fixed.	
	Methotrexate / 0.045			

Parameter	Deterministic Analysis		Probabilistic Analysis		Source
			Distribution	Parameters	
Treatment Response Parameters					
Treatment	**EULAR Response**	**Probability**			
Adalimumab.	Good; Moderate	0.252 0.448	Dirichlet	(0.252, 0.448, 0.3)	Systematic review and network meta-analysis in Stevenson et al. (2016)
Rituximab.	Good; Moderate	0.242 0.448		(0.242, 0.448, 0.31)	
Tocilizumab.	Good; Moderate	0.568 0.346		(0.568, 0.346, 0.086)	
Methotrexate.	Good; Moderate	0.094 0.357		(0.094, 0.357, 0.549)	
HAQ reduction following treatment response.	**EULAR Response**	**Mean (SE)**			
	Good; Moderate	-0.672 (0.112) -0.317 (0.048)	1-Gamma.	(231.04, 0.0072) (752.82, 0.0017)	Stevenson et al. (2016).
Test Parameters					
	Test Accuracy	**Mean (95% CI)**			
ADAb test.	Sensitivity Specificity	0.32 (0.21-0.46) 0.98 (0.93-0.99)	Beta.	(16.962, 35.714) (62.852, 1.283)	Jani et al. (2016).
Drug level test: *Full dose.*	Sensitivity Specificity	0.95 (0.85-0.98) 0.68 (0.28-0.92)	Multivariate normal†.	LogitSens(0.658) LogitSpec(1.488) Cov(0.388)	Hierarchical meta-analysis in *Section 6.2.3.1.11*.
Half dose.	Sensitivity Specificity	1 (Not reported) 0.93 (Not reported)	Beta.	(23, 0.01) (38.294, 2.706)	Chen et al. (2016).
HAQ multipliers for monitoring test.	**Test Outcome**	**Multiplier (Range)**			
	True-positive; False-positive	0.5 (0-1) 0.5 (0-1)	Uniform.	(0,1) (0,1)	Assumption.
Probability of adalimumab low drug levels in remission.	**Probability (95% CI)**				
	0.23 (0.16-0.32)		Beta.	(24.220, 81.084)	Kuijper et al. (2015).
HAQ increase due to flare.	**HAQ Increase (95% CI)**				
	0.250 (Not reported)		Gamma.	(1, 0.251)	Markusse et al. (2015).
Duration of flare.	One week.			Fixed.	Bykerk et al. (2014).

Note: †=Values for the multivariate normal distribution obtained from Cholesky decomposition of the variance-covariance matrix; ♦=Malottki et al. (2011) and Stevenson et al. (2016); Abbreviations: ONS=Office for National Statistics.

6.3.3.1.1. Time to Death

Each patient required an estimate of their time to death, in order to define their lifetime duration within the model. This estimate should be representative of the target population of patients with RA in England. All-cause mortality data were obtained from the most recent national life tables (for the years 2012 to 2014) for England published by the *Office for National Statistics* (Office for National Statistics, 2015). A parametric survival analysis was performed separately on the data for men and women (reported in full in *Appendix 31*).

Gompertz survival curves fit the observed mortality data best for both men and women according to the AIC and BIC test statistics (Akaike, 1974; Schwarz, 1978). The Gompertz function for (i) men was defined by a shape parameter of 0.103924 and rate parameter of 0.0000154, and for (ii) women was defined by a shape parameter of 0.1162922 and a rate parameter of 0.000004307. Both curves followed a similar trajectory but the male Gompertz curve was further to the left, which demonstrated a greater risk of all-cause mortality relative to women.

6.3.3.1.2. Mortality Adjustment for Rheumatoid Arthritis

Patients with RA have a greater risk of mortality than the general population (Kvien, 2004). Therefore, the time to death estimated from the general population all-cause mortality data required an adjustment to account for this increased mortality in the target population. The individual-level model-based economic evaluations identified by the systematic review in *Appendix 27* reported that this adjustment could be made by making mortality a function of the patient's HAQ score (see *Appendix 27; Mortality*).

Following the modelling approach of Stevenson et al. (2016), which provided evidence for the most recent NICE technology appraisal of TNFi therapies for RA, the RA-specific mortality was related to each patient's HAQ by using the evidence from Michaud et al. (2012). Michaud et al. (2012) used patient-level data from 10,319 patients with RA enrolled to the *National Data Bank for Rheumatic Diseases* longitudinal cohort, to estimate the predictive ability of their baseline HAQ (at study entry) on mortality. The results demonstrated that a worse (higher) baseline HAQ was associated with a greater risk of mortality.

Michaud et al. (2012) estimated their results as hazard ratios across seven categories of HAQ score, reported in Table 6.4. A *hazard ratio* provided the likelihood that an event (mortality) would have occured in an intervention group (in this case, defined by baseline HAQ score) relative to a control group (in this case, a HAQ equal to zero) (Blagoev et al., 2012).

The decision analytic model applied these hazard ratios to the all-cause mortality survival curves, conditional on the value of each individual patient's simulated baseline HAQ attribute (Caro et al., 2016b). The greater a patient's HAQ at model entry, the greater the hazard ratio applied to their survival curve, and the more the survival curve shifted to the left, resulting in a higher probability of earlier mortality. The implicit assumption of modelling RA-specific mortality in this way was that an improvement in HAQ over the duration of treatment did not affect the probability of death.

6.3.3.1.3. EULAR Response to Treatment

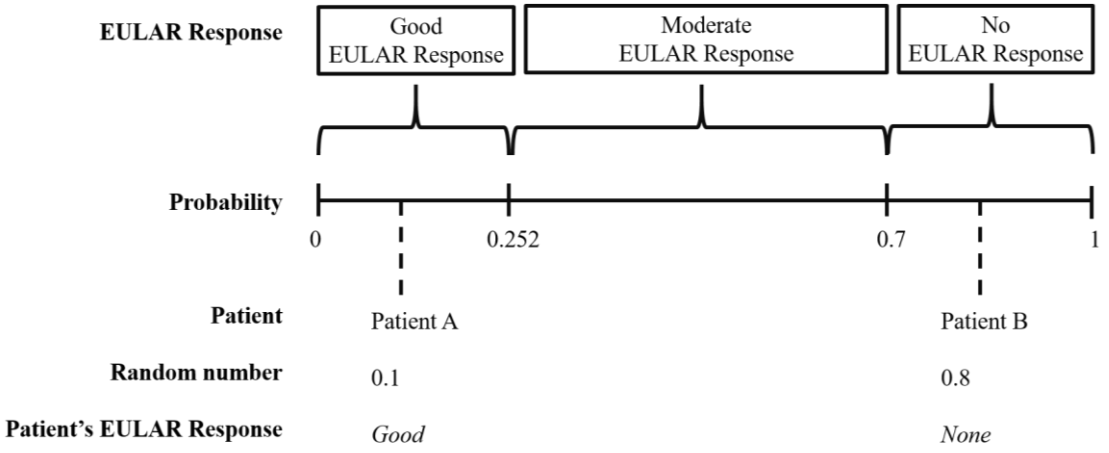
Response to all therapies was represented as a EULAR response that occurred six months from treatment initiation, consistent with current practice in England (National Institute for Health and Care Excellence, 2016a). Two published individual-level model-based economic evaluations for RA have modelled treatment response in terms of a EULAR response (see *Appendix 27; Initial Treatment Response*). Estimates of six-month treatment efficacy were derived from the independent systematic review and network meta-analysis of published RCT evidence, performed by Stevenson et al. (2016, p. 84, Table 31) for the most-recent technology appraisal of TNFi therapies for RA by NICE. A *network meta-analysis* is a method of evidence synthesis that uses all available trial evidence to facilitate comparisons between treatments that have not necessarily been evaluated in a head-to-head RCT, but which instead have a common comparator treatment (Jansen et al., 2011).

The network meta-analysis by Stevenson et al. (2016) was the most suitable source of treatment efficacy for this study, relative to the other relevant published network meta-analyses (Orme et al., 2012; Jansen et al., 2014; Buckley et al., 2015; Hazlewood et al., 2016; Singh et al., 2016a), because treatment efficacy was: (i) evaluated in terms of EULAR response rather than ACR response, which is used in clinical practice in England, and (ii) trials only included patients that had previously failed methotrexate, which was consistent with the target population for this economic evaluation. The probabilities of

achieving a *good* and *moderate* EULAR response to each treatment are reported in Table 6.4.

As multinomial probabilities, by definition, must sum to one, a random number was drawn from a uniform distribution, individually for each patient, to determine their response to each treatment in the model (Caro et al., 2016b). Figure 6.1 illustrates the simulation of a EULAR response to adalimumab plus methotrexate for two different patients using the values derived from the published network meta-analysis. The probabilities were arranged such that: a *good EULAR response* was associated with random numbers up to 0.252; a *moderate EULAR response* was associated with random numbers between 0.252 and 0.7; and *no EULAR response* was associated with random numbers greater than 0.7. In the example shown in Figure 6.1, *Patient A* simulated the number 0.1 from a uniform distribution and subsequently achieved a good EULAR response to therapy, whereas *Patient B* simulated the number 0.8 from the same uniform distribution and subsequently achieved no EULAR response.

Figure 6.1. *Example of simulating a EULAR response to adalimumab plus methotrexate.*



Random Number from Uniform Distribution (x)	EULAR Response
$x \leq 0.252$	Good
$0.252 < x \leq 0.7$	Moderate
$x > 0.7$	None

Source: Probability of a EULAR response (van Gestel et al., 1998) was derived from the network meta-analysis of Stevenson et al. (2016).

6.3.3.1.4. HAQ Improvement following EULAR Response

A EULAR response, six months after starting any treatment, was assumed to be associated with an improvement (reduction) in a patient's HAQ. Approximately ninety percent of the individual-level model-based economic evaluations for RA published previously have

assumed an immediate reduction in HAQ following treatment response (see *Appendix 27; Disease Progression - Initial Treatment Response*).

The observed HAQ improvement was estimated by Stevenson et al. (2016) using data from 10,186 patients with RA in England enrolled in the *BSRBR* register. The mean reduction in HAQ at six months following treatment initiation was estimated conditional on the following patient characteristics, measured at enrolment to the BSRBR: age; sex; disease duration; DAS28; and previous number of cDMARDs. The reduction in HAQ was estimated to be greater for patients with a *good EULAR response* (-0.672) compared with those with a *moderate EULAR response* (-0.317). Patients with *no EULAR response* were assumed to experience no change in their HAQ (Stevenson et al., 2016).

The two assumptions underlying treatment response were that: (i) no EULAR response did not directly worsen a patient's functional ability, and (ii) the improvement in HAQ did not vary between different treatments (for example, a good EULAR response to adalimumab was equivalent to a good EULAR response to rituximab). These assumptions have been made previously in published individual-level model-based economic evaluations for RA (see *Appendix 27; Disease Progression – Initial Treatment Response*).

6.3.3.1.5. Time to Treatment Failure

Patients that had a EULAR response to any bDMARD were assumed to respond to treatment for a finite period of time, to represent the occurrence of secondary non-response. The model therefore required an estimate of the time that each patient would fail each treatment. As a DES was used, the most appropriate way to estimate the time to treatment failure was to assign a time for each individual patient; an alternative approach, if the model had used time cycles, would have been to apply a probability of treatment failure in each cycle (see *Appendix 27; Treatment Failure*).

Souto et al. (2016) performed a systematic review and meta-analysis of all studies that estimated the discontinuation of bDMARDs in patients with RA using data from registry or health care databases. The analysis included ninety-eight studies, comprising over 200,000 unique patients with RA. The meta-analysis results reported the annual percentage of RA patients that discontinued bDMARD therapy over a four-year period.

This decision analytic model used the annual probabilities of treatment discontinuation that were estimated for all TNFi therapies. A parametric survival analysis (reported in *Appendix 32*) was performed on the data from Souto et al. (2016). A Weibull survival curve was chosen, on the basis of clinical plausibility, to model the time to treatment failure, defined by a shape parameter of 1.351142 and a scale parameter of 4.708286305. The model simulated each patient's time to treatment failure by: (i) drawing a random number from a uniform distribution and (ii) determining the corresponding time according to the Weibull survival curve. Due to limitations of data availability, the time to treatment failure for all bDMARDs were simulated from the same Weibull curve. Patients that had received all three bDMARDs in the treatment sequence were assumed to receive methotrexate for the remainder of their lives as palliative care.

6.3.3.1.6. Annual HAQ Progression

A patient's HAQ score was assumed to progress over time to represent the gradual worsening of RA (Madan et al., 2015). Following the modelling approach used in two previous NICE technology appraisals for RA (Malottki et al., 2011; Stevenson et al., 2016), each patient's HAQ score increased annually by: (i) 0.045 units when prescribed methotrexate only, and (ii) by zero units when prescribed any bDMARD. This modelling assumption implied that the progression of RA was prevented for the duration of treatment with a bDMARD.

6.3.3.1.7. Time to Developing ADAb against Adalimumab

A proportion of patients that received adalimumab were expected to develop ADAb against their treatment, which was a key biomarker of interest to stratify subsequent treatment decisions (see *Chapter 1*, Section 1.3.5). Therefore, the model required an estimate of the time to developing adalimumab ADAb.

A systematic review was conducted to identify studies that reported the time to developing adalimumab ADAb (reported in *Appendix 33*). The most appropriate study identified by this review was Bartelds et al. (2011), which assessed the development of ADAb in RA patients the received adalimumab over a duration of three years. ADAb were measured by radioimmunoassay and adalimumab drug levels were measured by a sandwich-ELISA. Patients were defined as ADAb-positive if their serum had: (i) ADAb titres greater than twelve arbitrary units per millilitre and (ii) serum drug levels lower than five milligrams

per litre (Bartelds et al., 2011). The thresholds used to define a positive ADAb test result were expressed as *arbitrary units* as a pragmatic solution, in the absence of published evidence on the correct cut-off threshold (Wadhwa et al., 2011).

The percentage of patients that developed adalimumab ADAb over the three-year study period was illustrated graphically (Bartelds et al., 2011, p.1463). Following the approach recommended by Guyot et al. (2012), the graphical data points were recreated using the *DigitizeIt* software to calculate the number of patients that developed ADAb at each follow-up time period (Bormann, 2016).

Parametric survival analysis was performed on the estimated data points (reported in *Appendix 33*). A log-normal survival curve fit the data best according to the AIC and BIC statistics (Akaike, 1974; Schwarz, 1978), defined by a mean of 1.146684 and a standard deviation of 0.7284289. The model simulated each patient's time to developing ADAb by: (i) drawing a random number from a uniform distribution, and (ii) determining the corresponding time according to the log-normal survival curve. For consistency with the data reported by Bartelds et al. (2011), all patients that developed adalimumab ADAb in the model were also assumed to have low drug levels.

6.3.3.1.8. Consequence of Developing ADAb against Adalimumab

The clinical literature demonstrated that patients who developed adalimumab ADAb were more likely to experience secondary non-response to treatment (described in Section 1.3.5). In the model, the time to adalimumab failure was reduced to occur earlier for patients that developed adalimumab ADAb, relative to those patients without ADAb. Therefore, the model required an estimate of the consequence of adalimumab ADAb on treatment failure.

Garcês et al. (2013) report a systematic review and meta-analysis of studies that estimated the impact of developing ADAb against a TNFi therapy on a patient's response to that TNFi. The meta-analysis estimated that RA patients with detectable ADAb against a TNFi had a reduced rate of response to treatment by fifty-three percent (relative risk = 0.47, 95% confidence interval = 0.33-0.65) compared with those patients without detectable ADAb (Garcês et al., 2013).

The model applied the RA-specific relative risk of treatment failure to the estimated time to adalimumab failure if the patient had developed adalimumab ADAb (Caro et al., 2016b). For example, a patient may have started the simulation with an estimated time to adalimumab failure of five years. However, if they developed ADAb over the duration of their adalimumab therapy, their time to adalimumab failure would have been updated to occur earlier at 2.35 years ($= 5 \times 0.47$).

6.3.3.1.9. Proportion of Patients with Low Adalimumab Drug Levels in Remission

The clinical evidence indicated that a proportion of patients may flare in disease activity after reducing the dose of their TNFi (Bykerk et al., 2016). The model therefore required an estimate of this proportion as an input parameter.

Kuijper et al. (2015) conducted a systematic review and meta-analysis of studies that investigated the risk of a flare in disease activity in patients with RA, following a de-escalation of TNFi therapy while in low disease activity or remission. The meta-analysis reported the results as the pooled flare rate per patient year of 0.26 (95% CI: 0.17-0.39) for good quality studies. The rate per patient year was converted into an annual probability by using the formula in Equation 6.1, reported in Briggs et al. (2006):

$$\text{Annual Probability} = 1 - Exp(-Rate\ per\ patent\ year) \qquad \textbf{(Equation 6.1)}$$

Using Equation 6.1 and the evidence synthesised by Kuijper et al. (2015), the probability of a disease flare, one year after TNFi dose reduction, was calculated as 23% (95% CI: 16% to 32%). The model assumed that patients only flared due to low drug levels, and by implication, twenty-three percent of patients had low adalimumab drug levels during remission.

6.3.3.1.10. Consequence of Flare in Remission

A flare in disease activity was assumed to have a detrimental impact on a patient's health, expressed as an increase in their HAQ for a short period of time; the model therefore required an estimate of: (i) the duration of a flare and (ii) the consequence of a flare on HAQ.

Markusse et al. (2015) estimated the impact of a flare in disease activity on the functional ability of patients with RA (expressed as a change in HAQ) recruited to the *BeSt* study in The Netherlands. A flare in disease activity was defined as an increase in DAS28 of greater than 0.6 units in patients that had a DAS28 score of at least 2.4. The HAQ of patients that flared, relative to those that did not, was estimated to have increased by 0.251 units (Markusse et al., 2015). Legitimate HAQ scores can only be expressed in multiples of 0.125 (Bruce et al., 2003), so the model therefore assumed that patients that flared after receiving reduced-dose adalimumab experienced a HAQ increase of 0.250. The model assumed, conservatively, that the duration of a flare was one week, based on the observational evidence reported by Bykerk et al. (2014).

6.3.3.1.11. Test Accuracy

The NICE DAP manual recommended explicitly that test accuracy should ideally be estimated from evidence identified by a systematic review (National Institute for Health and Care Excellence, 2011a). Test accuracy studies estimate the performance (sensitivity and specificity) of a test that measures a particular biomarker against the performance of a reliable reference standard using a receiver operating characteristic (ROC) curve (Macaskill et al., 2010). The model required estimates of test accuracy for three testing scenarios in RA patients receiving adalimumab: (i) accuracy of measuring ADAb by ELISA relative to radioimmunoassay; (ii) accuracy of measuring drug levels by ELISA to identify patients that would maintain response to full-dose adalimumab; and (iii) accuracy of measuring drug levels by ELISA to identify patients that would maintain response to reduced-dose adalimumab. The systematic review that identified these studies is reported in full in *Appendix 34*. The quality of test accuracy studies included in the review were appraised using the *Quality Assessment of Diagnostic Accuracy Studies – 2* (QUADAS-2) checklist (Whiting et al., 2011).

Accuracy of Adalimumab ADAb ELISA Relative to Radioimmunoassay

The time to developing adalimumab ADAb, estimated using evidence from Bartelds et al., (2011) in Section 6.3.3.1.7, measured ADAb with a radioimmunoassay. The systematic review (see *Appendix 34*) identified one study that assessed the concordance between measuring adalimumab ADAb by radioimmunoassay (the reference standard) and by ELISA. Jani et al. (2016b) measured adalimumab ADAb with a radioimmunoassay and ELISA from 159 serum samples of RA patients treated with adalimumab in England.

Following the manufacturer's instructions, the cut-points chosen to define ADAb positivity were 12 AU/mL for the radioimmunoassay (the same as in Bartelds et al. (2011)) and 3.5 AU/mL for the ELISA. Relative to detecting adalimumab ADAb with radioimmunoassay, the ELISA test had a sensitivity of 32.2% (95% CI: 20.6% to 45.6%) and a specificity of 98% (95% CI: 93% to 99%).

Accuracy of Drug Level ELISA and Response: Full Dose Adalimumab

The systematic review identified four studies that used a ROC analysis to estimate the accuracy of measuring adalimumab drug levels by ELISA to predict an observed treatment response (the reference standard) (Rosas et al., 2014; Chen et al., 2015a; Jani et al., 2015a; Pouw et al., 2015). The four studies were heterogeneous in the drug level cut-points assumed, measures of outcome, and patient populations. *Appendix 34* reports a bivariate meta-analysis that was used to synthesise these test accuracy data (Reitsma et al., 2005; Macaskill et al., 2010). The pooled estimate of sensitivity (the probability of a normal drug level for responders) was 95% (95% CI: 85% to 98%) and specificity (the probability of a low drug level for non-responders) was 68% (95% CI: 28% to 92%).

Accuracy of Drug Level ELISA and Response: Reduced Dose Adalimumab

The model required an estimate of the accuracy of measuring adalimumab drug levels by ELISA for patients in remission to distinguish between those patients that would, and would not, maintain response after receiving reduced dose adalimumab. The systematic review in *Appendix 34* identified one study that estimated a drug level cut-point by using a ROC analysis in patients with RA that received reduced-dose adalimumab. Chen et al. (2016) measured adalimumab drug levels by sandwich ELISA in twenty-five patients with RA who were (i) already in remission (DAS28 < 2.6) and (ii) were receiving half-dose adalimumab (40mg every month) plus methotrexate. The reference standard was whether patients remained in remission at twenty-four weeks after receiving half-dose adalimumab with methotrexate. Chen et al. (2016) estimated that a drug level cut point of 6.4ug/mL could predict persistent remission, with a sensitivity of 100% and a specificity of 93.4%. No standard error or confidence interval was reported for this published estimate.

6.3.3.1.12. Consequence of Treatment Decisions According to Test Results

The ADAb and drug level tests had the potential to report false-positive and false-negative results, given their imperfect accuracy. The model assumed that the consequences of treatment decisions based on the ELISA test results affected a patient's HAQ score. These assumed consequences are reported in Table 6.5.

Treatment Decisions following Routine Monitoring

The model assumed that when any treatment failed, the patient's HAQ increased by a magnitude equal to the initial reduction associated with the EULAR response (a *perfect rebound*). The benefit and harm of treatment decisions according to the routine testing for adalimumab ADAb and drug levels was assumed to affect the HAQ rebound by a multiplier, summarised in Table 6.5.

Table 6.5. *Consequence of stratified treatment decisions according to test results.*

Test Outcome	True Patient State	Treatment Decision	Health Consequence of Treatment Decision
Test: Routine Monitoring of Adalimumab ADAb to Inform Treatment Change			
True-positive.	ADAb positive; Low drug level.	Change treatment to rituximab.	HAQ rebound by: (HAQ Rebound) x α
False-positive.	ADAb negative; Normal drug level.	Change treatment to rituximab.	HAQ rebound by: (HAQ Rebound) + [(HAQ Rebound) x β]
True-negative.	ADAb negative; Normal drug level.	Continue adalimumab.	None.
False-negative.	ADAb positive; Low drug level.	Continue adalimumab.	None.
Test: Adalimumab Drug Levels in Remission to Inform Dose Reduction			
True-positive.	Normal drug level.	Half adalimumab dose.	None.
False-positive.	Low drug level.	Half adalimumab dose.	Flare.
True-negative.	Low drug level.	Continue adalimumab.	None.
False-negative.	Normal drug level.	Continue adalimumab.	None.

Note: α,β were multipliers, between zero and one, that adjusted the HAQ rebound upon changing treatment to to rituximab; Abbreviations: ADAb=anti-drug antibody.

Patients that tested positive for immunogenicity ((i) ADAb positive or (ii) ADAb positive and low drug levels) had their adalimumab therapy pre-emptively changed to rituximab. The use of rituximab was relevant to clinical practice in England, justified by the model conceptualisation procedure in *Chapter Five*.

- In patients with a true-positive test result (the patient truly was ADAb-positive and had low drug levels), the treatment change was assumed to be beneficial as secondary non-response to adalimumab was avoided. This benefit was represented by the multiplier $\alpha \in [0,1]$, such that the patient's HAQ rebound due to changing therapy was *less* than the initial reduction in HAQ.
- In patients with a false-positive test result (the patient truly was ADAb-negative and had normal drug levels), the treatment change was assumed to be inappropriate and subsequently caused harm. This harm was represented by the multiplier $\beta \in [0,1]$, such that the patient's HAQ rebound was *greater* than the initial reduction in HAQ.

The values of the HAQ multipliers were unknown. The base-case analysis used the values $\alpha = 0.5$ and $\beta = 0.5$ and subsequent sensitivity analyses varied these values extensively.

Patients that had a negative test result for immunogenicity were assumed to continue adalimumab therapy. Patients with a true-negative test result (the patient was truly ADAb-negative and had normal drug levels) were assumed to receive continued benefit from adalimumab therapy. Patients with a false-negative test result (the patient was truly ADAb-positive and had low drug levels) were assumed to experience earlier secondary failure of adalimumab, compared with those patients without adalimumab ADAb.

Treatment Decisions following Remission Testing

The drug level test in remission identified patients that could maintain remission following a reduction in adalimumab, according to whether their adalimumab drug levels were high. There was no QALY benefit assumed to be associated with reducing the dose of adalimumab; however, inappropriate dose-reduction was assumed to cause a flare in disease activity (Section 6.3.3.1.9). The consequences of treatment decisions based on remission testing are reported in Table 6.5.

The dose of adalimumab was assumed to be halved if testing reported drug levels to be high. Patients with a true-positive test result (the patient truly had high drug levels) experienced change in QALYs following adalimumab dose-reduction. In contrast, patients with a false-positive test result (the patient truly had low drug levels) experienced a flare and an adjustment in QALYs (see Section 6.3.3.1.10) because reduced-dose adalimumab was inappropriate. No adjustments were made to treatment if the test result reported that the patient had low adalimumab drug levels, resulting in no negative impact on health consequences. However, patients with a false-negative test result (the patient truly had high drug levels) could have been treated with reduced-dose adalimumab with no harmful health consequences.

6.3.3.2. Quality-adjusted Life Years

QALYs were calculated by estimating a patient's health-related quality of life, informed by the EQ-5D instrument (EuroQol Group, 1990), and were discounted at 3.5% per year, in accordance with the NICE Reference Case (National Institute for Health and Care Excellence, 2013a). The health-related quality of life weights were assumed to be a function of the HAQ score, following the approaches of similar individual-level model-based economic evaluations for patients with RA (see *Appendix 27; QALYs*). Therefore, the model required an equation, referred to as a *mapping algorithm*, to represent the relationship between HAQ and EQ-5D (Dakin, 2013).

There were many published mapping algorithms available to estimate EQ-5D from HAQ in patients with RA (Pennington et al., 2014). The base-case analysis of this model used the quadratic mapping algorithm estimated previously for the *NICE Technology Appraisal 195* by Malottki et al. (2011) which used: (i) data from patients with RA in the UK to measure health states, and (ii) UK-specific EQ-5D-3L tariff data to estimate the value of health states. This mapping algorithm is reported in Equation 6.2.

$$EQ5D = a - b_1 HAQ - b_2 HAQ^2 \qquad \textbf{(Equation 6.2)}$$

where a=0.804 (95% CI: 0.711-0.897), b_1=0.203 (95% CI: 0.054-0.351), and b_2=0.045 (95% CI: -0.007-0.096). Subsequently, by using this equation to estimate a patient's EQ-5D, as their HAQ increased (and became worse), their estimated QALY gain reduced.

6.3.3.3. Resource Use

The evidence used to estimate resource use is explained in the following three sections for: treatments (Section 6.3.3.3.1); hospitalisations (Section 6.3.3.3.2); and testing (Section 6.3.3.3.3). All resource use assumptions are reported in Table 6.6.

6.3.3.3.1. Treatments

All treatments were administered according to their recommended dose (see Table 6.6) in the *British National Formulary* (2016). All bDMARD therapies were assumed to be co-prescribed with methotrexate. Reduced-dose adalimumab assumed that the time between the administration of injections doubled (from every two weeks to every four weeks) consistent with recommendations by EULAR (Smolen et al., 2014). Patients that flared from reduced-dose adalimumab were assumed to receive one course of corticosteroid therapy (methylprednisolone), administered by intramuscular injection, prior to reverting their treatment to full-dose adalimumab.

Administration of therapies by intravenous infusion (rituximab, tocilizumab) was assumed to last for one hour. Following the evidence generated for *NICE Technology Appraisal 375* by Stevenson et al. (2016), administration of subcutaneous adalimumab was assumed to be performed by a nurse in ten percent of cases. The resources associated with the monthly monitoring of a patient's full blood count and biochemical profile were omitted from the analysis because they were common to every strategy in the model.

Table 6.6. *Model input parameter values for resource utilisation and unit costs.*

Treatment Cost (bDMARD, cDMARD, and Corticosteroid)

Treatment	Dose[a]	Unit Cost [a] (£; 2015/16)	Annual Cost (£; 2015/16)
Adalimumab† (Full dose).	40mg every two weeks, subcutaneous injection.	£352.14 per 40mg.	£9,155.64.
Adalimumab† (Half dose).	40mg every four weeks, subcutaneous injection.	£352.14 per 40mg.	£4,577.82.
Biosimilar adalimumab▲ (Full dose).	40mg every two weeks, subcutaneous injection.	£352.14 per 40mg.	£6103.76.
Biosimilar adalimumab▲ (Half dose).	40mg every four weeks, subcutaneous injection.	£352.14 per 40mg.	£3051.88.
Rituximab†.	2g every nine months, intravenous infusion.	£873.15 per 500mg.	£4,656.8.
Tocilizumab†.	800mg every four weeks, intravenous infusion.	£512 per 400mg.	£12,288.
Methotrexate♦.	20mg weekly, oral tablet.	£0.082 per 2.5mg.	£34.03.
Methylprednisolone♦.	120mg for one week, intramuscular injection.	£8.88 per 120mg.	£8.88.

Treatment Administration Cost

Resource	Quantity[b]	Unit Cost [b] (£; 2015/16)
Intravenous infusion.	One hour.	£154 per infusion.
Nurse-led subcutaneous injection.	Ten percent of injections.	£2.61 per injection.

ADAb and Drug Level ELISA Test Cost

Test Scenario	Unit Cost [c] (£; 2015/16)
Single test.	£133.78 per patient.
Concurrent testing.	£152.52 per patient.

Hospitalisation Cost

HAQ	Mean Days Hospitalised Per Year[d]	Annual Cost[e] (£; 2015/16)
0 < HAQ < 0.5	0.26	£160.12
0.6 < HAQ < 1	0.13	£80.06
1.1 < HAQ < 1.5	0.51	£314.07
1.6 < HAQ < 2	0.72	£443.40
2.1 < HAQ < 2.6	1.86	£1,145.44
2.6 < HAQ < 3	4.16	£2,561.85

Note: †=Branded pharmaceutical; ▲=One-third price discount for biosimilars assumed; ♦Generic pharmaceutical. Sources: [a]=British National Formulary (2016); [b]=Stevenson et al. (2016); [c]=Jani et al. (2016a); [d]=Roche (2006, p.110); [e]=Department of Health (2016b); In the PSA, resources required for testing and treatments were assumed to be fixed, and resources for hospitalisations were simulated from gamma distributions (described further in *Appendix 37*).

6.3.3.3.2. Hospitalisations

Hospitalisations were assumed to be positively related to a patient's HAQ score. Previous individual-level model-based economic evaluations for RA have also assumed a positive relationship between each patient's HAQ score and their frequency of hospitalisation (see *Appendix 27; Direct Medical Costs*). The mean annual days hospitalised per patient were sourced from the original submission of evidence by the manufacturer, *Roche*, during the NICE Single Technology Appraisal of rituximab (Roche., 2006). Estimated days of hospitalisation (Table 6.6) were based on data from the *Norfolk Arthritis Register* (Symmons et al., 2003) cohort, and were considered to be conservative estimates because only patients with a HAQ score greater than 2.1 experienced at least one day hospitalised per year. These values have been used in different published model-based economic evaluations for RA, including in the generation of evidence for the most recent NICE technology appraisal for TNFi therapies by Stevenson et al. (2016). A potential limitation of these data, however, was that the relationship between HAQ and the mean number of hospitalisations was not monotonic.

6.3.3.3.3. Testing

To identify and quantify the resources required for testing TNFi ADAb and drug levels by ELISA in routine clinical practice, a microcosting study was performed in collaboration with Dr. Meghna Jani (see *Appendix 35* for a version of the published manuscript). This microcosting study estimated the direct health care resources per patient that were required for testing, during: (i) the pre-testing phase, (ii) the analysis of samples, and (iii) the determination of the treatment decision (Jani et al., 2016a). The microcosting study produced essential evidence for the decision analytic model and was performed as a collaborative study to supplement the main content of this thesis.

6.3.3.4. Unit Costs

All unit costs were expressed in pound sterling (£) at a price year of 2015/16 and no price inflation was performed. Costs were discounted at 3.5%, consistent with the NICE reference case (National Institute for Health and Care Excellence, 2013a).

6.3.3.4.1. Treatments

The unit costs of all treatments were obtained from the *British National Formulary* (2016), reported in Table 6.6. Following the evidence generated by Stevenson et al. (2016) for the *NICE Technology Appraisal 375* for RA, a sixty minute intravenous infusion was assumed to cost £154 and the ten percent of nurse-administered subcutaneous injections was assumed to cost an average of £2.61 per injection.

6.3.3.4.2. Hospitalisations

The unit cost of one day hospitalised was £615.83. This value was the unit cost of a non-elective short stay in secondary care, identified in the *NHS National Schedule of Reference Costs 2015-2016* (Department of Health, 2016b)**.**

6.3.3.4.3. Testing

The unit cost of testing both ADAb and drug levels together by ELISA (concurrent testing) was £152.52 per patient. The unit cost of testing either the ADAb or drug level ELISA only was £133.78 per patient (Jani et al., 2016a).

6.3.4. Model Validation

Two approaches were taken to ensure the validity of the decision analytic model. The model was built by taking into account best-practice strategies for reducing variability in simulation analyses (see Section 6.3.4.1) and by assessing its internal validity (see Section 6.3.4.2).

6.3.4.1. Strategies to Reduce Variability

Variability in the chance occurrence of events within any simulation analysis may reduce the precision of the estimated outcomes (costs and QALYs) of interest (Caro et al., 2016b). Such variability in a decision analytic model, in particular, is unhelpful because it cannot be reduced by further collection of data, unlike the uncertainty in the values of input parameters (Briggs et al., 2006). The conventional approach to reduce undesirable variability in the outcomes of a DES model was to increase the number of individual patients simulated through the model (Caro et al., 2016b). However, a larger sample size

may have been computationally expensive and may have increased the model run time (Griffin et al., 2006). Therefore, best-practice variability-reduction strategies were implemented to increase the precision of the simulated expected outcomes, reduce the sample size required to achieve those expected outcomes, and reduce model run times (Karnon et al., 2012). This DES model used three strategies to reduce undesirable variability and improve the validity of the simulation:

(i) The same set of patients experienced every comparator strategy in Table 6.1. A particular comparator strategy may have otherwise appeared more cost-effective by chance if, for example, the simulated patients were healthier than those simulated for a different comparator strategy (Karnon et al., 2012; Caro et al., 2016b);

(ii) The generation of random numbers within the model was seeded, following best-practice recommendations for simulation analyses (Davis et al., 2014). Seeded random numbers ensured that the same set of random numbers were used if the model was run more than once, which eliminated any undesirable variability in outcomes between different runs of the model.

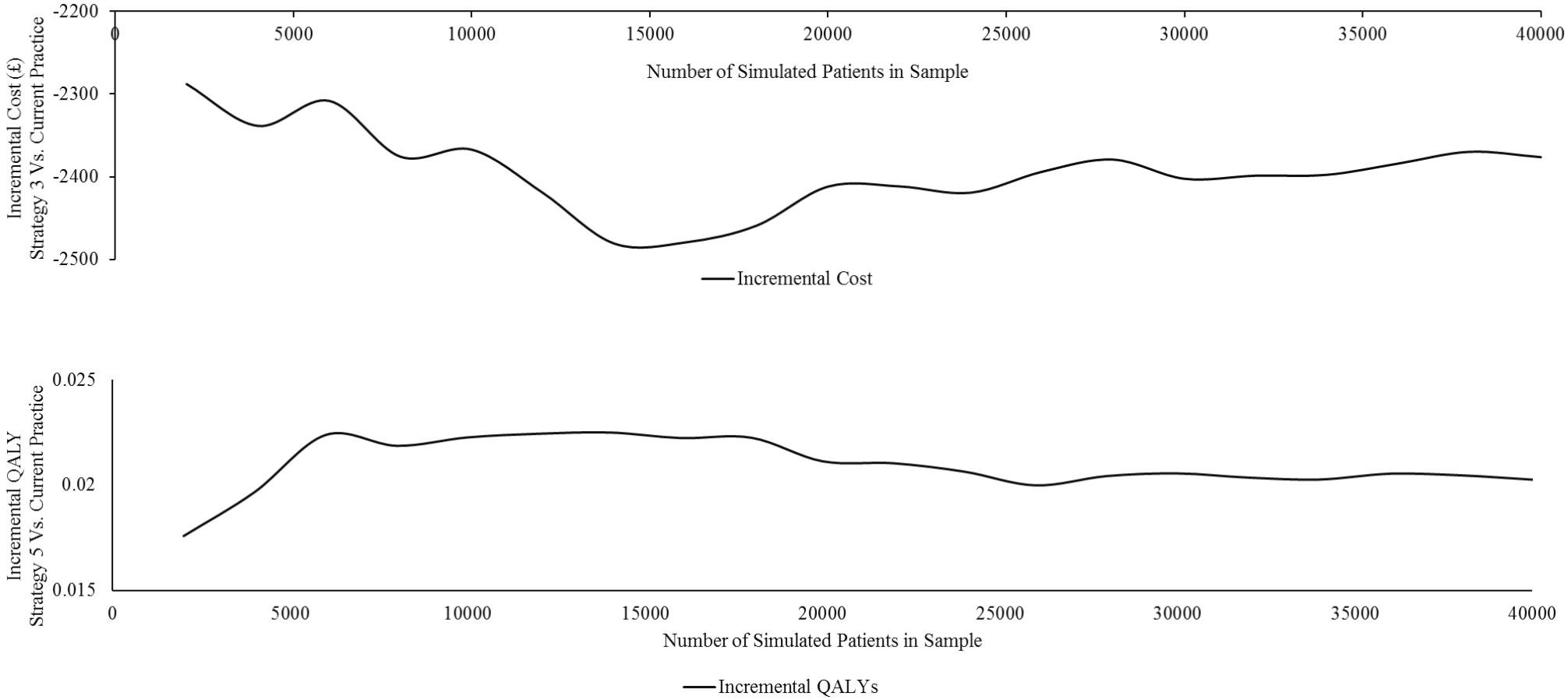
(iii) The probabilities associated with EULAR responses and event times were fixed in advance for each individual patient before model entry, and did not vary between the comparator strategies being simulated. This variability-reduction strategy is known as using *common random numbers* (Stout et al., 2008; Karnon et al., 2012; Murphy et al., 2013). QALY and cost differences between strategies may have otherwise occurred by chance due to a different random simulation of an individual-level probability. For example, a particular comparator strategy may have appeared more cost-effective if the random number used to simulate time to death from the Gompertz survival curve was more favourable (and increased the patient's lifespan) compared with the value simulated for the other strategies. Therefore, any estimated differences in costs and QALYs were only due to differences in the decisions to stratify treatment between the comparator strategies.

The sample size of patients that were simulated through the model was chosen to be 20,000 individuals. To justify this decision, a simulation was performed with 40,000 patients (run time = 24 hours) and the expected cost and QALYs were evaluated for the most complex strategy (*Strategy 3*) for an increasing sample size (Davis et al., 2014). Figure 6.2

illustrates the estimated expected incremental cost and QALYs for *Strategy 3* compared with current practice, as the sample size increased.

The difference in incremental costs as the sample size increased was approximately £35 (Incremental cost = -£2,411 for 20,000 patients; -£2,376 for 40,000 patients), and in incremental QALYs was approximately 0.0008 QALYs (Incremental QALYs = 0.021132 for 20,000 patients; 0.020260 for 40,000 patients). Therefore, the variation in incremental costs and QALYs appeared to have stabilised by 20,000 patients.

Figure 6.2. *Determine sample size: expected (top) incremental cost and (bottom) incremental QALY for Strategy 3 versus. Current Practice over increasing sample size.*



6.3.4.2. Internal Validity

Assessing the validity of a model may improve the credibility of the final estimated outcomes (Tappenden et al., 2014). A quantitative model that has included all relevant elements for a specific decision problem is said to have *face validity* (Caro et al., 2016b). *Internal validity* refers to whether a quantitative model is implemented correctly, for example, in terms of its internal logic, programming, and calculations (Caro et al., 2016b). The DES model in this study was implemented by running a computer code written in a general programming language (*R*). All computer codes have the potential for errors, known in the literature as *bugs* (Caro et al., 2016b). The following four debugging internal validity checks were performed to assess whether the model ran correctly:

(i) Set all unit costs equal to zero, to assess whether total costs over the simulation also equalled zero;
(ii) Set all parameters to calculate QALYs from a HAQ score equal to zero, to assess whether total QALYs over the simulation also equalled zero;
(iii) Increase the discount rate for costs and QALYs, to assess whether the total costs and QALYs reduced;
(iv) A histogram of 100,000 random draws from each survival curve within the model was produced to ensure that the simulation generated a clinically plausible distribution of times to each event.

The results of all internal validity checks are reported in *Appendix 36*.

6.3.5. Analysis

The analysis section describes how the results of the economic evaluation were obtained for the base-case (Section 6.3.5.1.), deterministic sensitivity analysis (Section 6.3.5.2), probabilistic sensitivity analysis (Section 6.3.5.3), and VOI analysis (Section 6.3.5.4).

6.3.5.1. Base-case Analysis

The base-case analysis simulated 20,000 patients individually through all treatment strategies reported in Table 6.1 to estimate the lifetime expected cost and QALYs associated with each strategy. The first stage of the base-case analysis compared all twelve intervention strategies with current practice as a common comparator. The second stage of

the base-case analysis conducted a fully incremental analysis by estimating ICERs between strategies that were not dominated or extendedly dominated (Drummond et al., 2015). Incremental net monetary benefits were then calculated for three pre-defined cost-effectiveness thresholds of £20,000, £30,000 and £50,000 per QALY gained. The expected cost of each alternative strategy was reported to two decimal places. The expected QALYs of each alternative strategy were reported to six decimal places because the magnitude of the incremental QALYs between strategies was anticipated to be small; for example, the harm from a flare in disease activity was assumed to last for one week only.

6.3.5.2. Deterministic Sensitivity Analyses

Ten one-way sensitivity analyses of the base-case results were performed, by varying the value of a single input parameter in the model (Briggs et al., 1999):

(i) The cost of all tests were reduced to £20 to represent the results of an economic evaluation that did not account for the additional resources required for testing. This cost was identified in the microcosting study as the price per ELISA test charged by a commercial manufacturer (see *Appendix 35*) (Jani et al., 2016a);

(ii) The HAQ multiplier for patients that changed treatment to rituximab after a true-positive monitoring test result (α) was omitted to represent a situation in which treatment stratification provided no health benefit;

(iii) The HAQ multiplier for patients that changed treatment to rituximab after a false-positive monitoring test result (β) was omitted to represent a situation in which treatment stratification provided no harm to health;

(iv) The probability of low adalimumab drug levels in remission was varied to the upper and lower values of the confidence interval estimated by Kuijper et al. (2015);

(v) The relative risk of losing response to adalimumab after developing ADAb was varied to the upper and lower values of the confidence interval estimated by Garcês et al. (2013);

(vi) The annual cost of adalimumab was reduced by one third to represent the anticipated price-reduction associated with biosimilar adalimumab. The price-

reduction of one third was consistent with the cost assumed by other studies that had investigated the use of biosimilar therapies (Grabowski, 2010).

(vii) The time to biologic treatment failure was estimated using a log-normal (rather than a Weibull) survival curve. The log-normal curve fit the data by Souto et al. (2016) best according to the AIC and BIC statistics, but had less clinical plausibility than the Weibull curve (see *Appendix 32* for survival analysis results).

(viii) The sensitivity of the results to the QALY mapping algorithm, used to obtain EQ-5D weights from HAQ scores, was assessed by using two different published mapping algorithms. The algorithms reported by Adams et al. (2011) and Barton et al. (2004b) were selected for the sensitivity analysis because they were both estimated by using data from patients with RA in the UK.

(ix) The annual HAQ progression of patients that received fourth-line methotrexate was set to zero to represent a situation in which disease was not assumed to worsen over time whilst receiveing any treatment.

(x) The rates of discounting were varied in line with the recommendations provided by the *NICE DAP Manual* (National Institute for Health and Care Excellence, 2011a). One sensitivity analysis presented the undiscounted outcomes and, following Stevenson et al. (2016), a second sensitivity analysis used the annual discount rates of 6% for costs and 1.5% for QALYs.

In addition, a two-way sensitivity analysis was performed by jointly omitting the benefit and harm associated with changing treatment to rituximab, according to the monitoring test result. This sensitivity analysis represented a situation in which stratified treatment decisions, informed by a monitoring test, had no positive or negative impact on health outcomes.

Lastly, a multiway sensitivity analysis was performed by assuming that all testing strategies were perfectly accurate. This sensitivity analysis represented a situation in which the relative cost-effectiveness of stratified medicine was not adversely affected by (i) failing to identify patients that were ADAb-positive with low drug levels; and (ii) inappropriately changing treatment in patients that were ADAb-negative.

6.3.5.3. Probabilistic Sensitivity Analysis

In accordance with best-practice and the requirements of the NICE Reference Case, a probabilistic sensitivity analysis (PSA) was performed to identify the impact of joint uncertainty in the model's input parameters on expected outcomes (Claxton et al., 2005; National Institute for Health and Care Excellence, 2013a). Distributions were assigned to all input parameters of the model, which is documented fully in *Appendix 37*. Correlation between input parameters was accounted for by Cholesky decomposition where a variance-covariance matrix was available (Briggs et al., 2006). The PSA simulated 1,000 patients across 100 PSA model simulations. PSA results were presented by plotting each strategy as a cost-effectiveness acceptability curve (giving the probability of cost-effectiveness at increasing cost-effectiveness thresholds) and as a cost-effectiveness acceptability frontier (giving the extent of uncertainty associated with making decisions based on expected values) (see *Appendix 5* for a description of the method).

6.3.5.4. Value of Information

Further prospective research into adalimumab ADAb and drug level testing, subsequent to the model-based cost-effectiveness analysis, may provide a benefit by reducing the uncertainty associated with making a decision to recommend a strategy based on current evidence. The PSA output was used to quantify this decision uncertainty by calculating the population EVPI (see *Appendix 6* for a description of the method) (Claxton et al., 2001; Mohiuddin et al., 2014; Wilson, 2015). Population EVPI required an estimate of (i) the annual incidence of eligible patients with RA which, according to the *NICE Rheumatoid Arthritis Commissioning Guide*, was assumed to be 1,750 patients (National Institute for Health and Care Excellence, 2013b); (ii) the anticipated product lifecycle of the ELISA tests, which was uncertain and assumed to be up to ten years; and (iii) the discount rate, which was assumed to be 3.5% (National Institute for Health and Care Excellence, 2013a).

6.4. Results

The results section presents findings for the deterministic analysis of the base-case (Section 6.4.1.), the deterministic sensitivity analyses (Section 6.4.2.), the probabilistic sensitivity analysis (Section 6.4.3.), and the VOI analysis (Section 6.4.4).

6.4.1. Deterministic Analysis of Base-case

The base-case results are presented for two analyses: Section 6.4.1.1 reports the results of comparing all twelve strategies (see Table 6.1) with a common comparator (current practice); Section 6.4.1.2 reports the results of comparing all strategies with each other in a fully incremental analysis.

6.4.1.1. Deterministic Analysis: Common Comparator (Current Practice)

The deterministic results, when all twelve comparator strategies were compared with current practice, are reported in Table 6.7. No strategy provided a positive net monetary benefit at a cost-effectiveness threshold of £20,000 or £30,000 per QALY. This set of results may suggest that, at the thresholds for cost-effectiveness conventionally assumed by NICE, the current provision of adalimumab therapy in England is unlikely to be a cost-effective use of health care resources. However, the incremental net monetary benefit of ten strategies, compared with current practice, were positive. Therefore, if a decision was made to recommend adalimumab in England (contrary to the economic evidence), adjustments to the administration of treatment in current practice (by dose-reduction or immunogenicity testing) may improve the relative cost-effectiveness of care.

Table 6.7. *Base-case results: deterministic analysis with common comparator (current practice).*

					Net Monetary Benefit (£)		Incremental Net Monetary Benefit (£)	
Comparator Strategy	**Expected Cost (£)**	**Expected QALY**	**Incremental Cost (£)**	**Incremental QALY**	**λ = £20,000 per QALY**	**λ = £30,000 per QALY**	**λ = £20,000 per QALY**	**λ = £30,000 per QALY**
Current Practice	114,001.57	2.740078	N/A	N/A	-59,200.01	-31,799.23	N/A	N/A
Strategy 1	114,611.27	2.839540	609.70	0.099462	-57,820.47	-29,425.07	1,379.54	2,374.16
Strategy 2	114,272.37	2.806897	270.80	0.066819	-58,134.44	-30,065.47	1,065.57	1,733.76
Strategy 3	111,589.70	2.761210	-2,411.88	0.021132	-56,365.50	-28,753.40	2,834.51	3,045.82
Strategy 4	112,951.32	2.791718	-1,050.26	0.051639	-57,116.96	-29,199.78	2,083.05	2,599.44
Strategy 5	111,064.18	2.740070	-2,937.39	-0.000009	-56,262.79	-28,862.10	2,937.21	2,937.13
Strategy 6	112,193.59	2.740073	-1,807.98	-0.000005	-57,392.13	-29,991.40	1,807.88	1,807.82
Strategy 7	114,702.42	2.730222	700.84	-0.009856	-60,097.98	-32,795.76	-897.97	-996.53
Strategy 8	114,390.58	2.761272	389.00	0.021194	-59,165.13	-31,552.41	34.88	246.82
Strategy 9	112,104.99	2.692252	-1,896.58	-0.047826	-58,259.96	-31,337.44	940.05	461.79
Strategy 10	113,383.81	2.703548	-617.76	-0.036530	-59,312.84	-32,277.36	-112.83	-478.13
Strategy 11	110,802.42	2.739957	-3,199.15	-0.000121	-56,003.29	-28,603.72	3,196.72	3,195.51
Strategy 12	112,079.83	2.739997	-1,921.75	-0.000082	-57,279.90	-29,879.93	1,920.11	1,919.30

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

The direction and magnitude of the incremental expected costs and QALYs in Table 6.7 were consistent with *a priori* reasoning regarding the anticipated outcomes from each strategy:

(i) *Strategy 11* and *Strategy 12* halved the dose of adalimumab in remission for all patients, which reduced the relative expected cost of treatment but also reduced the relative expected QALYs gained (because an increased proportion of patients experienced a flare in disease activity);

(ii) Dose-reduction strategies that occurred after three years had a greater expected cost than those that occurred after two years (because patients received full-dose adalimumab for a longer time period);

(iii) Dose-reduction strategies that were informed by drug level testing (*Strategy 5*, *Strategy 6*) had higher expected QALYs than dose-reduction strategies in all patients (*Strategy 11*, *Strategy 12*) because fewer patients experienced an inappropriate reduction of treatment intensity;

(iv) *Strategy 1* and *Strategy 2* routinely tested adalimumab ADAb and drug levels, whereas *Strategy 7* and *Strategy 8* only tested ADAb. As a consequence of greater test accuracy, the QALYs gained from *Strategy 1* and *Strategy 2* were greater than for *Strategy 7* and *Strategy 8*.

6.4.1.2. Deterministic Analysis: Incremental Analysis

The incremental analysis first ranked all strategies in ascending order of QALY gain (see Table 6.8). The incremental analysis was then performed in this sequence to identify the dominated and extendedly dominated strategies (shaded in grey). Seven strategies were dominated by another comparator strategy and were excluded from the incremental analysis. Current practice was dominated, providing further evidence that the conventional approach to adalimumab therapy was unlikely to be a relatively cost-effective use of health care resources in England. Three strategies were extendedly dominated by a linear combination of other non-dominated comparator strategies. Table 6.9 reports the fully incremental analysis of the three non-dominated comparator strategies.

Table 6.8. *Incremental base-case results: dominated and extendedly dominated strategies.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Dominated or Extendedly Dominated
Strategy 9	112,104.99	2.692252	Dominated by Strategy 11
Strategy 10	113,383.81	2.703548	Dominated by Strategy 11
Strategy 7	114,702.42	2.730222	Dominated by Strategy 11
Strategy 11	110,802.42	2.739957	
Strategy 12	112,079.83	2.739997	Dominated by Strategy 5
Strategy 5	111,064.18	2.740070	Extendedly dominated by 90% of Strategy 11 and 10% of Strategy 3
Strategy 6	112,193.59	2.740073	Dominated by Strategy 3
Current Practice	114,001.57	2.740078	Dominated by Strategy 3
Strategy 3	111,589.70	2.761210	
Strategy 8	114,390.58	2.761272	Dominated by Strategy 4
Strategy 4	112,951.32	2.791718	Extendedly dominated by 45% of Strategy 11 and 55% of Strategy 1
Strategy 2	114,272.37	2.806897	Extendedly dominated by 20% of Strategy 11 and 80% of Strategy 1
Strategy 1	114,611.27	2.839540	

Note: Grey shading=dominated or extendedly dominated strategy.

Table 6.9. *Incremental base-case results: relative cost-effectiveness.*

				Incremental Net Monetary Benefit (£)		
Comparator Strategy	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	787.27	0.021253	37,042.87	-362.21	-149.68	275.38
Strategy 1	3,021.58	0.078330	38,574.74	-1,454.97	-671.66	894.95

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

The three non-dominated strategies were: reducing adalimumab doses in all patients after two years of response (*Strategy 11*); testing adalimumab ADAb and drug levels every three months (*Strategy 1*); and testing adalimumab ADAb and drug levels every three months plus testing adalimumab drug levels in remission after two years (*Strategy 3*).

According to the conventional cost-effectiveness thresholds used by NICE (£20,000 to £30,000 per QALY gained), the use of immunogenicity testing by ELISA (*Strategy 3*,

Strategy 1) was not cost-effective relative to reducing adalimumab doses in all patients after two years. The ICER for *Strategy 3* compared with *Strategy 11* (£37,042 per QALY gained) and for *Strategy 1* compared with *Strategy 3* (£38,574 per QALY gained) exceeded the conventional cost-effectiveness threshold of NICE. Immunogenicity testing may provide a positive incremental net monetary benefit, relative to reducing adalimumab doses after two years, at higher thresholds for cost-effectiveness (for example, £50,000 per QALY gained; Table 6.9).

6.4.2. Deterministic Sensitivity Analyses

The results of all deterministic sensitivity analyses are reported in *Appendix 38*. The following summarises these results by the type of input parameter that was varied (*test characteristics, disease characteristics, treatments,* and *structural assumptions*).

Test Characteristics

The base-case results were sensitive to the unit cost of testing (*Appendix 38*; Table A38.1). The relative cost-effectiveness of *Strategy 3* and *Strategy 1* improved by reducing the cost of all tests to £20. In particular, *Strategy 3* (test ADAb and drug levels every three months, then test drug levels in remission after two years) became cost-effective, relative to *Strategy 11* (reduce doses in all patients at two years), for conventional thresholds of cost-effectiveness assumed by NICE (incremental net monetary benefit = £96 for a cost-effectiveness threshold = £20,000 per QALY gained). However by assuming a test cost of £20, all additional direct health care resources that were necessary to implement ADAb and drug level were (inappropriately) excluded from the analysis.

If patients with a true-positive ADAb and drug level test result were assumed to experience no benefit from the pre-emptive change in treatment to rituximab, the testing strategies (*Strategy 1* and *Strategy 3*) were dominated by *Strategy 11* (*Appendix 38*; Table A38.2). Therefore, in this scenario, reducing the dose of adalimumab in all patients after two years cost less, and caused less harm, than routinely testing patients for ADAb and drug levels.

Alternatively, by assuming no harm from inappropriately changing treatment to rituximab in patients with a false-positive monitoring test result provided only a modest improvement the relative cost-effectiveness of *Strategy 3* and *Strategy 1* (*Appendix 38*; Table A38.3). A

modest improvement was observed because only a minority of patients had a false-positive test result, due to the high accuracy of concurrent ADAb and drug level testing.

If the consequence of treatment stratification, according to the monitoring tests, was varied simultaneously in a two-way sensitivity analysis, such that no harm was caused by false-positive test results and no benefit was derived from true-positive test results, routine ADAb and drug level testing (*Strategy 1* and *Strategy 3*) was dominated by *Strategy 11* (*Appendix 38*; Table A38.4).

If all ELISA tests were assumed to be perfectly accurate (100% sensitivity and specificity) in a multiway sensitivity analysis, the relative cost-effectiveness of both *Strategy 3* and *Strategy 1* improved. In particular, *Strategy 3* had an ICER of £12,704 per QALY gained, relative to *Strategy 11*, and a positive incremental net monetary benefit over the range of cost-effectiveness thresholds conventionally assumed by NICE (*Appendix 38*; Table A38.5).

Disease Characteristics

The base-case estimates for the relative cost-effectiveness of the testing strategies **(***Strategy 3, Strategy 1*) were robust to changes in the probability of patients with low adalimumab drug levels in remission. The total expected QALYs derived from reducing the dose of adalimumab in all patients at two years (*Strategy 11*), relative to the base-case results, were (i) higher if fewer patients were assumed to have low adalimumab drug levels (*Appendix 38*; Table A38.6) and (ii) lower if more patients were assumed to have low adalimumab drug levels (*Appendix 38*; Table A38.7).

The base-case results were also robust to changes in the relative risk of adalimumab treatment failure after developing ADAb. The incremental net monetary benefits of *Strategy 3* and *Strategy 1* remained negative, according to the conventional thresholds for cost-effectiveness used by NICE, when this relative risk was lower (*Appendix 38*; Table A38.8) and higher (*Appendix 38*; Table A38.9) than the value assumed in the base-case.

The relative cost-effectiveness of all strategies improved if no HAQ progression was assumed during response to fourth-line methotrexate therapy (*Appendix 38*; Table A38.10). This assumption may not be plausible in reality, however, due to clinical evidence that

suggested the disease activity of patients with RA worsened over time while receiving cDMARDs, relative to bDMARDs (Madan et al., 2015; Stevenson et al., 2016).

Treatments

If the price of adalimumab was assumed to be reduced by one third, to represent the price of biosimilar adalimumab, *Strategy 3* became extendedly dominated by *Strategy 11* and *Strategy 1*. The relative cost-effectiveness of *Strategy 1*, compared with *Strategy 11*, was close to (but not within) the range of thresholds for cost-effectiveness conventionally assumed by NICE (*Appendix 38*; Table A38.11). The base-case result that favoured *Strategy 11* was therefore robust to the assumption that patients were prescribed biosimilar adalimumab.

Structural Assumptions

The base-case result was sensitive to the rate at which future costs and QALYs were discounted. *Strategy 1* generated the highest expected incremental net monetary benefit, at a cost-effectiveness threshold of £30,000, if no discounting was applied to future outcomes (*Appendix 38*; Table A38.12). In addition, both *Strategy 1* and *Strategy 3* generated a positive incremental net monetary benefit if outcomes were discounted at a differential rate (6% for costs; 1.5% for QALYs) (*Appendix 38*; Table A38.13).

The base-case results were robust when a log-normal survival curve was used to model the time to failure of each bDMARD therapy instead of a Weibull survival curve (*Appendix 38*; Table A38.14). The magnitude of QALY gains varied between the strategies when different algorithms were used to map between HAQ and EQ-5D; however, the relative cost-effectiveness of *Strategy 3* and *Strategy 1* did not improve, compared to the outcomes estimated in the base-case analysis (*Appendix 38*; Table A38.15; Table 38.16).

6.4.3. Probabilistic Sensitivity Analysis

The results of the PSA are presented as a CEAC in Figure 6.3. Strategies were excluded from the illustration of the CEAC if they had a zero probability of being cost-effective at all cost-effectiveness thresholds (*Strategy 6, Strategy 12*). *Strategy 11* had the highest probability of being cost-effective at the lower thresholds of relative cost-effectiveness. *Strategy 1* had the highest probability of being cost-effective for cost-effectiveness

thresholds greater than approximately £70,000 per QALY gained. The probability that *Strategy 11* was cost-effective was 64% for a threshold of £20,000 per QALY gained and 56% for a threshold of £30,000 per QALY gained.

Figure 6.4 presents a CEAF that plotted the probability the strategy with the greatest expected net monetary benefit was cost-effective, across a range of cost-effectiveness thresholds. The CEAF illustrated the decision uncertainty associated with making recommendations according to expected values based on current information (Fenwick et al., 2001; Fenwick et al., 2004). Three strategies provided the greatest net monetary benefit for cost-effectiveness thresholds up to £100,000 per QALY gained. There were two breaks in the CEAF, referred to as *discontinuities*, when the strategy with the highest expected net benefit changed. The value of the cost-effectiveness threshold, at which a discontinuity in the CEAF occurred, was equal to ICER between the two respective strategies (Fenwick et al., 2001). The discontinuities in the CEAF occurred at £41,625 (*Strategy 3* Vs. *Strategy 11*) and £46,171 (*Strategy 1* Vs. *Strategy 3*).

6.4.4. Value of Information Analysis

The EVPI per patient was estimated from the PSA simulation results by subtracting the expected net monetary benefit under current information from the expected net monetary benefit with perfect information (Claxton et al., 2001; Wilson, 2015). The EVPI per patient, reported in Table 6.10, was calculated for a range of cost-effectiveness thresholds (between £1,000 and £100,000 per QALY gained). At at cost-effectiveness threshold of £20,000 per QALY gained, the estimated EVPI per patient was £485.78.

The incidence of patients with RA in England was approximately 17,500 patients per year, and 10% of those patients were eligible for bDMARD therapy (National Institute for Health and Care Excellence, 2013b). The time horizon of the adalimumab ADAb and drug level ELISA test's product lifecycle was uncertain. The population EVPI was therefore calculated (see *Appendix 6* for method) over a range of time horizons (between zero and ten years), discounted at 3.5% per year, assuming a cost-effectiveness threshold of £20,000 per QALY gained (reported in Table 9.11). By assuming a ten-year time horizon, the eligible future population to benefit from further research (discounted at 3.5%) was estimated to be 14,554 patients. Therefore, at a cost-effectiveness threshold of £20,000 per QALY, the population EVPI was approximately £7,070,099.

Figure 6.3. *Cost-effectiveness acceptability curves.*

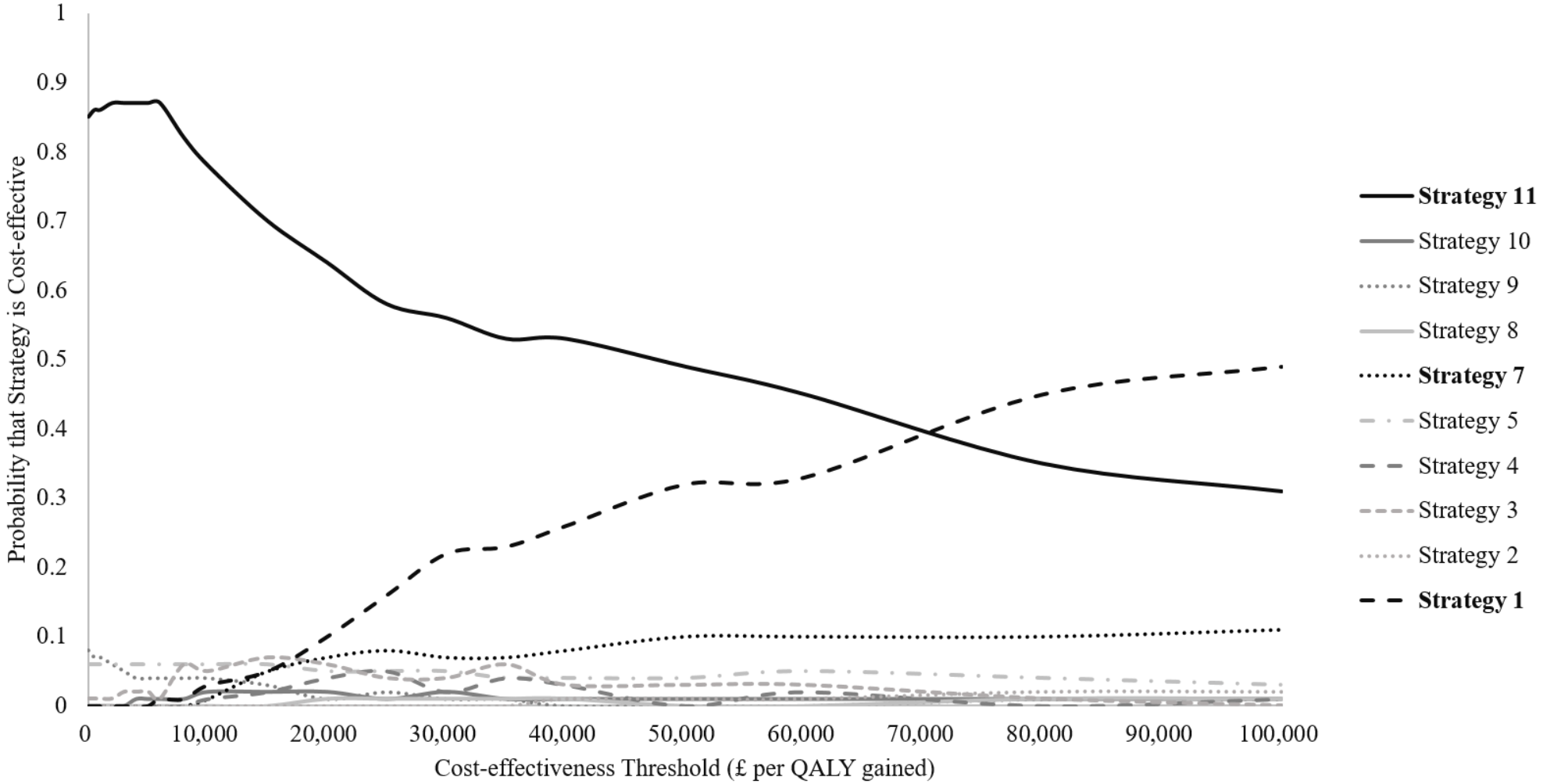


Figure 6.4. *Cost-effectiveness acceptability frontier.*

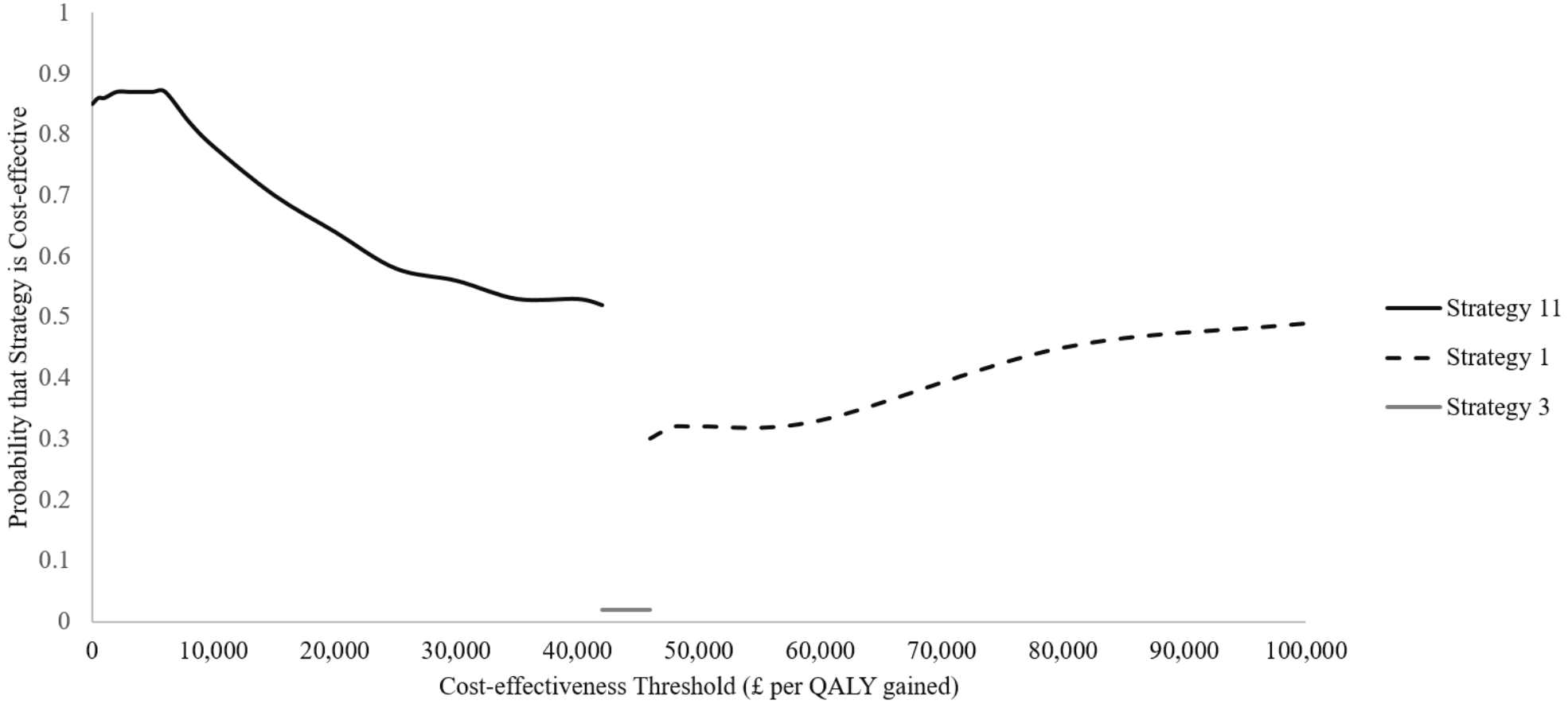


Table 6.10. *Estimation of EVPI per patient for a range of cost-effectiveness thresholds.*

	Maximum Expected Net Monetary Benefit (£)		
λ (£ per QALY gained)	**Current Information**	**Perfect Information**	**EVPI per patient (£)**
1,000	-99,469.54	-99,449.23	20.31
2,000	-96,717.11	-96,700.30	16.81
3,000	-93,964.67	-93,947.73	16.94
4,000	-91,212.24	-91,191.69	20.55
5,000	-88,459.80	-88,435.61	24.19
6,000	-85,707.37	-85,678.47	28.90
8,000	-80,202.50	-80,153.06	49.44
10,000	-74,697.63	-74,614.18	83.45
15,000	-60,935.46	-60,690.12	245.34
20,000	-47,173.29	-46,687.51	485.78
25,000	-33,411.13	-32,589.07	822.06
30,000	-19,648.96	-18,435.10	1,213.85
35,000	-5,886.79	-4,265.09	1,621.70
40,000	7,875.38	9,918.62	2,043.24
50,000	35,876.17	38,348.14	2,471.97
60,000	64,380.47	66,809.16	2,428.69
80,000	121,389.08	123,804.65	2,415.57
100,000	178,397.68	180,888.56	2,490.87

Abbreviations: EVPI=expected value of perfect information; λ=cost-effectiveness threshold.

Table 6.11. *Estimation of population EVPI over a range of product time horizons, assuming a cost-effectiveness threshold of £20,000 per QALY gained.*

Product Time Horizon (Years)	**Annual Incidence of bDMARD-Eligible Patients with RA**	**Discounted† Annual Incidence**	**Cumulative Annual Discounted† Eligible Population**	**Population EVPI (£)**
1	1,750	1,690.82	1,690.82	821,370.44
2	1,750	1,633.64	3,324.46	1,614,965.07
3	1,750	1,578.40	4,902.86	2,381,723.16
4	1,750	1,525.02	6,427.89	3,122,552.23
5	1,750	1,473.45	7,901.34	3,838,329.12
6	1,750	1,423.63	9,324.97	4,529,900.99
7	1,750	1,375.48	10,700.45	5,198,086.37
8	1,750	1,328.97	12,029.42	5,843,676.11
9	1,750	1,284.03	13,313.45	6,467,434.31
10	1,750	1,240.61	14,554.06	7,070,099.24

†: 3.5% annual discount rate was assumed.

6.5. Discussion

This study addressed the decision problem in Table 5.8 (Section 5.5) and aimed to determine whether adalimumab ADAb and drug level ELISA testing, to stratify treatment for patients with RA in England, was a relatively cost-effective use of health care resources. An early model-based cost-effectiveness analysis was performed to understand the indicative incremental costs and benefits, and the key drivers of relative cost-effectiveness, of stratifying treatment decisions for RA using commercial ELISA-based tests. A *de novo* decision analytic model was constructed (as a DES) that simulated individual patients (with characteristics derived from the BSRBR-RA cohort) across thirteen different test-and-treatment strategies (Table 6.1). The model was representative of current practice in England and was developed following extensive model conceptualisation (reported in *Chapter Five*). The values of all input parameters were estimated from relevant published sources and the model was evaluated probabilistically to quantify the decision uncertainty associated with recommending each strategy in the NHS in England.

The results indicated that the standard administration of adalimumab (40mg every two weeks with methotrexate until treatment failure) observed in current practice was unlikely to be relatively cost-effective at the conventional thresholds for cost-effectiveness assumed by NICE. Previously published economic evaluations, including those used as evidence for NICE technology appraisals, have also questioned the relative cost-effectiveness of TNFi therapies for RA (van der Velde et al., 2011; Joensuu et al., 2015; Stevenson et al., 2016). The most cost-effective treatment strategy in this study, relative to all other strategies, was to halve the dose of adalimumab in all patients after two years of responding (*Strategy 11*; Incremental net monetary benefit = £3,196 if the cost-effectiveness threshold was £20,000 per QALY gained).

All thirteen comparator strategies in this economic evaluation generated a negative net monetary benefit at the standard cost-effectiveness thresholds assumed by NICE. Negative net monetary benefits imply that a treatment strategy caused a net reduction in population health (Drummond et al., 2015); in the case of this study, the health forgone as a consequence of treating patients with adalimumab was greater than the health gained. However, if decision-makers choose to overlook the published economic evidence, and the evidence provided by this study, and decide to recommend adalimumab for RA, the

findings of this study suggested that adjustments may be possible when prescribing adalimumab to improve its relative cost-effectiveness, compared with the existing approach, in current practice.

The cost-effectiveness of testing adalimumab ADAb and drug levels was sensitive to the unit cost of testing. If all additional resources to perform testing were omitted from the analysis (such as the need for an additional patient appointment or the time taken to interpret the results), *Strategy 3* became cost-effective relative to a uniform dose reduction in all patients. Only one model-based economic evaluation of a stratified medicine for RA published previously had quantified all of the additional resources necessary to perform testing (Kowada, 2010). It was therefore possible that the remaining economic evaluations of a stratified medicine in RA (described in *Chapter Two*) had underestimated the (opportunity) cost of testing, and in turn, overestimated its relative cost-effectiveness of treatment stratification.

All testing strategies that measured adalimumab ADAb *without* measuring concurrent drug levels were dominated. This finding demonstrated the importance of testing both ADAb and drug levels to improve the accuracy of testing (Jani et al., 2016b). As a consequence, fewer patients experienced a false-positive ADAb test result, which mitigated the number of inappropriate changes to treatment and the subsequent loss of QALYs.

The base-case result was sensitive to the accuracy of testing; the relative cost-effectiveness of *Strategy 3* and *Strategy 1* improved substantially by assuming that the sensitivity and specificity of all tests were 100% (*Appendix 38*; Table A38.5). *Chapter Two* reported a systematic review of economic evaluations of stratified medicine in RA and found that forty percent of included studies assumed that testing was perfectly accurate. This assumption, to use the terminology of Phelps et al. (1988, p.284), although untenable in practice, was evidence that testing cleared the first "hurdle" of establishing relative cost-effectiveness. ELISA test manufacturers may now subsequently have an incentive to improve the accuracy of testing, given the knowledge that greater test accuracy may improve the relative cost-effectiveness of treatment stratification, *ceteris paribus*.

A uniform dose-reduction of adalimumab in all patients after two years of responding represented a *disinvestment* in health care resources. As a consequence, *Strategy 11* provided a positive incremental net monetary benefit (indicating relative cost-effectiveness) but also led to a reduction in costs *and* QALYs compared to current practice.

The QALY loss occurred due to the proportion of patients that flared after halving their adalimumab dose. The *STRASS* trial randomised patients with RA, treated with adalimumab or etanercept that were in remission, to either (i) a strategy that de-escalated the intensity of treatment or (ii) a strategy that maintained full-dose TNFi therapy; the results corroborated the findings of this model-based economic evaluation and concluded that dose-reduction strategies were relatively cost-effective but simultaneously reduced costs and QALYs (Vanier et al., 2017). Decision-makers, however, may choose to only recommend health technologies that are both effective *and* cost-effective (Dowie, 2004). As a consequence, failing to recommend a cost-effective technology on the grounds that it provides a health loss to identifiable patients (and a health gain to unidentifiable patients) will not allocate resources such that population health is maximised (Claxton et al., 2015b; Cookson, 2015).

The population EVPI estimated by the study was large (£7,070,099) and indicated that future prospective research, subsequent to this model-based cost-effectiveness analysis, would likely be of value to the NHS (Thorn et al., 2016). The cost of previous research projects in England, that have received funding from public resources, have been below this estimated population EVPI; for example, the *NIHR Programme Grants for Applied Research* can provide awards for up to £2 million for research into disease management in the NHS (National Institute for Health Research, 2012). This result may provide legitimacy to agenda for research set by EULAR, and the requests made by the rheumatologists in *Chapter Three*, for further research into the use of ADAb and drug level testing in routine clinical practice (Smolen et al., 2014).

There are limitations to the estimate of population EVPI in this study and to VOI analyses more generally. The magnitude of population EVPI is equivalent to the upper-bound on the value of further research to reduce parameter uncertainty; it does not necessarily inform whether more research is desirable (Claxton et al., 2004; Wilson, 2015). The complexity and run-time of the DES framework, however, precluded the ability to calculate the ENBS of different research designs. Further research also takes time to be generated and the population EVPI does not make statements about *when* such research should be produced; for example, a prospective RCT may take a number of years to complete, from ethical approval, to patient recruitment, to data collection, and to data analysis. Conditional on the product lifecycle of the health technology, independent research may be produced or wider organisational constraints may change, such that the relative cost-effectiveness of a study

designed to reduce parameter uncertainty may change temporally. In addition, no single programme of research will eliminate all parameter uncertainty and, therefore, health technology assessment should be considered as an iterative process by informing the design of, and analysing the impact of, further research over time (Sculpher et al., 1997). The estimate of parameter uncertainty from a decision analytic model, and the subsequent VOI analysis, are conditional on the structural assumptions of the model (Barton et al., 2004a). Alternative structural assumptions could have been made (for example, by including a specific health state for patients that undergo surgery) that may have affected the uncertainty associated with the relative cost-effectiveness of testing. Therefore, VOI analyses should be interpreted as a guide to inform future research decisions, rather than as a definitive estimate of the value of future research studies. Lastly, the lack of correlation between patient-level attributes, test sensitivity and specificity, and the use of an uninformative prior distribution to characterise the uncertainty associated with the HAQ multiplier may have inflated the parameter uncertainty within the model and, in turn, contributed to overestimating the population EVPI.

Limitations

One potential limitation of this study, common to all early model-based economic evaluations, was the limited evidence available to populate the input parameters (Annemans et al., 2000). The DES model in this study was developed using all available evidence, but that evidence may have been derived from a single source or may have been greatly uncertain. The economic evaluation of a new health technology is, however, an iterative process that can be updated as further research becomes available (Sculpher, 1997). One specific advantage of this study, therefore, was to provide evidence that suggested such further research may be valuable to the NHS in England.

Due to the evidence limitations in this study, the sensitivity and specificity of the adalimumab ADAb and drug level ELISA assays were assumed to be constant over time. This assumption may have overestimated the accuracy of testing because test performance may change depending on the prevalence of a biomarker in the patient population (Longo et al., 2014). However, it was unlikely that this assumption biased the results because, even under such favourable conditions for testing, *Strategy 11* (which didn't include a test) was estimated to be cost-effective relative to all other strategies.

A third potential limitation was that uncertainty in the HAQ rebound multipliers following true-positive and false-positive monitoring test results were sampled from a uniform distribution in the PSA. Alternatively, expert elicitation methods could have been used to characterise the distribution of this uncertainty (Iglesias et al., 2016), if all possible multiplier values between zero and one were not equally likely. However, it may have been a challenge to find appropriate experts for this study, given than the rheumatologists in *Chapter Three* were unsure of how to use the ELISA tests in routine clinical practice. It was therefore more appropriate, in absence of evidence, to include *some* uncertainty in the HAQ rebound multipliers, as an uninformative prior distribution, rather than to assume a fixed value (Boshuizen et al., 2009).

A final practical limitation of performing a DES model was that each simulation of 20,000 patients required an approximate run time of eight hours per comparator strategy. This computational demand limited the ability to perform extensive deterministic sensitivity analyses (Griffin et al., 2006). However, a deterministic sensitivity analysis was conducted for all parameters that related to testing and a PSA was performed to propagate the joint uncertainty through all input parameters of the model, as required by the NICE Reference Case (National Institute for Health and Care Excellence, 2013a).

Implications for Future Research

The accuracy of testing, and the health consequence from stratified treatment decisions according to true-positive and false-positive monitoring test results, were both key drivers of the relative cost-effectiveness of adalimumab ADAb and drug level testing. Future prospective research on these parameters may be justified, given the population EVPI estimated by the study, to reduce the decision uncertainty associated with recommending ADAb and drug level ELISA testing in current practice. However, as the population EVPI may be overestimated, such future research studies should be deliverable within a short time period whilst also taking the cost of the study into consideration.

The introduction of new bDMARD therapies and the *treat-to-target* paradigm has changed how patients with RA are treated in routine clinical practice substantially (Smolen et al., 2016). It may be possible that these new approaches to patient management have reduced the average number of days that patients require hospitalisation, compared with the evidence that was used in this study (Roche., 2006). A potential study for future research, therefore, could estimate an updated relationship between the mean days hospitalised per

year and the HAQ score, by using a sample of patients in England that have been treated according to a more-recent therapeutic paradigm.

Since the results of this study were estimated, clinical evidence has started to be produced that investigated the use of testing ADAb and drug levels in patients with RA that received a different TNFi therapy (certolizumab pegol) (Jani et al., 2017). A future model-based economic evaluation could therefore estimate the relative cost-effectiveness of testing ADAb and drug levels with ELISA tests, to stratify a subsequent treatment decision, in patients that received different monoclonal TNFi therapies (infliximab, golimumab, certolizumab pegol) as the supporting clinical evidence base develops.

Adalimumab is licenced not just for patients with RA, but also for patients with psoriasis, psoriatic arthritis, axial spondyloarthritis, and Crohn's disease (European Medicines Agency, 2017). Between 2015 and 2015, the total expenditure on adalimumab in secondary care within the NHS in England was higher than the expenditure on any other pharmaceutical treatment across all indications (approximately £371 million) (Health and Social Care Information Centre, 2015). Given that dose-reduction strategies were found to be relatively cost-effective for patients with RA in this study, future research could investigate the relative cost-effectiveness of reduced-dose adalimumab across the spectrum of diseases defined in the product licence. In light of the magnitude of NHS resources that are allocated to full-dose adalimumab at present, such research may have the potential to benefit population health substantially.

6.6. Conclusion

This study conducted an early model-based economic evaluation of testing adalimumab ADAb and drug levels by ELISA to stratify treatment for patients with RA in England. A DES model simulated 20,000 hypothetical patients, representative of the patient population with RA in England, individually through thirteen different alternative test-and-treatment strategies over a lifetime time horizon. The results suggested that, based on current evidence, the use of adalimumab ADAb and drug level testing by ELISA was unlikely to be a relatively cost-effective use of health care resources, compared with halving the dose of adalimumab in all patients after responding for two years. There was, however, substantial decision uncertainty associated with the estimates of relative cost-effectiveness. The population EVPI was estimated to be greater than £7 million, which may justify the calls for further prospective research into ADAb and drug level testing, to reduce the

decision uncertainty associated with recommending the ELISA-based testing strategies in routine practice.

Chapter 7
Discussion and Conclusion

Chapter Seven presents a discussion of the thesis in terms of the specific contributions to knowledge (Section 7.1), the potential implications for different stakeholders (Section 7.2), the limitations (Section 7.3), and the topics for further research (Section 7.4). Section 7.5 presents the final concluding remarks.

The overall aim of this thesis was to provide evidence for the relative cost-effectiveness of a biomarker test to stratify treatment for patients with RA, consistent with the requirements of decision-makers for the NHS in England. The thesis began, for the purpose of orientation, by juxtaposing a background of *a priori* investment and health policy promise in stratified medicine against the observation that relatively few examples of stratified medicine had translated into routine clinical practice. It was posited that the production of economic evidence, in particular, was essential to advance the use of biomarker tests to stratify treatments in the NHS which, in turn, provided the policy-relevance that motivated the research in this thesis.

The thesis focussed on evaluating a specific case study of stratified medicine for RA, the measurement of adalimumab ADAb and drug levels by commercial ELISA-based assays (Section 1.3.5), that was an emerging health technology at the inception of this thesis and was considered by the international EULAR organisation as a high-priority for clinically-relevant research (Smolen et al., 2014). The thesis addressed three related research questions regarding (i) the existing economic evidence for stratified medicine in RA; (ii) the treatment decisions made in current practice for patients with RA in England; and (iii) whether adalimumab ADAb and drug level testing to stratify treatment was a relatively cost-effective use of health care resources. Each research question was answered using specific methods (systematic reviews, qualitative thematic framework analysis, quantitative econometric analysis, and decision analytic modelling), presented across five

chapters that, when considered sequentially, advanced the economic evidence base for stratified medicine in RA.

7.1. Contributions to Knowledge

The research in this thesis provided seven clear contributions to knowledge that are now summarised, within and beyond the context of the thesis. A strategy to disseminate these contributions as academic peer-reviewed publications is described in *Appendix 1*.

The first contribution to knowledge, reported in *Chapter Two*, was a systematic review of model-based economic evaluations of stratified medicine in RA. Previously published systematic reviews that had identified economic evaluations of stratified medicine had, in general, focussed on only one specific type of health technology to stratify treatment, such as genomic testing to predict treatment response (Phillips et al., 2004; Vegter et al., 2008; Beaulieu et al., 2010; Vegter et al., 2010; Wong et al., 2010; Hatz et al., 2014; Berm et al., 2016; Plumpton et al., 2016). The study in *Chapter Two* had a broader focus by arguing, consistently with the NICE DAP, that different types of health technologies could be used to stratify treatment decisions, such as imaging and an invasive biopsy (National Institute for Health and Care Excellence, 2011a). The contribution of the systematic review in *Chapter Two* was therefore novel, not only because it identified all existing economic evidence for stratified medicine in RA, but also because it provided an in-depth survey of existing economic evidence for *all* potential ways to stratify treatment in a *single* disease. The potential limitations of the ten economic evaluations identified by the systematic review (for example, a poor characterisation of current practice; insufficient comparator strategies; limited quantification of the resources required to implement testing; insufficient reporting of strategies to identify evidence for test accuracy; and no VOI analyses) established a set of relevant topics for further research that were addressed in the subsequent chapters of the thesis.

The second contribution to knowledge, reported in *Chapter Three*, was an exploratory qualitative analysis of the factors that influenced treatment decisions for patients with RA in England. This study had a practical value for the thesis, to inform the design of the subsequent *de novo* model-based economic evaluation in *Chapter Six*, by identifying a relevant care pathway for patients with RA according to a sample of consultant rheumatologists distributed across England. Prior evidence, however, indicated that

treatment decisions for RA were variable across England (Tugnet et al., 2013; Blake et al., 2014; The British Society for Rheumatology, 2015) and, to date, no evidence had attempted to explain the potential sources of this variability. The results, which suggested that treatment decisions were influenced by factors at three levels (the external environment; internal to the hospital; at the level of the individual rheumatologist) was therefore a novel contribution towards the developing literature regarding treatment variation in rheumatology practice (Tugnet et al., 2013; Blake et al., 2014; The British Society for Rheumatology, 2015).

The third contribution to knowledge, also reported in *Chapter Three*, was the identification of four potential barriers to using the ADAb and drug level ELISA tests in routine practice (the recognition of a clinical problem; understanding the role of testing; lack of evidence to support testing; capacity and resource constraints), as perceived by the sample of consultant rheumatologists distributed across England. The effective application of any stratified medicine in routine practice is reliant on the behaviour of clinicians to prescribe the correct treatment according to the test result (Garrison et al., 2006; Annemans et al., 2013; Buchannan et al., 2013). However, pragmatic trials of stratified medicine have demonstrated that clinicians may not always comply with the appropriate treatment decision (Thompson et al., 2014). The four barriers were found to be similar to those reported by clinicians with respect to using different types of medical test in routine practice (Jones et al., 2013; Raghavan et al., 2014).

The fourth contribution to knowledge, reported in *Chapter Four*, was the quantitative econometric analysis of the patient-level factors that influenced TNFi prescribing decisions observed in routine practice for patients with RA in England. The results of three related studies, which analysed TNFi prescribing decisions observed for patients with RA in North America, may have been subject to endogeneity (DeWitt et al., 2006; Carter et al., 2012; Zhang et al., 2013). The study in *Chapter Four* built on the findings of *Chapter Three* by attempting to control for the unobservable factors that may have influenced treatment decisions within the external environment and internal to the hospital, to mitigate the influence of endogeneity through omitted variable bias. The results found that patient-level characteristics (age; concomitant methotrexate; socioeconomic characteristics) had a significant influence on the choice of TNFi prescribed, which may have indicated that implicit treatment stratification occurred within routine clinical practice in England.

The fifth contribution to knowledge, reported in *Chapter Five*, was a systematic review of RA-specific prescribing algorithms that included TNFi ADAb and drug level testing. No previous study had synthesised the different recommendations for TNFi ADAb and drug level testing that were designed specifically for patients with RA. The appropriate use for testing was unknown *a priori*, as evidenced by the potential barriers to testing (perceived by the rheumatologists) in *Chapter Three*. The results of this study were used to inform the relevant comparator strategies for the subsequent economic evaluation in *Chapter Six*.

The sixth contribution to knowledge, reported in *Chapter Five*, was the development of a novel algebraic conceptualisation technique that facilitated the early identification of potentially relevant comparator strategies (from a larger set of plausible candidate strategies) during the development of a *de novo* model-based economic evaluation. A common challenge when evaluating the relative cost-effectiveness of a new medical test to stratify a treatment, early in its product lifecycle, was that the appropriate purpose of testing may be uncertain (Buisman et al., 2016) A limited clinical evidence base may increase the number of plausible candidate testing strategies; however, constraints on a decision analyst's time may reduce the feasibility of comparing all plausible strategies in a full economic evaluation (Owens et al., 2017). The algebraic conceptualisation technique was applied to the results of the systematic review of RA-specific prescribing algorithms and identified two broad approaches to stratify treatment that were deemed to be potentially relevant comparator strategies for the decision problem in this thesis.

The seventh contribution to knowledge, reported in *Chapter Six*, was the early model-based economic evaluation of adalimumab ADAb and drug level testing to stratify treatment for patients with RA in England. This study was the first model-based economic evaluation of stratified medicine for RA that was designed specifically for decision-makers in England. An extensive model conceptualisation and development process (*Chapter Five*) ensured that the economic evaluation included all comparator strategies that were directly relevant to clinical practice in England. The results suggested that, based on current evidence, stratified medicine according to adalimumab ADAb and drug level ELISA testing was not likely to be a relatively cost-effective use of health care resources in England. The accompanying VOI analysis, however, indicated that further prospective research into the ELISA-based testing strategies had the potential to be worthwhile, in order to reduce the decision uncertainty associated with recommending a stratified testing strategy in routine clinical practice.

7.2. Thesis Implications

The research presented within this thesis had implications for different stakeholders in the health care system. These implications are now reported for commercial medical test manufacturers (Section 7.2.1), population health care decision-makers (Section 7.2.2), and and decision analysts (Section 7.2.3).

7.2.1. Commercial Medical Test Manufacturers

Commercial manufacturers that have produced a new medical test to stratify a treatment face an immediate challenge to demonstrate the value of testing to both (i) decision-makers, by communicating how treatment stratification could be consistent with the objectives of the health care system (for example, population health maximisation), and to (ii) clinicians, by explaining how testing may be useful in routine clinical practice (Faulkner et al., 2012). For example, the ELISA-based tests evaluated in this thesis were an example of a complementary diagnostic (Milne et al., 2015), with a lack of end-to-end evidence and data to support the clinical utility of testing (Rogowski et al., 2009; National Institute for Health and Care Excellence, 2011a; Phillips et al., 2013).

The current regulatory standards for licencing a new medical test in Europe require evidence of a test's accuracy and safety only (Payne, 2009; Frueh, 2013; Payne et al., 2013a), which may decrease the incentive for test manufacturers to produce evidence of clinical utility during product development. Pharmaceutical manufacturers may also be unwilling to generate evidence for the clinical utility of testing if treatment stratification reduces their treatment's market share (Towse et al., 2013). Therefore, a practical implication for commercial medical test manufacturers was that an early model-based cost-effectiveness analysis could be incorporated during the development of a new test to (i) ensure that a test to stratify treatment is developed in alignment with the values of decision-makers in the health care system (Sculpher, 1997; Annemans et al., 2000; Buisman et al., 2016) and (ii) to provide evidence to clinicians regarding the purpose of testing in routine clinical practice.

7.2.2. Population Health Care Decision-makers

NICE has a mandate to maximise population health subject to the prevailing budget constraint for health care, and it achieves this objective by making resource allocation

decisions after evaluating the relative effectiveness *and* cost-effectiveness of new health technologies (National Institute for Health and Care Excellence, 2013a). This evaluative approach by NICE may be sufficient for health technologies that require an investment in health care resources; however, the demonstration of effectiveness may not be appropriate for health technologies that require a *disinvestment* in health care resources (Bryan et al., 2014; Scotland et al., 2017). For example, the most cost-effective strategy estimated by the economic evaluation in *Chapter Six* was to reduce adalimumab doses in all patients after two years of responding; this strategy represented a disinvestment in health care resources (compared with current practice) and simultaneously (i) maximised population health; (ii) was relatively cost-effective; but (iii) was *not* relatively effective because a proportion of patients with reduced-dose adalimumab experienced a flare in disease activity.

If objections are made on the grounds that it is unethical to reduce the health of known patients (to release resources for other unidentifiable patients in the health care system) then, for consistency, it must also be equally unethical to reduce the health of those unidentified patients by treating a known population of patients with a strategy that is not relatively cost-effective (Dowie, 2004). Failure to recommend a strategy that is relatively cost-effective, but not effective, is inconsistent with population health maximisation and implies that identifiable lives are worth more than unidentifiable lives (Claxton et al., 2015b; Cookson, 2015). The practical implication of the results in this thesis for NICE was to recommend consulting their *Citizens Council* to improve the transparency in the social value judgements made with respect to disinvestment decisions (Rawlins, 2005; Culyer et al., 2006; Shah et al., 2013).

7.2.3. Decision Analysts

There were three implications for decision analysts who may conduct a future model-based economic evaluation of stratified medicine; these implications are now described.

The scientific literature regarding a new medical test to stratify treatment may develop rapidly over a short period of time (Buisman et al., 2016). For example, there were three major developments in the literature over the duration of this thesis, which subsequently affected the model-based economic evaluation reported in *Chapter Six*: (i) in 2013, Krieckaert et al. (2015) published an *early-view* version of their cost-effectiveness analysis of adalimumab drug level testing for patients with RA in The Netherlands; (ii) in 2014, the NICE DAP published the scope of a (now complete) technology appraisal that included

TNFi ADAb and drug level ELISA testing for patients with Crohn's disease (National Institute for Health and Care Excellence, 2016c); and (iii) in 2015, the (now complete) NICE multiple technology appraisal for RA was updated, mid-appraisal, to include biosimilar therapies as relevant comparator treatments (National Institute for Health and Care Excellence, 2016a). The implication for decision analysts, based on the experience of this thesis, was that new model-based economic evaluations of stratified medicine must be expectant of, and be adaptable to, similar rapid developments in the existing evidence base.

A second implication for decision analysts was that economic evaluations of stratified medicine must appropriately account for (i) all direct health care resources associated with testing and (ii) the accuracy of testing. The economic evaluations of stratified medicine in RA that were published previously, identified in *Chapter Two*, were found to have inadequately characterised the values of (and uncertainty in) test accuracy and the resources necessary for testing. Failing to do so may have underestimated the relative cost-effectiveness of treatment stratification and biased an economic evaluation in favour of the testing strategy (National Institute for Health and Care Excellence, 2011a; Miguel et al., 2015). For example, in the economic evaluation presented in *Chapter Six*, the values of both test accuracy and the cost of testing were found to be key drivers of the relative cost-effectiveness of ADAb and drug level testing.

A third implication of this thesis for decision analysts was that a thorough model conceptualisation process is essential for the evaluation of new examples of stratified medicine that are characterised by a lack of end-to-end evidence. For example, the conceptualisation process in this thesis (*Chapter Five*) was used to (i) identify relevant comparator strategies, (ii) define the decision problem, (iii) characterise disease progression, and (iv) characterise the care pathways that included testing. In addition, the involvement of clinical experts in the conceptualisation process was invaluable. The responses provided by the rheumatologists in *Chapter Three*, for example, led to the inclusion of dose-reduction strategies in the decision problem for the model-based economic evaluation in *Chapter Five*, which were subsequently found to be cost-effective, relative to the testing strategies, in *Chapter Six*.

7.3. Limitations

The scope of this thesis encompassed the generation of economic evidence that conformed to the pre-specified requirements of the decision-makers responsible for health care resource allocation in England, in order to be relevant to health care policy and decision-making. The specific limitations of the individual studies in the thesis were reported within their respective chapters and are not repeated here. There were, however, two potential broader limitations to the analytic approach of the thesis, related to the omission of non-health benefits (Section 7.3.1) and the capacity to implement stratified medicine (Section 7.3.2), that are now discussed.

7.3.1. The Non-health Benefits of Stratified Medicine

The first potential limitation was that the diagnostic information obtained from any test result may have had an intrinsic value to both clinicians and patients, but the prevailing paradigm for the economic evaluation of health technologies in England was concerned with the maximisation of population health only. For example, the methods used by the NICE DAP to evaluate a stratified medicine linked the information derived from a test result to an appropriate treatment decision, which was then linked to a final health outcome (Byron et al., 2014); the *benefit* of testing, therefore, was framed entirely in terms of the consequential impact on the net QALYs gained or lost (National Institute for Health and Care Excellence, 2011a).

The limitations of using QALYs as a measure of benefit were not unique to stratified medicine; for example, relevant dimensions of health benefit in some diseases (such as those characterised by aural or visual problems) may not be reflected by the EQ-5D instrument adequately (Longworth et al., 2014), and the evaluation of public health interventions may utilise a broader set of outcome measures (Edwards et al., 2013; Brazier et al., 2015). The diagnostic information derived from the result of a medical test, an essential element of a stratified medicine, may have incurred (non-health) benefits that were not quantified within the QALY outcome measure (Payne et al., 2013c). The intrinsic value of such diagnostic information derived from a test result has been referred to as the *value of knowing* (Zamora et al., 2016).

In this thesis, one candidate way of using the ELISA-based testing strategies was excluded from the decision problem because it was expected to impose additional costs without an accompanying QALY gain (Section 5.4; testing patients after they had already lost response to treatment). This testing strategy would have provided diagnostic information (a patient's prevailing adalimumab drug levels and ADAb status) that may have been valued by clinicians and patients. Complex interventions (such as clinical genetic services) that produce no direct health benefit have been funded previously by the NHS in England, which may imply that the non-health benefit derived from diagnostic information had *some* value within the health care system (Payne et al., 2013b). However it was not clear how, or if, those non-health benefits could have been incorporated within the prevailing paradigm for the economic evaluation of health technologies in England.

7.3.2. Capacity to Implement Stratified Medicine

The second potential limitation was that an implicit assumption was made, within the generation of economic evidence in this thesis, that the capacity to implement the ELISA-based testing strategies would be available immediately in the NHS. However, the qualitative study in *Chapter Three* revealed that some rheumatology units across England did not have the resources or laboratory requirements available at present to perform routine ELISA testing in clinical practice.

Cost-effectiveness analyses presuppose, to use the economic term, a *first-best* world in which all necessary assumptions hold and the results, when applied to decision-making, enable a movement towards the allocative efficiency of population health care resources (Weinstein et al., 1973). The scenario in which a decision is actually made, however, may be characterised as *second-best* if those necessary assumptions do not hold in practice (Culyer, 2016). For example, there may be heterogeneity in the size of local budgets for health care that affects the opportunity cost of displaced health technologies (Gafni et al., 2006; Eckermann et al., 2014; Paulden et al., 2014a), and health technologies may not be perfectly divisible or exhibit constant returns to scale (Weinstein et al., 1973; Birch et al., 1993). The practical implication of applying first-best solutions in a second-best world, the *theory of the second-best* (Lipsey et al., 1956), is that resource allocation decisions may move outcomes away from maximising the objective function of population health (Birch et al., 1992; Sculpher et al., 2005b; Culyer, 2016).

The capacity to implement a testing strategy is instrumental to the application of a stratified medicine in practice, and its absence is therefore indicative of a second-best environment. The capacity to use a medical test may be affected by a complex series of constraints, including the laboratory facilities available to analyse samples from a test (Cree et al., 2014), the time taken to collect and process samples (Grasas et al., 2014), and the availability of a computerised system to present and/or interpret the results of a test within a routine consultation (van Rooij et al., 2012). Such capacity constrains may impose an opportunity cost on population health (Jahn et al., 2010) which, if unaccounted for, may have overestimated the relative cost-effectiveness of stratified medicine in this thesis.

7.4. Future Research

The findings presented in this thesis, and their respective limitations, stimulated a number of questions that required further research. The topics for further research, which were not subsequently addressed by the thesis, are now summarised by the chapter in which they were presented.

Chapter Three

(i) To understand the decision-making process and evidence requirements of local decision-makers when recommending a new testing strategy to stratify treatment in clinical practice, by conducting a qualitative interview-based study with members of different CCGs in England;

(ii) To conduct a qualitative interview-based study with a sample of rheumatology nurse specialists in England, in order to explore their perceived influence on routine prescribing decisions for patients with RA.

Chapter Four

(i) To replicate the results of the econometric study in *Chapter Four* by using a larger sample of patients with RA in England from the BSRBR register;

(ii) To utilise the data from patients that were prescribed biosimilar TNFi therapies in England, that were collected recently by the BSRBR register, to estimate the factors that influenced their selection in routine prescribing decisions;

(iii) To estimate the magnitude of influence that specific hospital-level factors (for example, size, quality, or teaching status) had on the choice of TNFi prescribed

to patients with RA in England, to partially explain the unobservable between-hospital heterogeneity that was present in *Chapter Four*.

Chapter Five

(i) To extend the algebraic early model-based conceptualisation technique (Section 5.4) by describing the profile of other potentially relevant outcomes over time (depending on the perspective of the analysis);

Chapter Six

(i) To conduct further prospective research on the accuracy of adalimumab ADAb and drug level ELISA testing in a sample of patients with RA, in order to reduce the parameter uncertainty that was present in the decision analytic model;

(ii) To conduct further prospective research on the health consequences (as a change in HAQ score) associated with using the ELISA-based testing strategies to stratify treatment for patients with RA, in order to reduce the parameter uncertainty that was present in the decision analytic model;

(iii) To update the statistical relationship between the HAQ score and the annual mean days hospitalised using a sample of patients with RA in England that were prescribed treatments according to a more-recent therapeutic paradigm.

(iv) To investigate the relative cost-effectiveness of using ADAb and drug level testing to stratify treatment for patients with RA that were treated with different monoclonal TNFi therapies (infliximab, golimumab, certolizumab pegol) as the clinical evidence base develops;

(v) To estimate the relative cost-effectiveness of prescribing reduced-dose adalimumab for a range of different diseases, including psoriasis, Crohn's disease, and psoriatic arthritis.

The two broader limitations of the thesis, described in Section 7.3, also identified topics that may be addressed by larger programmes of subsequent research. These potential research programmes are now described.

Programme One: The Non-health Benefits of Stratified Medicine

Given the objective of population health maximisation and the fundamental principal that the opportunity cost of any incremental health care expenditure will inevitably fall on

health itself, the incorporation of non-health benefits in the decision-making process (such as the value of diagnostic information derived from a test) has the potential to reduce overall population health (Sculpher et al., 2017). Four relevant topics for further research were therefore:

(i) To investigate the normative case for whether the diagnostic information derived from a medical test ought to be an appropriate benefit by which to allocate population health care resources;
(ii) To explore how the value to patients and clinicians, derived from knowing specific diagnostic information, may be quantified;
(iii) To quantify the trade-off between health outcomes and the value derived from diagnostic information, acceptable to decision-makers, under a scenario that knowing the result of a test will not improve health outcomes;
(iv) To quantify the opportunity cost (in terms of population health forgone) of allocating health care resources towards an exemplar medical test that provides no direct health benefit, but provides value from diagnostic information only.

Programme Two: Capacity to Implement Stratified Medicine

The capacity to implement a testing strategy is essential to realise the (clinical and economic) benefits of a stratified medicine in routine practice. However, the evidence generated by an economic evaluation conventionally assumes the absence of any constraint on capacity (Crane et al., 2013). Three relevant topics for further research were therefore:

(i) To investigate if, and how, published model-based economic evaluations of stratified medicine have parametrised capacity constraints within the structure of their decision analytic models;
(ii) To review the health technologies recommended by previous NICE DAP appraisals in terms of the pre-existing capacity available to implement the testing strategies nationally;
(iii) To estimate the opportunity cost associated with the imperfect implementation of an exemplar testing strategy to stratify treatment, and how dynamic changes to capacity constraints over time may affect the relative cost-effectiveness of testing.

7.5 Conclusion

The *personalisation* of health care, by tailoring treatment decisions to specific known characteristics of patients, through the means of stratified medicine, has been expressed as a high-priority for health policy in England. The successful introduction of stratified medicine into routine clinical practice is reliant on the development of economic evidence to advance its uptake within the NHS in England.

This thesis aimed to provide evidence for the relative cost-effectiveness of a biomarker test to stratify treatment for patients with RA, consistent with the requirements of decision-makers for the NHS in England. A specific example of stratified medicine for patients with RA, by testing adalimumab ADAb and drug levels with commercial ELISA-based health technologies, was used as a case study. The thesis utilised a mixed methods approach and produced seven clear contributions to advance knowledge on the economics of stratified medicine in RA.

The distinct challenges associated with producing economic evidence early in the product lifecycle of a new medical test were identified throughout the thesis. For example, such challenges included (i) handling a lack of end-to-end evidence, (ii) uncertainty over how to use the ELISA tests in practice, (iii) uncertainty in the accuracy of testing, (iv) uncertainty in the health outcomes and resources associated with testing, and (v) the potential barriers to using the tests to stratify treatment in routine practice.

The results of the economic evaluation suggested that, based on current evidence, the use of adalimumab ADAb and drug level testing to stratify treatment was unlikely to be a relatively cost-effective use of health care resources in England, according to the conventional thresholds for cost-effectiveness assumed by NICE. There was considerable decision uncertainty associated with this result and further prospective research, in particular relating to the accuracy of testing and the health outcomes associated with treatment stratification, was determined to be potentially valuable to the NHS.

The key contributions of this thesis were in advancing the economic evidence base for stratified medicine in RA and by demonstrating the value of producing relevant economic evidence for decision-makers early within the product lifecycle of a new medical test to stratify a subsequent treatment decision.

References

Aaltonen, K., Virkki, L., Malmivaara, A., Konttinen, Y., Nordström, D., & Blom, M. (2012). "Systematic Review and Meta-Analysis of the Efficacy and Safety of Existing TNF Blocking Agents in Treatment of Rheumatoid Arthritis", *PLoS One*, Vol. 7, 1(e30275), pp. 1-14.

Adams, R., Craig, B., Veale, D., Bresnihan, B., Walsh, C., FitzGerald, O., & Barry, M. (2011). "The Impact of a Revised EQ-5D Population Scoring on Preference-based Utility Scores in an Inflammatory Arthritis Cohort", *Value in Health*, Vol. 14, 6, pp. 921-927.

Aggarwal, R., Ringold, S., Khanna, D., Neogi, T., Johnson, S., Miller, A., Brunner, H., Ogawa, R., Felson, D., Ogdie, A., Aletaha, D., & Feldman, B. (2015). "Distinctions between Diagnostic and Classification Criteria?", *Arthritis Care & Research*, Vol. 67, 7, pp. 891-897.

Akaike, H. (1974). "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, Vol. 19, 6, pp. 716-723.

Akehurst, R. (2003). "Making Decisions on Technology Availability in the British National Health Service - Why we Need Reliable Models", *Value in Health*, Vol. 6, 1, pp. 3-5.

Albano, S., Santana-Sahagun, E., & Weisman, M. (2001). "Cigarette Smoking and Rheumatoid Arthritis", *Seminars in Arthritis and Rheumatism*, Vol. 31, 3, pp. 146-159.

Aletaha, D., Neogi, T., Silman, A., Funovits, J., Felson, D., Bingham, C., Birnbaum, G., Burmester, G., Bykerk, V., Cohen, M., Combe, B., Costenbader, K., Dougados, M., Emery, P., Ferraccioli, G., Hazes, J., Hobbs, K., Huizinga, T., Kavanaugh, A., Kay, J., Kvien, T., Laing, T., Mease, P., Ménard, H., Moreland, L., Naden, R., Pincus, T., Smolen, J., Stanislawska-Biernat, E., Symmons, D., Tak, P., Upchurch, K., Vencovský, J., Wolfe, F., & Hawker, G. (2010). "2010 Rheumatoid Arthritis Classification Criteria: An American College of Rheumatology/European League Against Rheumatism Collaborative Initiative", *Annals of the Rheumatic Diseases*, Vol. 62, 9, pp. 2569-2581.

Anderson, R. (2010). "Systematic Reviews of Economic Evaluations: Utility or Futility?", *Health Economics*, Vol. 19, 3, pp. 350-364.

Annemans, L., Genesté, B., & Jolain, B. (2000). "Early Modelling for Assessing Health and Economic Outcomes of Drug Therapy", *Value in Health*, Vol. 3, 6, pp. 427-434.

Annemans, L., Redekop, W., & Payne, K. (2013). "Current Methodological Issues in the Economic Assessment of Personalized Medicine", *Value in Health*, Vol. 16, S6, pp. S20-S26.

Arnett, F., Edworthy, S., Bloch, D., McShane, D., Fries, J., Cooper, N., Healey, L., Kaplan, S., Liang, M., Luthra, H., Medsger, T., Mitchell, D., Neustadt, D., Pinals, R., Schaller, J., Sharp, J., Wilder, R., & Hunder, G. (1988). "The American Rheumatism Association 1987 Revised Criteria for the Classification of Rheumatoid Arthritis", *Arthritis & Rheumatology*, Vol. 31, 3, pp. 315-324.

Arrow, K. (1963). "Uncertainty and the Welfare Economics of Medical Care", *The American Economic Review*, Vol. 53, 5, pp. 941-973.

Arrow, K., & Lind, A. (1970). "Uncertainty and the Evaluation of Public Investment Decisions", *The American Economic Review*, Vol. 60, 3, pp. 364-378.

Ayres, L. (2008). "Semi-structured Interview". In: Given, L. (Ed.) *The SAGE Encyclopedia of Qualitative Research Methods*. (1 ed.) pp. 810-811. London: SAGE Publications Ltd.

Baillet, A., Gossec, L., Carmona, L., de Wit, M., Eijk-Hustings, Y., Bertheussen, H., Alison, K., Toft, M., Kouloumas, M., Ferreira, R., Oliver, S., Rubbert-Roth, A., van Assen, S., Dixon, W., Finckh, A., Zink, A., Kremer, J., Kvien, T., Nurmohamed, M., van der Heijde, D., & Dougados, M. (2016). "Points to Consider for Reporting, Screening for and Preventing Selected Comorbidities in Chronic Inflammatory Rheumatic Diseases in Daily Practice: A EULAR Initiative", *Annals of the Rheumatic Diseases*, Vol. 75, 6, pp. 965-973.

Baio, G., & Dawid, A. (2015). "Probabilistic Sensitivity Analysis in Health Economics", *Statistical Methods in Medical Research*, Vol. 24, 6, pp. 615-634.

Bansback, N., Ara, R., Karnon, J., & Anis, A. (2008). "Economic Evaluations in Rheumatoid Arthritis: A Critical Review of Measures Used to Define Health States", *Pharmacoeconomics*, Vol. 26, 5, pp. 395-408.

Bansback, N., Regier, D., Ara, R., Brennan, A., Shojania, K., Esdaile, J., Anis, A., & Marra, C. (2005). "An Overview of Economic Evaluations for Drugs Used in Rheumatoid Arthritis: Focus on Tumour Necrosis Factor-α Antagonists", *Drugs*, Vol. 65, 4, pp. 473-496.

Bartelds, G., Krieckaert, C., Nurmohamed, M., van Schouwenburg, P., Lems, W., Twisk, J., Dijkmans, B., Aarden, L., & Wolbink, G. (2011). "Development of Antidrug Antibodies

Against Adalimumab and Association with Disease Activity and Treatment Failure During Long-term Follow-up", *Journal of the American Medical Association*, Vol. 305, 14, pp. 1460-1468.

Barton, A., & Pitzalis, C. (2016). "Stratified Medicine in Rheumatoid Arthritis - The MATURA Programme: Targeted Treatment for Patients", *Rheumatology*, Vol. [Epub ahead of print], DOI: 10.1093/rheumatology/kew369, pp. 1-4.

Barton, G., Briggs, A., & Fenwick, E. (2008). "Optimal Cost-effectiveness Decisions: The Role of the Cost-effectiveness Acceptability Curve (CEAC), the Cost-effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfect Information (EVPI)", *Value in Health*, Vol. 11, 5, pp. 886-897.

Barton, P. (2011). "Development of the Birmingham Rheumatoid Arthritis Model: Past, Present and Future Plans", *Rheumatology*, Vol. 50, S4, pp. iv32-iv38.

Barton, P., Bryan, S., & Robinson, S. (2004a). "Modelling in the Economic Evaluation of Health Care: Selecting the Appropriate Approach", *Journal of Health Services Research & Policy*, Vol. 9, 2, pp. 110-118.

Barton, P., Jobanputra, P., Wilson, J., Bryan, S., & Burls, A. (2004b). "The use of Modelling to Evaluate New Drugs for Patients with a Chronic Condition: the Case of Antibodies against Tumour Necrosis Factor in Rheumatoid Arthritis", *Health Technology Assessment*, Vol. 8, 11, pp. 1-91.

Basu, A., & Meltzer, D. (2007). "Value of Information on Preference Heterogeneity and Individualized Care", *Medical Decision Making*, Vol. 27, 2, pp. 112-127.

Baxter, K., Weiss, M., & Le Grand, J. (2008). "The Dynamics of Commissioning Across Organisational and Clinical Boundaries", *Journal of Health Organization and Management*, Vol. 22, 2, pp. 111-128.

Beaulieu, M., de Denus, S., & Lachaine, J. (2010). "Systematic Review of Pharmacoeconomic Studies of Pharmacogenomic Tests", *Pharmacogenomics*, Vol. 11, 11, pp. 1573-1590.

Beckert, W., Christensen, M., & Collyer, K. (2012). "Choice of NHS-funded Hospital Services in England", *The Economic Journal*, Vol. 122, 560, pp. 400-417.

Becla, L., Lunshof, J., Gurwitz, D., den Bäumen, T., Westerhoff, H., Lange, B., & Brand, A. (2011). "Health Technology Assessment in the Era of Personalized Health Care", *International Journal of Technology Assessment in Health Care*, Vol. 27, 2, pp. 118-126.

Bendtzen, K. (2011). "Is There a Need for Immunopharmacologic Guidance of Anti–Tumor Necrosis Factor Therapies?", *Arthritis & Rheumatology*, Vol. 63, 4, pp. 867-870.

Benzeval, M., Kumari, M., & Jones, A. (2016). "How do Biomarkers and Genetics Contribute to Understanding Society", *Health Economics*, Vol. 25, 10, pp. 1219-1222.

Bergquist, S., Felson, D., Prashker, M., & Freedberg, K. (1995). "The Cost-effectiveness of Liver Biopsy in Rheumatoid Arthritis Patients Treated with Methotrexate", *Arthritis & Rheumatology*, Vol. 38, 3, pp. 326-333.

Berm, E., de Looff, M., Wilffert, B., Boersma, C., Annemans, L., Vegter, S., van Boven, J., & Postma, M. (2016). "Economic Evaluation of Pharmacogenetic and Pharmacogenomic Screening Tests: A Systematic Review. Second Update of the Literature", *PLoS One*, Vol. 11, 1, pp. 1-22.

Birch, S., & Gafni, A. (1992). "Cost Effectiveness/utility Analyses. Do Current Decision Rules Lead us to where we want to be?", *Journal of Health Economics*, Vol. 11, 3, pp. 279-296.

Birch, S., & Gafni, A. (1993). "Changing the Problem to Fit the Solution: Johannesson and Weinstein's (Mis) Application of Economics to Real World Problems", *Journal of Health Economics*, Vol. 12, 4, pp. 469-476.

Blagoev, K., Wilkerson, J., & Fojo, T. (2012). "Hazard Ratios in Cancer Clinical Trials - A Primer", *Nature Reviews Clinical Oncology*, Vol. 9, 3, pp. 178-183.

Blair, E., Clarke, B., & O'Neill, T. (2011). "Sustaining Development of Stratified Medicines in the UK Healthcare System: A Commentary", *Personalized Medicine*, Vol. 8, 5, pp. 517-521.

Blake, T., Rao, V., Hashmi, T., Erb, N., O'Reilly, S., Shaffu, S., Obrenovic, K., & Packham, J. (2014). "The Perplexity of Prescribing and Switching of Biologic Drugs in Rheumatoid Arthritis: a UK Regional Audit of Practice", *BMC Musculoskeletal Disorders*, Vol. 15, 290, pp. 1-8.

Blomqvist, A. (1991). "The Doctor as Double Agent: Information Asymmetry, Health Insurance, and Medical Care", *Journal of Health Economics*, Vol. 10, 4, pp. 411-432.

Bloss, G., & Haaga, J. (2013). "Economics of Personalized Health Care and Prevention: Introduction", *Forum for Health Economics and Policy*, Vol. 16, 2, pp. S1-S11.

Boers, M. (2003). "Understanding the Window of Opportunity Concept in Early Rheumatoid Arthritis", *Arthritis & Rheumatology*, Vol. 48, 7, pp. 1771-1774.

Bojke, L., Claxton, K., Sculpher, M., & Palmer, S. (2009). "Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods", *Value in Health*, Vol. 12, 5, pp. 739-749.

Bormann, I. (2016). "DigitizeIt: Digitizer Software". *Version 2.2.*, Braunschweig: DigitizeIt.

Boshuizen, H., & van Baal, P. (2009). "Probabilistic Sensitivity Analysis: Be a Bayesian", *Value in Health*, Vol. 12, 8, pp. 1210-1214.

Braun, A., & Clarke, V. (2006). "Using Thematic Analysis in Psychology", *Qualitative Research in Psychology*, Vol. 3, 2, pp. 77-101.

Brazier, J., Deverill, M., Green, C., Harper, R., & Booth, A. (1999). "A Review of the use of Health Status Measures in Economic Evaluation", *Health Technology Assessment*, Vol. 3, 9, pp. 1-166.

Brazier, J., Roberts, J., & Deverill, M. (2002). "The Estimation of a Preference-based Measure of Health from the SF-36", *Journal of Health Economics*, Vol. 21, 2, pp. 271-292.

Brazier, J., & Tsuchiya, A. (2015). "Improving Cross-Sector Comparisons: Going Beyond the Health-Related QALY", *Applied Health Economics and Health Policy*, Vol. 13, 6, pp. 557-565.

Brennan, A., & Akehurst, R. (2000). "Modelling in Health Economic Evaluation. What is its Place? What is its Value?", *Pharmacoeconomics*, Vol. 17, 5, pp. 445-459.

Brennan, A., Chick, S., & Davies, R. (2006). "A Taxonomy of Model Structures for Economic Evaluation of Health Technologies", *Health Economics*, Vol. 15, 12, pp. 1295-1310.

Briggs, A. (1999). "A Bayesian Approach to Stochastic Cost-effectiveness Analysis", *Health Economics*, Vol. 8, 3, pp. 257-261.

Briggs, A. (2000). "Handling Uncertainty in Cost-effectiveness Models", *Pharmacoeconomics*, Vol. 17, 5, pp. 479-500.

Briggs, A., & Gray, A. (1999). "Handling Uncertainty when Performing Economic Evaluation of Healthcare Interventions", *Health Technology Assessment*, Vol. 3, 2, pp. 1-134.

Briggs, A., & O'Brien, B. (2001). "The Death of Cost-minimization Analysis?", *Health Economics*, Vol. 10, 2, pp. 179-184.

Briggs, A., & Sculpher, M. (1998). "An Introduction to Markov Modelling for Economic Evaluation", *Pharmacoeconomics*, Vol. 13, 4, pp. 397-409.

Briggs, A., Sculpher, M., & Claxton, K. (2006). *Decision Modelling for Health Economic Evaluation*. (1 ed.) Oxford: Oxford University Press.

British National Formulary. (2016). "British National Formulary: Accessed via MedicinesComplete". [Online] Available at: http://www.medicinescomplete.com/mc/bnf/current/ [Accessed 8 April 2016].

Brockis, E., Marsden, G., Cole, A., & Devlin, N. (2016). "A Review of NICE Methods Across Health Technology Assessment Programmes: Differences, Justifications and Implications". *Research Paper 16/03*, London: Office of Health Economics.

Brouwer, W., Culyer, A., van Exel, N., & Rutten, F. (2008). "Welfarism Vs. Extra-welfarism", *Journal of Health Economics*, Vol. 27, 2, pp. 325-338.

Bruce, B., & J., F. (2003). "The Stanford Health Assessment Questionnaire: Dimensions and Practical Applications", *Health and Quality of Life Outcomes*, Vol. 1, 20, pp. 1-6.

Bryan, S., Mitton, C., & Donaldson, C. (2014). "Breaking the Addiction to Technology Adoption", *Health Economics*, Vol. 23, 4, pp. 379-383.

Bryan, S., Williams, I., & McIver, S. (2007). "Seeing the NICE Side of Cost-effectiveness Analysis: A Qualitative Investigation of the use of CEA in NICE Technology Appraisals", *Health Economics*, Vol. 16, 2, pp. 179-193.

Buchannan, J., Wordsworth, S., & Schuh, A. (2013). "Issues Surrounding the Health Economic Evaluation of Genomic Technologies", *Pharmacogenomics*, Vol. 14, 15, pp. 1833-1847.

Buckley, F., Finckh, A., Huizinga, T., Dejonckheere, F., & Jansen, J. (2015). "Comparative Efficacy of Novel DMARDs as Monotherapy and in Combination with Methotrexate in Rheumatoid Arthritis Patients with Inadequate Response to Conventional DMARDs: A Network Meta-Analysis", *Journal of Managed Care & Speciality Pharmacy*, Vol. 21, 5, pp. 409-423.

Buisman, L., Rutten-van Mölken, M., Postmus, D., Luime, J., Uyl-de Groot, C., & Redekop, W. (2016). "The Early Bird Catches the Worm: Early Cost-effectiveness Analysis of New Medical Tests", *International Journal of Technology Assessment in Health Care*, Vol. 32, 1-2, pp. 46-53.

Bukhari, M., Abernethy, R., Deighton, C., Ding, T., Hyrich, K., Lunt, M., Lugmani, R., Kiely, P., Bosworth, A., Ledingham, J., Ostör, A., Gadsby, K., McKenna, F., Finney, D., Dixey, J., & BSR and BHPR Standards Guidelines and Audit Working Group. (2011). "BSR and BHPR Guidelines on the use of Rituximab in Rheumatoid Arthritis", *Rheumatology*, Vol. 50, 11, pp. 2311-2313.

Buxton, M. (2006). "Economic Evaluation and Decision Making in the UK", *Pharmacoeconomics*, Vol. 24, 11, pp. 1133-1142.

Buxton, M., Drummond, M., van Hout, B., Prince, R., Sheldon, T., Szucs, T., & Vray, M. (1997). "Modelling in Economic Evaluation: An Unavoidable Fact of Life", *Health Economics*, Vol. 6, 3, pp. 217-227.

Bykerk, V., Bingham, C., Choy, E., Lin, D., Alten, R., Christensen, R., Furst, D., Hewlett, S., Leong, A., March, L., Woodworth, T., Boire, G., Haraoui, B., Hitchon, C., Jamal, S., Keystone, E., Pope, J., Tin, D., Thorn, J., & Bartlett, S. (2016). "Identifying Flares in Rheumatoid Arthritis: Reliability and Construct Validation of the OMERACT RA Flare Core Domain Set", *RMD Open*, Vol. 2, e000225, pp. 1-10.

Bykerk, V., Shadick, N., Frits, M., Bingham, C., Jeffery, I., Iannaccone, C., Weinblatt, M., & Solomon, D. (2014). "Flares in Rheumatoid Arthritis: Frequency and Management. A Report from the BRASS Registry", *The Journal of Rheumatology*, Vol. 41, 2, pp. 227-234.

Byron, S., Crabb, N., George, E., Marlow, M., & Newland, A. (2014). "The Health Technology Assessment of Companion Diagnostics: Experience of NICE ", *Clinical Cancer Research*, Vol. 20, 6, pp. 1469-1476.

Cameron, A., & Miller, D. (2015). "A Practitioner's Guide to Cluster-Robust Inference", *The Journal of Human Resources*, Vol. 50, 2, pp. 317-372.

Caro, J. (2005). "Pharmacoeconomic Analysis using Discrete Event Simulation", *Pharmacoeconomics*, Vol. 23, 4, pp. 323-332.

Caro, J., & Möller, J. (2016a). "Advantages and Disadvantages of Discrete-event Simulation for Health Economic Analyses", *Expert Review of Pharmacoeconomics & Outcomes Research*, Vol. 16, 3, pp. 327-329.

Caro, J., Möller, J., Karnon, J., Stahl, J., & Ishak, J. (2016b). *Discrete Event Simulation for Health Technology Assessment*. (1 ed.) Boca Raton: CRC Press.

Carter, C., Changolkar, A., & McKenzie, R. (2012). "Adalimumab, Etanercept, and Infliximab Utilization Patterns and Drug Costs among Rheumatoid Arthritis Patients", *Journal of Medical Economics*, Vol. 15, 2, pp. 332-339.

Casey, D. (2016). "Key Strategic Factors for Stakeholders in the Current Global Biosimilar Market", *Drug Discovery Today*, Vol. 21, 2, pp. 208-211.

Centre for Reviews and Dissemination. (2009). *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. (1 ed.) York: Centre for Reviews and Dissemination.

Chen, D., Chen, Y., Hsieh, T., Hung, W., Hsieh, C., Chen, H., Tang, K., & Lan, J. (2016). "Drug Trough Levels Predict Therapeutic Responses to Dose Reduction of Adalimumab for Rheumatoid Arthritis Patients During 24 Weeks of Follow-up", *Rheumatology*, Vol. 55, 1, pp. 143-148.

Chen, D., Chen, Y., Tsai, W., Tseng, J., Chen, Y., Hsieh, C., Hung, W., & Lan, J. (2015a). "Significant Associations of Antidrug Antibody Levels with Serum Drug Trough Levels and Therapeutic Response of Adalimumab and Etanercept Treatment in Rheumatoid Arthritis", *Annals of the Rheumatic Diseases*, Vol. 74, 3, pp. 1-9.

Chen, D., Chen, Y., Tsai, W. T., J., Chen, Y., Hsieh, C., Hung, W., & Lan, J. (2015b). "Significant Associations of Antidrug Antibody Levels with Serum Drug Trough Levels and Therapeutic Response of Adalimumab and Etanercept Treatment in Rheumatoid Arthritis", *Annals of the Rheumatic Diseases*, Vol. 74, 3, p. e16.

Chilcott, J., Brennan, A., Booth, A., Karnon, J., & Tappenden, P. (2003). "The Role of Modelling in Prioritising and Planning Clinical Trials", *Health Technology Assessment*, Vol. 7, 23, pp. 1-138.

Chilcott, J., Tappenden, P., Rawdin, A., Johnson, M., Kaltenthaler, E., Paisley, S., Papaioannou, D., & Shippam, A. (2010). "Avoiding and Identifying Errors in Health Technology Assessment Models: Qualitative Study and Methodological Review", *Health Technology Assessment*, Vol. 14, 25, pp. 1-135.

Choy, E., Aletaha, D., Behrens, F., Finckh, A., Gomez-Reino, J., Gottenberg, J., Schuch, F., & Rubbert-Roth, A. (2016). "Monotherapy with Biologic Disease-modifying Anti-rheumatic Drugs in Rheumatoid Arthritis", *Rheumatology*, Vol. [Epub ahead of print], DOI: 10.1093/rheumatology/kew271, pp. 1-9.

Claxton, K. (1999a). "Bayesian Approaches to the Value of Information: Implications for the Regulation of New Pharmaceuticals", *Health Economics*, Vol. 8, 3, pp. 269-274.

Claxton, K. (1999b). "The Irrelevance of Inference: A Decision-making Approach to the Stochastic Evaluation of Health Care Technologies", *Journal of Health Economics*, Vol. 18, 3, pp. 341-364.

Claxton, K. (2008). "Exploring Uncertainty in Cost-effectiveness Analysis", *Pharmacoeconomics*, Vol. 26, 9, pp. 781-798.

Claxton, K., Ginnelly, L., Sculpher, M., Philips, Z., & Palmer, S. (2004). "A Pilot Study on the use of Decision Theory and Value of Information Analysis as Part of the NHS Technology Assessment Programme", *Health Technology Assessment*, Vol. 8, 31, pp. 1-118.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P., & Sculpher, M. (2015a). "Methods for the Estimation of the National Institute for Health and Care Excellence Cost-effectiveness Threshold", *Health Technology Assessment*, Vol. 19, 14, pp. 1-503.

Claxton, K., Neumann, P., Araki, S., & Weinstein, M. (2001). "Bayesian Value-of-information Analysis. An Application to a Policy Model of Alzheimer's Disease", *International Journal of Technology Assessment in Health Care*, Vol. 17, 1, pp. 38-55.

Claxton, K., Paulden, M., Gravelle, H., Brouwer, W., & Culyer, A. (2011). "Discounting and Decision Making in the Economic Evaluation of Health-care Technologies", *Health Economics*, Vol. 20, 1, pp. 2-15.

Claxton, K., & Posnett, J. (1996). "An Economic Approach to Clinical Trial Design and Research Priority-setting", *Health Economics*, Vol. 5, 6, pp. 513-524.

Claxton, K., & Sculpher, M. (2006a). "Using Value of Information Analysis to Prioritise Health Research: Some Lessons from Recent UK Experience", *Pharmacoeconomics*, Vol. 24, 11, pp. 1055-1068.

Claxton, K., Sculpher, M., Culyer, A., McCabe, C., Briggs, A., Akehurst, R., Buxton, M., & Brazier, J. (2006b). "Discounting and Cost-effectiveness in NICE - Stepping Back to Sort Out a Confusion", *Health Economics*, Vol. 15, 1, pp. 1-4.

Claxton, K., Sculpher, M., & Drummond, M. (2002). "A Rational Framework for Decision Making by the National Institute for Clinical Excellence (NICE)", *The Lancet*, Vol. 360, 9334, pp. 711-715.

Claxton, K., Sculpher, M., McCabe, C., Briggs, A., Akehurst, R., Buxton, M., Brazier, J., & O'Hagen, T. (2005). "Probabilistic Sensitivity Analysis for NICE Technology Assessment: Not an Optional Extra", *Health Economics*, Vol. 14, 4, pp. 339-347.

Claxton, K., Sculpher, M., Palmer, D., & Culyer, A. (2015b). "Causes for Concern: Is NICE Failing to Uphold its Responsibilities to all NHS Patients?", *Health Economics*, Vol. 24, 1, pp. 1-7.

Claxton, K., Walker, S., Palmer, S., & Sculpher, M. (2010). "Appropriate Perspectives for Health Care Decisions". *CHE Research Paper 54*, York: University of York.

Coast, J. (1999). "The Appropriate uses of Qualitative Methods in Health Economics", *Health Economics*, Vol. 8, 4, pp. 345-353.

Coast, J., McDonald, R., & Baker, R. (2004). "Issues Arising from the use of Qualitative Methods in Health Economics", *Journal of Health Services Research & Policy*, Vol. 9, 3, pp. 171-176.

Coast, J., Smith, R., & Lorgelly, P. (2008). "Welfarism, Extra-welfarism and Capability: The Spread of Ideas in Health Economics", *Social Science & Medicine*, Vol. 67, 7, pp. 1190-1198.

Collins, F., & Varmus, H. (2015). "A New Initiative on Precision Medicine", *The New England Journal of Medicine*, Vol. 372, 9, pp. 793-795.

Cookson, R. (2015). "Justice and the NICE Approach", *Journal of Medical Ethics*, Vol. 41, 1, pp. 99-102.

Cookson, R., McDaid, D., & Maynard, A. (2001). "Wrong SIGN, NICE Mess: Is National Guidance Distorting Allocation of Resources", *British Medical Journal*, Vol. 323, 7315, pp. 743-745.

Coyle, D., Buxton, M., & O'Brien, B. (2003). "Stratified Cost-effectiveness Analysis: A Framework for Establishing Efficient Limited use Criteria", *Health Economics*, Vol. 12, 5, pp. 421-427.

Coyle, D., Lee, K., & Cooper, N. (2010). "Chapter 9: Use of Evidence in Decision Models". In: Shemilt, I., Mugford, M., Vale, L., Marsh, K., & Donaldson, C. (Eds.) *Evidence-based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice*. (2 ed.) pp. 106-113. Oxford: Wiley-Blackwell.

Craig, D., Rice, S., Aguiar-Ibáñez, R., Glanville, J., Kleijnen, J., & Drummond, M. (2007). "Chapter 3: Guidance for Writing NHS EED Abstracts". In: Craig, D., & Rice, S. (Eds.) *NHS Economic Evaluations Database Handbook*. (3 ed.) pp. 13-64. York: Centre for Reviews and Dissemination.

Crane, G., Kymes, S., Hiller, J., Casson, R., Martin, A., & Karnon, J. (2013). "Accounting for Costs, QALYs, and Capacity Constraints: Using Discrete-event Simulation to Evaluate Alternative Service Delivery and Organizational Scenarios for Hospital-based Glaucoma Services", *Medical Decision Making*, Vol. 33, 8, pp. 986-997.

Cree, I., Deans, Z., Ligtenberg, M., Normanno, N., Edsjö, A., Rouleau, E., Solé, F., Thunnissen, E., Timens, W., Schuuring, E., Dequeker, E., Murray, S., Dietel, M., Groenen, P., & van Krieken, J. (2014). "Guidance for Laboratories Performing Molecular Pathology for Cancer Patients", *Journal of Clinical Pathology*, Vol. 67, 11, pp. 923-931.

Crofford, L. (2013). "Use of NSAIDs in Treating Patients with Arthritis", *Arthritis Research & Therapy*, Vol. 15, S3, pp. 1-10.

Culyer, A. (1997). "The Rationing Debate: Maximising the Health of the Whole Community", *British Medical Journal*, Vol. 314, 7081, pp. 667-669.

Culyer, A. (2016). "Cost-effectiveness Thresholds in Health Care: A Bookshelf Guide to their Meaning and use", *Health Economics, Policy and Law*, Vol. 11, 4, pp. 415-432.

Culyer, A., & Lomas, J. (2006). "Deliberative Processes and Evidence-informed Decision Making in Healthcare: Do they Work and how Might we Know?", *Evidence & Policy*, Vol. 2, 3, pp. 357-371.

Cuppen, B., Welsing, P., Sprengers, J., Bijlsma, J., Marijnissen, A., van Laar, J., Lafeber, F., & Nair, S. (2016). "Personalized Biological Treatment for Rheumatoid Arthritis: A Systematic Review with a Focus on Clinical Applicability", *Rheumatology*, Vol. 55, 5, pp. 826-839.

Currier, K. (2000). *Comparative Statics Analysis in Economics*. (1 ed.) London: World Scientific Publishing.

Curtis, J., & Singh, J. (2011). "Use of Biologics in Rheumatoid Arthritis: Current and Emerging Paradigms of Care", *Clinical Therapeutics*, Vol. 33, 6, pp. 679-707.

Cylus, J., Richardson, E., Findley, L., Longley, M., O'Neill, C., & Steel, D. (2015). "United Kingdom: Health System Review", *Health Systems in Transition*, Vol. 17, 5, pp. 1-126.

Daïen, C., Daïen, V., Parussini, E., Dupuy, A., Combe, B., & Morel, J. (2012). "Etanercept Concentration in Patients with Rheumatoid Arthritis and Its Potential Influence on Treatment Decisions: A Pilot Study", *The Journal of Rheumatology*, Vol. 39, 8, pp. 1533-1538.

Dakin, H. (2013). "Review of Studies Mapping from Quality of Life or Clinical Measures to EQ-5D: An Online Database", *Health and Quality of Life Outcomes*, Vol. 11, 151, pp. 1-6.

Dakin, H., Devlin, N., Feng, Y., Rice, N., O'Neill, P., & Parkin, D. (2015). "The Influence of Cost-effectiveness and other Factors on NICE Decisions", *Health Economics*, Vol. 24, 10, pp. 1256-1271.

Davis, J., Furstenthal, L., Desai, A., Norris, T., Sutaria, S., Fleming, E., & Ma, P. (2009). "The Microeconomics of Personalized Medicine: Today's Challenge and Tomorrow's Promise", *Nature Reviews Drug Discovery*, Vol. 8, 4, pp. 279-286.

Davis, S., Stevenson, M., Tappenden, P., & Wailoo, A. (2014). "Cost-effectiveness Modelling using Patient-level Simulation". *NICE DSU Technical Support Document 15*, Sheffield: Decision Support Unit.

den Broeder, A., van der Maas, A., & van den Bemt, B. (2013). "Antidrug Antibodies to Tumour Necrosis Factor Specific Neutralising Agents in Chronic Inflammatory Diseases: A Real Issue, a Clinical Perspective; Comment on the Article by Vincent et al", *Annals of the Rheumatic Diseases*, Vol. 72, 7, e14, pp. 1-2.

Department of Health. (2014). "Framework Agreement between the Department of Health and the National Institute for Health and Care Excellence". London: Department of Health.

Department of Health. (2016a). "The Government's Mandate to NHS England for 2016-17". London: Department of Health.

Department of Health. (2016b). "NHS Reference Costs: 2015 to 2016". London: Department of Health.

Devlin, N., & Parkin, D. (2004). "Does NICE have a Cost-effectiveness Threshold and what other Factors Influence its Decisions? A Binary Choice Analysis", *Health Economics*, Vol. 13, 5, pp. 437-452.

DeWitt, E., Glick, H., Albert, D., Joffe, M., & Wolfe, F. (2006). "Medicare Coverage of Tumor Necrosis Factor Alpha Inhibitors as an Influence on Physicians' Prescribing Behavior", *Archives of Internal Medicine*, Vol. 166, 1, pp. 57-63.

Ding, T., Ledingham, J., Luqmani, R., Westlake, S., Hyrich, K., Lunt, M., Kiely, P., Bukhari, M., Abernethy, R., Bosworth, A., Ostor, A., Gadsby, K., McKenna, F., Finney, D., Dixey, J., & Deighton, C. (2010). "BSR and BHPR Rheumatoid Arthritis Guidelines on Safety of Anti-TNF Therapies", *Rheumatology*, Vol. 49, 11, pp. 2217-2219.

Dinnes, J., Deeks, J., Kirby, J., & Roderick, P. (2005). "A Methodological Review of how Heterogeneity has been Examined in Systematic Reviews of Diagnostic Test Accuracy", *Health Technology Assessment*, Vol. 8, 12, pp. 1-113.

Dörner, T., Strand, V., Cornes, P., Gonçalves, J., Gulácsi, L., Kay, J., Kvien, T., Smolen, J., Tanaka, Y., & Burmester, G. (2016). "The Changing Landscape of Biosimilars in Rheumatology", *Annals of the Rheumatic Diseases*, Vol. 75, 6, pp. 974-982.

Doubilet, P., Begg, C., Weinstein, M., Braun, P., & McNeil, B. (1985). "Probabilistic Sensitivity Analysis using Monte Carlo Simulation: A Practical Approach", *Medical Decision Making*, Vol. 5, 2, pp. 157-177.

Dougados, M. (2016). "Comorbidities in Rheumatoid Arthritis", *Current Opinion in Rheumatology*, Vol. 28, 3, pp. 282-288.

Dowie, J. (2004). "Why Cost-Effectiveness should Trump (Clinical) Effectiveness: The Ethical Economics of the South West Quadrant", *Health Economics*, Vol. 13, 5, pp. 453-459.

Dranove, D. (2012). "Health Care Markets, Regulators, and Certifiers". In: Pauly, M., McGuire, T., & Barros, P. (Eds.) *Handbook of Health Economics: Volume 2*. (1 ed.) pp. 639-690. Oxford: Elsevier.

Drummond, M., Barbieri, M., & Wong, J. (2005). "Analytic Choices in Economic Models of Treatments for Rheumatoid Arthritis: What Makes a Difference?", *Medical Decision Making*, Vol. 25, 5, pp. 520-533.

Drummond, M., Brandt, A., Luce, B., & Rovira, J. (1993). "Standardizing Methodologies for Economic Evaluations in Health Care: Practice, Problems, and Potential", *International Journal of Technology Assessment in Health Care*, Vol. 9, 1, pp. 26-36.

Drummond, M., Sculpher, M., Claxton, K., Stoddart, G., & Torrance, G. (2015). *Methods for the Economic Evaluation of Health Care Programmes*. (4 ed.) Oxford: Oxford University Press.

Duffy, S., & Christie, J. (2007). "Chapter 2: NHS EED Process". In: Craig, D., & Rice, S. (Eds.) *NHS Economic Evaluation Database Handbook*. (3 ed.) pp. 5-12. York: Centre for Reviews and Dissemination.

Dwyer, S., & Buckle, J. (2009). "The Space Between: On Being an Insider-Outsider in Qualitative Research", *International Journal of Qualitative Methods*, Vol. 8, 1, pp. 55-63.

Eckermann, S., Karnon, J., & Willan, A. (2010). "The Value of Value of Information: Best Informing Research Design and Prioritization using Current Methods", *Pharmacoeconomics*, Vol. 28, 9, pp. 699-709.

Eckermann, S., & Pekarsky, B. (2014). "Can the Real Opportunity Cost Stand Up: Displaced Services, the Straw Man Outside the Room", *Pharmacoeconomics*, Vol. 32, 4, pp. 319-325.

Edwards, R., Charles, J., & Lloyd-Williams, H. (2013). "Public Health Economics: A Systematic Review of Guidance for the Economic Evaluation of Public Health Interventions and Discussion of Key Methodological Issues", *BMC Public Health*, Vol. 13, 1001, pp. 1-13.

Emery, P. (2012). "Optimizing Outcomes in Patients with Rheumatoid Arthritis and an Inadequate Response to Anti-TNF Treatment", *Rheumatology*, Vol. 51, s5, pp. v22-v30.

Emery, P., & Dörner, T. (2011). "Optimising Treatment in Rheumatoid Arthritis: A Review of Potential Biomarkers of Response", *Annals of the Rheumatic Diseases*, Vol. 70, 12, pp. 2063-2070.

Erreygers, G., & van Ourti, T. (2011). "Measuring Socioeconomic Inequality in Health, Health Care and Health Financing by Means of Rank-dependent Indices: A Recipe for Good Practice", *Journal of Health Economics*, Vol. 30, 4, pp. 685-694.

Espinoza, M., Manca, A., Claxton, K., & Sculpher, M. (2014a). "The Value of Heterogeneity for Cost-effectiveness Subgroup Analysis: Conceptual Framework and Application", *Medical Decision Making*, Vol. 34, 8, pp. 951-964.

Espinoza, M., Sculpher, M., Manca, A., & Basu, A. (2014b). "Analysing Heterogeneity to Support Decision Making". In: Culyer, A. (Ed.) *Encyclopedia of Health Economics Vol. 1*. (1 ed.) pp. 71-76. London: Elsevier.

European Commission. (2010). "Biomarkers for Patient Stratification". Brussels: European Commission.

European Medicines Agency. (2012). "EPAR Summary for the Public: Remicade (Infliximab)". *EMA/76495/2012 EMEA/H/C/000240* London: European Medicines Agency.

European Medicines Agency. (2014). "EPAR Summary for the Public: Enbrel (Etanercept)". *EMA/398383/2014 EMEA/H/C/000262*, London: European Medicines Agency.

European Medicines Agency. (2016a). "EPAR Summary for the Public: Cimzia (Certolizumab Pegol)". *EMA/831922/2015 EMEA/H/C/001037*, London: European Medicines Agency.

European Medicines Agency. (2016b). "EPAR Summary for the Public: Simponi (Golimumab)". *EMA/412395/2016 EMEA/H/C/000992*, London: European Medicines Agency.

European Medicines Agency. (2017). "EPAR Summary for the Public: Humira (Adalimumab)". *EMA/807118/2016 EMEA/H/C/000481*, London: European Medicines Agency.

EuroQol Group. (1990). "EuroQol - A New Facility for the Measurement of Health-related Quality of Life", *Health Policy*, Vol. 16, 3, pp. 199-208.

Faulkner, E., Annemans, L., Garrison, L., Helfand, M., Holtorf, A., Hornberger, J., Hughes, D., Li, T., Malone, D., Payne, K., Siebert, U., Towse, A., Veenstra, D., & Watkins, J. (2012). "Challenges in the Development and Reimbursement of Personalized Medicine - Payer and Manufacturer Perspectives and Implications for Health Economics and Outcomes Research: A Report for the ISPOR Personalized Medicine Special Interest Group", *Value in Health*, Vol. 15, 8, pp. 1162-1171.

Fautrel, B., & den Broeder, A. (2015). "De-intensifying Treatment in Established Rheumatoid Arthritis (RA): Why, how, when, and in whom can DMARDs be Tapered?", *Best Practice & Research Clinical Rheumatology*, Vol. 29, 4-5, pp. 550-565.

Feldmann, M. (2002). "Development of Anti-TNF Therapy for Rheumatoid Arthritis", *Nature Reviews Immunology*, Vol. 2, 5, pp. 364-371.

Feldmann, M., Brennan, F., Foxwell, B., Taylor, P., Williams, R., & Maini, R. (2005). "Anti-TNF Therapy: Where have we got to in 2005?", *Journal of Autoimmunity*, Vol. 25, S, pp. 26-28.

Feldmann, M., & Maini, R. (2010). "Anti-TNF Therapy, from Rationale to Standard of Care: What Lessons has it Taught us?", *The Journal of Immunology*, Vol. 185, 2, pp. 791-794.

Fenwick, E. (2014). "Economic Evaluation, Uncertainty in". In: Culyer, A. (Ed.) *Encyclopedia of Health Economics Vol. 1*. (1 ed.) pp. 224-231. London: Elsevier.

Fenwick, E., & Bryford, S. (2005). "A Guide to Cost-effectiveness Acceptability Curves", *The British Journal of Psychiatry*, Vol. 187, 2, pp. 106-108.

Fenwick, E., Claxton, K., & Sculpher, M. (2001). "Representing Uncertainty: The Role of Cost-effectiveness Acceptability Curves", *Health Economics*, Vol. 10, 8, pp. 779-787.

Fenwick, E., Marshall, D., Levy, A., & Nichol, G. (2006). "Using and Interpreting Cost-effectiveness Acceptability Curves: An Example using Data from a Trial of Management Strategies for Atrial Fibrillation", *BMC Health Services Research*, Vol. 6, 52, pp. 1-8.

Fenwick, E., O'Brien, B., & Briggs, A. (2004). "Cost-effectiveness Acceptability Curves - Facts, Fallacies and Frequently Asked Questions", *Health Economics*, Vol. 13, 5, pp. 405-415.

Ferguson, B. (1985). "Physician Objectives and Resource Allocation", *Journal of Health Economics*, Vol. 4, 1, pp. 35-42.

Firestein, G. (2003). "Evolving Concepts of Rheumatoid Arthritis", *Nature*, Vol. 423, 6937, pp. 356-361.

FitzGerald, C., & Hurst, S. (2017). "Implicit Bias in Healthcare Professionals: A Systematic Review", *BMC Medical Ethics*, Vol. 18, 19, pp. 1-18.

Fleeman, N., Payne, K., Newman, W., Howell, S., Boland, A., Oyee, J., Saborido, C., Santander, A., & Dickson, R. (2013). "Are Health Technology Assessments of Pharmacogenetic Tests Feasible? A Case Study of CYP2D6 Testing in the Treatment of Breast Cancer with Tamoxifen", *Personalized Medicine*, Vol. 10, 6, pp. 601-611.

Fleurbaey, M., & Schokkaert, E. (2009). "Unfair Inequalities in Health and Health Care", *Journal of Health Economics*, Vol. 27, 1, pp. 73-90.

Frick, K. (2009). "Microcosting Quantity Data Collection Methods", *Medical Care*, Vol. 47, 7S1, pp. S76-S81.

Friedman, H., & Greenblatt, D. (1986). "Rational Therapeutic Drug Monitoring", *Journal of the American Medical Association*, Vol. 256, 16, pp. 2227-2233.

Frueh, F. (2013). "Regulation, Reimbursement, and the Long Road of Implementation of Personalized Medicine - A Perspective from the United States", *Value in Health*, Vol. 16, 6, pp. S27-S31.

Fugel, H., Nuijten, M., & Faulkner, E. (2014). "The Application of Economics Concepts to Stratified Medicine - Use of Health Economics Data to Support Market Access for Stratified Medicine Interventions", *Journal of Medical Economics*, Vol. 5, 305-311,

Fugel, H., Nuijten, M., Postma, M., & Redekop, W. (2016). "Economic Evaluation in Stratified Medicine: Methodological Issues and Challenges", *Frontiers in Pharmacology*, Vol. 7, 113, pp. 1-7.

Gafni, A., & Birch, S. (2006). "Incremental Cost-effectiveness Ratios (ICERs): The Silence of the Lambda", *Social Science & Medicine*, Vol. 62, 9, pp. 2091-2100.

Gale, N., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). "Using the Framework Method for the Analysis of Qualitative Data in Multi-disciplinary Health Research", *BMC Medical Research Methodology*, Vol. 13, 117, pp. 1-8.

Ganz, M., Hansen, B., Valencia, X., & Strandberg-Larsen, M. (2015). "Key Data Elements for use in Cost-utility Modeling of Biological Treatments for Rheumatoid Arthritis", *Journal of Medical Economics*, Vol. 18, 5, pp. 366-375.

Garcês, S., Antues, M., Benito-Garcia, E., de Silva, J., Aarden, L., & Demengeot, J. (2014). "A Preliminary Algorithm Introducing Immunogenicity Assessment in the Management of Patients with RA Receiving Tumour Necrosis Factor Inhibitor Therapies", *Annals of the Rheumatic Diseases*, Vol. 73, 6, pp. 1138-1143.

Garcês, S., Demengeot, J., & Benito-Garcia, E. (2013). "The Immunogenicity of Anti-TNF Therapy in Immune-Mediated Inflammatory Diseases: A Systematic Review of the Literature with a Meta-Analysis", *Annals of the Rheumatic Diseases*, Vol. 73, 1, pp. 1947-1955.

Garrison, L., & Austin, M. (2006). "Linking Pharmacogenetics-based Diagnostics and Drugs for Personalized Medicine", *Health Affairs*, Vol. 25, 5, pp. 1281-1290.

Garrison, L., & Austin, M. (2007). "The Economics of Personalized Medicine: A Model of Incentives for Value Creation and Capture", *Therapeutic Innovation & Regulatory Science*, Vol. 41, 4, pp. 501-509.

Gavan, S., Daker-White, G., Barton, A., & Payne, K. (2016a). "OP0198-HPR Exploring Factors Which Influence Anti-TNF Treatment Decisions for Rheumatoid Arthritis in England – A Qualitative Analysis", *Annals of the Rheumatic Diseases*, Vol. 75, S2, pp. 131-132.

Gavan, S., Harrison, M., Barton, A., Hyrich, K., Isaacs, J., Morgan, A., Wilson, A., & Payne, K. (2016b). "PMS76 – Identifying Factors which Influence the Selection of Anti-TNF Treatment in Patients with Rheumatoid Arthritis in England", *Value in Health*, Vol. 19, 3, pp. A238-A238.

Gavan, S., Harrison, M., Iglesias, C., Barton, A., Manca, A., & Payne, K. (2014). "Economics of Stratified Medicine in Rheumatoid Arthritis", *Current Rheumatology Reports*, Vol. 16, 12(468), pp. 1-11.

Genomics England. (2015). "The 100,000 Genomes Project Protocol". London: Genomics England.

Gibson, D., Rooney, M., Finnegan, S., Qiu, J., Thompson, D., LaBaer, J., Pennington, S., & Duncan, M. (2012). "Biomarkers in Rheumatology, Now and in the Future", *Rheumatology*, Vol. 2012, 51, pp. 423-433.

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). "Methods of Data Collection in Qualitative Research: Interviews and Focus Groups.", *British Dental Journal*, Vol. 204, 6, pp. 291-295.

Glaser, B., & Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. (1 ed.) Chicago: Aldine Publishing Company.

Glick, H., Doshi, J., Sonnad, S., & Polsky, D. (2015). *Economic Evaluation in Clinical Trials*. (2 ed.) Oxford: Oxford University Press.

Grabowski, H. (2010). "Chapter 2: Biologics and Data Exclusivity: Balancing Incentives for Innovation and Cost Savings". In: Mattison, N., Mestre-Ferrandiz, J., & Towse, A. (Eds.) *Biosimilars: How much will Entry and Price Competition Result?* (1 ed.) pp. 8-24. London: Office of Health Economics.

Grabowski, H., Long, G., & Mortimer, R. (2014). "Biosimilars". In: Culyer, A. (Ed.) *Encyclopedia of Health Economics*. (1 ed.) pp. 86-97. Oxford: Elsevier.

Grasas, A., Ramalhinho, H., Pessoa, L., Resende, M., Caballé, I., & Barba, N. (2014). "On the Improvement of Blood Sample Collection at Clinical Laboratories", *BMC Health Services Research*, Vol. 14, 12, pp. 1-9.

Gravelle, H., Morris, S., & Sutton, M. (2006). "Economic Studies of Equity in the Consumption of Health Care". In: Jones, A. (Ed.) *The Elgar Companion to Health Economics*. (1 ed.) pp. 193-204. Cheltenham: Edward Elgar Publishing Limited.

Gravelle, H., & Rees, R. (2004). *Microeconomics*. (3 ed.) Harlow: Pearson Education Limited.

Gravelle, H., Sutton, M., & Morris, S. (2003). "Modelling Supply and Demand Influences on the use of Health Care: Implications for Deriving a Needs-based Capitation Formula", *Health Economics*, Vol. 12, 12, pp. 985-1004.

Greene, W. (2012). *Econometric Analysis*. (7 ed.) London: Prentice Hall.

Griffin, S., Claxton, K., Hawkins, N., & Sculpher, M. (2006). "Probabilistic Analysis and Computationally Expensive Models: Necessary and Required?", *Value in Health*, Vol. 9, 4, pp. 244-252.

Griffin, S., Claxton, K., Palmer, S., & Sculpher, M. (2011). "Dangerous Omissions: The Consequences of Ignoring Decision Uncertainty", *Health Economics*, Vol. 20, 2, pp. 212-224.

Groot, B., Stijnen, T., Weinstein, M., & Hunink, M. (2011). "The Combined Analysis of Uncertainty and Patient Heterogeneity in Medical Decision Models", *Medical Decision Making*, Vol. 31, 4, pp. 650-661.

Groot, B., Weinstein, M., Stijnen, T., Heijenbrok-Kal, M., & Hunink, M. (2010). "Uncertainty and Patient Heterogeneity in Medical Decision Models", *Medical Decision Making*, Vol. 30, 2, pp. 194-205.

Grossman, M. (1972). "On the Concept of Health Capital and the Demand for Health ", *Journal of Political Economy*, Vol. 80, 2, pp. 223-255.

Gunasekara, F., Carter, K., & Blakely, T. (2008). "Glossary for Econometrics and Epidemiology", *Journal of Epidemiology & Community Health*, Vol. 62, 10, pp. 858-861.

Guyot, P., Ades, A., Ouwens, M., & Welton, N. (2012). "Enhancing Secondary Analysis of Survival Data: Reconstructing the Data from Published Kaplan-Meier Survival Curves", *BMC Medical Research Methodology*, Vol. 12, 9, pp. 1-13.

Harbord, R., Whiting, P., Sterne, J., Egger, M., Deeks, J., Shang, A., & Bachmann, L. (2008). "An Empirical Comparison of Methods for Meta-analysis of Diagnostic Accuracy Showed Hierarchical Models are Necessary", *Journal of Clinical Epidemiology*, Vol. 61, 11, pp. 1095-1103.

Hatz, M., Schremser, K., & Rogowski, W. (2014). "Is Individualized Medicine More Cost-effective? A Systematic Review", *Pharmacoeconomics*, Vol. 32, 5, pp. 443-455.

Hazlewood, G., Barnabe, C., Tomlinson, G., Marshall, D., Devoe, D., & Bombardier, C. (2016). "Methotrexate Monotherapy and Methotrexate Combination Therapy with Traditional and Biologic Disease Modifying Antirheumatic Drugs for Rheumatoid

Arthritis: Abridged Cochrane Systematic Review and Network Meta-analysis", *British Medical Journal*, Vol. 353, i1777 pp. 1-14.

Health and Social Care Information Centre. (2015). "Prescribing Costs in Hospitals and the Community: England 2014-15". London: Health and Social Care Information Centre.

Heather, E., Payne, K., Harrison, M., & Symmons, D. (2014). "Including Adverse Drug Events in Economic Evaluations of Anti-Tumour Necrosis Factor-α Drugs for Adult Rheumatoid Arthritis: A Systematic Review of Economic Decision Analytic Models", *Pharmacoeconomics*, Vol. 32, 2, pp. 109-134.

Holland, C., Paulden, M., Pechlivanoglou, P., & McCabe, C. (2016). "A Comparison of Four Software Programs for Implementing Decision Analytic Cost Effectiveness Models". *PACEOMICS Working Paper 2016_01*, Edmonton: PACEOMICS.

Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). "The Health Utilities Index (HUI): Concepts, Measurement Properties and Applications", *Health and Quality of Life Outcomes*, Vol. 1, 54, pp. 1-13.

Hudson, M., & Suissa, S. (2010). "Avoiding Common Pitfalls in the Analysis of Observational Studies of New Treatments for Rheumatoid Arthritis", *Arthritis Care & Research*, Vol. 62, 6, pp. 805-810.

Humphreys, J., Verstappen, S., Hyrich, K., Chipping, J., Marshall, T., & Symmons, D. (2013). "The Incidence of Rheumatoid Arthritis in the UK: Comparisons using the 2010 ACR/EULAR Classification Criteria and the 1987 ACR Classification Criteria. Results from the Norfolk Arthritis Register", *Annals of the Rheumatic Diseases*, Vol. 72, 8, pp. 1315-1320.

Husbands, S., Jowett, S., Barton, P., & Coast, J. (2017). "How Qualitative Methods Can be Used to Inform Model Development", *Pharmacoeconomics*, Vol. Epub ahead of print, pp. DOI: 10.1007/s40273-40017-40499-z.

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A., Mauskopf, J., & Loder, E. (2013). "Consolidated Health Economic Evaluation Reporting Standards (CHEERS) - Explanation and Elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force", *Value in Health*, Vol. 16, 2, pp. 231-250.

Husereau, D., Marshall, D., Levy, A., Peacock, S., & Hoch, J. (2014). "Health Technology Assessment and Personalized Medicine: Are Economic Evaluation Guidelines Sufficient to

Support Decision Making", *International Journal of Technology Assessment in Health Care*, Vol. 30, 2, pp. 179-187.

Hutton, J. (2012). "'Health Economics' and the Evolution of Economic Evaluation of Health Technologies", *Health Economics*, Vol. 21, 1, pp. 13-18.

Hyrich, K., Watson, K., Lunt, M., & Symmons, D. (2011). "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*, Vol. 50, 1, pp. 117-123.

Iglesias, C., Thompson, A., Rogowski, W., & Payne, K. (2016). "Reporting Guidelines for the use of Expert Judgement in Model-based Economic Evaluations", *Pharmacoeconomics*, Vol. 34, 11, pp. 1161-1172.

Innovate UK. (2016). "Precision Medicine Catapult: World-leading Innovation Centre for Precision Medicine". London: Innovate UK.

Isaacs, J., & Ferraccioli, G. (2011). "The Need for Personalised Medicine for Rheumatoid Arthritis", *Annals of the Rheumatic Diseases*, Vol. 70, 1, pp. 4-7.

Jahn, B., Pfeiffer, K., Theurl, E., Tarride, J., & Goeree, R. (2010). "Capacity Constraints and Cost-effectiveness: A Discrete Event Simulation for Drug-eluting Stents.", *Medical Decision Making*, Vol. 30, 1, pp. 16-28.

Jalal, H., Pechlivanoglou, P., Krijkamp, E., Alarid-Escudero, F., Enns, E., & Hunink, M. (2017). "An Overview of R in Health Decision Sciences", *Medical Decision Making*, Vol. [Epub ahead of print], DOI: 10.1177/0272989X16686559, pp. 1-12.

Jamnitski, A., Bartelds, G., Nurmohamed, M., Van Schouwenburg, P., Van Schaardenburg, D., Stapel, S., Dijkmans, B., Aarden, L., & Wolbink, G. (2011). "The Presence or Absence of Antibodies to Infliximab or Adalimumab Determines the Outcome of Switching to Etanercept", *Annals of the Rheumatic Diseases*, Vol. 70, 2, pp. 284-288.

Jani, M., Barton, A., Warren, R., Griffiths, C., & Chinoy, H. (2014). "The Role of DMARDs in Reducing the Immunogenicity of TNF Inhibitors in Chronic Inflammatory Diseases", *Rheumatology*, Vol. 53, 2, pp. 213-222.

Jani, M., Chinoy, H., Warren, R., Griffiths, C., Plant, D., Fu, B., Morgan, A., Wilson, A., Isaacs, J., Hyrich, K., Barton, A., & BRAGGSS. (2015a). "Clinical Utility of Random Anti-tumor Necrosis Factor Drug-level Testing and Measurement of Antidrug Antibodies

on the Long-term Treatment Response in Rheumatoid Arthritis", *Arthritis & Rheumatology*, Vol. 67, 8, pp. 2011-2019.

Jani, M., Gavan, S., Dixon, W., Harrison, B., Moran, A., Barton, A., & Payne, K. (2016a). "A Microcosting Study of Immunogenicity and TNFi Drug Level Tests for Therapeutic Monitoring in Clinical Practice", *Rheumatology*, Vol. 55, 12, pp. 2131-2137.

Jani, M., Isaacs, J., Morgan, A., Wilson, A., Plant, D., Hyrich, K., Chinoy, H., & Barton, A. (2016b). "Detection of Anti-drug Antibodies using a Bridging ELISA Compared with Radioimmunoassay in Adalimumab-treated Rheumatoid Arthritis Patients with Random Drug Levels", *Rheumatology*, Vol. 55, 11, pp. 2050-2055.

Jani, M., Isaacs, J., Morgan, A., Wilson, A., Plant, D., Hyrich, K., Chinoy, H., & Barton, A. (2017). "High Frequency of Antidrug Antibodies and Association of Random Drug Levels with Efficacy in Certolizumab Pegol-Treated Patients with Rheumatoid Arthritis: Results from the BRAGGSS Cohort", *Annals of the Rheumatic Diseases*, Vol. 76, 1, pp. 208-213.

Jani, M., Reid, V., Parker, B., Chinoy, H., & Bruce, I. (2015b). *Harmonised Biologics Pathway for Rheumatoid Arthritis*. (3 ed.) Manchester: Manchester Academic Health Science Centre.

Jansen, J., Buckley, F., Dejonckheere, F., & Ogale, S. (2014). "Comparative Efficacy of Biologics as Monotherapy and in Combination with Methotrexate on Patient Reported Outcomes (PROs) in Rheumatoid Arthritis Patients with an Inadequate Response to Conventional DMARDs - A Systematic Review and Network Meta-analysis", *Health and Quality of Life Outcomes*, Vol. 12, 102, pp. 1-15.

Jansen, J., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., Lee, K., Boersma, C., Annemans, L., & Cappelleri, J. (2011). "Interpreting Indirect Treatment Comparisons and Network Meta-analysis for Health-care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1", *Value in Health*, Vol. 14, 4, pp. 417-428.

Janssen, K., Donders, R., Harrell Jr, F., Vergouwe, Y., Chen, Q., Grobbee, D., & Moons, K. (2010). "Missing Covariate Data in Medical Research: to Impute is Better than to Ignore", *Journal of Clinical Epidemiology*, Vol. 63, 7, pp. 721-727.

Jobanputra, P., Barton, P., Bryan, S., & Burls, A. (2002). "The Effectiveness of Infliximab and Etanercept for the Treatment of Rheumatoid Arthritis: a Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 6, 21, pp. 1-110.

Joensuu, T., Huoponen, S., Aaltonen, K., Konttinen, Y., Nordström, D., & Blom, M. (2015). "The Cost-Effectiveness of Biologics for the Treatment of Rheumatoid Arthritis: A Systematic Review", *PLoS One*, Vol. 17, 10(e0119683), pp. 1-27.

Johannesson, M., Jönsson, B., & Karlsson, G. (1996). "Outcome Measurement in Economic Evaluation", *Health Economics*, Vol. 5, 4, pp. 279-296.

Jones, C., Howick, J., Roberts, N., Price, C., Heneghan, C., Plüddermann, A., & Thompson, M. (2013). "Primary Care Clinicians' Attitudes towards Point-of-care Blood Testing: A Systematic Review of Qualitative Studies", *BMC Family Practice*, Vol. 14, 117, pp. 1-9.

Kalkan, A., Roback, K., Hallert, E., & Carlsson, P. (2014). "Factors Influencing Rheumatologists' Prescription of Biologic Treatment in Rheumatoid Arthritis: An Interview Study", *Implementation Science*, Vol. 9, A153, pp. 1-12.

Kaltenthaler, E., Tappenden, P., & Paisley, S. (2013). "Reviewing the Evidence to Inform the Population of Cost-Effectiveness Models within Health Technology Assessments", *Value in Health*, Vol. 16, 5, pp. 830-836.

Kaltenthaler, E., Tappenden, P., Paisley, S., & Squires, H. (2011). "Identifying and Reviewing Evidence to Inform the Conceptualisation and Population of Cost-effectiveness Models". *NICE DSU Technical Support Document*, No. 13. Sheffield: Decision Support Unit.

Kaplan, W., Wirtz, V., Mantel-Teeuwisse, A., Stolk, P., Duthey, B., & Liaing, R. (2013). *Priority Medicines for Europe and the World 2013 Update*. (1 ed.) Geneva: World Health Organization.

Karlsson, G., & Johannesson, M. (1996). "The Decision Rules of Cost-effectiveness Analysis", *Pharmacoeconomics*, Vol. 9, 2, pp. 113-120.

Karnon, J., Stahl, J., Brennan, A., Caro, J., Mar, J., & Möller, J. (2012). "Modeling using Discrete Event Simulation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4", *Value in Health*, Vol. 15, 6, pp. 821-827.

Karsdal, M., Bay-Jensen, A., Henriksen, K., Henriksen, K., Christiansen, C., Genant, H., Chamberlin, C., & Platt, A. (2014). "Rheumatoid Arthritis: A Case for Personalized Health Care?", *Arthritis Care & Research*, Vol. 66, 9, pp. 1273-1280.

Kee, F., Sheehy, N., O'Hare, L., Bane, C., Bell, A., Dempster, M., & FitzGerald, O. (2005). "Rheumatologists' Judgements about the Efficacy of Anti-TNF Therapy in Two Neighbouring Regions", *Rheumatology*, Vol. 44, 11, pp. 1407-1413.

Kelly, C., & Mir, F. (2009). "Economics of Biological Therapies", *British Medical Journal*, Vol. 339, 7722, pp. 666-669.

Kent, S., Briggs, A., Eckermann, S., & Berry, C. (2013). "Are Value of Information Methods Ready for Prime Time? An Application to Alternative Treatment Strategies for NSTEMI Patients.", *International Journal of Technology Assessment in Health Care*, Vol. 29, 4, pp. 435-442.

Kiely, P. (2016). "Biologic Efficacy Optimization - A Step Towards Personalized Medicine", *Rheumatology*, Vol. 55, 5, pp. 780-788.

Kievit, W., van Hulst, L., van Riel, P., & Fraenkel, L. (2010). "Factors That Influence Rheumatologists' Decisions to Escalate Care in Rheumatoid Arthritis: Results from a Choice-Based Conjoint Analysis", *Arthritis Care & Research*, Vol. 62, 6, pp. 842-847.

Kim, S., Jun, J., El-Sohemy, A., & Bae, S. (2006). "Cost-effectiveness Analysis of MTHFR Polymorphism Screening by Polymerase Chain Reaction in Korean Patients with Rheumatoid Arthritis Receiving Methotrexate", *The Journal of Rheumatology*, Vol. 33, 7, pp. 1266-1274.

Kjellsson, G., Gerdtham, U., & Petrie, D. (2015). "Lies, Damned Lies, and Health Inequality Measurements: Understanding the Value Judgments", *Epidemiology*, Vol. 26, 5, pp. 673-680.

Konnopka, A., Conrad, K., Baerwald, C., & König, H. (2008). "Cost Effectiveness of the Determination of Autoantibodies against Cyclic Citrullinated Peptide in the Early Diagnosis of Rheumatoid Arthritis ", *Annals of the Rheumatic Diseases*, Vol. 67, 10, pp. 1399-1405.

Kowada. (2010). "Cost Effectiveness of Interferon-gamma Release Assay for Tuberculosis Screening of Rheumatoid Arthritis Patients Prior to Initiation of Tumor Necrosis Factor-α Antagonist Therapy", *Molecular Diagnosis & Therapy*, Vol. 14, 6, pp. 367-373.

Krieckaert, C., Nair, S., Nurmohamed, M., van Dongen, C., Lems, W., Lafeber, F., Bijlsma, J., Koffijberg, H., Wolbink, G., & Welsing, P. (2015). "Personalised Treatment using Serum Drug Levels of Adalimumab in Patients with Rheumatoid Arthritis: An

Evaluation of Costs and Effects ", *Annals of the Rheumatic Diseases*, Vol. 74, 2, pp. 361-368.

Krieckaert, C., Rispens, T., & Wolbink, G. (2012). "Immunogenicity of Biological Therapeutics: From Assay to Patient", *Current Opinion in Rheumatology*, Vol. 24, 3, pp. 306-311.

Kuijper, T., Lamers-Karnebeek, F., Jacobs, J., Hazes, J., & Luime, J. (2015). "Flare Rate in Patients with Rheumatoid Arthritis in Low Disease Activity or Remission When Tapering or Stopping Synthetic or Biologic DMARD: A Systematic Review", *The Journal of Rheumatology*, Vol. 42, 11, pp. 2012-2022.

Kvien, T. (2004). "Epidemiology and Burden of Illness of Rheumatoid Arthritis", *Pharmacoeconomics*, Vol. 22, S1, pp. 1-12.

Lafond, S., Charlesworth, A., & Roberts, A. (2016). "A Perfect Storm: An Impossible Climate for NHS Providers' Finances?". London: The Health Foundation.

Lahiri, M., & Dixon, W. (2015). "Risk of Infection with Biologic Antirheumatic Therapies in Patients with Rheumatoid Arthritis", *Best Practice & Research Clinical Rheumatology*, Vol. 29, 2, pp. 290-305.

Laudicella, M., Siciliani, L., & Cookson, R. (2012). "Waiting Times and Socioeconomic Status: Evidence from England", *Social Science & Medicine*, Vol. 74, 9, pp. 1331-1341.

Lee, J., Tollefson, E., Daly, M., & Kielb, E. (2013). "A Generalized Health Economic and Outcomes Research Model for the Evaluation of Companion Diagnostics and Targeted Therapies", *Expert Review of Pharmacoeconomics & Outcomes Research*, Vol. 13, 3, pp. 361-370.

Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ionnidis, J., Clarke, M., Devereaux, P., Kleijnen, J., & Moher, D. (2009). "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Healthcare Interventions: Explanation and Elaboration", *British Medical Journal*, Vol. 339, b2700, pp. 1-27.

Lindstrom, T., & Robinson, W. (2010). "Biomarkers for Rheumatoid Arthritis: Making it Personal", *Scandinavian Journal of Clinical and Laboratory Investigation*, Vol. 70, S242, pp. 79-84.

Lipsey, R., & Lancaster, K. (1956). "The General Theory of Second Best", *The Review of Economic Studies*, Vol. 24, 1, pp. 11-32.

Little, R., & Rubin, D. (1987). *Statistical Analysis with Missing Data*. (1 ed.) New York: John Wiley & Sons.

Long, J., & Freese, J. (2012). *Regression Models for Categorical Dependent Variables using Stata*. (3 ed.) Texas: Stata Press.

Longo, R., Baxter, P., Hall, P., Hewison, J., Afshar, M., Hall, G., & McCabe, C. (2014). "Methods for Identifying the Cost-effective Case Definition Cut-off for Sequential Monitoring Tests: An Extension of Phelps and Mushlin", *Pharmacoeconomics*, Vol. 32, 4, pp. 327-343.

Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernández Alava, M., Mukuria, C., Rowen, D., Tosh, J., Tsuchiya, A., Evans, P., Devianee Keetharuth, A., & Brazier, J. (2014). "Use of Generic and Condition-specific Measures of Health-related Quality of Life in NICE Decision-Making: A Systematic Review, Statistical Modelling and Survey", *Health Technology Assessment*, Vol. 18, 9, pp. 1-224.

Lourenco, T., Grant, A., Burr, J., & Vale, L. (2011). "The Introduction of New Interventional Procedures in the British National Health Service - A Qualitative Study", *Health Policy*, Vol. 100, 1, pp. 35-42.

Luce, B., & Claxton, K. (1999). "Redefining the Analytical Approach to Pharmacoeconomics", *Health Economics*, Vol. 8, 3, pp. 187-189.

Luce, B., & Elixhauser, A. (1990). "Estimating Costs in the Economic Evaluation of Medical Technologies", *International Journal of Technology Assessment in Health Care*, Vol. 6, 1, pp. 57-75.

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. (2010). "Chapter 10: Analysing and Presenting Results". In: Deeks, J., Bossuyt, P., & Gatsonis, C. (Eds.) *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. (1 ed.) pp. 1-61. London: The Cochrane Collaboration.

Madan, J., Ades, A., & Welton, N. (2011). "An Overview of Models used in Economic Analyses of Biologic Therapies for Arthritis - from Current Diversity to Future Consensus", *Rheumatology*, Vol. 50, S4, pp. iv10-iv18.

Madan, J., Ades, T., Barton, P., Bojke, L., Choy, E., Helliwell, P., Jobanputra, P., Stein, K., Stevens, A., Tosh, J., Verstappen, S., & Wailoo, A. (2015). "Consensus Decision

Models for Biologics in Rheumatoid and Psoriatic Arthritis: Recommendations of a Multidisciplinary Working Party", *Rheumatology and Therapy*, Vol. 2, 2, pp. 113-125.

Madsen, O. (2013). "Agreement between the DAS28-CRP Assessment with 3 and 4 Variables in Patients with Rheumatoid Arthritis Treated with Biological Agents in the Daily Clinic", *The Journal of Rheumatology*, Vol. 40, 4, pp. 379-385.

Malaviya, A., Ledingham, J., Bloxham, J., Bosworth, A., Buch, M., Choy, E., Cope, A., Isaacs, J., Marshall, D., Wright, D., Ostör, A., & BSR Clinical Affairs Committee & Standards Audit and Guidelines Working Group. (2014). "The 2013 BSR and BHPR Guideline for the use of Intravenous Tocilizumab in the Treatment of Adult Patients with Rheumatoid Arthritis", *Rheumatology*, Vol. 53, 7, pp. 1344-1346.

Malottki, K., Barton, P., Tsourapas, A., Uthman, A., Lui, Z., Routh, K., Cannock, M., Jobanputra, P., Moore, D., Fry-Smith, A., & Chen, Y. (2011). "Adalimumab, Etanercept, Infliximab, Rituximab and Abatacept for the Treatment of Rheumatoid Arthritis after the Failure of a Tumour Necrosis Factor Inhibitor: a Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 15, 14, pp. 1-278.

Mark, T., & Swait, J. (2004). "Using Stated Preference and Revealed Preference Modeling to Evaluate Prescribing Decisions", *Health Economics*, Vol. 13, 6, pp. 563-573.

Markusse, I., Dirven, L., Gerards, A., van Groenendal, J., Ronday, H., Kerstens, P., Lems, W., Huizinga, T., & Allaart, C. (2015). "Disease Flares in Rheumatoid Arthritis are Associated with Joint Damage Progression and Disability: 10-year Results from the BeSt Study", *Arthritis Research & Therapy*, Vol. 17, 232, pp. 1-9.

Marra, C., Esdaile, J., & Anis, A. (2002). "Practical Pharmacogenetics: the Cost Effectiveness of Screening for Thiopurine S-methyltransferase Polymorphisms in Patients with Rheumatological Conditions Treated with Azathioprine", *The Journal of Rheumatology*, Vol. 29, 12, pp. 2507-2512.

Mauskopf, J., Paul, J., Grant, D., & Stergachis, A. (1998). "The Role of Cost-consequences Analysis in Healthcare Decision-making", *Pharmacoeconomics*, Vol. 13, 3, pp. 277-288.

Maxwell, J., Potter, C., Hyrich, K., BRAGGSS., Barton, A., Worthington, J., Isaacs, J., Moran, A., & Wilson, A. (2008). "Association of the Tumour Necrosis Factor-308 Variant with Differential Response to Anti-TNF Agents in the Treatment of Rheumatoid Arthritis", *Human Molecular Genetics*, Vol. 17, 22, pp. 3532-3538.

Maynard, A. (1979). "Pricing, Demanders, and the Supply of Health Care", *International Journal of Health Services*, Vol. 9, 1, pp. 121-133.

McCabe, C., Claxton, K., & Culyer, A. (2008). "The NICE Cost-effectiveness Threshold: What it is and what that Means", *Pharmacoeconomics*, Vol. 26, 9, pp. 733-744.

McCarthy, M. (2015). "Obama Seeks $213m to Fund "Precision Medicine"", *British Medical Journal*, Vol. 350, h587, pp. 1-1.

McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior". In: Zarembka, P. (Ed.) *Frontiers of Econometrics*. (1 ed.) pp. 105-142. New York: Academic Press.

McInnes, I., & Schett, G. (2011). "The Pathogenesis of Rheumatoid Arthritis", *The New England Journal of Medicine*, Vol. 365, 23, pp. 2205-2219.

McNamme, C. (2016). "UK Pharmacogenetics and Stratified Medicine Network", *Personalized Medicine*, Vol. 13, 2, pp. 107-112.

McPherson, K. (1989). "International Differences in Medical Care Practices", *Health Care Financing Review*, Vol. 1989, S1, pp. 9-20.

Meadows, N., Morrison, A., Brindley, D., Schuh, A., & Barker, R. (2015). "An Evaluation of Regulatory and Commercial Barriers to Stratified Medicine Development and Adoption", *The Pharmacogenomics Journal*, Vol. 15, 1, pp. 6-12.

Medical Research Council. (2016). "Guidance for Outline Stage: Stratified Medicine Consortium Applicants". London: Medical Research Council.

Merlin, T., Lehman, S., Hiller, J., & Ryan, P. (2013). "The "Linked Evidence Approach" to Assess Medical Tests: A Critical Analysis", *International Journal of Technology Assessment in Health Care*, Vol. 29, 3, pp. 343-350.

Meune, C., Touzé, E., Trinquart, L., & Allanore, Y. (2009). "Trends in Cardiovascular Mortality in Patients with Rheumatoid Arthritis over 50 Years: A Systematic Review and Meta-analysis of Cohort Studies", *Rheumatology*, Vol. 48, 10, pp. 1309-1313.

Michaud, K., Vera-Llonch, M., & Oster, G. (2012). "Mortality Risk by Functional Status and Health-Related Quality of Life in Patients with Rheumatoid Arthritis", *The Journal of Rheumatology*, Vol. 39, 1, pp. 54-59.

Michaud, K., Wallenstein, G., & Wolfe, F. (2011). "Treatment and Non-treatment Predictors of HAQ Disability Progression in RA: a Longitudinal Study of 18,485 Patients", *Arthritis Care & Research*, Vol. 63, 3, pp. 366-372.

Miguel, L., & Hulstaert, F. (2015). "The Importance of Test Accuracy in Economic Evaluations of Companion Diagnostics", *Journal of Comparative Effectiveness Research*, Vol. 4, 6, pp. 569-577.

Miller, C. (1995). "In-depth Interviewing by Telephone: Some Practical Considerations", *Evaluation & Research in Education*, Vol. 9, 1, pp. 29-38.

Miller, I., Ashton-Chess, J., Spolders, H., Fert, V., Ferrara, J., Kroll, W., Askaa, J., Larcier, P., Terry, P., Bruinvels, A., & Huriez, A. (2011). "Market Access Challenges in the EU for High Medical Value Diagnostic Tests", *Personalized Medicine*, Vol. 8, 2, pp. 137-148.

Milne, C., Bryan, C., Garafalo, S., & McKiernan, M. (2015). "Complementary Versus Companion Diagnostics: Apples and Oranges?", *Biomarkers in Medicine*, Vol. 9, 1, pp. 25-34.

Mire-Sluis, A., Barrett, Y., Devanarayan, V., Koren, E., Liu, H., Maia, M., Parish, T., Scott, G., Shankar, G., Shores, E., Swanson, S., Taniguchi, G., Wierda, D., & Zuckerman, L. (2004). "Recommendations for the Design and Optimization of Immunoassays used in the Detection of Host Antibodies against Biotechnology Products", *Journal of Immunological Methods*, Vol. 289, 1-2, pp. 1-16.

Mócsai, A., Kovács, L., & Gergely, P. (2014). "What is the Future of Targeted Therapy in Rheumatology: Biologics or Small Molecules?", *BMC Medicine*, Vol. 12, 43, pp. 1-9.

Mohiuddin, S., Payne, K., & Fenwick, E. (2014). "Use of Value of Information in UK Health Technology Assessments", *International Journal of Technology Assessment in Health Care*, Vol. 30, 6, pp. 553-570.

Mok, C., Tsai, W., Chen, D., & Wei, J. (2016). "Immunogenicity of Anti-TNF Biologic Agents in the Treatment of Rheumatoid Arthritis", *Expert Opinion on Biologic Therapy*, Vol. 16, 2, pp. 201-211.

Monaco, C., Nanchahal, J., Taylor, P., & Feldmann, M. (2015). "Anti-TNF Therapy: Past, Present and Future", *International Immunology*, Vol. 27, 1, pp. 55-62.

Mooney, G., & Ryan, M. (1993). "Agency in Health Care: Getting Beyond First Principles", *Journal of Health Economics*, Vol. 12, 2, pp. 125-135.

Morgan, D. (2008). "Sampling Frame". In: Given, L. (Ed.) *The SAGE Encyclopedia of Qualitative Research Methods*. (1 ed.) pp. 800-801. London: SAGE Publications Ltd.

Morris, S., Devlin, N., Parkin, D., & Spencer, A. (2012). *Economic Analysis in Health Care*. (2 ed.) Chichester: John Wiley & Sons, Ltd.

Morris, S., Sutton, M., & Gravelle, H. (2005). "Inequity and Inequality in the use of Health Care in England: an Empirical Investigation", *Social Science & Medicine*, Vol. 60, 6, pp. 1251-1266.

Mulleman, D., Ducourau, E., Paintaud, G., Ternant, D., Watier, H., & Goupille, P. (2012). "Should Anti-TNF-α Drug Levels and/or Anti-Drug Antibodies be Assayed in Patients Treated for Rheumatoid Arthritis?", *Joint Bone Spine*, Vol. 79, 2, pp. 109-112.

Mulleman, D., Méric, J., Paintaud, G., Ducourau, E., Magdelaine-Beuzelin, C., Valat, J., Goupille, P., & Centre National de la Recherche Scientifique UMR 6239 GICC. (2009). "Infliximab Concentration Monitoring Improves the Control of Disease Activity in Rheumatoid Arthritis", *Arthritis Research & Therapy*, Vol. 11, 6, p. R178.

Murphy, D., Klein, R., Smolen, L., Klein, T., & Roberts, S. (2013). "Using Common Random Numbers in Health Care Cost-Effectiveness Simulation Modeling", *Health Services Research*, Vol. 48, 4, pp. 1508-1525.

Mushlin, A. (1999). "Challenges and Opportunities in Economic Evaluations of Diagnostic Tests and Procedures", *Academic Radiology*, Vol. 6, S1, pp. S128-S131.

Nair, S., Welsing, P., Jacobs, J., van Larr, J., Rensen, W., Ardine de Wit, G., Bijlsma, J., & Lafeber, F. (2015). "Economic Evaluation of a Tight-control Treatment Strategy using an Imaging Device (Handscan) for Monitoring Joint Inflammation in Early Rheumatoid Arthritis", *Clinical and Experimental Rheumatology*, Vol. 33, 6, pp. 831-838.

Nam, J., Ramiro, S., Gaujoux-Viala, C., Takase, K., Leon-Garcia, M., Emery, P., Gossec, L., Landewe, R., Smolen, J., & Buch, M. (2014). "Efficacy of Biological Disease-modifying Antirheumatic Drugs: A Systematic Literature Review Informing the 2013 Update of the EULAR Recommendations for the Management of Rheumatoid Arthritis", *Annals of the Rheumatic Diseases*, Vol. 73, 3, pp. 516-528.

National Audit Office. (2009). *Services for People with Rheumatoid Arthritis*. (1 ed.) London: National Audit Office, .

National Institute for Health and Care Excellence. (2009). "Rheumatoid Arthritis in Adults: Management". *NICE Clinical Guideline*, CG79. London: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence. (2010). "Adalimumab, Etanercept, Infliximab, Rituximab and Abatacept for the Treatment of Rheumatoid Arthritis after the Failure of a TNF Inhibitor". *NICE Technology Appraisal Guidance*, TA195. London: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence. (2011a). "Diagnostics Assessment Programme Manual". Manchester: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence. (2011b). "Medical Technologies Evaluation Programme: Methods Guide". London: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence. (2013a). "Guide to the Methods of Technology Appraisal". *NICE Process and Methods*, PMG9. London: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence. (2013b). "Rheumatoid Arthritis: Commissioning Guide". Manchester: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence. (2014). "Diagnostics Assessment Programme - Crohn's Disease: Tests for Therapeutic Monitoring of TNF Inhibitors (LISA-TRACKER ELISA Kits, TNFα-Blocker ELISA Kits, and Promonitor ELISA Kits)". *Final Scope*, Manchester: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence. (2016a). "Adalimumab, Etanercept, Infliximab, Certolizumab Pegol, Golimumab, Tocilizumab, and Abatacept for Rheumatoid Arthritis not Previously Treated with DMARDs or after Conventional DMARDs only have Failed". *NICE Technology Appraisal Guidance*, TA375. London: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence. (2016b). "National Institute for Health and Care Excellence: NICE Charter". [Online] Available at: https://www.nice.org.uk/Media/Default/About/Who-we-are/NICE_Charter.pdf [Accessed 27 March 2017].

National Institute for Health and Care Excellence. (2016c). "Therapeutic Monitoring of TNF-Alpha Inhibitors in Crohn's Disease (LISA-TRACKER ELISA Kits, IDKmonitor ELISA Kits, and Promonitor ELISA Kits)". *NICE Diagnostics Guidance*, DG22. London: National Institute for Health and Care Excellence.

National Institute for Health Research. (2012). *Programme Grants for Applied Research: Six Competitions, 100 Awards*. (1 ed.) London: National Institute for Health Research.

National Institute for Health Research. (2014). "NIHR Stratified Medicine Capabilities: Supporting Innovation and Expertise in Stratified Medicine". London: National Institute for Health Research.

Neyt, M., & van Brabandt, H. (2011). "The Importance of the Comparator in Economic Evaluations: Working on the Efficiency Frontier", *Pharmacoeconomics*, Vol. 29, 11, pp. 913-916.

Norum, K. (2008). "Reality and Multiple Realities". In: Given, L. (Ed.) *The SAGE Encyclopedia of Qualitative Research Methods*. (1 ed.) pp. 736-739. London: SAGE Publications Ltd.

O'Brien, B., Harris, I., Beckman, T., Reed, D., & Cook, D. (2014). "Standards for Reporting Qualitative Research: A Synthesis of Recommendations", *Academic Medicine*, Vol. 89, 9, pp. 1245-1251.

O'Hagan, T., & Luce, B. (2003). *A Primer on Bayesian Statistics in Health Economics and Outcomes Research*. (1 ed.) Bethesda: Medtap International Inc.

O'Hagan, T., McCabe, C., Akehurst, R., Brennan, A., Briggs, A., Claxton, K., Fenwick, E., Fryback, D., Sculpher, M., Spiegelhalter, D., & Willian, A. (2005). "Incorporation of Uncertainty in Health Economic Modelling Studies", *Pharmacoeconomics*, Vol. 23, 6, pp. 529-536.

O'Mahony, J., Naber, S., Normand, C., Sharp, J., O'Leary, J., & de Kok, I. (2015a). "Beware of Kinked Frontiers: A Systematic Review of the Choice of Comparator Strategies in Cost-Effectiveness Analyses of Human Papillomavirus Testing in Cervical Screening", *Value in Health*, Vol. 18, 8, pp. 1138-1151.

O'Mahony, J., Newall, A., & van Rosmalen, J. (2015b). "Dealing with Time in Health Economic Evaluation: Methodological Issues and Recommendations for Practice", *Pharmacoeconomics*, Vol. 33, 12, pp. 1255-1268.

O'Mahony, J., & Normand, C. (2015c). "HIQA's CEA of Breast Screening: Pragmatic Policy Recommendations are Welcome, but ACERs Reported as ICERs are Not", *Value in Health*, Vol. 18, 8, pp. 941-945.

O'Mahony, J., & Paulden, M. (2014). "NICE's Selective Application of Differential Discounting: Ambiguous, Inconsistent, and Unjustified", *Value in Health*, Vol. 17, 5, pp. 493-496.

Obermann, K., Scheppe, J., & Glazinski, B. (2013). "More than Figures? Qualitative Research in Health Economics", *Health Economics*, Vol. 22, 3, pp. 253-257.

Office for National Statistics. (2015). "National Life Tables: England (2012-2014)". Newport: Office for National Statistics.

Oh, K., Anis, A., & Bae, S. (2004). "Pharmacoeconomic Analysis of Thiopurine Methyltransferase Polymorphism Screening by Polymerase Chain Reaction for Treatment with Azathioprine in Korea", *Rheumatology*, Vol. 43, 2, pp. 156-163.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., K., I., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., Graham, R., Manoharan, A., Ortmann, W., Bhangale, T., Denny, J., Carroll, R., Eyler, A., Greenberg, J., Kremer, J., Pappas, D., Jiang, L., Yin, J., Ye, L., Su, D., Yang, J., Xie, G., Keystone, E., Westra, H., Esko, T., Metspalu, A., Zhou, X., Gupta, N., Mirel, D., Stahl, E., Diogo, D., Cui, J., Liao, K., Guo, M., Myouzen, K., Kawaguchi, T., Coenen, M., van Riel, P., van de Laar, M., Guchelaar, H., Huizinga, T., Dieudé, P., Mariette, X., Bridges, S., Zhernakova, A., Toes, R., Tak, P., Miceli-Richard, C., Bang, S., Lee, H., Martin, J., Gonzalez-Gay, M., Rodriguez-Rodriguez, L., Rantapää-Dahlqvist, S., Arlestig, L., Choi, H., Kamatani, Y., Galan, P., Lathrop, M., RACI consortium, GARNET consortium, Eyre, S., Bowes, J., Barton, A., de Vries, N., Moreland, L., Criswell, L., Karlson, E., Taniguchi, A., Yamada, R., Kubo, M., Liu, J., Bae, S., Worthington, J., Padyukov, L., Klareskog, L., Gregersen, P., Raychaudhuri, S., Stranger, B., De Jager, P., Franke, L., Visscher, P., Brown, M., Yamanaka, H., Mimori, T., Takahashi, A., Xu, H., Behrens, T., Siminovitch, K., Momohara, S., Matsuda, F., Yamamoto, K., & Plenge, R. (2014). "Genetics of Rheumatoid Arthritis Contributes to Biology and Drug Discovery", *Nature*, Vol. 506, 7488, pp. 376-381.

Oosterhoff, M., van der Maas, M., & Steuten, L. (2016). "A Systematic Review of Health Economic Evaluations of Diagnostic Biomarkers", *Applied Health Economics and Health Policy*, Vol. 14, 1, pp. 51-65.

Orme, M., Macgilchrist, K., Mitchell, S., Spurden, D., & Bird, A. (2012). "Systematic Review and Network Meta-analysis of Combination and Monotherapy Treatments in

Disease-modifying Antirheumatic Drug-experienced Patients with Rheumatoid Arthritis: Analysis of American College of Rheumatology Criteria Scores 20, 50, and 70", *Biologics: Targets and Therapy*, Vol. 2012, 6, pp. 429-464.

Owens, D., Siegel, J., Sculpher, M., & Salomon, J. (2017). "Chapter Four: Designing a Cost-effectiveness Analysis". In: Neumann, P., Sanders, G., Russell, L., Siegel, J., & Ganiats, T. (Eds.) *Cost-effectiveness in Health and Medicine*. (2 ed.) pp. 75-104. New York: Oxford University Press.

Palmer, D., & El Miedany, Y. (2010). "Biological Nurse Specialist: Goodwill to Good Practice", *British Journal of Nursing*, Vol. 19, 8, pp. 477-480.

Palmer, S., & Raftery, J. (1999a). "Economics Notes: Opportunity Cost", *British Medical Journal*, Vol. 318, 7197, pp. 1551-1552.

Palmer, S., & Torgerson, D. (1999b). "Economics Notes: Definitions of Efficiency", *British Medical Journal*, Vol. 318, 7191, pp. 1136-1136.

Patton, M. (2003). *Qualitative Research & Evaluation Methods*. (3 ed.) London: Sage Publications Ltd.

Paulden, M., & Claxton, K. (2012). "Budget Allocation and the Revealed Social Rate of Time Preference for Health", *Health Economics*, Vol. 21, 5, pp. 612-618.

Paulden, M., McCabe, C., & Karnon, J. (2014a). "Achieving Allocative Efficiency in Healthcare: Nice in Theory, not so NICE in Practice?", *Pharmacoeconomics*, Vol. 32, 4, pp. 315-318.

Paulden, M., O'Mahony, J., Culyer, A., & McCabe, C. (2014b). "Some Inconsistencies in NICE's Consideration of Social Values", *Pharmacoeconomics*, Vol. 32, 11, pp. 1043-1053.

Payne, K. (2008). "Towards an Economic Evidence Base for Pharmacogenetics: Consideration of Outcomes is Key", *Pharmacogenomics*, Vol. 9, 1, pp. 1-4.

Payne, K. (2009). "Fish and Chips all Round? Regulation of DNA-Based Genetic Diagnostics", *Health Economics*, Vol. 18, 11, pp. 1233-1236.

Payne, K., & Annemans, L. (2013a). "Reflections on Market Access for Personalized Medicine: Recommendations for Europe", *Value in Health*, Vol. 16, 6, pp. S32-S38.

Payne, K., McAllister, M., & Davies, L. (2013b). "Valuing the Economic Benefits of Complex Interventions: When Maximising Health is not Sufficient", *Health Economics*, Vol. 22, 3, pp. 258-271.

Payne, K., & Shabaruddin, F. (2010). "Cost-effectiveness Analysis in Pharmacogenomics", *Pharmacogenomics*, Vol. 11, 5, pp. 643-646.

Payne, K., & Thompson, A. (2013c). "Economics of Pharmacogenomics: Rethinking Beyond QALYs?", *Current Pharmacogenomics and Personalized Medicine*, Vol. 11, 3, pp. 187-195.

Pearson, E. (2016). "Personalized Medicine in Diabetes: The Role of 'Omics' and Biomarkers", *Diabetic Medicine*, Vol. 33, 6,

Pennington, B., & Davis, S. (2014). "Mapping from the Health Assessment Questionnaire to the EQ-5D: The Impact of Different Algorithms on Cost-effectiveness Results", *Value in Health*, Vol. 17, 8, pp. 762-771.

Phelps, C., & Mushlin, A. (1988). "Focusing Technology Assessment using Medical Decision Theory", *Medical Decision Making*, Vol. 8, 4, pp. 279-289.

Phillips, K., Sakowski, J., Liang, S., & Ponce, N. (2013). "Economic Perspectives on Personalized Health Care and Prevention", *Forum for Health Economics and Policy*, Vol. 16, 2, pp. S23-S52.

Phillips, K., & van Bebber, S. (2004). "A Systematic Review of Cost-effectiveness Analyses of Pharmacogenomic Interventions", *Pharmacogenomics*, Vol. 5, 8, pp. 1139-1149.

Pincus, T., Gibson, K., & Castrejón, I. (2013). "Update on Methotrexate as the Anchor Drug for Rheumatoid Arthritis", *Bulletin of the Hospital for Joint Diseases*, Vol. 71, S1, pp. S9-S19.

Pineda, C., Hernández, G., Jacobs, I., Alvarez, D., & Carini, C. (2016). "Assessing the Immunogenicity of Biopharmaceuticals", *BioDrugs*, Vol. 30, 3, pp. 195-206.

Plumpton, C., Roberts, D., Pirmohamed, M., & Hughes, D. (2016). "A Systematic Review of Economic Evaluations of Pharmacogenetic Testing for Prevention of Adverse Drug Reactions", *Pharmacoeconomics*, Vol. 34, 8, pp. 771-793.

Pokorska-Bocci, A., Stewart, A., Sagoo, G., Hall, A., Kroese, M., & Burton, H. (2014). "'Personalized Medicine': What's in a Name?", *Personalized Medicine*, Vol. 11, 2, pp. 197-210.

Poland, B. (1995). "Transcription Quality as an Aspect of Rigor in Qualitative Research", *Qualitative Inquiry*, Vol. 1, 3, pp. 290-310.

Poland, B. (2008). "Transcription". In: Given, L. (Ed.) *The SAGE Encyclopedia of Qualitative Research Methods*. (1 ed.) pp. 884-886. London: SAGE Publications Ltd.

Pouw, M., Krieckaert, C., Nurmohamed, M., van der Kleij, D., Aarden, L., Rispens, T., & Wolbink, G. (2015). "Key Findings Towards Optimising Adalimumab Treatment: the Concentration-effect Curve", *Annals of the Rheumatic Diseases*, Vol. 74, 3, pp. 513-518.

Prevoo, M., van't Hof, M., Kuper, H., van Leeuwen, M., van de Putte, L., & van Riel, P. (1995). "Modified Disease Activity Scores that Include Twenty-eight-joint Counts. Development and Validation in a Prospective Longitudinal Study of Patients with Rheumatoid Arthritis", *Arthritis & Rheumatology*, Vol. 38, 1, pp. 44-48.

Propper, C. (1995). "Agency and Incentives in the NHS Internal Market", *Social Science & Medicine*, Vol. 40, 12, pp. 1683-1690.

Propper, C., Eachus, J., Chan, P., Pearson, N., & Smith, G. (2005). "Access to Health Care Resources in the UK: the Case of Care for Arthritis", *Health Economics*, Vol. 14, 4, pp. 391-406.

R Core Team. (2015). "R: A Language and Environment for Statistical Computing". Vienna, Austria: R Foundation for Statistical Computing.

Rabe-Hesketh, S., & Skrondal, A. (2005). *Multilevel and Longitudinal Modeling using Stata*. (2 ed.) College Station: Stata Press.

Radner, H., Smolen, J., & Aletaha, D. (2010). "Impact of Comorbidity on Physical Function in Patients with Rheumatoid Arthritis", *Annals of the Rheumatic Diseases*, Vol. 69, 3, pp. 536-541.

Radstake, T., Svenson, M., Eijsbouts, A., van den Hoogen, F., Enevold, C., van Riel, P., & Bendtzen, K. (2009). "Formation of Antibodies Against Infliximab and Adalimumab Strongly Correlates with Functional Drug Levels and Clinical Responses in Rheumatoid Arthritis", *Annals of the Rheumatic Diseases*, Vol. 68, 11, pp. 1739-1745.

Raganath, V., Maranian, P., Elashoff, D., Woodworth, T., Khanna, D., Hahn, T., Sarkisian, C., Kremer, J., Furst, D., & Paulus, H. (2013). "Comorbidities are Associated with Poorer Outcomes in Community Patients with Rheumatoid Arthritis", *Rheumatology*, Vol. 52, 10, pp. 1809-1817.

Raghavan, S., & Vassy, J. (2014). "Do Physicians think Genomic Medicine will be useful for Patient Care?", *Personalized Medicine*, Vol. 11, 4, pp. 424-433.

Ramaekers, B., Joore, M., & Grutters, J. (2013). "How Should we Deal with Patient Heterogeneity in Economic Evaluation: A Systematic Review of National Pharmacoeconomic Guidelines", *Value in Health*, Vol. 16, 5, pp. 855-862.

Rawlins, M. (2005). "Pharmacopolitics and Deliberative Democracy", *Clinical Medicine*, Vol. 5, 5, pp. 471-475.

Rawlins, M., & Culyer, A. (2004). "National Institute for Clinical Excellence and its Value Judgments", *British Medical Journal*, Vol. 329, 7459, pp. 224-227.

Redekop, W., & Mladsi, D. (2013). "The Faces of Personalized Medicine: A Framework for Understanding its Meaning and Scope", *Value in Health*, Vol. 16, S6, pp. S4-S9.

Reitsma, J., Glas, A., Rutjes, A., Scholten, R., Bossuyt, P., & Zwinderman, A. (2005). "Bivariate Analysis of Sensitivity and Specificity Produces Informative Summary Measures in Diagnostic Reviews", *Journal of Clinical Epidemiology*, Vol. 58, 10, pp. 982-990.

Rezvan, P., Lee, K., & Simpson, J. (2015). "The Rise of Multiple Imputation: A Review of the Reporting and Implementation of the Method in Medical Research", *BMC Medical Research Methodology*, Vol. 15, 30, pp. 1-14.

Rice, N., & Jones, A. (1997). "Multilevel Models and Health Economics", *Health Economics*, Vol. 6, 6, pp. 561-575.

Ritchie, J., & Spencer, L. (2001). "Qualitative Data Analysis for Applied Policy Research". In: Bryman, A., & Burgess, R. (Eds.) *Analyzing Qualitative Data*. (6 ed.) pp. 173-194. London: Routledge.

Roberts, M., Russell, L., Paltiel, D., Chambers, M., McEwan, P., & Krahn, M. (2012). "Conceptualizing a Model: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2", *Value in Health*, Vol. 15, 6, pp. 804-811.

Robinson, S. (2008). "Conceptual Modelling for Simulation Part I: Definition and Requirements", *The Journal of the Operational Research Society*, Vol. 59, 3, pp. 278-290.

Roche. (2006). "MabThera (rituximab) NICE STA Submission". [Online] Available at: https://www.nice.org.uk/guidance/ta126/resources/roche-products-ltd2 [Accessed 14 July 2016].

Rogowski, W., Grosse, S., & Khoury, M. (2009). "Challenges of Translating Genetic Tests into Clinical and Public Health Practice", *Nature Reviews Genetics*, Vol. 10, 7, pp. 489-495.

Rogowski, W., Hartz, S., & John, J. (2008). "Clearing up the Hazy Road from Bench to Bedside: A Framework for Integrating the Fourth Hurdle into Translational Medicine", *BMC Health Services Research*, Vol. 8, 194, pp. 1-12.

Rogowski, W., Payne, K., Schnell-Inderst, P., Manca, A., Rochau, U., Jahn, B., Alagoz, O., Leidl, R., & Siebert, U. (2015). "Concepts of 'Personalization' in Personalized Medicine: Implications for Economic Evaluation", *Pharmacoeconomics*, Vol. 33, 1, pp. 49-59.

Roland, M., & Torgerson, D. (1998). "Understanding Controlled Trials: What are Pragmatic Trials?", *British Medical Journal*, Vol. 316, 2127:285, pp. 1-1.

Romaniuk, H., Patton, G., & Carlin, J. (2014). "Multiple Imputation in a Longitudinal Cohort Study: a Case Study of Sensitivity to Imputation Methods", *American Journal of Epidemiology*, Vol. 180, 9, pp. 920-932.

Rosas, J., Llinares-Tello, F., de la Torre, I., Santos-Ramírez, C., Senabre-Gallego, J., Valor, L., Barber-Vallés, X., Hernández-Flórez, D., Santos-Soler, G., Salas-Heredia, E., Carreño, L., & AIRE-MB Group. (2014). "Clinical Relevance of Monitoring Serum Levels of Adalimumab in Patients with Rheumatoid Arthritis in Daily Practice", *Clinical and Experimental Rheumatology*, Vol. 32, 6, pp. 942-948.

Royston, P. (2009). "Multiple Imputation of Missing Values: Further Update of ice, with an Emphasis on Categorical Variables", *The Stata Journal*, Vol. 9, 3, pp. 466-477.

Royston, P., & White, I. (2011). "Multiple Imputation by Chained Equations (MICE): Implementation in STATA", *Journal of Statistical Software*, Vol. 45, 4, pp. 1-20.

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. (1 ed.) New York: John Wiley & Sons.

Russell, A. (2008). "Quality-of-Life Assessment in Rheumatoid Arthritis", *Pharmacoeconomics*, Vol. 26, 10, pp. 831-846.

Ryan, M. (1994). "Agency in Health Care: Lessons for Economists from Sociologists", *The American Journal of Economics and Sociology*, Vol. 53, 2, pp. 207-217.

Sandelowski, M. (1995). "Sample Size in Qualitative Research", *Research in Nursing & Health*, Vol. 18, 2, pp. 179-183.

Sanders, G., Neuman, P., Basu, A., Brock, D., Feeny, D., Krahn, M., Kuntz, K., Meltzer, D., Owens, D., Prosser, L., Salomon, J., Sculpher, M., Trikalinos, T., Russell, L., Siegel, J., & Ganiats, T. (2016). "Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-effectiveness in Health and Medicine", *Journal of the American Medical Association*, Vol. 316, 10, pp. 1093-1103.

Schaeverbeke, T., Truchetet, M., Kostine, M., Barnetche, T., Bannwarth, B., & Richez, C. (2016). "Immunogenicity of Biologic Agents in Rheumatoid Arthritis Patients: Lessons for Clinical Practice", *Rheumatology*, Vol. 55, 2, pp. 210-220.

Schleidgen, S., Klinger, C., Bertram, T., Rogowski, W., & Marckmann, G. (2013). "What is Personalized Medicine: Sharpening a Vague Term Based on a Systematic Literature Review", *BMC Medical Ethics*, Vol. 15, 55, pp. 1-12.

Schoenbaum, S., & Gottlieb, L. (1990). "Algorithm Based Improvement of Clinical Quality", *British Medical Journal*, Vol. 301, 6765, pp. 1374-1376.

Scholz, S., & Mittendorf, T. (2014). "Modeling Rheumatoid Arthritis using Different Techniques - a Review of Model Construction and Results", *Health Economics Review*, Vol. 4, 18, pp. 1-16.

Schwarz, G. (1978). "Estimating the Dimension of a Model", *The Annals of Statistics*, Vol. 6, 2, pp. 461-464.

Scotland, G., & Bryan, S. (2017). "Why Do Health Economists Promote Technology Adoption Rather Than the Search for Efficiency? A Proposal for a Change in Our Approach to Economic Evaluation in Health Care", *Medical Decision Making*, Vol. 37, 2, pp. 139-147.

Scott, D., Wolfe, F., & Huizinga, T. (2010). "Rheumatoid Arthritis", *The Lancet*, Vol. 376, 9746, pp. 1094-1108.

Sculpher, M. (1997). "The Iterative use of Economic Evaluation as Part of the Process of Health Technology Assessment", *Journal of Health Services Research & Policy*, Vol. 2, 1, pp. 26-30.

Sculpher, M. (2008). "Subgroups and Heterogeneity in Cost-effectiveness Analysis", *Pharmacoeconomics*, Vol. 26, 9, pp. 799-806.

Sculpher, M., & Claxton, K. (2005a). "Establishing the Cost-effectiveness of New Pharmaceuticals under Conditions of Uncertainty - When is there Sufficient Evidence?", *Value in Health*, Vol. 8, 4, pp. 433-446.

Sculpher, M., Claxton, K., & Akehurst, R. (2005b). "It's Just Evaluation for Decision-making: Recent Developments in, and Challenges for, Cost-effectiveness Research". In: Smith, P., Ginnelly, L., & Sculpher, M. (Eds.) *Health Policy and Economics: Opportunities and Challenges*. ed.) pp. 8-41. Maidenhead: Open University Press.

Sculpher, M., Claxton, K., Drummond, M., & McCabe, C. (2006a). "Whither Trial-based Economic Evaluation for Health Care Decision Making?", *Health Economics*, Vol. 15, 7, pp. 677-687.

Sculpher, M., Claxton, K., & Pearson, S. (2017). "Developing a Value Framework: The Need to Reflect the Opportunity Costs of Funding Decisions", *Value in Health*, Vol. 20, 2, pp. 234-239.

Sculpher, M., & Drummond, M. (2006b). "Analysis Sans Frontières: Can we ever Make Economic Evaluations Generalisable Across Jurisdictions?", *Pharmacoeconomics*, Vol. 24, 11, pp. 1087-1099.

Shabaruddin, F., Fleeman, N., & Payne, K. (2015). "Economic Evaluations of Personalized Medicine: Existing Challenges and Current Developments", *Pharmacogenomics and Personalized Medicine*, Vol. 2015, 8, pp. 115-128.

Shah, K., Cookson, R., Culyer, A., & Littlejohns, P. (2013). "NICE's Social Value Judgements about Equity in Health and Health Care", *Health Economics, Policy and Law*, Vol. 8, 2, pp. 145-165.

Sheldon, T., Cullum, N., Dawson, D., Lankshear, A., Lowson, K., Watt, I., West, P., Wright, D., & Wright, J. (2004). "What's the Evidence that NICE Guidance has been Implemented? Results from a National Evaluation using Time Series Analysis, Audit of Patients' Notes, and Interviews", *British Medical Journal*, Vol. 329, 7473, pp. 1-8.

Shemilt, I., Mugford, M., Vale, L., Marsh, K., Donaldson, C., & Drummond, M. (2010). "Evidence Synthesis, Economics and Public Policy", *Research Synthesis Methods*, Vol. 1, 2, pp. 126-135.

Shiell, A., Donaldson, C., Mitton, C., & Currie, G. (2002). "Health Economic Evaluation", *Journal of Epidemiology & Community Health*, Vol. 56, 2, pp. 85-88.

Shih, Y. (2003). "Bayesian Approach in Pharmacoeconomics: Relevance to Decision-makers", *Expert Review of Pharmacoeconomics & Outcomes Research*, Vol. 3, 3, pp. 237-250.

Shortell, S., Waters, T., Clarke, K., & Budetti, P. (1998). "Physicians as Double Agents: Maintaining Trust in an Era of Multiple Accountabilities", *Journal of the American Medical Association*, Vol. 280, 12, pp. 1102-1108.

Siebert, S., Lyall, D., Mackay, D., Porter, D., McInnes, I., Sattar, N., & Pell, J. (2016). "Characteristics of Rheumatoid Arthritis and its Association with Major Comorbid Conditions: Cross-sectional Study of 502 649 UK Biobank Participants ", *RMD Open*, Vol. 2, e000267, pp. 1-10.

Silberberg, E. (1974). "A Revision of Comparative Statics Methodology in Economics, or, how to do Comparative Statics on the Back of an Envelope", *Journal of Economic Theory*, Vol. 7, 2, pp. 159-172.

Silverman, D., & Marvasti, A. (2008). *Doing Qualitative Research: A Comprehensive Guide*. (1 ed.) London: Sage Publications Ltd.

Singh, J., Hossain, A., Tanjong Ghogomu, E., Kotb, A., Christensen, R., Mudano, A., Maxwell, L., Shah, N., Tugwell, P., & Wells, G. (2016a). "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 of Systematic Reviews*, Vol. 5, CD012183, pp. 1-75.

Singh, J., Saag, K., Bridges, S., Akl, E., Bannuru, R., Sullivan, M., Vaysbrot, E., McNaughton, C., Osani, M., Shmerling, R., Curtis, J., Furst, D., Parks, D., Kavanaugh, A., O'Dell, J., King, C., Leong, A., Matteson, E., Schousboe, J., Drevlow, B., Ginsberg, S.,

Grober, J., St Clair, E., Tindall, E., Miller, A., & McAlindon, T. (2016b). "2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis", *Arthritis & Rheumatology*, Vol. 68, 1, pp. 1-26.

Smith, A., Sutton, A., & Shinkins, B. (2016). "Early Cost-effectiveness Analysis of New Medical Tests: Response", *International Journal of Technology Assessment in Health Care*, Vol. 32, 4, pp. 324-325.

Smolen, J., & Aletaha, D. (2015). "Rheumatoid Arthritis Therapy Reappraisal: Strategies, Opportunities and Challenges", *Nature Reviews Rheumatology*, Vol. 11, 5, pp. 276-289.

Smolen, J., Breedveld, F., Burmester, G., Bykerk, V., Dougados, M., Emery, P., Kvien, T., Navarro-Compán, M., Oliver, S., Schoels, M., Scholte-Voshaar, M., Stamm, T., Stoffer, M., Takeuchi, T., Aletaha, D., Andreu, J., Aringer, M., Bergman, M., Betteridge, N., Bijlsma, H., Burkhardt, H., Cardiel, M., Combe, B., Durez, P., Fonseca, J., Gibofsky, A., Gomez-Reino, J., Graninger, W., Hannonen, P., Haraoui, B., Kouloumas, M., Landewe, R., Martin-Mola, E., Nash, P., Ostergaard, M., Östör, A., Richards, P., Sokka-Isler, T., Thorne, C., Tzioufas, A., van Vollenhoven, R., de Wit, M., & van der Heijde, D. (2016). "Treating Rheumatoid Arthritis to Target: 2014 Update of Recommendations of an International Task Force", *Annals of the Rheumatic Diseases*, Vol. 75, 1, pp. 3-15.

Smolen, J., Landewé, R., Breedveld, F., Buch, M., Burmester, G., Dougados, M., Emery, P., Gaujoux-Viala, C., Gossec, L., Nam, J., Ramiro, S., Winthrop, K., de Wit, M., Aletaha, D., Betteridge, N., Biklsma, J., Boers, M., Buttgereit, F., Combe, B., Cutolo, M., Damjanov, N., Hazes, J., Kouloumas, M., Kvien, T., Mariette, X., Pavelka, K., van Riel, P., Rubbert-Roth, A., Scholte-Voshaar, M., Scott, D., Sokka-Isler, T., Wong, J., & van der Heijde, D. (2014). "EULAR Recommendations for the Management of Rheumatoid Arthritis with Synthetic and Biological Disease-modifying Antirheumatic Drugs: 2013 Update", *Annals of the Rheumatic Diseases*, Vol. 73, 3, pp. 492-509.

Snape, D., & Spencer, L. (2003). "The Foundations of Qualitative Research". In: Ritchie, J., & Lewis, J. (Eds.) *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. (1 ed.) pp. 1-23. London: SAGE Publications Ltd.

Solomon, D., & Kuntz, K. (2000). "Should Postmenopausal Women with Rheumatoid Arthritis who are Starting Corticosteroid Treatment be Screened for Osteoporosis?: A Cost-effectiveness Analysis", *Arthritis & Rheumatology*, Vol. 43, 9, pp. 1967-1975.

Sonnenberg, F., & Beck, J. (1993). "Markov Models in Medical Decision Making: A Practical Guide", *Medical Decision Making*, Vol. 13, 4, pp. 322-338.

Souto, A., Maneiro, J., & Gómez-Reino, J. (2016). "Rate of Discontinuation and Drug Survival of Biologic Therapies in Rheumatoid Arthritis: A Systematic Review and Meta-Analysis of Drug Registries and Health Care Databases", *Rheumatology*, Vol. 55, 3, pp. 523-534.

StataCorp. (2013). "Stata Statistical Software". *Release 13*, College Station, TX: StataCorp LP.

Sterne, J., White, I., Carlin, J., Spratt, M., Royston, P., Kenward, M., Wood, A., & Carpenter, J. (2009). "Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls", *British Medical Journal*, Vol. 338, p. b2393.

Steuten, L., & Ramsey, S. (2014). "Improving early cycle economic evaluation of diagnostic technologies", *Expert Review of Pharmacoeconomics & Outcomes Research*, Vol. 14, 4, pp. 491-498.

Steuten, L., van de Wetering, G., Groothuis-Oudshoorn, K., & Retèl, V. (2013). "A Systematic and Critical Review of the Evolving Methods and Applications of Value of Information in Academia and Practice", *Pharmacoeconomics*, Vol. 31, 1, pp. 25-48.

Stevens, W., & Normand, C. (2004). "Optimisation Versus Certainty: Understanding the Issue of Heterogeneity in Economic Evaluation", *Social Science & Medicine*, Vol. 58, 2, pp. 315-320.

Stevenson, M., Archer, R., Tosh, J., Simpson, E., Everson-Hock, E., Stevens, J., Hernandez-Alava, M., Paisley, S., Dickinson, K., Scott, D., Young, A., & Wailoo, A. (2016). "Adalimumab, Etanercept, Infliximab, Certolizumab Pegol, Golimumab, Tocilizumab, and Abatacept for the Treatment of Rheumatoid Arthritis not Previously Treated with Disease-Modifying Antirheumatic Drugs and After Failure of Conventional Disease-Modifying Antirheumatic Drugs Only: Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 20, 35, pp. 1-610.

Stevenson, M., Tappenden, P., & Squires, H. (2014). "Methods for Handling Uncertainty within Pharmaceutical Funding Decisions", *International Journal of Systems Science*, Vol. 45, 1, pp. 60-68.

Stinnett, A., & Mullahy, J. (1998). "Net Health Benefits: a New Framework for the Analysis of Uncertainty in Cost-effectiveness Analysis", *Medical Decision Making*, Vol. 18, 2, pp. s68-s80.

Stokoe, H., & Deady, P. (2011). "Purchasing Medicines". In: Stephens, M. (Ed.) *Hospital Pharmacy*. (2 ed.) pp. 27-39. London: Pharmaceutical Press.

Stout, N., & Goldie, S. (2008). "Keeping the Noise Down: Common Random Numbers for Disease Simulation Modeling", *Health Care Management Science*, Vol. 11, 4, pp. 399-406.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). "UK Biobank: An Open Access Resource for Identifying the Cause of a Wide Range of Complex Diseases of Middle and Old Age", *PLoS Medicine*, Vol. 12, 3, pp. 1-10.

Suter, L., Fraenkel, L., & Braithwaite, S. (2011). "Cost-effectiveness of Adding Magnetic Resonance Imaging to Rheumatoid Arthritis Management", *JAMA Internal Medicine*, Vol. 171, 7, pp. 657-667.

Symmons, D. (2002). "Epidemiology of Rheumatoid Arthritis: Determinants of Onset, Persistence and Outcome", *Best Practice & Research Clinical Rheumatology*, Vol. 16, 5, pp. 707-722.

Symmons, D., & Silman, A. (2003). "The Norfolk Arthritis Register (NOAR)", *Clinical and Experimental Rheumatology*, Vol. 21, 5,S31, pp. S94-S99.

Symmons, D., Turner, G., Webb, R., Asten, P., Barrett, E., Lunt, M., Scott, D., & Silman, A. (2002). "The Prevalence of Rheumatoid Arthritis in the United Kingdom: New Estimates for a New Century", *Rheumatology*, Vol. 41, 7, pp. 793-800.

Tak, P. (2012). "A Personalized Medicine Approach to Biologic Treatment of Rheumatoid Arthritis: A Preliminary Treatment Algorithm", *Rheumatology*, Vol. 51, 4, pp. 600-609.

Tappenden, P. (2014). "Problem Structuring for Health Economic Model Development". In: Culyer, A. (Ed.) *Encyclopedia of Health Economics Vol. 3*. (1 ed.) pp. 168-179. Oxford: Elsevier.

Tappenden, P., & Chilcott, J. (2014). "Avoiding and Identifying Errors and Other Threats to the Credibility of Health Economic Models", *Pharmacoeconomics*, Vol. 32, 10, pp. 967-979.

The Academy of Medical Sciences. (2013). "Realising the Potential of Stratified Medicine". London: The Academy of Medical Sciences.

The British Society for Rheumatology. (2015). "National Clinical Audit for Rheumatoid and Early Inflammatory Arthritis: 1st Annual Report 2015". London: The British Society for Rheumatology.

The British Society for Rheumatology. (2016). *Biosimilars and the BSR Biologics Registers for RA*. (1 ed.) London: The British Society for Rheumatology, .

Thompson, A., Newman, W., Elliott, R., Roberts, S., Tricker, K., & Payne, K. (2014). "The Cost-Effectiveness of a Pharmacogenetic Test: A Trial-Based Evaluation of TPMT Genotyping for Azathioprine", *Value in Health*, Vol. 17, 1, pp. 22-33.

Thompson, M. (1981). "Decision-analytic Determination of Study Size: The Case of Electronic Fetal Monitoring", *Medical Decision Making*, Vol. 1, 2, pp. 165-179.

Thorn, J., Coast, J., & Andronis, L. (2016). "Interpretation of the Expected Value of Perfect Information and Research Recommendations: A Systematic Review and Empirical Investigation", *Medical Decision Making*, Vol. 36, 3, pp. 285-295.

Tintner, G. (1953). "The Definition of Econometrics", *Econometrica*, Vol. 21, 1, pp. 31-40.

Tosh, J., Brennan, A., Wailoo, A., & Bansback, N. (2011). "The Sheffield Rheumatoid Arthritis Health Economic Model", *Rheumatology*, Vol. 50, S4, pp. iv26-iv31.

Tosh, J., Stevenson, M., & Akehurst, R. (2014). "Health Economic Modelling of Treatment Sequences for Rheumatoid Arthritis: A Systematic Review", *Current Rheumatology Reports*, Vol. 16, 447, pp. 1-19.

Tosh, J., & Wailoo, A. (2008). "Review of Software for Decision Modelling". *Decision Support Unit Report*, Sheffield: Decision Support Unit.

Towse, A., & Garrison, L. (2013). "Economic Incentives for Evidence Generation: Promoting an Efficient Path to Personalized Medicine", *Value in Health*, Vol. 16, S6, pp. S39-S43.

Trusheim, M., & Berndt, E. (2015). "The Clinical Benefits, Ethics, and Economics of Stratified Medicine and Companion Diagnostics", *Drug Discovery Today*, Vol. 20, 12, pp. 1439-1450.

Trusheim, M., Berndt, E., & Douglas, F. (2007). "Stratified Medicine: Strategic and Economic Implications of Combining Drugs and Clinical Biomarkers", *Nature Reviews Drug Discovery*, Vol. 6, 4, pp. 287-293.

Tsao, N., Bansback, N., Shojania, K., & Marra, C. (2012). "The Issue of Comparators in Economic Evaluations of Biologic Response Modifiers in Rheumatoid Arthritis", *Best Practice & Research Clinical Rheumatology*, Vol. 26, 5, pp. 659-676.

Tugnet, N., Pearce, F., Tosounidou, S., Obrenovic, K., Erb, N., Packham, J., & Sandhu, R. (2013). "To what Extent is NICE Guidance on the Management of Rheumatoid Arthritis in Adults being Implemented in Clinical Practice? A Regional Survey", *Clinical Medicine*, Vol. 13, 1, pp. 42-46.

UK Genetic Testing Network. (2016). "Proposal Form for the Evaluation of a Genetic Test for NHS Service Gene Dossier/Additional Provider". London: UK Genetic Testing Network.

UK National Screening Committee. (2016). "Screening in the UK: Making Effective Recommendations - 2015 to 2016". London: UK National Screening Committee.

Upchurch, K., & Kay, J. (2012). "Evolution of Treatment for Rheumatoid Arthritis", *Rheumatology*, Vol. 51, s6, pp. vi28-vi36.

Vale, L. (2010). "Health Technology Assessment and Economic Evaluation: Arguments for a National Approach", *Value in Health*, Vol. 13, 6, pp. 859-861.

Vallejo-Torres, L., & Morris, S. (2014). "Measuring Vertical Inequity in the Delivery of Healthcare". In: Culyer, A. (Ed.) *Encyclopedia of Health Economics Vol. 2*. (1 ed.) pp. 247-254. Oxford: Elsevier.

Vallejo-Torres, L., Steuten, L., Buxton, M., Girling, A., Lilford, R., & Young, T. (2008). "Integrating Health Economics Modelling in the Product Development Cycle of Medical Devices: A Bayesian Approach", *International Journal of Technology Assessment in Health Care*, Vol. 24, 4, pp. 459-464.

van den Broek, M., Visser, K., Allaart, C., & Huizinga, T. (2013). "Personalized Medicine: Predicting Responses to Therapy in Patients with RA", *Current Opinion in Pharmacology*, Vol. 13, 3, pp. 463-469.

van der Gaag, J., & Wolfe, B. (1991). "Estimating Demand for Medical Care: Health as a Critical Factor for Adults and Children". In: Duru, G., & Paelinck, J. (Eds.) *Econometrics of Health Care*. (1 ed.) pp. 31-58. London: Kluwer Academic Publishers.

van der Velde, G., Machado, M., Ieraci, L., Witteman, W., Bombardier, C., & Krahn, M. (2011). "Cost-effectiveness of Biologic Response Modifiers Compared to Disease-modifying Antirheumatic Drugs for Rheumatoid Arthritis: A Systematic Review", *Arthritis Care & Research*, Vol. 63, 1, pp. 65-78.

van Doorslaer, E., Koolman, X., & Jones, A. (2004). "Explaining Income-related Inequalities in Doctor Utilisation in Europe", *Health Economics*, Vol. 13, 7, pp. 629-647.

van Gestel, A., Grutters, J., Schouten, J., Webers, C., Beckers, H., Joore, M., & Severens, J. (2012). "The Role of the Expected Value of Individualized Care in Cost-effectiveness Analysis and Decision Making", *Value in Health*, Vol. 15, 1, pp. 13-21.

van Gestel, A., Haagsma, C., & van Riel, P. (1998). "Validation of Rheumatoid Arthritis Improvement Criteria that Include Simplified Joint Counts", *Arthritis & Rheumatology*, Vol. 41, 10, pp. 1845-1850.

van Gestel, A., Prevoo, M., van't Hof, M., van Rijswijk, M., van de Putte, L., & Van Riel, P. (1996). "Development and Validation of the European League Against Rheumatism Response Criteria for Rheumatoid Arthritis. Comparison with the Preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria", *Arthritis & Rheumatology*, Vol. 39, 1, pp. 34-40.

van Nies, J., Krabben, A., Schoones, J., Huizinga, T., Kloppenburg, M., & van der Helm-van Mil, A. (2014). "What is the Evidence for the Presence of a Therapeutic Window of Opportunity in Rheumatoid Arthritis? A Systematic Literature Review", *Annals of the Rheumatic Diseases*, Vol. 73, 5, pp. 861-870.

van Rooij, T., Wilson, D., & Marsh, S. (2012). "Personalized Medicine Policy Challenges: Measuring Clinical Utility at Point of Care", *Expert Review of Pharmacoeconomics & Outcomes Research*, Vol. 12, 3, pp. 289-295.

van Schouwenburg, P., Rispens, T., & Wolbink, G. (2013). "Immunogenicity of Anti-TNF Biologic Therapies for Rheumatoid Arthritis", *Nature Reviews Rheumatology*, Vol. 9, 3, pp. 164-172.

Vanier, A., Mariette, X., Tubach, F., & Fautrel, B. (2017). "Cost-Effectiveness of TNF-Blocker Injection Spacing for Patients with Established Rheumatoid Arthritis in

Remission: An Economic Evaluation from the Spacing of TNF-Blocker Injections in Rheumatoid Arthritis Trial", *Value in Health*, Vol. [Epub ahead of print]. p. DOI: 10.1016/j.jval.2017.1001.1005.

Vegter, S., Boersma, C., Rozenbaum, M., Wilffert, B., Navis, G., & Postma, M. (2008). "Pharmacoeconomic Evaluations of Pharmacogenetic and Genomic Screening Programmes: A Systematic Review on Content and Adherence to Guidelines", *Pharmacoeconomics*, Vol. 26, 7, pp. 569-587.

Vegter, S., Jansen, E., Postma, M., & Boersma, C. (2010). "Economic Evaluations of Pharmacogenetic and Genomic Screening Programs: Update of the Literature", *Drug Development Research*, Vol. 71, 8, pp. 492-501.

Verstappen, S., Jacobs, J., van der Veen, M., Heurkens, A., Schenk, Y., ter Borg, E., Blaauw, A., & Bijlsma, J. (2007). "Intensive Treatment with Methotrexate in Early Rheumatoid Arthritis: Aiming for Remission. Computer Assisted Management in Early Rheumatoid Arthritis (CAMERA, an Open-label Strategy Trial)", *Annals of the Rheumatic Diseases*, Vol. 66, 11, pp. 1443-1449.

Vick, S., & Scott, A. (1998). "Agency in Health Care. Examining Patients' Preferences for Attributes of the Doctor-patient Relationship", *Journal of Health Economics*, Vol. 17, 5, pp. 587-605.

Vincent, F., Morand, E., Murphy, K., Mackay, F., Mariette, X., & Marcelli, C. (2013). "Antidrug Antibodies (ADAb) to Tumour Necrosis Factor (TNF)-Specific Neutralising Agents in Chronic Inflammatory Diseases: a Real Issue, a Clinical Perspective", *Annals of the Rheumatic Diseases*, Vol. 72, 2, pp. 165-178.

Wadhwa, M., & Thorpe, R. (2011). "Guidance on Immunogenicity Assessment of Biologically Derived Therapeutic Proteins: A European Perspective". In: Tovey, M. (Ed.) *Detection and Quantification of Antibodies to Biopharmaceuticals: Practical and Applied Considerations*. (1 ed.) pp. 37-56. Hoboken: John Wiley & Sons.

Wagstaff, A. (2015). "Commentary: Value Judgments in Health Inequality Measurement", *Epidemiology*, Vol. 26, 5, pp. 670-672.

Wagstaff, A., van Doorslaer, E., & Paci, P. (1991). "On the Measurement of Horizontal Inequity in the Delivery of Health Care", *Journal of Health Economics*, Vol. 10, 2, pp. 169-205.

Walker, S., Palmer, S., & Sculpher, M. (2007). "The Role of NICE Technology Appraisals in NHS Rationing", *British Medical Bulletin*, Vol. 81-82, 1, pp. 51-64.

Wardman, M. (1988). "A Comparison of Revealed Preference and Stated Preference Models of Travel Behaviour", *Journal of Transport Economics and Policy*, Vol. 22, 1, pp. 71-91.

Weinstein, M. (1990). "Principles of Cost-effective Resource Allocation in Health Care Organizations", *International Journal of Technology Assessment in Health Care*, Vol. 6, 1, pp. 93-103.

Weinstein, M., Siegel, J., Gold, M., Kamlet, M., & Russell, L. (1996). "Recommendations of the Panel on Cost-effectiveness in Health and Medicine", *Journal of the American Medical Association*, Vol. 276, 15, pp. 1253-1258.

Weinstein, M., & Stason, W. (1977). "Foundations of Cost-effectiveness Analysis for Health and Medical Practices", *The New England Journal of Medicine*, Vol. 296, 13, pp. 716-721.

Weinstein, M., & Zeckhauser, R. (1973). "Critical Ratios and Efficient Allocation", *Journal of Public Economics*, Vol. 2, 2, pp. 147-157.

Westhoff, G., Rau, R., & Zink, A. (2008). "Rheumatoid Arthritis Patients who Smoke have a Higher Need for DMARDs and Feel Worse, but they do not have More Joint Damage than Non-smokers of the Same Serological Group", *Rheumatology*, Vol. 47, 6, pp. 849-854.

White, I., Royston, P., & Wood, A. (2011). "Multiple Imputation using Chained Equations: Issues and Guidance for Practice", *Statistics in Medicine*, Vol. 30, 4, pp. 377-399.

Whiting, P., Rutjes, A., Westwood, M., Mallett, S., Deeks, J., Reitsma, J., Leeflang, M., Sterne, J., & Bossuyt, P. (2011). "QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies", *Annals of Internal Medicine*, Vol. 115, 8, pp. 529-536.

Williams, A. (1988). "Priority Setting in Public and Private Health Care. A Guide Through the Ideological Jungle", *Journal of Health Economics*, Vol. 7, 2, pp. 173-183.

Williams, R. (2000). "A Note on Robust Variance Estimation for Cluster-correlated Data", *Biometrics*, Vol. 56, 2, pp. 645-646.

Wilson, E. (2015). "A Practical Guide to Value of Information Analysis", *Pharmacoeconomics*, Vol. 33, 2, pp. 105-121.

Wong, W., Carlson, J., Thariani, R., & Veenstra, D. (2010). "Cost Effectiveness of Pharmacogenomics: A Critical and Systematic Review", *Pharmacoeconomics*, Vol. 28, 11, pp. 1001-1013.

Wood, A., White, E., & Royston, P. (2008). "How should Variable Selection be Performed with Multiply Imputed Data?", *Statistics in Medicine*, Vol. 27, 17, pp. 3227-3246.

Woodcock, J. (2007). "The Prospects for "Personalized Medicine" in Drug Development and Drug Therapy", *Clinical Pharmacology & Therapeutics*, Vol. 81, 2, pp. 164-169.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. (2 ed.) London: The MIT Press.

Woolf, S., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). "Potential Benefits, Limitations, and Harms of Clinical Guidelines", *British Medical Journal*, Vol. 318, 7182, pp. 527-530.

World Health Organization. (2000). "Obesity: Preventing and Managing the Global Epidemic". *WHO Technical Report Series*, 894. Geneva: World Health Organization.

Zamora, B., Mestre-Ferrandiz, J., Rejon-Parilla, J., & Garrison, L. (2016). "Phase I: Complementary Diagnostics: A Literature Review on the Value of Knowing". *OHE Research Paper*, London: The Office for Health Economics.

Zechmeister-Koss, I., Schnell-Inderst, P., & Zauner, G. (2014). "Appropriate Evidence Sources for Populating Decision Analytic Models within Health Technology Assessment (HTA): A Systematic Review of HTA Manuals and Health Economic Guidelines", *Medical Decision Making*, Vol. 34, 3, pp. 288-299.

Zhang, J., Xie, F., Delzell, E., Chen, L., Kilgore, M., Yun, H., Saag, K., Lewis, J., & Curtis, J. (2013). "Trends in the use of Biologic Agents among Rheumatoid Arthritis Patients Enrolled in the US Medicare Program", *Arthritis Care & Research*, Vol. 65, 11, pp. 1743-1751.

Zhang, W., & Anis, A. (2011). "The Economic Burden of Rheumatoid Arthritis: Beyond Health Care Costs", *Clinical Rheumatology*, Vol. 30, S1, pp. 25-32.

Zohoori, N., & Savitz, D. (1997). "Econometric Approaches to Epidemiologic Data: Relating Endogeneity and Unobserved Heterogeneity to Confounding", *Annals of Epidemiology*, Vol. 7, 4, pp. 251-257.

Zuvekas, S. (2014). "Empirical Determinants of Health Care Demand". In: Culyer, A. (Ed.) *Encyclopedia of Health Economics Vol. 1*. (1 ed.) pp. 343-354. Oxford: Elsevier.

Appendix 1: Strategy to Disseminate Research

This appendix reports the proposed publication strategy to disseminate the research presented in this thesis. Seven clear outputs are described in terms of their (i) proposed title; (ii) the relevant section from the thesis; and (iii) a description of the manuscript.

Publication One

Title: *Economics of Stratified Medicine in Rheumatoid Arthritis.*
Relevant Section from Thesis: *Chapter Two*.
Description: A systematic review of model-based economic evaluations of stratified medicine in rheumatoid arthritis was reported in *Chapter Two*. The study synthesised existing economic evidence for stratified medicine in rheumatoid arthritis and was published in *Current Rheumatology Reports* in December 2014 as the initial output from this thesis.

Publication Two

Title: *Exploring the Factors that Influenced Treatment Decisions with Anti-TNF Therapies for Patients with Rheumatoid Arthritis in England: A Qualitative Investigation*.
Relevant Section from Thesis: *Chapter Three*.
Description: A qualitative study was reported in *Chapter Three* which explored the factors that influenced specific prescribing decisions for patients with rheumatoid arthritis, with a sample of consultant rheumatologists across England. This study is proposed for dissemination as a published manuscript, of national interest, because of (i) growing evidence of regional variation in rheumatology practice in England, and (ii) a limited existing evidence base on how routine treatment decisions are actually made for patients with rheumatoid arthritis in England.
Relevant Audience: Rheumatologists, patients, and regional decision-makers in England.
Target Journal: This study had specific relevance to current practice for rheumatoid arthritis in England. Therefore, the most suitable target journal for the manuscript was *Rheumatology*, which is the official journal of the *British Society for Rheumatology*.

Publication Three

Title: *Barriers to Stratified Medicine Perceived by Rheumatologists in England: The Case of Routine Anti-TNF Immunogenicity Testing.*

Relevant Section from Thesis: *Chapter Three*.

Description: The study in *Chapter Three* also reported the potential barriers to using a test (anti-TNF anti-drug antibody and drug level testing) to stratify treatment, perceived by the sample of rheumatologists in England. This set of results addressed an explicit research objective and was therefore deemed to be applicable for dissemination as an independent manuscript.

Relevant Audience: Product manufacturers (tests and treatments), rheumatologists, and local and national decision-makers.

Target Journal: This study contributed to the broader literature regarding the potential barriers to stratified medicine in routine practice, and has a specific application to the practice of rheumatology. Therefore, the most suitable target journal for this manuscript was *Rheumatology*.

Publication Four

Title: *Estimating the Factors that Influenced the choice of Anti-TNF Prescribed to Patients with Rheumatoid Arthritis in England*.

Relevant Section from Thesis: *Chapter Four*.

Description: The study in *Chapter Four* estimated the patient-level factors that influenced the choice of anti-TNF therapy prescribed to patients with rheumatoid arthritis, based on treatment decisions observed in routine practice in England. The study addressed similar topics as the qualitative study in *Chapter Three* (regional variation in health care; uncertainty over how treatment decisions were made in practice) but from a quantitative perspective. The two studies were therefore complementary to each other.

Relevant Audience: Rheumatologists, patients, and regional decision-makers in England.

Target Journal: This study addressed a similar topic, for a similar audience, as those addressed by *Publication Two*. Therefore, the most suitable target journal for this study was also deemed to be *Rheumatology*.

Publication Five

Title: *Systematic Review of Prescribing Algorithms for Anti-TNF Immunogenicity Assessment in Rheumatoid Arthritis*.

Relevant Section from Thesis: *Section 5.3; Chapter Five*.

Description: A novel systematic review of prescribing algorithms was performed in *Section 5.3* of this thesis, which synthesised all published recommendations on how to use anti-TNF anti-drug antibody and drug level testing in routine practice. This study was timely because these tests were at an early stage of their product lifecycle, with limited clinical (and economic) evidence to describe how they could be used to inform treatment decisions.

Relevant Audience: Product manufacturers (tests and treatments), rheumatologists, local and national decision-makers, and decision analysts.

Target Journal: This study had relevance for the clinical practice of rheumatology globally because, to date, no study had synthesised the published recommendations, specifically for patients with rheumatoid arthritis, on how to use anti-TNF anti-drug antibody and drug level testing. Therefore, the most suitable target journal for this manuscript was the *Annals of the Rheumatic Diseases*, which is the official journal of the *European League Against Rheumatism*.

Publication Six

Title: *Early Model-based Conceptualisation Technique to Identify Potentially Relevant Comparators when Multiple Candidate Strategies Exist*.

Relevant Section from Thesis: *Section 5.4; Chapter Five*.

Description: A novel algebraic conceptualisation technique, to identify potentially relevant comparator strategies when performing an early economic evaluation, was developed in *Section 5.4*. The identification of relevant comparators is a challenge during the early evaluation of new test-based strategies to stratify treatment, in particular, because a limited clinical evidence base may increase the number of plausible candidate strategies in which the test could be used in practice.

Relevant Audience: Decision analysts with an interest in the early economic evaluation of health technologies.

Target Journal: This study had specific relevance to decision analysts and developed a method that was subsequently applied to the test evaluated in this thesis. Therefore, the most suitable target journal for this manuscript was *Medical Decision Making*, which

publishes (amongst other topics) novel methods for health technology assessment and decision-making.

Publication Seven

Title: *Economic Evaluation of Routine Immunogenicity Monitoring, Remission Drug Level Assessment, and Uniform Dose-reduction Strategies in Patients with Rheumatoid Arthritis that Received Adalimumab*.

Relevant Section from Thesis: *Chapter Six*.

Description: The study in *Chapter Six* presents a *de novo* model-based economic evaluation of stratified medicine in RA. The economic evaluation provided evidence for using anti-TNF anti-drug antibody and drug level testing in multiple ways to stratify treatment and for reducing the dose of anti-TNF therapies in all patients. The study also estimated the value of conducting further prospective research to reduce the uncertainty associated with making decisions based on current evidence. The study was a substantial and novel contribution to the literature, with specific relevance to decision-makers in England, and a general relevance to an international audience interested in using the tests in routine practice.

Relevant Audience: Rheumatologists, health care decision-makers, health care payers, patients, and decision analysts.

Target Journal: One suitable target journal for this manuscript, given the relevance of this economic evaluation to an international audience, was the *Annals of the Rheumatic Diseases*. The *European League Against Rheumatism* have expressed that research into the usefulness of anti-drug antibody and drug level testing is a high-priority, and the study in *Chapter Six* contributed to this agenda for research. Alternatively, the manuscript could be submitted to a general medical journal; for example, the *Journal of the American Medical Association* would also be a suitable target journal because a number of the model's input parameters were estimated from studies that were published previously in this journal.

Appendix 2: The NICE Reference Case

This appendix reports the details of the NICE Reference Case (Table A2.1). The NICE Reference Case specifies the methods, considered to be appropriate by NICE, for producing economic evidence to inform population health care resource allocation decisions in the NHS. Consistency in the methods of conducting an economic evaluation facilitates NICE to make comparisons between the outcomes of different technology appraisals (National Institute for Health and Care Excellence, 2013).

Table A2.1. *Summary of the NICE Reference Case.*

Component of Economic Evaluation	NICE Reference Case
Defining the decision problem.	Scoping phase by NICE.
Comparators.	Defined in the scope. All relevant comparators.
Perspective on outcomes.	All direct health effects for patients or, where relevant, for carers.
Perspectives on costs.	NHS and Personal & Social Services.
Type of economic evaluation.	Cost-utility analysis with fully incremental analysis.
Time horizon.	Long enough to reflect all important differences in costs or outcomes between technologies being compared.
Synthesis of evidence on health effects.	Based on systematic review.
Measuring and valuing health effects.	Health effects expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life.
Source of data for measurement of health-related quality of life.	Reported directly by patients and/or carers.
Sources of preference data for valuation of changes in health-related quality of life.	Representative sample of the UK population.
Equity considerations.	An additional QALY has the same weight regardless of who receives the health benefit.
Evidence on resource use and costs.	Costs should relate to NHS and PSS resources and should be valued using prices relevant to NHS and PSS.
Discounting.	The same annual rate for both costs and health effects (3.5%).
Handling of Uncertainty.	Probabilistic sensitivity analysis is preferred.

Source: National Institute for Health and Care Excellence (2013, pp. 31-57).

A2. References

National Institute for Health and Care Excellence. (2013). "Guide to the Methods of Technology Appraisal". *NICE Process and Methods, PMG9*. London: National Institute for Health and Care Excellence.

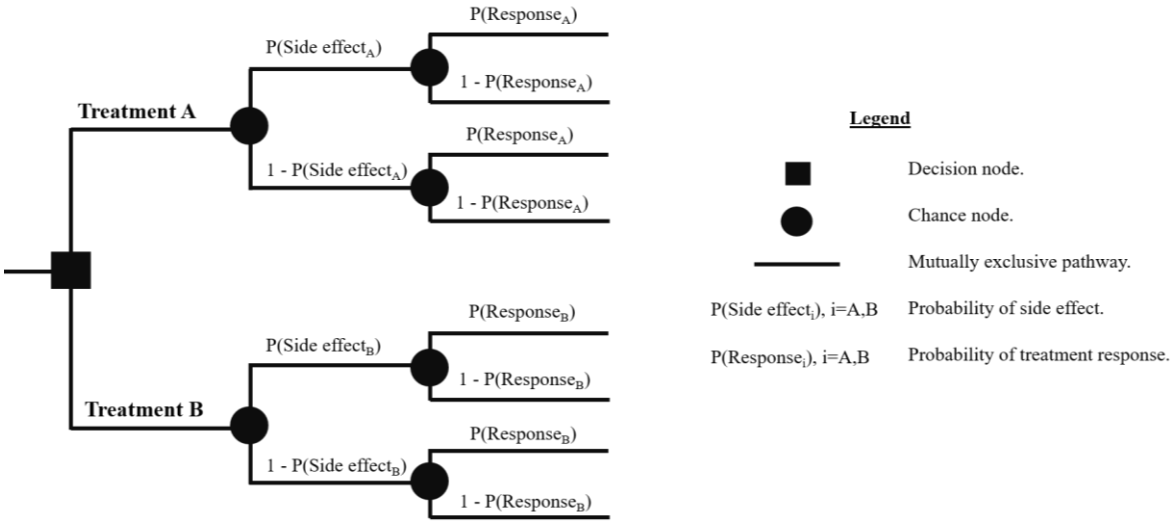
Appendix 3: Description of Three Types of Decision Analytic Model

This appendix describes three different types of decision analytic model (decision tree, Markov model, and discrete event simulation) that may be used to conduct a model-based economic evaluation in order to estimate the expected costs and health outcomes associated with different comparator health technologies. The model types each use a different approach to estimate expected outcomes. Each type of model is explained in terms of its key design features, advantages, and disadvantages. The choice of model should be appropriate to address the decision problem at hand, which is explained further during the conceptualisation of the *de novo* decision analytic model in *Chapter Five*.

A3.1. Decision Trees

A decision tree represents a decision (for example, a treatment decision) and the subsequent logical flow of chance events that may occur to a patient over time (Keeler, 1995). An illustrative example of a decision tree is provided in Figure A3.1 which represented a decision between Treatment A and Treatment B.

Figure A3.1. *Illustrative example of a decision tree.*



The design of any decision tree is characterised by three key elements: (i) a *decision node* represents the decision to be made, and is conventionally depicted by a solid square; (ii) a *chance node* represents an uncertain event that may occur after a decision, and is conventionally depicted by a solid circle; and (iii) a *branch* is the link between a decision node and the flow of chance nodes, and is conventionally depicted by a solid line (Drummond et al., 2015). The branches of a decision tree form mutually exclusive

pathways that a patient may follow (Kuntz et al., 2013). Terminal nodes, which represent outcomes of interest (such as costs or QALYs), are attached to the end of each mutually exclusive pathway. Probabilities associated with each chance node can be estimated from the clinical literature (for example, the probability of a side-effect).

The most common way to estimate the expected costs and QALYs, associated with each treatment alternative, in a decision tree is by calculating an analytic solution (Brennan et al., 2006). The probability that a patient may experience a specific mutually exclusive pathway can be found by multiplying the probabilities associated with each chance event along that pathway (Drummond et al., 2015). The expected outcomes associated with a specific pathway can then be found by multiplying the pathway-specific probability by the values of the pathway-specific terminal nodes. The expected outcomes of a particular treatment alternative can finally be estimated by summing the expected outcomes of each mutually exclusive pathway associated with that treatment. This whole analytic process is often referred to as *rolling-back the tree* (Briggs et al., 2006).

The principal advantages of decision trees are that they are relatively simple to construct and comprehend. Decision trees, however, have a number of disadvantages. Firstly, they may become difficult to handle as the number of mutually exclusive pathways increases (which is often referred to as a *bushy* decision tree) (Sonnenberg et al., 1993). Secondly, given that expected outcomes are estimated analytically, all chance outcomes are assumed to occur at the same time, which may not be justifiable for chronic diseases that are characterised by clinical events that occur over a long time horizon (Karnon et al., 2014). Finally, by implication of the first two limitations, it is difficult to use a decision tree to characterise clinical events that recur over time (such as periods of intermittent remission or regular monitoring of treatment).

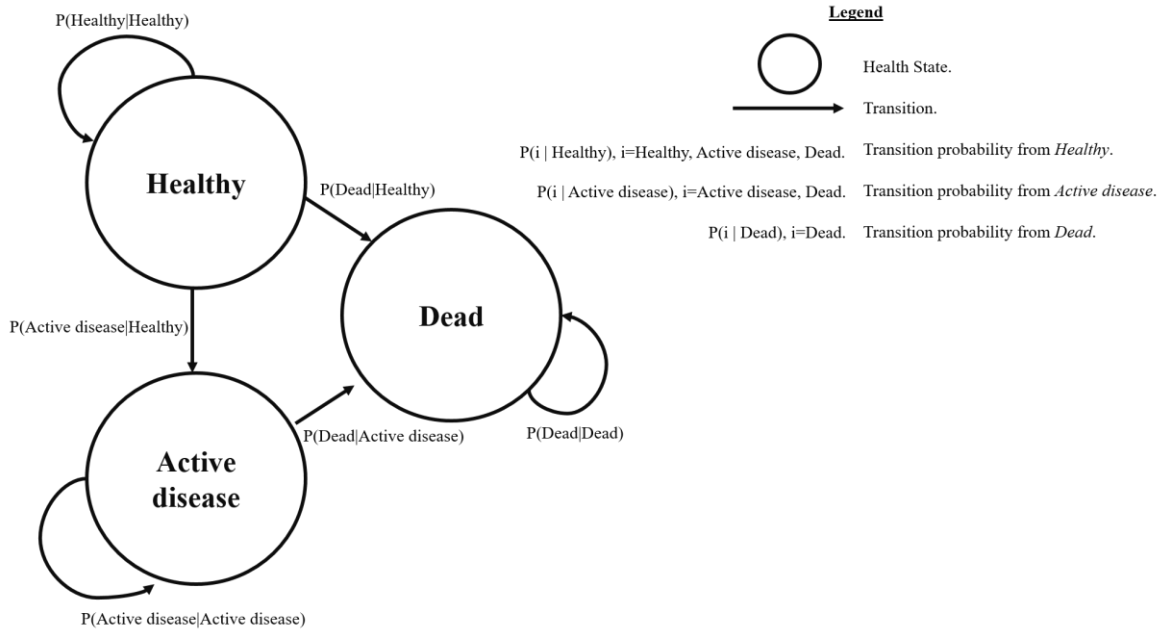
A3.2. Markov Models

A Markov model represents the movement of patients between mutually exclusive health states over time (Briggs et al., 2006). An illustrative example of a three-state Markov model is provided in Figure A3.2.

The design of a Markov model is characterised by three key elements: (i) a finite set of mutually exclusive *health states*, relevant to the decision problem, that describe the progression of disease (Briggs et al., 1998). Patients must reside in one health state at any point in time; (ii) *transition probabilities* that define the probability of moving between each health state per unit of time. The set of transition probabilities from one health state to

another must sum to one (Briggs et al., 1998); and (iii) time, which is divided into discrete *cycles* of equal length (Sonnenberg et al., 1993; Standfield et al., 2014).

Figure A3.2. *Illustrative example of a Markov model.*



The most common way to estimate the expected costs and QALYs associated with each alternative strategy in a Markov model is to simulate a cohort of identical patients between the health states over time. Costs and QALYs are assigned to each health state. Patients accrue costs and QALYs according to the duration of time they reside in each health state (Sonnenberg et al., 1993). Expected outcomes are estimated by dividing the total costs and QALYs (accrued by the whole cohort) by the number of patients within the cohort.

The advantage of Markov models is that, by explicitly accounting for time, recurring or remitting clinical events can be incorporated into the analysis (Kuntz et al., 2013). However, there are three potential limitations to Markov models. Firstly, Markov models are characterised by the *Markovian assumption* which assumes that the probability of transitioning between health states depends only on the current health state (Briggs et al., 1998). The practical implication of the Markovian assumption is that Markov models do not have any memory regarding the states that each patient has previously occupied (Sonnenberg et al., 1993; Briggs et al., 1998). The Markovian assumption may be too restrictive if future clinical events may be affected by events previously experienced by a patient. A second potential limitation is that, given patients can only reside in one state at any point in time, the model may need (i) many states to characterise the progression of disease sufficiently, which may not be computationally efficient, or (ii) potentially restrictive assumptions with respect to the choice of health states (Caro et al., 2005).

Lastly, patients in reality transition between health states in continuous time rather than in discrete cycles. Therefore, a *half-cycle correction* is often applied to Markov models, which adjusts the results by assuming that transitions occur, on average, half-way through each cycle (Sonnenberg et al., 1993; Briggs et al., 1998).

A3.3. Discrete Event Simulation

A discrete event simulation (DES) represents the experience of individual patients over time in terms of the specific events that may occur. Decision analytic models that are evaluated at the level of the individual patient may also be referred to as *individual sampling models* or *microsimulation* models (Kuntz et al., 2013).

There are four key elements that characterise the design of a DES: (i) *entities* are the objects being modelling that experience events over time. In the case of the economic evaluation of health technologies, entities are conventionally defined as individual patients (Caro et al., 2016a); (ii) *attributes* are the specific characteristics that describe an entity, such as the age and sex of a patient. Each patient may have their own unique value for an attribute, unlike with a Markov model which assumes a homogeneous cohort. The values of a patient's attributes can be updated at any time (Caro et al., 2016a); (iii) *events* are characterised broadly as something that can happen to a patient over time. For example, events may comprise disease progression or the administration of a treatment or test (Caro et al., 2016a); and (iv) *time* is advanced in a DES model according to when the next event is scheduled to occur (Caro et al., 2016a).

DES models estimate expected costs and QALYs by simulating a cohort of patients individually over a predetermined period of time. A DES begins when a single patient enters the model and their attributes are sampled from a distribution representative of a wider population. The patient's times to specific events are then sampled from survival curves described using time-to-event data. The events are scheduled in ascending order of time and the patient experiences each event sequentially. The time to specific events can be updated, if required, during the simulation (Caro et al., 2016b). The patient's own costs and QALYs derived over time are stored as attributes. The process is repeated as each individual patient leaves the model until all patients in the cohort have been simulated through the DES. The expected costs and QALYs are finally calculated, similarly to a Markov model, by dividing the total costs and QALYs for the sample by the number of patients within the sample.

A key advantage of DES models is that the simulation can remember the history of each individual patient by storing information within the patient-level attributes. The ability to remember each patient's history may be valuable if the likelihood of future clinical events may be influenced by the events previously experienced by a patient (Caro et al., 2016b). A notable disadvantage, however, is that DES models can be computationally demanding and may require thousands of patients to be simulated in order to estimate stable results (Griffin et al., 2006). The computational demand may be increased further by conducting a probabilistic sensitivity analysis (PSA) because two loops must be simulated (an outer-loop to simulate the PSA model parameters and an inner-loop to simulate the individual patients) (Griffin et al., 2006). An additional factor that may limit the use of a DES model is whether time-to-event data are available to simulate event times for each patient (Caro, 2005).

A3. References

Brennan, A., Chick, S., & Davies, R. (2006). "A Taxonomy of Model Structures for Economic Evaluation of Health Technologies", *Health Economics*, Vol. 15, 12, pp. 1295-1310.

Briggs, A., & Sculpher, M. (1998). "An Introduction to Markov Modelling for Economic Evaluation", *Pharmacoeconomics*, Vol. 13, 4, pp. 397-409.

Briggs, A., Sculpher, M., & Claxton, K. (2006). "Decision Modelling for Health Economic Evaluation". (1 ed.) Oxford: Oxford University Press.

Caro, J. (2005). "Pharmacoeconomic Analysis using Discrete Event Simulation", *Pharmacoeconomics*, Vol. 23, 4, pp. 323-332.

Caro, J., & Möller, J. (2016b). "Advantages and Disadvantages of Discrete-event Simulation for Health Economic Analyses", *Expert Review of Pharmacoeconomics & Outcomes Research*, Vol. 16, 3, pp. 327-329.

Caro, J., Möller, J., Karnon, J., Stahl, J., & Ishak, J. (2016a). "Discrete Event Simulation for Health Technology Assessment". (1 ed.) Boca Raton: CRC Press.

Drummond, M., Sculpher, M., Claxton, K., Stoddart, G., & Torrance, G. (2015). "Methods for the Economic Evaluation of Health Care Programmes". (4 ed.) Oxford: Oxford University Press.

Griffin, S., Claxton, K., Hawkins, N., & Sculpher, M. (2006). "Probabilistic Analysis and Computationally Expensive Models: Necessary and Required?", *Value in Health*, Vol. 9, 4, pp. 244-252.

Karnon, J., & Haji Ali Afzali, H. (2014). "When to use Discrete Event Simulation (DES) for the Economic Evaluation of Health Technologies? A Review and Critique of the Costs and Benefits of DES", *Pharmacoeconomic*s, Vol. 32, 6, pp. 547-558.

Keeler, E. (1995). “Chapter 9: Decision Trees and Markov Models in Cost-effectiveness Research”. In: Sloan, F. (Ed.) *Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies* (1 ed.) pp. 185-205. Cambridge: Cambridge University Press.

Kuntz, K., Sainfort, F., Butler, M., Taylor, B., Kulasingam, S., Gregory, S., Mann, E., Anderson, J., & Kane, R. (2013). "Decision and Simulation Modeling in Systematic Reviews". *AHRQ Publication No. 11(13)-EHC037-EF*. Rockville: Agency for Healthcare Research and Quality.

Sonnenberg, F., & Beck, J. (1993). "Markov Models in Medical Decision Making: A Practical Guide", *Medical Decision Making*, Vol. 13, 4, pp. 322-338.

Standfield, L., Comans, T., & Scuffham, P. (2014). "Markov Modelling and Discrete Event Simulation in Health Care: A Systematic Comparison", *International Journal of Technology Assessment in Health Care*, Vol. 30, 2, pp. 165-172.

Appendix 4: Parametric Distributions for a Probabilistic Sensitivity Analysis

Statistical distributions must be assigned to a decision analytic model's input parameters in order to conduct a PSA. Monte Carlo simulation can then sample from these distributions, by using a random number generator, to select a value for each input parameter. The use of distributions, rather than point estimates, reflects the uncertainty in the input parameters' values appropriately. In practice, there are only a few candidate distributions that can be used to characterise the uncertainty of a model's parameters. This appendix documents the candidate distributions that can be used to characterise uncertainty in the different types of parameter, and how these distributions are defined statistically. This appendix summarises the exposition of parametric distributions provided by Briggs et al. (2006, pp.84-93).

Beta Distribution

Beta distributions are suitable for data that are bound between zero and one, such as probabilities. Beta distributions are defined by two parameters: α (the number of events) and β (the sample size-α). If only summary data are available, beta distributions can be fit by method of moments. For example, if a sample mean $\bar{\mu}$ and variance s^2 are known, then:

$$\alpha + b = \frac{\bar{\mu} \times (1 - \bar{\mu})}{s^2} - 1$$

$$\alpha = \bar{\mu}(\alpha + \beta)$$

$$\beta = \alpha(1 - \bar{\mu})/\bar{\mu}$$

Dirichlet Distribution

Dirichlet distributions are suitable for multinomial outcome data. The number of parameters to define a Dirichlet distribution is equal to the number of outcomes. For example, a EULAR response (see *Appendix 7*) to treatment for patients with RA is a multinomial outcome (the patient can have either a *good*, *moderate*, or *no response*) that can be defined by a three-parameter Dirichlet distribution.

Practically, the Dirichlet distribution can be implemented as a series of conditional beta distributions. For example, if there are *k* multinomial outcomes, where α_k defines the number of events observed for each *k*, then:

Step 1: Sample a value, π_1, from $Beta \sim (\alpha_1, \sum_{j=2}^{k} \alpha_j)$;

Step 2: Sample a value, ϕ_j, from $Beta \sim (\alpha_j, \sum_{i=j+1}^{k} \alpha_i) \quad \forall\, j = 2, \ldots, k-1$;

Step 3: Set $\pi_j = (1 - \sum_{i=1}^{j-1} \pi_i)\phi_j \quad \forall\, j = 2, \ldots, k-1$;

Step 4: Set $\pi_k = 1 - \sum_{i=1}^{k-1} \pi_i$

Lognormal Distribution

The lognormal distribution may be used for relative risks as their confidence intervals are calculated on a log-odds scale. Practically, assigning a lognormal distribution can be achieved by:

Step 1: Calculate the natural log of the relative risk point estimate ($LogRR$) and confidence interval (CI);

Step 2: Calculate the log-standard error (SE) by: $SE(\text{LogRR}) = \frac{Upper\ CI - Lower\ CI}{2\ x\ 1.96}$;

Step 3: Sample from a normal distribution $Normal \sim (LogRR, SE(LogRR))$;

Step 4: Calculate the exponential of the sampled value.

Gamma Distribution

A gamma distribution is bound between zero and positive infinity. Therefore, given the skew of a gamma distribution, it can be used to characterise uncertainty in resource use (cost data are conventionally positively-skewed because (i) they cannot be negative and (ii) a small proportion of patients will often disproportionately require a large quantify of health care resources). Gamma distributions are defined by two parameters (α, β) and can be fit by method of moments. For example, if the sample mean $\bar{\mu}$ and variance s^2 are known, then:

$$\alpha = \frac{\bar{\mu}^2}{s^2}$$

$$\beta = s^2/\bar{\mu}$$

Normal Distribution

The normal distribution is a candidate distribution for any parameter because, with a sufficient sample size, the sampling distribution of the mean will be normally distributed irrespective of the data's underlying distribution (by the central limit theorem). A normal distribution is characterised by a mean and standard error.

Multivariate Normal Distribution

It is possible to correlate the values sampled for two parameters if their variance-covariance relationship is known. A Cholesky decomposition can be performed on the

variance-covariance matrix to enable the two (correlated) parameters to be sampled from a multivariate normal distribution. The following section uses matrix algebra to explain the process of performing a Cholesky decomposition, to induce correlation in the sampled values of two parameters *(x_1, x_2)*, in three steps:

Step 1: Define the variance-covariance matrix *(V)* between the two parameters:

$$V = \begin{bmatrix} Var(x_1) & Cov(x_1, x_2) \\ Cov(x_1, x_2) & Var(x_2) \end{bmatrix}$$

where *Var(x_i)* is the variance of parameter x_i, and *Cov(x_1, x_2)* is the covariance between the two parameters.

The Cholesky decomposition of *(V)* is a lower-triangular matrix (the bottom-left half of the matrix is non-zero), defined by *(T)*, such that, if multiplied by its transpose *(T')* will equal *(V)*:

$$TT' = V$$

This expression can therefore be written as:

$$\begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \times \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} = \begin{bmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{bmatrix} = \begin{bmatrix} Var(x_1) & Cov(x_1, x_2) \\ Cov(x_1, x_2) & Var(x_2) \end{bmatrix}$$
$$(T \quad \times \quad T') \quad = \qquad V \qquad == \qquad\qquad V$$

Step 2: Given that the variance-covariance matrix *(V)* is known, it is possible to calculate the values of *a*, *b*, and *c* within the Cholesky decomposition *(T)*:

$$T = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} = \begin{bmatrix} \sqrt{Var(x_1)} & 0 \\ Cov(x_1, x_2)/a & \sqrt{Var(x_2) - b^2} \end{bmatrix}$$

Step 3: To sample correlated values for all x_i, draw a value from a standard normal distribution for each parameter input *(z_i)* and multiply it by the Cholesky decomposition matrix. Finally, add the parameter-specific mean to this random variate *(μ_i):*

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} * \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \mu_1 + (a * z_1) \\ \mu_2 + (b * z_1) + (c * z_2) \end{bmatrix}$$

A4 References

Briggs, A., Claxton, K., & Sculpher, M. (2006). *Decision Modelling for Health Economic Evaluation*. (1 Ed.). Oxford: Oxford University Press.

Appendix 5: Cost-effectiveness Curves and Frontiers: Estimation and Interpretation

This appendix explains the general method for presenting uncertainty in the output of a model-based economic evaluation by using CEACs and CEAFs. The explanation describes how to calculate net benefits from the PSA output (Section A5.1), and how to present and interpret these data as a CEAC (Section A5.2) and CEAF (Section A5.3).

A5.1. Calculating Net Benefits

The calculation of each alternative strategy's net benefit is essential to the production of a CEAC and CEAF. A decision analytic model will have estimated, for each alternative strategy *(j)*, within each PSA simulation *(i)*, a different expected cost (C_j^i) and a different expected QALY (Q_j^i). The net (monetary) benefit (NB_j^i) of each alternative *(j)*, within each PSA simulation *(i)*, can be calculated by multiplying the expected QALY by a cost-effectiveness threshold *(λ)* and subtracting the expected cost, as stated in Equation A5.1:

$$NB_j^i = (\lambda \times Q_j^i) - C_j^i \qquad \textbf{(Equation A5.1)}$$

For example, Table A5.1 illustrates the calculation of the net benefits for one alternative strategy *(j=1)* over fives PSA simulations *(i=5)*, assuming a cost-effectiveness threshold *(λ)* of £20,000 per QALY gained. The example in Table A5.1 illustrates that as the values of the input parameters change between PSA simulations, the expected costs, QALYs, and net benefits of the alternative also change. A CEAC can be used to graphically represent this parameter uncertainty for multiple alternative strategies.

Table A5.1. *Illustrative example of calculating the net benefit of one strategy over five PSA simulations; cost-effectiveness threshold = £20,000 per QALY gained.*

PSA Simulation *(i)*	Expected Cost (£) (C_1^i)	Expected QALY (Q_1^i)	Net Benefit† (£) (NB_1^i)
PSA_1	£10,000	3 QALYs	£50,000
PSA_2	£50,000	10 QALYs	£150,000
PSA_3	£70,000	6 QALYs	£50,000
PSA_4	£30,000	3 QALYs	£30,000
PSA_5	£15,000	4 QALYs	£65,000

Note: †=Calculated using Equation A5.1.

A5.2. Cost-effectiveness Acceptability Curves

A CEAC plots the probability that each alternative strategy is relatively cost-effective (Y-axis) over a range of cost-effectiveness thresholds (X-axis) (Fenwick et al., 2001; Fenwick et al., 2004; Fenwick et al., 2005). The probability that an alternative strategy is cost-effective, for a given cost-effectiveness threshold, is equal to the proportion of Monte Carlo simulations in the PSA where its net benefit is the largest (Fenwick et al., 2001). At each cost-effectiveness threshold, the probability of cost-effectiveness for all alternatives must therefore sum to one. Figure A5.1 illustrates a CEAC for three hypothetical alternative strategies (*Strategy 1, Strategy 2, Strategy 3*).

The shape of a CEAC depends on the joint density of costs and QALYs derived from the PSA output (Fenwick et al., 2004). CEACs can only inform statements of probability; in the case of more than two alternatives, the alternative with the highest probability of cost-effectiveness may not necessarily have the highest expected net benefit (Fenwick et al., 2001). In the example CEAC presented in Figure A5.1, *Strategy 1* had the highest probability of being cost-effective (approximately 60%) for a range of cost-effectiveness thresholds, up to approximately £43,000 per QALY gained. *Strategy 2* had the highest probability of being cost-effective for cost-effectiveness thresholds greater than approximately £43,000 per QALY gained.

A5.3. Cost-effectiveness Acceptability Frontier

A CEAF represents the decision uncertainty associated with making recommendations according to the expected net benefit of each alternative (Fenwick et al., 2006). For each value of the cost-effectiveness threshold (X-axis), a CEAF plots the probability that the alternative with the highest expected net benefit is relatively cost-effective (Y-axis) (Fenwick et al., 2001; Barton et al., 2008); alternatives with the highest expected net benefit may not necessarily have the highest probability of being cost-effective (Barton et al., 2008). Figure A5.2 illustrates a CEAF for the same three hypothetical alternative strategies as in Figure A5.1.

The points at which a break in the CEAF occurs are referred to as "discontinuities" (Fenwick et al., 2014, p. 228); discontinuities arise when the alternative strategy with the highest expected net benefit changes. The value of the cost-effectiveness threshold at which a discontinuity in the CEAF occurs is equal to the ICER between the two alternatives (Fenwick et al., 2001; Barton et al., 2008).

In the example CEAF presented in Figure A5.2, *Strategy 1* had the highest expected net benefit up to a cost-effectiveness threshold of approximately £25,000 per QALY gained; at which point, *Strategy 2* had the highest expected net benefit. *Strategy 3* did not have the highest expected net benefit over the range of cost-effectiveness thresholds and was therefore not presented on the CEAF. The cost-effectiveness threshold at which the discontinuity in the CEAF occurred (£25,000 per QALY gained) was equal to the ICER between *Strategy 2* and *Strategy 1*.

Figure A5.1. *Illustrative example of a CEAC.*

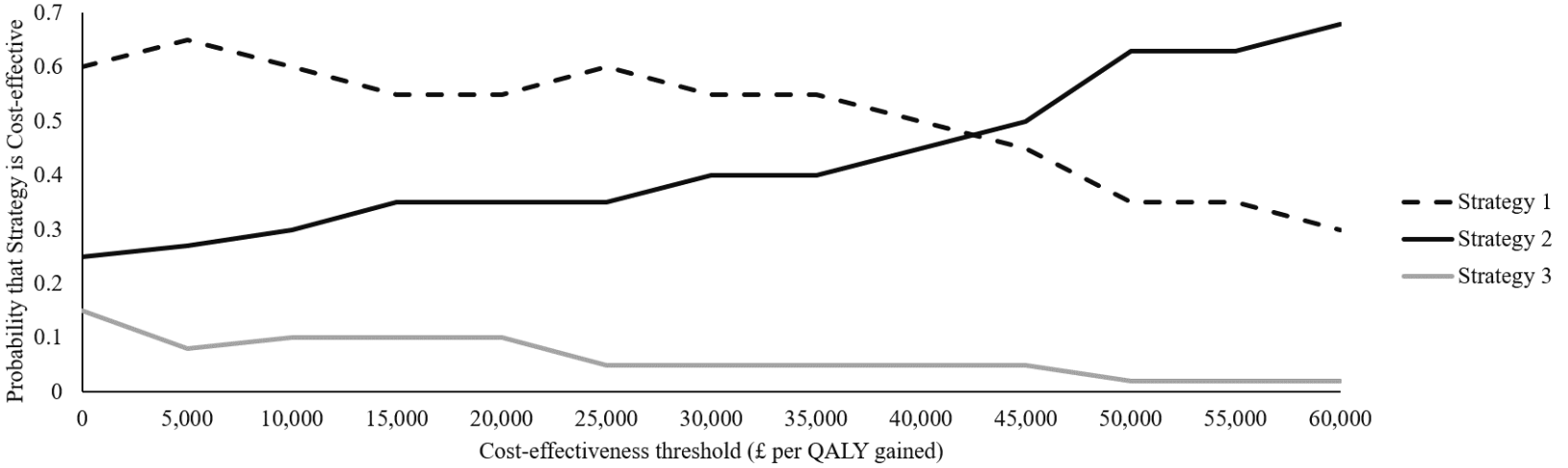
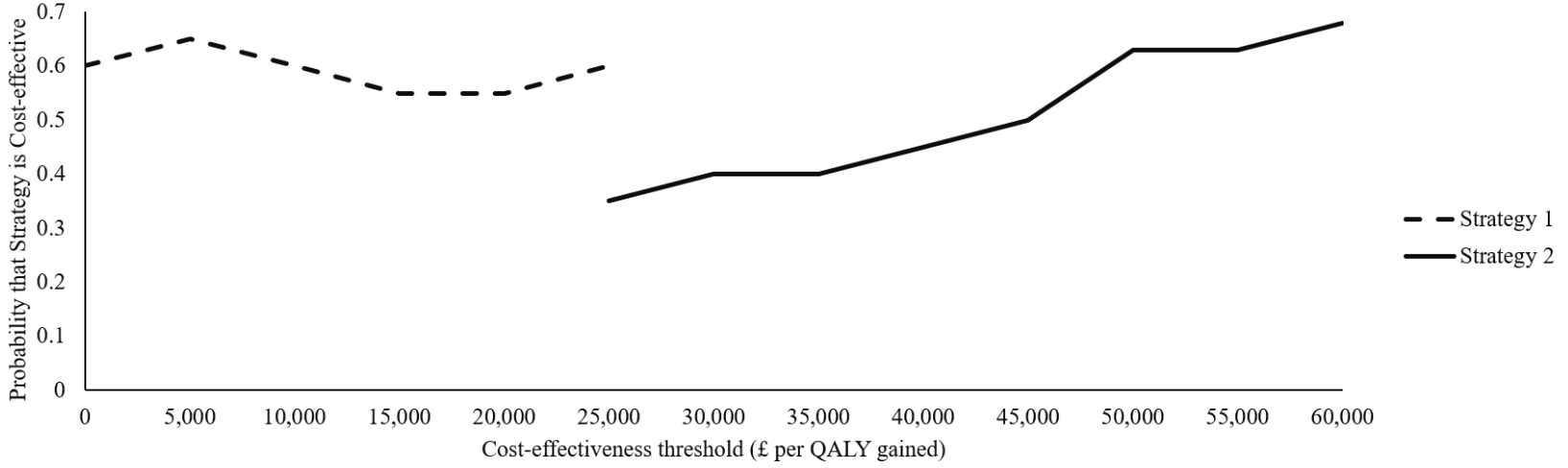


Figure A5.2. *Illustrative example of a CEAF.*



A5. References

Barton, G., Briggs, A., & Fenwick, E. (2008). "Optimal Cost-effectiveness Decisions: The Role of the Cost-effectiveness Acceptability Curve (CEAC), the Cost-effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfect Information (EVPI)", *Value in Health*, Vol. 11, 5, pp. 886-897.

Fenwick, E. (2014). “Economic Evaluation, Uncertainty in”. In: Culyer, A. (Ed.) Encyclopaedia of Health Economics. (1 ed.) pp. 224-231. Oxford: Elsevier.

Fenwick, E., & Bryford, S. (2005). "A Guide to Cost-effectiveness Acceptability Curves", *The British Journal of Psychiatry*, Vol. 187, 2, pp. 106-108.

Fenwick, E., Claxton, K., & Sculpher, M. (2001). "Representing Uncertainty: The Role of Cost-effectiveness Acceptability Curves", *Health Economics*, Vol. 10, 8, pp. 779-787.

Fenwick, E., Marshall, D., Levy, A., & Nichol, G. (2006). "Using and Interpreting Cost-effectiveness Acceptability Curves: An Example using Data from a Trial of Management Strategies for Atrial Fibrillation", *BMC Health Services Research*, Vol. 6, 52, pp. 1-8.

Fenwick, E., O'Brien, B., & Briggs, A. (2004). "Cost-effectiveness Acceptability Curves - Facts, Fallacies and Frequently Asked Questions", *Health Economics*, Vol. 13, 5, pp. 405-415.

Appendix 6: An Illustrative Example of Estimating Population EVPI

This appendix provides an illustrative example of how to estimate the population EVPI from the PSA output of a hypothetical decision analytic model. The appendix summarises the techniques described in Wilson et al. (2015) and Claxton et al. (2001).

Background

In this hypothetical example, decision-makers had to recommend one of two different treatments (*Treatment A* or *Treatment B*) for a particular disease. A decision analytic model was structured and synthesised all available evidence to inform their decision. The relative cost-effectiveness of each treatment was uncertain; a PSA was performed by simulating patients through the model five times to estimate the net monetary benefit (NMB) of each treatment. The NMB of each treatment varied over each PSA simulation, according to the different values that were sampled for the input parameters. The results of these five hypothetical PSA simulations are reported in the *Current Information* section of Table A6.1.

Table A6.1. *Estimating EVPI per patient from PSA simulation output.*

	Current Information		**Perfect Information**	
	Net Monetary Benefit (£)			
PSA Simulation	**Treatment A**	**Treatment B**	**Treatment with Highest Net Benefit**	**Maximum Net Monetary Benefit (£)**
PSA_1	500	200	Treatment A	500
PSA_2	250	450	Treatment B	450
PSA_3	450	600	Treatment B	600
PSA_4	300	100	Treatment A	300
PSA_5	800	300	Treatment A	800
Expected Outcome†	£460	£330	**Expected Outcome†**	£530

Note: †=Expected outcomes were derived by calculating the mean outcome over all PSA simulations.

By propagating parameter uncertainty through the model and averaging over all five PSA simulations, the expected NMB for *Treatment A* was £460 and for *Treatment B* was £330. Based on current information, the decision-makers recommended *Treatment A* because it

provided the greatest expected NMB. However, in 40% of the PSA simulations (n=2; PSA_2 and PSA_3), *Treatment A* did not produce the greatest NMB. Therefore, making decisions based on expected values with current information was subject to uncertainty.

Calculating EVPI per Patient

The expected NMB from recommending *Treatment A*, in this example, was equivalent to the maximum expected net benefit of making a decision based on current information. Written using general notation, if a decision-maker was faced with *(j)* alternative treatments, and the net benefits derived from those *(j)* alternatives *(NB(.))* were conditional on the values taken by a set of uncertain input parameters *(θ)*, then the decision-maker would recommend the alternative that had the greatest expected net benefit based on current information $[\max_j E_\theta NB(j,\theta)]$.

If the decision-maker had perfect information instead, there would be no parameter uncertainty and they would recommend the alternative *(j)* with the highest NMB in each PSA simulation. For example, in Table A6.1, if the PSA_1 parameters were observed with certainty, the decision-maker would recommend *Treatment A*; if the PSA_2 parameters were observed with certainty instead, the decision-maker would recommend *Treatment B*. However the *true* values of the model's input parameters were not known. Therefore, the expected NMB from a making a recommendation with perfect information is estimated by averaging over the maximum possible NMB in each of the PSA simulations $[E_\theta \max_j NB(j,\theta)]$ (Claxton et al., 2001; Wilson et al., 2015). For example, in Table A6.1, the expected NMB with perfect information was estimated to be £530.

The EVPI per patient is the difference between the expected NMB based on perfect information and the expected NMB based on current information (see Equation A6.1):

$$EVPI_{Patient} = \left[E_\theta \max_j NB(j,\theta)\right] - \left[\max_j E_\theta NB(j,\theta)\right] \quad \textbf{(Equation A6.1)}$$

The expected NMB based on current and perfect information, reported in Table A6.1, can be substituted into this formula to estimate the EVPI per patient:

$$EVPI_{Patient} = £530 - £460$$

$EVPI_{Patient} = £70$

Therefore, the EVPI per patient for the hypothetical example in Table A6.1 was equal to £70.

Calculating Population EVPI

Information derived from research has public good properties (non-rival, non-excludable) and its benefits can therefore be applied to the whole population of eligible patients, expressed as the population EVPI (Claxton et al., 2001; 1999). The population EVPI establishes an upper-bound on the value of (and rational maximum cost of) additional research to reduce parameter uncertainty (Claxton et al., 2001).

Population EVPI can be estimated by multiplying the EVPI per patient by the (discounted) effective patient population size (Wilson et al., 2015). The effective patient population size is a function of the effective product lifetime of the treatments *(T)*, the incidence of patients with the disease *(I_t)* in a given time period *(t)*, and the annual discount rate *(r)*. The duration of the intervention's product lifetime will be finite, and may depend on factors such as patent expiry, the production of new research from an unrelated research project, or the introduction of new comparator technologies (Phillips et al. 2008).

In the hypothetical example described in this appendix, for illustrative purposes, it was assumed that (i) the treatments had an effective product lifetime of ten years, (ii) there were 100 new patients every year, and (iii) the annual discount rate was 3.5%. Table A6.2 demonstrates that, under these conditions, the total discounted effective population size was 831.65 patients.

The population EVPI was calculated by multiplying the EVPI per patient (£70) by the (discounted) effective patient population size (831.65 patients) (see Equation A6.2):

$$EVPI_{Population} = EVPI_{Patient} \times \sum_{t=1}^{T} \frac{I_t}{(1+r)^t} \quad \textbf{(Equation A6.2)}$$

$EVPI_{Population} = 70 \times 831.65$

$EVPI_{Population} = £58{,}215.50$

In this example, the population EVPI was estimated to be £58,215.50. The population EVPI therefore sets a necessary condition for future research to be cost-effective: in this hypothetical example, if future research to reduce parameter uncertainty cost more than £58,215.50, the maximum possible benefit derived from that research would be lower than the amount it would cost to produce it. In contrast, future research would be potentially cost-effective if it cost less than £58,215.50 to produce.

Table A6.2. *Estimating the discounted effective population of patients*.

Year	Incidence of New Patients	Discounted Incidence of New Patients
1	100	96.62
2	100	93.35
3	100	90.19
4	100	87.14
5	100	84.20
6	100	81.35
7	100	78.60
8	100	75.94
9	100	73.37
10	100	70.89
	Total	831.65 patients

Note: Annual discount rate assumed = 3.5%

A6. References

Claxton, K. (1999). "The Irrelevance of Inference: A Decision-making Approach to the Stochastic Evaluation of Health Care Technologies", *Journal of Health Economics*, Vol. 18, 3, pp. 341-364.

Claxton, K., Neumann, P., Araki, S., & Weinstein, M. (2001). "Bayesian Value-of-information Analysis. An Application to a Policy Model of Alzheimer's Disease", *International Journal of Technology Assessment in Health Care*, Vol. 17, 1, pp. 38-55.

Philips, Z., Claxton, K., & Palmer, S. (2008). "The Half-life of Truth: What are Appropriate Time Horizons for Research Decisions?", *Medical Decision Making*, Vol. 28, 3, pp. 287-299.

Wilson, E. (2015). "A Practical Guide to Value of Information Analysis", *Pharmacoeconomics*, Vol. 33, 2, pp. 105-121.

Appendix 7: Condition-specific Outcome Measures for Rheumatoid Arthritis

This appendix provides details about the condition-specific outcome measures for RA that were used within this thesis. The outcome measures describe the classification of disease (Section A7.1 and A7.2), functional ability (Section A7.3), disease activity (Section A7.4), and treatment response (Section A7.5).

A7.1. Disease Classification: 1987 ACR Criteria

The instrument to classify patients with RA has undergone gradual adjustments since its initial inception (Arnett et al., 1988; Aletaha et al., 2010). The primary purpose of the RA classification criteria is to ensure that clinical studies include a homogenous sample of patients with RA (Symmons, 2002; Aggarwal et al., 2015).

The *American College of Rheumatology* (ACR) *1987 Classification Criteria* is reported in Table A7.1 and remains widely used in empirical research studies. Patients are classified as having RA if they have at least four of the seven criteria in Table A7.1 (Arnett et al., 1988).

A7.2. Disease Classification: 2010 ACR/EULAR Criteria

The ACR and *European League Against Rheumatism* (EULAR) amended the classification criteria to include patients with RA at an earlier stage of the disease (Aletaha et al., 2010). The *2010 ACR/EULAR Classification Criteria* is reported in Table A7.2. A total score between zero and ten is calculated by summing the scores associated with the response to the four criteria. Patients with a total score of at least six are classified as having RA (Aletaha et al., 2010).

A7.3. Functional Ability: HAQ-DI

The *Health Assessment Questionnaire* is an instrument that assesses functional ability, originally designed in rheumatology but now widely used in other diseases (Bruce et al., 2003a). The shorter *Disability Index* (HAQ-DI) version (reported in Table A7.3) assesses physical movement in twenty questions across eight categories: dressing, rising, eating, walking, hygiene, reach, grip, and usual activities (Bruce et al., 2003a). The highest score in each category is used to calculate the total HAQ-DI score and each category has equal weight. The total HAQ-DI score is between zero (no disability) and three (complete

disability), and there are twenty-five possible numerical outcomes in increments of 0.125. Patients can be interpreted as having *mild-moderate* (HAQ-DI=0-1), *moderate-severe* (HAQ-DI=1-2), or *severe-very severe* (HAQ-DI=2-3) disability (Bruce et al., 2003a).

Table A7.1. *ACR 1987 Classification Criteria.*

Criterion	Definition
Morning stiffness.	In/around the joints, lasting for one hour before maximal improvement.
Arthritis of three or more joint areas.	At least three joint areas simultaneously have soft tissue swelling or fluid.
Arthritis of hand joints.	At least one area swollen in a wrist, MCP or PIP joint.
Symmetric joints.	Simultaneous involvement of the same joint areas on both sides of the body.
Rheumatoid nodules.	Subcutaneous nodules.
Serum rheumatoid factor.	Abnormal amount of serum rheumatoid factor.
Radiographic changes.	Radiographic changes typical of RA on hand and wrist radiographs.

Source: Arnett et al. (1988, p.319); Abbreviations: MCP=Metacarpophalangeal joints; PIP=Proximal Interphalangeal joints.

Table A7.2. *ACR/EULAR 2010 Classification Criteria.*

Criterion	Score
1. Joint involvement:	
1 large joint	0
2-10 large joints	1
1-3 small joints	2
4-10 small joints	3
>10 joints (at least 1 small joint)	5
2. Serology:	
Negative RF and negative ACPA	0
Low-positive RF or low-positive ACPA	2
High-positive RF or high-positive ACPA	3
3. Acute-phase reactants:	
Normal CRP and normal ESR	0
Abnormal CRP or abnormal ESR	1
4. Duration of symptoms:	
<6 weeks	0
≥6 weeks	1

Source: Aletaha et al. (2010, p.1583).

Table A7.3. *HAQ-DI instrument to assess functional ability.*

Are you able to:	Without ANY difficulty Score = 0	With SOME difficulty Score = 1	With MUCH difficulty Score = 2	UNABLE to do Score = 3
1. Dressing & Grooming				
a. Dress yourself including tying shoelaces and doing buttons?				
b. Shampoo your hair?				
2. Rising				
a. Stand up from an armless straight chair?				
b. Get in and out of bed?				
3. Eating				
a. Cut your meat?				
b. Lift a cup of glass to your mouth?				
c. Open a new carton of milk (or soap powder)?				
4. Walking				
a. Walk outdoors on flat ground?				
b. Climb up five steps?				
5. Hygiene				
a. Wash and dry your entire body?				
b. Take a bath?				
c. Get on and off the toilet?				
6. Reach				
a. Reach and get a 5lb objected (eg. a bag of potatoes) from above your head?				
b. Bend down to pick up clothing from the floor?				
7. Grip				
a. Open car doors?				
b. Open jars which have previously been opened?				
c. Turn taps on and off?				
8. Activities				
a. Run errands and shop?				
b. Get in and out of a car?				
c. Do chores such as vacuuming, housework, or light gardening?				

Source: Bruce et al. (2003a; 2003b).

A7.4 Disease Activity: DAS28

The *Disease Activity Score-28 Joint Count* (DAS28) is a composite measure of disease activity that comprises a count of tender and swollen joints, and an assessment of the patient's erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP) (Prevoo et al.,

1995; Madsen, 2013). The DAS28 can also be measured with an optional global health assessment. Table A7.4 reports four different methods to calculate the DAS28.

Table A7.4. *Calculation of DAS28.*

DAS28 Version	Formula
ESR with global health.	$0.56\sqrt{Tender} + 0.28\sqrt{Swollen} + 0.7Ln(ESR) + 0.014Global$
ESR without global health.	$[0.56\sqrt{Tender} + 0.28\sqrt{Swollen} + 0.7Ln(ESR)] \times 1.08 + 0.16$
CRP with global health.	$0.56\sqrt{Tender} + 0.28\sqrt{Swollen} + 0.36Ln(CRP + 1) + 0.96Global$
CRP without global health.	$[0.56\sqrt{Tender} + 0.28\sqrt{Swollen} + 0.36Ln(CRP + 1)] \times 1.10 + 1.15$

Source: Prevoo et al. (1995, p.46) and Madsen (2013, p.380). Abbreviations: *Tender* = 28 tender joint count; *Swollen* = 28 swollen joint counts; *Global* = global health assessment.

A DAS28 score is bound between zero and 9.4. Disease activity is classified as *low* (DAS28≤3.2), *moderate* (3.2<DAS28≤5.1), or *high* (DAS28>5.1) (van Gestel et al., 1998).

A7.5. Treatment Response: EULAR Response

A EULAR response is an outcome that classifies a patient's response to treatment and is used within clinical practice for patients with RA in England. Response is characterised by the absolute change in DAS28 score and the level of disease activity experienced by the patient (reported in Table A7.5).

Table A7.5. *EULAR response.*

	Change in DAS28		
Disease Activity	**>1.2**	**>0.6 and ≤1.2**	**≤0.6**
Low disease activity DAS28≤3.2	Good	Moderate	No Response
Moderate disease activity 3.2<DAS28≤5.1	Moderate	Moderate	No Response
High disease activity DAS28>5.1	Moderate	No Response	No Response

Source: Van Gestel et al. (1998, p.1846).

A reduction in DAS28 of at least 1.2 and the presence of low disease activity is equivalent to a *good EULAR response* (van Gestel et al., 1998).

A7. References

Aggarwal, R., Ringold, S., Khanna, D., Neogi, T., Johnson, S., Miller, A., Brunner, H., Ogawa, R., Felson, D., Ogdie, A., Aletaha, D., & Feldman, B. (2015). "Distinctions between Diagnostic and Classification Criteria?", *Arthritis Care & Research*, Vol. 67, 7, pp. 891-897.

Aletaha, D., Neogi, T., Silman, A., Funovits, J., Felson, D., Bingham, C., Birnbaum, G., Burmester, G., Bykerk, V., Cohen, M., Combe, B., Costenbader, K., Dougados, M., Emery, P., Ferraccioli, G., Hazes, J., Hobbs, K., Huizinga, T., Kavanaugh, A., Kay, J., Kvien, T., Laing, T., Mease, P., Ménard, H., Moreland, L., Naden, R., Pincus, T., Smolen, J., Stanislawska-Biernat, E., Symmons, D., Tak, P., Upchurch, K., Vencovský, J., Wolfe, F., & Hawker, G. (2010). "2010 Rheumatoid Arthritis Classification Criteria: An American Collect of Rheumatology/European League Against Rheumatism Collaborative Initiative", *Annals of the Rheumatic Diseases*, Vol. 62, 9, pp. 2569-2581.

Arnett, F., Edworthy, S., Bloch, D., McShane, D., Fries, J., Cooper, N., Healey, L., Kaplan, S., Liang, M., Luthra, H., Medsger, T., Mitchell, D., Neustadt, D., Pinals, R., Schaller, J., Sharp, J., Wilder, R., & Hunder, G. (1988). "The American Rheumatism Association 1987 Revised Criteria for the Classification of Rheumatoid Arthritis", *Arthritis & Rheumatology*, Vol. 31, 3, pp. 315-324.

Bruce, B., & Fries, J. (2003a). "The Stanford Health Assessment Questionnaire: Dimensions and Practical Applications", *Health and Quality of Life Outcomes*, Vol. 1, 20, pp. 1-6.

Bruce, B., & Fries, J. (2003b). “The Stanford Health Assessment Questionnaire: A Review of its History, Issues, Progress, and Documentation”, *The Journal of Rheumatology*, Vol. 30, 1, pp. 167-178.

Madsen, O. (2013). "Agreement between the DAS28-CRP Assessment with 3 and 4 Variables in Patients with Rheumatoid Arthritis Treated with Biological Agents in the Daily Clinic", *The Journal of Rheumatology*, Vol. 40, 4, pp. 379-385.

Prevoo, M., van't Hof, M., Kuper, H., van Leeuwen, M., van de Putte, L., & van Riel, P. (1995). "Modified Disease Activity Scores that Include Twenty-eight-joint Counts. Development and Validation in a Prospective Longitudinal Study of Patients with Rheumatoid Arthritis", *Arthritis & Rheumatology*, Vol. 38, 1, pp. 44-48.

Symmons, D. (2002). "Epidemiology of Rheumatoid Arthritis: Determinants of Onset, Persistence and Outcome", *Best Practice & Research Clinical Rheumatology*, Vol. 16, 5, pp. 707-722.

van Gestel, A., Haagsma, C., & van Riel, P. (1998). "Validation of Rheumatoid Arthritis Improvement Criteria that Include Simplified Joint Counts", *Arthritis & Rheumatology*, Vol. 41, 10, pp. 1845-1850.

Appendix 8: Publication Version of Chapter Two

This appendix presents the published version of the systematic review of economic evaluations of stratified medicine that was reported in *Chapter Two*. The systematic review was published in *Current Rheumatology Reports* in December 2014. The content reported in *Chapter Two* has been updated since the systematic review was published (in December 2016).

The appropriate citation for the study is:

- **Gavan, S**., Harrison, M., Iglesias, C., Barton, A., Manca, A., & Payne, K. (2014). "Economics of Stratified Medicine in Rheumatoid Arthritis", *Current Rheumatology Reports*, Vol. 16, 12(468), pp. 1-11.

Curr Rheumatol Rep (2014) 16:468
DOI 10.1007/s11926-014-0468-x

HEALTH ECONOMICS AND QUALITY OF LIFE (M HARRISON, SECTION EDITOR)

Economics of Stratified Medicine in Rheumatoid Arthritis

Sean Gavan · Mark Harrison · Cynthia Iglesias · Anne Barton · Andrea Manca · Katherine Payne

Published online: 4 November 2014

Abstract Clinically relevant examples of stratified medicine are available for patients with rheumatoid arthritis (RA). The aim of this study was to understand the current economic evidence for stratified medicine in RA. Two systematic reviews were conducted to identify: (1) all economic evaluations of stratified treatments for rheumatoid arthritis, or those which have used a subgroup analysis, and (2) all stated preference studies of treatments for rheumatoid arthritis. Ten economic evaluations of stratified treatments for RA, 38 economic evaluations including with a subgroup analysis and eight stated preference studies were identified. There was some evidence to support that stratified approaches to treating a patient with RA may be cost-effective. However, there remain key gaps in the economic evidence base needed to support the introduction of stratified medicine in RA into healthcare systems and considerable uncertainty about how proposed stratified approaches will impact future patient preferences, outcomes and costs when used in routine practice.

Keywords Economic evaluation · Stated preference · Rheumatoid arthritis · Stratified medicine

This article is part of the Topical Collection on *Health Economics and Quality of Life*

Electronic supplementary material The online version of this article (doi:10.1007/s11926-014-0468-x) contains supplementary material, which is available to authorized users.

S. Gavan · K. Payne (✉)
Manchester Centre for Health Economics, Institute of Population Health, The University of Manchester, Oxford Road, Manchester M13 9PL, UK
e-mail: Katherine.payne@manchester.ac.uk

M. Harrison
Faculty of Pharmaceutical Sciences, The University of British Columbia, Vancouver, Canada

C. Iglesias
Department of Health Sciences, University of York, York, UK

A. Barton
Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, The University of Manchester, Manchester, UK

A. Manca
Centre for Health Economics, University of York, York, UK

Introduction

Examples of stratified medicine, in which a companion diagnostic test, applying genetic or biomarker information, is used to target treatments to subgroups of patients, are now emerging into clinical practice. A number of synonymous terms and concepts, such as personalised medicine, pharmacogenetics and pharmacogenomics, are used in the literature [1•, 2]. To date, however, there are few examples of personalised medicine. Many of the existing clinical applications involve a stratified approach to medicine, in which a pretreatment aim is to identify those patients who are most likely to effectively and safely respond. This article focuses on the potential for stratified medicine in rheumatoid arthritis, from an economic perspective.

Rheumatoid arthritis (RA) is a common autoimmune condition affecting up to 1 % of the population and is characterised by inflammation of synovial joints, which can lead to irreversible joint damage and disability. National Audit Office figures indicate that 45 % of RA patients are of working age and within 1 year of diagnosis, around 30 % of patients are unemployed because of their disease (www.nao.org.uk). It is estimated that RA costs the NHS in England approximately £560 million annually in direct healthcare costs and up to £4.8 billion per year in work-related disability. The biggest modifier of prognosis is treatment and the introduction of early, effective therapy has consistently been shown to

improve long-term outcomes, including the degree of joint damage, disability and unemployment [3–5].

The treatment of RA is informed through European-produced clinical guidelines [6]. In England, the National Institute for Health and Care Excellence (NICE) offers treatment guidelines for RA [7, 8]. However, even with the availability of such guidelines, clinicians still have a number of explicit choices to make when defining a treatment pathway for an individual patient with RA. Methotrexate, a disease-modifying anti-rheumatic drug (DMARD), is recommended by NICE as the first-line treatment for RA and has very modest drug costs (£75 per year, per patient); however, only 55 % of patients remain on this relatively inexpensive therapy 2 years post diagnosis. Those patients who fail to respond to methotrexate and at least one other DMARD become eligible for the more expensive biologic treatments (costing approximately £8–10,000 per patient, per year). These biologic therapies include a number of treatment options such as TNF-pathway blocking drugs (anti-TNFs); the CTLA4 analogue, abatacept; an IL6 pathway-blocking drug, tocilizumab and the B cell-depleting therapy, rituximab. There is a significant non-response rate to all of these treatments. For example, around one quarter of patients fail to respond to anti-TNF biological drugs. The inability to predict which patients will, or will not, respond to a particular therapy before a treatment has started results in a major unmet medical need.

Stratified medicine in RA potentially allows for the identification of safe and effective response predictors to inform the development of companion diagnostic tests to guide treatment selection. Using such a companion diagnostic test would facilitate the allocation of patients to strata, defined by the therapy they are most likely to respond to, early in the disease process. RA provides an ideal setting in which to introduce a stratified approach to medicine. To prevent irreversible joint damage, a stratified approach would promote early identification and fast-tracking of the strata of patients destined to require biological therapy. Getting the treatment right the first time is the key to improve short- and long-term health and non-health outcomes for patients with RA [9]. Furthermore, achieving a safe and good response early in the treatment pathway may result in a cost-effective use of healthcare resources and reduce the economic burden of RA to society. The earlier a patient with RA is treated with the most appropriate therapy, the more likely they are to experience a sustained improvement in health-related quality of life and the less likely the health resources forgone will be used to fund an ineffective treatment.

There are a number of potential options for stratified medicine in the context of RA [10, 11•]. Figure 1 provides a simplified overview of the diagnostic and treatment pathway for RA and shows some potential applications of a stratified approach to predicting, diagnosing and treating RA [12, 13]. The implementation of a stratified approach to diagnose and manage RA, as illustrated in Fig. 1, raises additional challenges and research questions that require a robust evidence base. For instance, the point at which the predictor biomarkers should be tested will be biomarker dependant. Genetic biomarkers for treatment response are not only considered to be ideal as they are stable and reliably tested, but also raise issues about data storage and confidentiality. In contrast, other predictive markers such as anti-drug antibodies can only be tested once a patient has started therapy, and the test results need to be generated close to when the sample is taken. A key unknown is if, and how, the result of the biomarker test will influence the prescribing and treating behaviour of a physician. Clearly, patients are key stakeholders in the use of stratified medicine. However, the influence of an accompanying diagnostic test to determine the response to treatment of the patient, and their subsequent behaviour, is unknown.

The above issues highlight the need for a robust clinical and economic evidence base, before stratified medicine for RA can be introduced into routine practice. Methods of economic evaluation, such as cost-effectiveness analysis, can generate information on the relative costs and benefits of using a stratified approach to medicine, compared to the conventional non-stratified approach. Despite the potential advantages of conducting a formal economic evaluation, many countries do not use the results of economic analyses as a tool to inform healthcare decision-making in the context of stratified medicine [14]. Similarly, stated preference methods (such as discrete choice experiments) can be used to identify which characteristics of a proposed new technology drive preferences and can potentially be used to inform health service developments underpinned by the views of patients, clinicians and the general public [15].

There are 12 published systematic reviews of the current economic evidence base, to support the use of approaches to personalised or stratified medicine, in terms of cost-effectiveness [16–22, 23•, 24–27]. De Bekker-Grob et al. [15] have also produced a review of all published discrete choice experiments in healthcare. None of these reviews, however, have explicitly focussed on the implications of stratified medicine in the context of the emerging paradigm of stratified medicine in RA.

The aim of this study is to understand the current economic evidence base for stratified medicine in RA, by evaluating the following: (i) published economic evaluations of stratified approaches to treating RA, and the use of subgroups in economic analyses, and (ii) stated preference studies that aim to elicit preferences for stratified approaches to treatments for RA.

Methods

Two systematic reviews were conducted to identify the economic evidence for stratified approaches in RA, by finding (i)

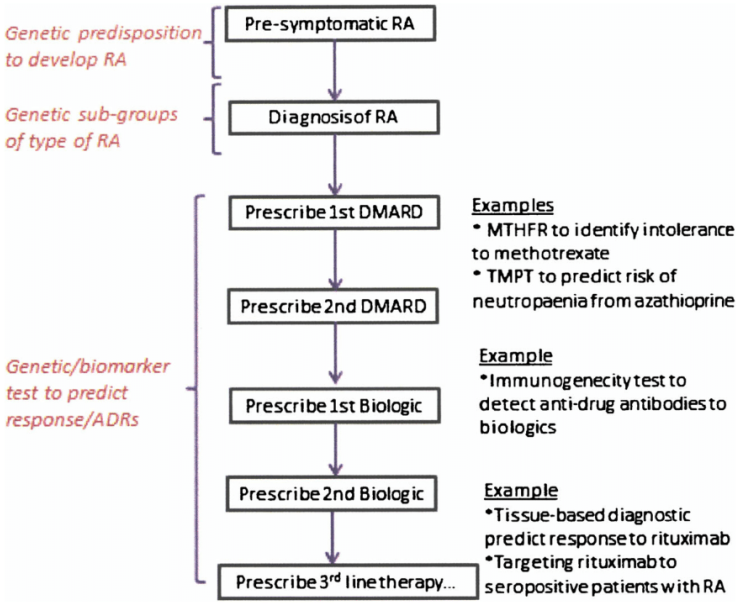


Fig. 1 An overview of potential applications of stratified medicine in RA

all published economic evaluations of treatments for RA that focussed on a stratified approach, or had included a subgroup analysis within the economic evaluation, and (ii) all published studies that had used a stated preference experiment, including discrete choice experiments (DCE), conjoint analysis (CA), adaptive conjoint analysis (ACA) or best-worst scaling (BWS) methods, to elicit preferences for treatment options in RA. The methods used to identify these two sets of literature are described below.

Identifying Economic Evaluations of Stratified Approaches in RA

We identified all economic evaluations of treatments for RA, which were published between January 1990 and April 2014. This included trial-based, observational-based and model-based economic evaluations published in peer-reviewed journals. From this complete list, economic evaluations were identified which had either (i) evaluated a stratified medicine that used an explicit targeted approach to defining the study population or (ii) included a subgroup analysis as part of the economic evaluation. An explicit targeted approach was defined as when a clinical decision is made for a subset of the patient population according to the outcome of a diagnostic test. A subgroup analysis was defined as the situation in which cost-effectiveness results are presented for subgroups of the larger patient population. The subgroup analyses could have been either explicitly defined as a *subgroup analysis* in the methodology or implicitly treated as such within a deterministic sensitivity analysis of a single model parameter value to generate the subgroups.

Supplementary Appendix 1 summarises the search strategies used. The search strategy was run in April 2014 in the following electronic databases: Ovid MEDLINE (full version, in-process and other non-indexed citations' version and daily update version), Ovid Embase, Web of Science and the NHS Economic Evaluations Database (accessed via the Centre for Reviews and Dissemination databases).

Identifying Stated Preference Studies of Stratified Approaches in RA

Stated preference studies (published between 1990 and April 2014) relevant to treatments for RA were identified by using a previously published systematic review strategy (see Supplementary Appendix 2) to identify DCEs in healthcare [28], in combination with condition-specific terms for RA. The following databases were searched on 27 May 2014: Web of Science, MEDLINE, Embase, PsycINFO and EconLit.

The Study Inclusion Process

Two reviewers independently screened all retrieved titles and abstracts to identify the studies eligible for inclusion in the systematic reviews. Conference abstracts, opinion/review/protocols and non-English articles were excluded. Following the initial screen of abstracts, full copies of papers were obtained and read in full, for the final review. Economic evaluations (as defined by Drummond et al. [29]) and stated preference studies relevant to treatments in RA were included if they had met the inclusion criteria that are detailed in Supplementary Appendix 3.

Within the total sample of identified economic evaluations of treatments for RA, those which involved a targeted approach or a subgroup analysis were identified at the data extraction stage by one reviewer (SG).

Data Collection and Extraction

Data were extracted from the identified studies by two reviewers (SG: economic evaluations; MH: stated preference studies) using a structured data collection form relevant to each study design (economic evaluation or stated or preference study). Results from both reviews were tabulated and then summarised in a narrative synthesis.

Results

This section summarises the main results from the systematic reviews of (i) economic evaluations of explicit targeted approaches to treatment for RA, and those which had included a subgroup analysis, and (ii) stated preference studies of stratified approaches to RA treatments.

Economic Evaluations of Stratified Approaches in RA

The systematic search strategy identified a total of 128 economic evaluations of treatments for RA, which were published between January 1990 and April 2014. Supplementary Appendix 4 summarises the study identification and inclusion process.

Half of the 128 economic evaluations had focussed exclusively on biologic treatments for RA, whilst approximately 30 % of the economic evaluations had solely evaluated either DMARD (n=15) or NSAID (n=24) therapy. Thirty eight economic evaluations included some form of subgroup analysis, and ten involved evaluating an explicit targeted approach to treatment. Of the 38 economic evaluations with a subgroup analysis, nine of the analyses had presented results exclusively by age and/or gender, whilst 25 evaluations had considered age and/or gender as potential subgroups amongst other patient characteristics. Such patient characteristics, by which cost-effectiveness results were reported, were used to create subgroups relating to, for example, smoking status, baseline disease severity, history of previous adverse events and the use of drug treatments additional to that being evaluated. Four of the studies with subgroup analyses were based on observational data [30–33], and one study combined an RCT with a model-based analysis [34]. The summaries of all 38 studies are available in Supplementary Appendix 5.

Table 1 provides an overview of the ten studies that had included an explicit targeted approach to treatment for RA. Supplementary Appendix 5 provides a more detailed summary of these ten studies. Nine of the studies were model-based economic evaluations. The general focus of the targeted approach was to use a companion diagnostic to predict the risk of specific adverse drug-related events (n=7 studies [35, 36, 38, 40–42, 44]). Two studies evaluated the optimisation of treatment(s) according to the result of a diagnostic test [39, 43]. One study evaluated the use of a diagnostic test to identify and treat RA patients from a wider set of patients with undifferentiated arthritis [37]. In general, the results from these evaluations suggested that the stratified approach was a cost-effective use of resources; however, there were key uncertainties in the data used to populate the models. The single trial-based study that we identified here as considering an explicit targeted approach to treatment in RA also reported extensive variation around the estimated mean costs and quality-adjusted life years [44].

Stated Preference Studies of Stratified Approaches in RA

Eight studies using stated preference methods in the context of RA were identified. Supplementary Appendix 6 shows the study inclusion process. Table 2 summarises the identified stated preference studies. The eight papers included three DCEs [45, 50, 51], four ACA [46–49], and one used a contingent ranking exercise [52]. Seven papers elicited preferences from people with RA and one from physicians; all studies involved eliciting preferences for drug treatment options, either non-biologic DMARDs or biologic DMARDs. None of the identified studies set out to explicitly assess preferences for stratified medicine, but one study included attributes which helped to identify physician preferences for treating people with RA according to clinical features [49]. Physicians who were asked whether they would escalate care (initiating a new biologic treatment) on the basis of six attributes, which represented characteristics that could be observed in a clinical examination, were found to place the greatest importance on disease activity, age and joint damage.

Three studies used subgroup analyses to provide evidence of potential differences in the treatment preferences of different groups of individuals, which could be informative for stratified medicine. Fraenkel [48] discussed findings of greater risk aversion towards drug toxicity in older patients being a factor to consider, when selecting treatments for people with RA. Similarly, two papers reporting the same ACA find differences according to ethnicity, between groups of people with RA, for the risk/benefit profile of treatments which they receive, which may explain the differential use of certain drugs [46, 47]. African-American people placed the most importance on the risk of cancer, whilst white patients placed the most importance on the likelihood of remission [46]. There was also a difference in preferences for aggressive treatment, with 50 % of white patients (compared with 16 %

Table 1 Summary of identified economic evaluations of explicit targeted approaches

Author (year) country	Intervention, comparator and population	Evaluation vehicle	Model type (as reported)	Analysis	Description of stratified approach
Bergquist et al. [35] (1995) USA	Intervention: cease methotrexate treatment, subject to liver biopsy to detect cirrhosis Comparator: no liver biopsy Population: White women with RA, aged 50, who take methotrexate	Model	Model type: decision tree Time horizon: 10 years Perspective: not reported Study type: CUA, CEA	Incremental analysis reported: yes PSA: no Other sensitivity analysis: one-way	Result of a liver biopsy: if cirrhosis is detected, methotrexate treatment is discontinued
Kim et al. [36] (2006) Korea	Intervention: methotrexate dosage based upon MTHFR genotype Comparator: conventional methotrexate dosing strategy Population: patients with RA (unspecified severity) in Korea	Model	Model type: decision tree Time horizon: 12 months Perspective: societal Study type: CEA	Incremental analysis reported: yes PSA: no Other sensitivity analysis: one-way	Polymorphism screening test: if patient has a mutant genotype, they receive a lower starting and maximum dose of methotrexate to reduce toxicity
Konnopka et al. [37] (2008) Germany	Intervention: diagnosis of RA by testing for antibodies against cyclic citrullinated peptides Comparator: diagnosis of RA by the ACR criteria Population: patients with early RA, currently classified with undifferentiated arthritis	Model	Model type: decision tree and Markov model Time horizon: 10 years Perspective: not reported Study type: CUA	Incremental analysis reported: yes PSA: yes Other sensitivity analysis: one-way	Result of the aCCP test, which if positive, the patient is treated as an RA patient. The comparator uses the ACR criteria to diagnose the patient with RA
Kowada [38] (2010) Japan	Intervention: interferon-gamma release assay for tuberculosis screening before anti-TNF treatment (QTF) Comparator: tuberculin skin test for tuberculosis screening before anti-TNF treatment (TST) Population: patients with RA, aged 40	Model	Model type: decision tree and Markov model Time horizon: lifetime Perspective: societal Study type: CUA	Incremental analysis reported: yes PSA: yes Other sensitivity analysis: one-way and two-way	Treatment with anti-TNF therapy is stratified according to a screening test for tuberculosis. If positive, the patient follows a pathway for treating tuberculosis
Krieckaert et al. [39] (2013) Netherlands	Intervention: personalised adalimumab dosage according to drug levels identified by ELISA and EULAR response Comparator: usual care Population: patients with RA starting ADL treatment	Model	Model type: Markov model Time horizon: 3 years Perspective: societal Study type: CUA	Incremental analysis reported: yes PSA: yes Other sensitivity analysis: scenario analyses and one-way	ADL treatment is delivered to patients according to decision rules based upon (1) EULAR response and (2) measured drug levels
Marra et al. [40] (2002) Canada	Intervention: polymerase chain reaction testing before azathioprine, resulting in dosage reduction in cases of reduced TMPT activity/ deficiency Comparator: no testing, with full usual dose of azathioprine Population: patients with rheumatological conditions (mainly RA and systemic lupus erythematosus)	Model	Model type: decision tree Time horizon: 6 months Perspective: third-party payer Study type: CEA	Incremental analysis reported: yes PSA: no Other sensitivity analysis: one-way	Treatment with azathioprine is stratified according to TPMT polymorphisms. If the patient has a reduced/ deficient TPMT activity, they are prescribed a lower dose of azathioprine

Table 1 (continued)

Author (year) country	Intervention, comparator and population	Evaluation vehicle	Model type (as reported)	Analysis	Description of stratified approach
Oh et al. [41] (2004) Korea	Intervention: determine TMPT activity using polymerase chain reaction testing for azathioprine dosing Comparator: standard weight-based dosing for azathioprine Population: patients with moderate-severe RA or systemic lupus erythematosus	Model	Model type: decision tree Time horizon: 12 months Perspective: societal Study type: CEA	Incremental analysis reported: yes PSA: no Other sensitivity analysis: one-way	Treatment with azathioprine is stratified according to TPMT polymorphism (wild type, heterozygous, homozygous)
Soloman et al. [42] (2000) USA	Intervention: etidronate and alendronate (1) screen and treat selectively based on bone mineral density (BMD) *T* score of <−1.0; (2) treat all (treat empirically without BMD testing) Comparator: watchful waiting Population: 55-year-old postmenopausal women with RA, starting corticosteroids	Model	Model type: Markov model Time horizon: lifetime Perspective: societal Study type: CUA	Incremental analysis reported: yes PSA: no Other sensitivity analysis: two-way	Treatment for corticosteroid-induced osteoporosis is stratified according to a bone mineral density test
Suter et al. [43] (2011) USA	Intervention: (1) Magnetic resonance imaging + standard tests and (2) treat all with tier 2 strategy initially Comparator: standard risk stratification tests (rheumatoid factor and/or anti-cyclic citrullinated peptide antibody tests) Population: patients with a new diagnosis of RA, aged 45	Model	Model type: decision tree Time horizon: lifetime Perspective: societal Study type: CUA	Incremental analysis reported: yes PSA: yes Other sensitivity analysis: one-way	Treatment is stratified according to diagnostic tests which aim to identify patients at risk of severe RA. Patients with less severe RA are given less intensive treatment
Thompson et al. [44] (2014) UK	Intervention: TPMT genotype test and azathioprine genotype test Comparator: current prescribing practice for azathioprine Population: patients with inflammatory disease (including RA and gastroenterology patients)	RCT	Model type: not appropriate Time horizon: 4 months Perspective: UK Health Service Study type: CUA	Incremental analysis reported: yes PSA: not appropriate Other sensitivity analysis: one-way	The dosage of azathioprine is stratified between patients according to the result of the TPMT genotype test

Perspectives are *as reported* within each economic evaluation

CUA cost-utility analysis, *CEA* cost-effectivenessanalysis, *PSA* probabilistic sensitivity analysis

Table 2 Summary of identified stated preference studies

Author, (year) country	Type	Treatment	Study sample	Attributes (number of levels)	Survey administration	Stratified or subgroup
Augustovski et al. [45] (2013) Argentina	DCE	Biologic agents	Study sample: (n=240) People with RA	1. Disease activity (3) 2. Mode of administration (3) 3. Frequency of dose (4) 4. Local adverse events (3) 5. Generalised adverse events (3) 6. Serious infections (2) 7. Costs (3)	Interview-completed survey, face-to-face	Stratified medicine: none Subgroup: none
Constantinescu et al. [46] (2009) USA	CA (adaptive)	Treatments (DMARD like)	Study sample: (n=136) People with RA	1. Remission (3) 2. Improvement (3) 3. Radiographic progression (3) 4. Route (3) 5. Injection reaction (3) 6. Reversible AEs (2) 7. Risk of lung injury (2) 8. Risk of TB (2) 9. Extremely rare adverse events (2) 10. Risk of cancer (2)	Interview-completed survey, face-to-face	Stratified medicine: none Subgroup: race
Constantinescu et al. [47] (2009) USA	CA (adaptive)	Treatments representing TNF inhibitors and methotrexate	Study sample: (n=136) People with RA	1. Remission (3) 2. Improvement (3) 3. Radiographic progression (3) 4. Route (3) 5. Injection reaction (3) 6. Reversible AEs (2) 7. Risk of lung injury (2) 8. Risk of TB (2) 9. Extremely rare adverse events (2) 10. Risk of cancer (2)	Interview-completed survey, face-to-face	Stratified medicine: none Subgroup: race
Fraenkel et al. [48] (2004) USA	CA (adaptive)	DMARDs	Study sample: (n=120) People with RA	1. Route (3) 2. Physician experience (2) 3. Onset (3) 4. Chance of benefit (3) 5. Bone erosions (2) 6. Injection site reaction (2) 7. Rash (3) 8. Oral ulcers (2) 9. Alopecia (2) 10. Nausea/vomiting (3) 11. Diarrhoea (3) 12. Cancer (2) 13. Nephrotoxicity (2) 14. Hepatotoxicity (2) 15. Pneumonitis (3) 16. Cost (4)	Interview-completed survey, face-to-face	Stratified medicine: treatment selection based on risk preference Subgroup: none

Table 2 (continued)

Author, (year) country	Type	Treatment	Study sample	Attributes (number of levels)	Survey administration	Stratified or subgroup
Kievit et al. [49] (2010) The Netherlands	CA (adaptive)	Escalation of DMARD (TNF inhibitor) therapy to control disease activity	Study sample: (n=135) Rheumatologists	1. Age (3) 2. Disease duration (3) 3. Clinical symptoms (3) 4. Joint damage (3) 5. Disease activity (3) 6. Current treatment (3)	Interactive computer-based	Stratified medicine: none Subgroup: Low/mod/high DAS Age Joint damage
Ozdemir et al. [50] (2009) USA	DCE	Primary RA treatment	Study sample: (n=463) People with RA	1. Chance medicine will work (4) 2. If it works, how long it takes (4) 3. Way you take medicine (5) 4. How long injection site is irritated (3) 5. Chance of serious infection (2) 6. Cost (5)	Online interactive computer-based	Stratified medicine: none Subgroup: none
Skjoldborg et al. [51] (2009) Denmark	DCE		Study sample: (n=178) People with RA	Numerous attributes not explicitly reported, but covered: 1. Morning stiffness (6) 2. Pain level (6) 3. Swollen joint count (6) 4. Fatigue (2) 5. Risk of minor infection (2) 6. Out of pocket payment for arthritis medication (18)	Interview-completed survey, face-to-face	Stratified medicine: none Subgroup: none
Slothuus et al. [52] (2002) Denmark	Contingent ranking	Treatment: anti-rheumatic drugs	Study sample: (n=115) People with RA	Numerous attributes not explicitly reported, but covered: 1. Morning stiffness (4) 2. Pain level (4) 3. Swollen joint count (4) 4. Side effects (4) 5. Monthly payment for anti-rheumatic drugs (4)	Interview-completed survey, face-to-face	Stratified medicine: none Subgroup: none

of African-American respondents) wanting this approach to managing their RA [47].

Discussion

This review has identified an emerging economic evidence base to support the use of stratified medicine in patients with RA. To date, however, there is limited evidence to understand some key aspects of the potential added value of proposed applications of stratified medicine in RA, including the correct approach to the sequencing of available biologic drugs to achieve a rapid, effective and safe response early in the care pathway. A related issue that is not yet supported by an economic evidence base is the potential use of panels of genetic and/or biomarker tests as part of the diagnostic care pathway to inform the subsequent sequencing of treatment. This panel-based approach to testing introduces the need to understand when in the diagnostic pathway to test patients, how to report test results and how to store the test results for subsequent use in a timely and effective manner. These issues, when combined, raise significant capacity issues for constrained healthcare systems. The current cost-effectiveness evidence base is dominated by Markov model-based evaluations, which are useful in the context of identifying incremental costs and benefits of new technologies at the population level. Alternative model-based approaches, such as discrete event simulation, would allow an assessment of the impact of the capacity of the health service and workforce, to deliver timely and effective stratified approaches to medicine for patients with RA [53, 54]. Alternative analytic approaches are also needed to understand the presence and impact of heterogeneity in the context of stratified approaches. A practical challenge for all economic evaluations of stratified medicine in RA will be, however, the gaps in the clinical evidence base needed to populate robust model-based evaluations, which could be addressed by using validated expert elicitation methods (for example, see Soares et al. [55]).

Methods of economic evaluation, in general, and cost-effectiveness and cost-utility analysis, specifically, are necessary but not sufficient on their own to provide decision-makers allocating scarce healthcare resources with information to support the introduction of stratified medicine in RA. There are two key stakeholders whose preferences and views will drive the eventual success of using stratified medicine in RA: clinicians and patients. Thompson et al. [44] suggested that clinicians may not always use the results from a thiopurine s-methyltransferase (TMPT) genetic test to inform subsequent prescribing behaviour, which undermines the potential value of the test to improve patient outcomes. Companion diagnostic tests, such as TMPT, offer a potential added value in terms of risk reduction by being able to identify patients at risk of adverse events before they occur. However, for this potential risk reduction to be observed in practice, clinicians (and patients) must perceive that such an added value exists and trust the evidence base supporting the use of the test. Stated preference studies could be used to elicit clinicians', and patients', perceptions of the added value of using companion diagnostics to identify and reduce the risk associated with taking medicines for RA. Similarly, preferences for the potential value of using a stratified approach to improve short- and long-term outcomes, and response rates, to medicines for RA could also be quantified using stated preference studies. A logical progression from this idea is that a patient's perceptions of the value of a stratified approach could affect their adherence to medicines, when guided by a stratified approach. There was some evidence identified in this review that particular groups of patients may have different preferences for the balance between risk and benefit of a treatment, and therefore, in theory, the additional information from a stratified approach could provide reassurance and affect adherence of these groups to treatment [46–48]. To date, there is no evidence to support the hypothesis that stratified medicine does improve patient adherence. These areas provide ample scope for future research, to generate a robust economic evidence base to support the introduction of stratified medicine in RA.

Conclusion

This review has identified some key gaps in the economic evidence base needed to support the introduction of stratified medicine in RA into healthcare systems. Stratified approaches to treating a patient with RA may be cost-effective, as demonstrated by the few economic evaluations already in existence. However, in the absence of robust clinical evidence, there is likely to be much uncertainty in how a proposed stratified approach will impact future patient outcomes and costs when used in routine practice. Preferences towards a stratified approach may also affect the viability of its introduction in practice. The value of a diagnostic test to stratify treatment is lost if a physician fails to follow the action indicated by the test result. Conversely, adherence to a given treatment may improve, if a patient's treatment is stratified according to their benefit/risk preference. Appropriately defined research questions using robust methods of economic evaluation and/or stated preference studies could meet the need for further evidence.

Acknowledgments This study was supported by a grant from innovate UK.

We would like the thank Cheryl Jones, Stuart Wright and Caroline Vass for their assistance in second screening the abstracts for the systematic review.

Compliance with Ethics Guidelines

Conflict of Interest Sean Gavan reports that he conducted this work as part of his Ph.D. at The University of Manchester funded by an NIHR Musculoskeletal Biomedical Research Unit (BRU) Ph.D. studentship (UK).

Mark Harrison and Cynthia Iglesias declare that they have no conflicts of interest.

Anne Barton reports the receipt of consulting fees from Eli Lilly and research grants from Pfizer.

Andrea Manca reports the receipt of educational speaker's fees and reimbursement of travel fees from CPR-Sante.

Katherine Payne reports that she has a research programme of work, supported by various public funding bodies, on the economics of stratified and personalised medicines and screening programmes which include the research topics addressed here.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance

1.• Redekop KW, Mladsi D. The faces of personalized medicine: a framework for understanding its meaning and scope. Value Health. 2013;16(6):S4–9. *This paper offers a framework for understanding the concept of personalised medicine.*
2. Schleidgen S, Klinger C, Bertram T, Rogowski WH, Marckmann G. What is personalized medicine: sharpening a vague term based on a systematic literature review. BMC Med Ethics. 2013;14(55):1–12.
3. Jayakumar K, Norton S, Dixey J, James D, Gough A, Williams P, et al. Sustained clinical remission in rheumatoid arthritis: prevalence and prognostic factors in an inception cohort of patients treated with conventional DMARDS. Rheumatol. 2012;51(1):169–75.
4. Scirè CA, Verstappen SM, Mirjafari H, Bunn DK, Lunt M, Montecucco C, et al. Reduction of long-term disability in inflammatory polyarthritis by early and persistent suppression of joint inflammation: results from the Norfolk Arthritis Register. Arthritis Care Res. 2011;63(7):945–52.
5. Nikiphorou E, Guh D, Bansback N, Zhang W, Dixey J, Williams P, et al. Work disability rates in RA. Results from an inception cohort with 24 years follow-up. Rheumatol. 2012;51(2):385–92.
6. Smolen J et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Ann Rheum Dis. 2010;69(6):964–75.
7. National Collaborating Centre for Chronic Conditions. Rheumatoid arthritis: national clinical guideline for management and treatment in adults. London: Royal College of Physicians; 2009.
8. National Institute for Health and Care Excellence. Adalimumab, etanercept, infliximab, rituximab and Abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor. NICE Technol Apprais 195. London: The National Institute for Health and Care Excellence; 2010.
9. Smolen JS, Aletaha D, Bijlsma JWJ, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis. 2010;69(4):631–7.
10. Littman BH. Translational strategies to implement personalized medicine: rheumatoid arthritis examples. Pers Med. 2009;6(4):429–37.
11.• Isaacs JD, Ferracciolo G. The need for personalised medicine for rheumatoid arthritis. Ann Rheum Dis. 2011;70(1):4–7. *This paper summarises some of the key approaches to personalised medicine in RA.*
12. Hughes LB, Danila MI, Bridges SL. Recent advances in personalizing rheumatoid arthritis therapy and management. Pers Med. 2009;6(2):159–70.
13. Ranganathan P. Pharmacogenomics of tumor necrosis factor antagonists in rheumatoid arthritis. Pharmacogenomics. 2005;6(5):481–90.
14. Faulkner E, Annemans L, Garrison L, et al. Challenges in the development and reimbursement of personalized medicine: payer and manufacturer perspectives and implications for health outcomes and economics research: a report of the ISPOR personalized medicine special interest group. Value Health. 2012;15(8):1162–71.
15. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145–72.
16. Beaulieu M, de Denus S, Lachaine J. Systematic review of pharmacoeconomic studies of pharmacogenomic tests. Pharmacogenomics. 2010;11(11):1573–90.
17. Carlson JJ, Henrikson NB, Veenstra DL, Ramsey SD. Economic analyses of human genetics services: a systematic review. Genet Med. 2005;7(8):519–23.
18. Djalalov S, Musa Z, Mendelson M, Siminovitch K, Hoch J. A review of economic evaluations of genetic testing services and interventions (2004–2009). Genet Med. 2011;13(2):89–94.
19. Giacomini M, Miller F, O'Brien B. Economic considerations for health insurance coverage of emerging genetic tests. Community Genet. 2003;6(2):61–73.
20. Hatz MH, Schremser K, Rogowski WH. Is individualized medicine more cost-effective? A systematic review. Pharmacoeconomics. 2014;32(5):443–55.
21. Jarrett J, Mugford M. Genetic health technology and economic evaluation: a critical review. Appl Health Econ Health Policy. 2006;5(1):27–35.
22. Phillips KA, van Bebber SL. A systematic review of cost-effectiveness analyses of pharmacogenomic interventions. Pharmacogenomics. 2004;5(8):1139–49.
23.• Phillips KA, Ann Sakowski J, Trosman J, Douglas MP, Liang SY, Neumann P. The economic value of personalized medicine tests: what we know and what we need to know. Genet Med. 2014;16(3):251–7. *This paper is the most recent previous systematic review of economic evaluations of personalised medicine.*
24. Rogowski W. Genetic screening by DNA technology: a systematic review of health economic evidence. Int J Technol Assess Health Care. 2006;22(3):327–37.
25. Vegter S, Boersma C, Rozenbaum M, Wilffert B, Navis G, Postma M. Pharmacoeconomic evaluations of pharmacogenetic and genomic screening programmes: a systematic review on content and adherence to guidelines. Pharmacoeconomics. 2008;26(7):569–87.
26. Vegter S, Jansen E, Postma MJ, Boersma C. Economic evaluations of pharmacogenetic and genomic screening programs: update of the literature. Drug Dev Res. 2010;71(8):492–501.

27. Wong WB, Carlson JJ, Thariani R, Veenstra DL. Cost effectiveness of pharmacogenomics: a critical and systematic review. Pharmacoeconomics. 2010;28(11):1001–13.
28. Harrison M, Rigby D, Vass C, Flynn T, Louviere J, Payne K. Risk as an attribute in discrete choice experiments: a systematic review of the literature. Patient. 2014;7(2):151–70.
29. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press; 2005.
30. Benucci M, Saviola G, Baiardi P, Manfredi M. Cost-effectiveness treatment with rituximab in patients with rheumatoid arthritis in real life. Rheumatol Int. 2011;31(11):1465–9.
31. Hoving J, Bartelds GM, Sluiter JK, Sadiraj K, Groot I, Lems WF, et al. Perceived work ability, quality of life, and fatigue in patients with rheumatoid arthritis after a 6-month course of TNF inhibitors: prospective intervention study and partial economic evaluation. Scand J Rheumatol. 2009;38(4):246–50.
32. Kobelt G, Eberhardt K, Geborek P. TNF inhibitors in the treatment of rheumatoid arthritis in clinical practice: costs and outcomes in a follow up study of patients with RA treated with etanercept or infliximab in southern Sweden. Ann Rheum Dis. 2004;63(1):4–10.
33. Virkki LM, Konttinen YT, Peltomaa R, Suontama K, Saario R, Immonen K, et al. Cost-effectiveness of infliximab in the treatment of rheumatoid arthritis in clinical practice. Clin Exp Rheumatol. 2008;26(6):1059–66.
34. Davey PJ, Meyer E. The cost effectiveness of misoprostol prophylaxis alongside long term nonsteroidal anti-inflammatory drugs: implications of the MUCOSA trial. Pharmacoeconomics. 2000;17 (3):295–304.
35. Bergquist SR, Felson DT, Prashker MJ, Freedberg KA. The cost-effectiveness of liver biopsy in rheumatoid arthritis patients treated with methotrexate. Arthritis Rheum. 1995;38(3):326–33.
36. Kim SK, Jun JB, El-Sohemy A, Bae SC. Cost-effectiveness analysis of MTHFR polymorphism screening by polymerase chain reaction in Korean patients with rheumatoid arthritis receiving methotrexate. J Rheumatol. 2006;33(7):1266–74.
37. Konnopka A, Conrad K, Baerwald C, König HH. Cost effectiveness of the determination of autoantibodies against cyclic citrullinated peptide in the early diagnosis of rheumatoid arthritis. Ann Rheum Dis. 2008;67(10):1399–405.
38. Kowada A. Cost effectiveness of interferon-gamma release assay for tuberculosis screening of rheumatoid arthritis patients prior to initiation of tumor necrosis factor-α antagonist therapy. Mol Diagn Ther. 2010;14(6):367–73.
39. Krieckaert CL, Nair SC, Nurmohamed MT, van Dongen CJ, Lems WF, Lafeber FP, et al. Personalised treatment using serum drug levels of adalimumab in patients with rheumatoid arthritis: an evaluation of costs and effects. Ann Rheum Dis. 2013. doi:10.1136/annrheumdis-2013-204101.
40. Marra CA, Esdaile JM, Anis AH. Practical pharmacogenetics: the cost effectiveness of screening for thiopurine s-methyltransferase polymorphisms in patients with rheumatological conditions treated with azathioprine. J Rheumatol. 2002;29(12):2507–12.
41. Oh KT, Anis AH, Bae SC. Pharmacoeconomic analysis of thiopurine methyltransferase polymorphism screening by polymerase chain reaction for treatment with azathioprine in Korea. Rheumatol. 2004;43(2):156–63.
42. Soloman DH, Kuntz KM. Should postmenopausal women with rheumatoid arthritis who are starting corticosteroid treatment be screened for osteoporosis? A cost-effectiveness analysis. Arthritis Rheum. 2000;43(9):1967–75.
43. Suter LG, Fraenkel L, Braithwaite RS. Cost-effectiveness of adding magnetic resonance imaging to rheumatoid arthritis management. Arch Intern Med. 2011;171(7):657–67.
44. Thompson AJ, Newman WG, Elliott RA, Roberts SA, Tricker K, Payne K. The cost-effectiveness of a pharmacogenetic test: a trial-based evaluation of TPMT genotyping for azathioprine. Value Health. 2014;17(1):22–33.
45. Augustovski F, Beratarrechea A, Irazola V, Rubinstein F, Tesolin P, Gonzalez J, et al. Patient preferences for biologic agents in rheumatoid arthritis: a discrete-choice experiment. Value Health. 2013;16(2):385–93.
46. Constantinescu F, Goucher S, Weinstein A, Smith W, Fraenkel L. Understanding why rheumatoid arthritis patient treatment preferences differ by race. Arthritis Rheum. 2009;61(4):413–8.
47. Constantinescu F, Goucher S, Weinstein A, Fraenkel L. Racial disparities in treatment preferences for rheumatoid arthritis. Med Care. 2009;47(3):350–5.
48. Fraenkel L, Bogardus S, Concato J, Felson D, Wittink D. Patient preferences for treatment of rheumatoid arthritis. Ann Rheum Dis. 2004;63(11):1372–8.
49. Kievit W, van Hulst L, van Riel P, Fraenkel L. Factors that influence rheumatologists' decisions to escalate care in rheumatoid arthritis: results from a choice-based conjoint analysis. Arthritis Care Res. 2010;62(6):842–7.
50. Ozdemir S, Johnson FR, Hauber AB. Hypothetical bias, cheap talk, and stated willingness to pay for health care. J Health Econ. 2009;28(4):894–901.
51. Skjoldborg US, Lauridsen J, Junker P. Reliability of the discrete choice experiment at the input and output level in patients with rheumatoid arthritis. Value Health. 2009;12(1):153–8.
52. Slothuus U, Larsen ML, Junker P. The contingent ranking method —a feasible and valid method when eliciting preferences for health care? Soc Sci Med. 2002;54(10):1601–9.
53. Karnon J. Alternative decision modelling techniques for the valuation of health care technologies: Markov processes versus discrete event simulation. Health Econ. 2003;12(10):837–48.
54. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM modeling good research practices task force-4. Value Health. 2012;15(6):821–7.
55. Soares MO, Bojke L, Dumville J, Iglesias C, Cullum N, Claxton K. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Stat Med. 2011;30 (19):2363–80.

Appendix 9: Chapter Two - Complete PRISMA Statement Checklist

This appendix provides the completed *PRISMA* checklist for the systematic review in *Chapter Two*. The checklist ensured that the systematic review was reported transparently and completely in accordance with the standards recommended for best-practice (Liberati et al., 2009).

Table A9.1. *Complete PRISMA Statement checklist for systematic review in Chapter Two*.

Section/Topic	Item No.	Checklist Item	Evidence
Title			
Title.	1	Identify the report as a systematic review, meta-analysis, or both.	Not applicable for PhD thesis.
Abstract			
Abstract.	2	Provide a structured summary.	Not applicable for PhD thesis.
Introduction			
Rationale.	3	Describe the rationale for the review in the context of what is already known.	Section 2.1.
Objectives.	4	Provide an explicit statement of questions being addressed with reference to PICOS.	Section 2.2.
Methods			
Protocol and registration.	5	Indicate if a review protocol exists, if and where is can be accessed, and, if applicable, provide registration number information.	No registration number exists. Review followed protocol as written in methods section.
Eligibility criteria.	6	Specify study characteristics and report characteristics used as criteria for eligibility, giving rationale.	Table 2.1. and Section 2.3; *Study Selection.*
Information sources.	7	Describe all information sources in the search and date last searched.	Section 2.3.
Search.	8	Present full electronic search strategy for at least one database.	Appendix 10.
Study selection.	9	State the process for selecting studies.	Section 2.3; *Study Selection*.
Data collection process.	10	Describe method of data extraction from reports.	Section 2.3; *Data Extraction and Analysis.*
Data items.	11	List and define variables for which data were sought.	Section 2.3; *Data Extraction and Analysis.*

Section/Topic	Item No	Checklist Item	Evidence
Risk of bias in individual studies.	12	Describe methods used for assessing risk of bias of individual studies.	Not applicable. Purpose of study was to critically appraise individual economic evaluations (and not to synthesise results of a common point estimate).
Summary measures.	13	State the principal summary measures.	Section 2.3; *Data Extraction and Analysis.*
Synthesis of results.	14	Describe the methods used for handling data and combining results of studies.	Section 2.3; *Data Extraction and Analysis.*
Risk of bias across studies.	15	Specify any assessment of risk of bias that may affect the cumulative evidence.	No applicable. Purpose of study was to critically appraise individual economic evaluations.
Additional analyses.	16	Describe methods of additional analyses.	Not applicable.
Results			
Study selection.	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusion at each state, ideally with a flow diagram.	Figure 2.1.
Study characteristics.	18	For each study, present characteristics for which data were extracted and provide citations.	Table 2.2; Appendix 11.
Risk of bias within studies.	19	Present data on risk of bias of each study and, if available, any outcome assessment.	Not applicable; potential risk of bias discussed throughout results (Section 2.4).
Results of individual studies.	20	For all outcomes, present for each study, (a) a simple summary data for each intervention group and (b) effect estimates and confidence intervals.	Results of individual economic evaluations are reported in Appendix 11.
Synthesis of results.	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Not applicable; Narrative synthesis throughout Section 2.4.
Risk of bias across studies.	22	Present results of any assessment of risk of bias across studies (see item 15).	Not applicable; Purpose of study was to critically appraise individual economic evaluations.

Section/Topic	Item No	Checklist Item	Evidence
Additional analyses.	23	Give results of additional analyses, if done.	Not applicable.
Discussion			
Summary of evidence.	24	Summarise the main findings including the strengths of evidence for each main outcome.	Section 2.5.
Limitations.	25	Discuss limitations at study and outcome level and at review level.	Section 2.5; *Limitations*.
Conclusions.	26	Provide a general interpretation of the results in the context of other evidence and implications for future research.	Section 2.5; *Implications for Future Research*; Section 2.6.
Funding			
Funding.	27	Describe sources of funding for the systematic review and other support, and the role of funding for the systematic review.	Not applicable for individual PhD chapter. Funders acknowledged at the start of the thesis.

Source: Liberati et al. (2009, p. 18).

A9. References

Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ionnidis, J., Clarke, M., Devereaux, P., Kleijnen, J., & Moher, D. (2009). "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Healthcare Interventions: Explanation and Elaboration", *British Medical Journal*, Vol. 339, b2700, pp. 1-27.

Appendix 10: Chapter Two – Search Strategies

This appendix reports the search strategies that were used in *Chapter Two* to identify all published economic evaluations of stratified medicine in RA, within (i) *Medline,* (ii) *Embase,* (iii) *Web of Science,* and (iv) *the Centre for Reviews and Dissemination Database*.

(i) Medline

1 economics/
2 exp "costs and cost analysis"/
3 economics, dental/
4 exp "economics, hospital"/
5 economics, medical/
6 economics, nursing/
7 economics, pharmaceutical/
8 (economic$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab.
9 (expenditure$ not energy).ti,ab.
10 value for money.ti,ab.
11 budget$.ti,ab.
12 (model or modelling).mp.
13 or/1-12
14 ((energy or oxygen) adj cost).ti,ab,rn.
15 (metabolic adj cost).ti,ab.
16 ((energy or oxygen) adj expenditure).ti,ab.
17 or/14-16
18 13 not 17
19 letter.pt.
20 editorial.pt.
21 historical article.pt.
22 or/19-21
23 18 not 22
24 Animals/
25 Humans/
26 24 not (24 and 25)
27 23 not 26

28 exp Arthritis, Rheumatoid/

29 rheumatoid arthritis.ti,ab,rn.

30 28 or 29

31 30 and 27

32 (1990$ or 1991$ or 1992$ or 1993$ or 1994$ or 1995$ or 1996$ or 1997$ or 1998$ or 1999$ or 2000$ or 2001$ or 2002$ or 2003$ or 2004$ or 2005$ or 2006$ or 2007$ or 2008$ or 2009$ or 2010$ or 2011$ or 2012$ or 2013$ or 2014$).ed.

33 31 and 32

34 remove duplicates from 33

(ii) Embase

1 health-economics/

2 exp economic-evaluation/

3 exp health-care-cost/

4 pharmacoeconomics/

5 1 or 2 or 3 or 4

6 (econom$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab.

7 (expenditure$ not energy).ti,ab.

8 (value adj2 money).ti,ab.

9 budget$.ti,ab.

10 6 or 7 or 8 or 9

11 5 or 10

12 letter.pt.

13 editorial.pt.

14 note.pt.

15 12 or 13 or 14

16 11 not 15

17 (metabolic adj cost).ti,ab.

18 ((energy or oxygen) adj cost).ti,ab.

19 ((energy or oxygen) adj expenditure).ti,ab.

20 17 or 18 or 19

21 16 not 20

22 exp animal/

23 exp animal-experiment/

24 nonhuman/

25 (rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep).ti,ab,sh.

26 22 or 23 or 24 or 25

27 exp human/

28 exp human-experiment/

29 27 or 28

30 26 not (26 and 29)

31 21 not 30

32 Rheumatoid Arthritis/

33 rheumatoid arthritis.ti,ab.

34 32 or 33

35 31 and 34

36 (1990$ or 1991$ or 1992$ or 1993$ or 1994$ or 1995$ or 1996$ or 1997$ or 1998$ or 1999$ or 2000$ or 2001$ or 2002$ or 2003$ or 2004$ or 2005$ or 2006$ or 2007$ or 2008$ or 2009$ or 2010$ or 2011$ or 2012$ or 2013$ or 2014$).em.

37 35 and 36

38 remove duplicates from 37

(iii) Web of Science

1. TI=((rheumat* same arthrit*) or ("rheumatoid arthritis"))
2. TS=(econom* or cost or costs or costly or costing or price or process or pricing or pharmacoeconom* or budget*)
3. #1 and #2
4. TS=(animal or animals or dog or dogs or hamster* or mice or mouse or rat or rats or bovine or sheep or guinea*)
5. #3 not #4

(iv) Centre for Reviews and Dissemination Database

Manually searched the NHS EED and HTA databases after using the MeSH term "Arthritis, Rheumatoid"

Appendix 11: Chapter Two – Full Data Extraction

This appendix presents the complete data extraction tables for the ten studies that were included in the systematic review of model-based economic evaluations of stratified medicine in RA (reported in *Chapter Two*).

The data extraction tables were designed according to the *Centre for Reviews and Dissemination*'s recommendations for appraising published economic evaluations (Craig et al., 2007).

The data that were extracted are presented over the following ten pages, in alphabetical order according to the lead authors' surname

Study: Bergquist et al. (1995)
Purpose of Stratified Approach: Detect cirrhosis by a liver biopsy to stratify continuation of treatment with methotrexate; avoid adverse drug reaction.

Study Design	Evaluation Characteristics	Data Sources	Test Characteristics	Analysis	Results
Intervention: Liver biopsy to stratify whether methotrexate should be continued. *Comparator:* No liver biopsy (continue methotrexate). *Population:* White women, aged 50, with RA receiving methotrexate. *Country:* USA.	*Evaluation vehicle:* Decision analytic model. *Model type:* Decision tree *Time horizon:* 5 and 10 years. *Perspective:* Not reported. *Evaluation method:* Cost-effectiveness analysis; Cost-utility analysis. *Benefits:* Life expectancy; QALYs *Direct costs included:* Biopsy, complications, treatments. *Indirect costs included:* None.	*Probabilities & outcomes:* National Centre for Health Statistics; Non-systematic review of literature. *Health-related quality of life:* Published values from the literature. *Resource use:* Charge data from Boston University Medical Centre. *Unit cost:* Charge data from Boston University Medical Centre. Test cost was reported to be based on typical utilisation of resources for outpatient liver biopsy. *Price year (currency):* 1992 ($ USA).	*Type of test:* Biopsy. *Test accuracy evidence:* Perfect sensitivity and specificity assumed. *Timing of testing:* 5 or 10 years after commencing methotrexate. *Consequence of testing:* Probability of harm (complication; death).	*Deterministic sensitivity analysis:* One-way sensitivity analysis (baseline probabilities, costs, QALYs). *Probabilistic sensitivity analysis:* None. *Value of information:* None.	*Base-case results:* Stratified approach (liver biopsy) was dominated. *Key drivers of relative cost-effectiveness:* Prior prevalence of liver cirrhosis.

Abbreviations: QALY=quality-adjusted life year; RA=rheumatoid arthritis.

Study: Kim et al. (2006).

Purpose of Stratified Approach: Test to stratify starting dose of methotrexate according to a genetic polymorphism ; avoid adverse drug reaction.

Study Design	Evaluation Characteristics	Data Sources	Test Characteristics	Analysis	Results
Intervention: MTHFR genotype test to stratify starting and maximum dose of methotrexate. *Comparator:* No testing (conventional methotrexate dose). *Population:* Patients with RA. *Country:* Korea	*Evaluation vehicle:* Decision analytic model. *Model type:* Decision tree. *Time horizon:* 12 months. *Perspective:* Societal. *Evaluation method:* Cost-effectiveness analysis. *Benefits:* Probability of continuing methotrexate therapy. *Direct costs included:* Treatments and monitoring, testing, hospitalisation. *Indirect costs included:* Out-of-pocked prescription fee.	*Probabilities & outcomes:* Retrospective analysis of accompanying patient data. Test accuracy assumed as perfect. *Health-related quality of life:* Not applicable. *Resource use:* Retrospective analysis of accompanying patient data. *Unit cost:* Mean price of products sold in Korean pharmacies during retrospective study; Observed hospitalisation costs; Unit cost of testing from Hanyang University Hospital. *Price year (currency):* 2004 (Korean won).	*Type of test:* Genetic test. *Test accuracy evidence:* Perfect sensitivity and specificity assumed. *Timing of testing:* Before commencing methotrexate. *Consequence of testing:* None reported.	*Deterministic sensitivity analysis:* One-way sensitivity analysis (prevalence of polymorphism, incidence of toxicity from methotrexate with polymorphism, cost of testing, cost of hospitalisation). *Probabilistic sensitivity analysis:* None. *Value of information:* None.	*Base-case results:* Stratified approach (genotype test) was dominant. *Key drivers of relative cost-effectiveness:* The incidence of methotrexate toxicity in patients with the genetic polymorphism.

Abbreviations: MTHFR= methylenetetrahydrofolate reductase; QALY=quality-adjusted life year; RA=rheumatoid arthritis.

Study: Konnopka et al. (2008).
Purpose of Stratified Approach: Stratify treatment by anti-CCP test to detect patients with RA earlier than conventional diagnosis; improve treatment effectiveness.

Study Design	Evaluation Characteristics	Data Sources	Test Characteristics	Analysis	Results
Intervention: Early diagnosis and treatment of RA using anti-CCP test. *Comparator:* Annual conventional diagnosis of RA with the ACR criteria. *Population:* Patients with early RA, currently classified with undifferentiated arthritis. *Country:* Germany	*Evaluation vehicle:* Decision analytic model. *Model type:* Decision tree and Markov model. *Time horizon:* 10 years. *Perspective:* Not reported. *Evaluation method:* Cost-utility analysis. *Benefits:* QALYs. *Direct costs included:* Testing, inpatient and outpatient treatments, pharmacological treatments. *Indirect costs included:* Out-of-pocket expenses.	*Probabilities & outcomes:* Non-systematic review of literature. *Health-related quality of life:* Published mapping algorithm from HAQ to EQ-5D. *Resource use:* Previously published studies. *Unit cost:* Previously published studies. Unit cost of testing obtained from Swiss data. *Price year (currency):* Year not reported (€).	*Type of test:* Molecular biochemical assay. *Test accuracy evidence:* Single published source. *Timing of testing:* At clinical presentation with arthritic pain. *Consequence of testing:* False-negative test result leads to a more rapid annual disease progression.	*Deterministic sensitivity analysis:* One-way sensitivity analysis (prevalence of polymorphism, incidence of toxicity from methotrexate with polymorphism, cost of testing, cost of hospitalisation). *Probabilistic sensitivity analysis:* Yes. *Value of information:* None.	*Base-case results:* The estimated ICER for testing was €930 per QALY gained. *Key drivers of relative cost-effectiveness:* The worse the consequence of a false-negative test result.

Abbreviations: ACR=American College of Rheumatology; Anti-CCP= antibodies against cyclic citrullinated peptides; EQ-5D=EuroQol-5 Dimension; HAQ=Health Assessment Questionnaire; ICER=Incremental cost-effectiveness ratio; QALY=quality-adjusted life year; RA=rheumatoid arthritis.

Study: Kowada et al. (2010).
Purpose of Stratified Approach: Stratify treatment for tuberculosis prior to TNFi initiation according to a test for latent tuberculosis; avoid adverse drug reaction.

Study Design	Evaluation Characteristics	Data Sources	Data Sources	Analysis	Results
Intervention: Interferon-gamma release assay (QuantiFERON-TB Gold In-Tube). *Comparator:* Tuberculin skin test. *Population:* Patients with RA, aged 40 years, before commencing TNFi therapy. *Country:* Japan.	*Evaluation vehicle:* Decision analytic model. *Model type:* Decision tree and Markov model. *Time horizon:* Lifetime. *Perspective:* Societal. *Evaluation method:* Cost-utility analysis. *Benefits:* QALYs. *Direct costs included:* Testing included: drawing blood, the test, one physician visit, and technician. Treatments. *Indirect costs included:* Productivity loss.	*Probabilities & outcomes:* Review of the published literature. *Health-related quality of life:* Published estimates. *Resource use:* Published economic evaluation. *Unit cost:* Government sources, Published economic evaluation. *Price year (currency):* 2009 (Yen).	*Type of test:* Molecular biochemical assay. *Test accuracy evidence:* Published systematic review and meta-analysis. *Timing of testing:* Before commencing TNFi therapy to detect latent tuberculosis. *Consequence of testing:* None reported.	*Deterministic sensitivity analysis:* One-way sensitivity analysis on all variables. *Probabilistic sensitivity analysis:* Yes. *Value of information:* No.	*Base-case results:* Intervention test was dominant compared to the tuberculin skin test. *Key drivers of relative cost-effectiveness:* Low incidence of tuberculosis in patients with RA.

Abbreviations: QALY=quality-adjusted life year; RA=rheumatoid arthritis; TNFi= tumour necrosis factor-α inhibitor.

Study: Krieckaert et al. (2015).
Purpose of Stratified Approach: Stratify adalimumab treatment by EULAR response and testing drug levels; reduce unnecessary health care resources.

Study Design	Evaluation Characteristics	Data Sources	Test Characteristics	Analysis	Results
Intervention: Treatment algorithm for adalimumab therapy based on EULAR response and a test for drug levels. *Comparator:* Usual care with adalimumab. *Population:* Patients with RA receiving adalimumab. *Country:* The Netherlands.	*Evaluation vehicle:* Decision analytic model. *Model type:* Markov model. *Time horizon:* 3 years. *Perspective:* Societal. *Evaluation method:* Cost-utility analysis. *Benefits:* QALYs. *Direct costs included:* Direct costs not specified, Testing, Treatment. *Indirect costs included:* Productivity costs.	*Probabilities & outcomes:* Published cohort study; Regression analysis of cohort treated with intervention and comparator. *Health-related quality of life:* Published cohort study (EQ-5D per health state). *Resource use:* Published cohort study. *Unit cost:* Published cohort study. *Price year (currency):* Year not reported (€).	*Type of test:* Molecular biochemical assay. *Test accuracy evidence:* Not reported; estimated from accompanying patient-level data. *Timing of testing:* 28 weeks after commencing adalimumab. *Consequence of testing:* None reported.	*Deterministic sensitivity analysis:* Scenario analysis (change response measure, different second-line bDMARD, different drug level cut-offs, cost and QALY assumptions of non-TNFi bDMARDs). *Probabilistic sensitivity analysis:* Yes. *Value of information:* No.	*Base-case results:* Intervention treatment algorithm was dominant. *Key drivers of relative cost-effectiveness:* EULAR response within the algorithm.

Abbreviations: bDMARD= biologic disease-modifying antirheumatic drug; EQ-5D= EuroQol-5 Dimension; EULAR=European League Against Rheumatism; QALY=quality-adjusted life year; RA=rheumatoid arthritis; TNFi= tumour necrosis factor-α inhibitor.

Study: Marra et al. (2002).
Purpose of Stratified Approach: Stratify azathioprine dose according to genetic test; avoid adverse drug reaction.

Study Design	Evaluation Characteristics	Data Sources	Test Characteristics	Analysis	Results
Intervention: Polymerase chain reaction test of TPMT activity to stratify azathioprine doses. *Comparator:* No testing – full dose of azathioprine in usual care. *Population:* Patients with rheumatic conditions (mainly RA and systemic lupus erythematosus). *Country:* Canada.	*Evaluation vehicle:* Decision analytic model. *Model type:* Decision tree. *Time horizon:* 6 months. *Perspective:* Third-party payer. *Evaluation method:* Cost-effectiveness analysis. *Benefits:* Number of adverse events avoided. *Direct costs included:* Testing, treatments, Hospitalisations, dispensing fee. *Indirect costs included:* None.	*Probabilities & outcomes:* Systematic review of published literature. *Health-related quality of life:* Not applicable. *Resource use:* Correspondence with experts and assumption. *Unit cost:* Estimated the cost of testing by proxy according to the cost of other clinically available polymerase chain reaction tests. Published cost model. *Price year (currency):* 1999 ($ CAN).	*Type of test:* Genetic test. *Test accuracy evidence:* Single published source. *Timing of testing:* Before commencing azathioprine therapy. *Consequence of testing:* False-positive result (inappropriately reduce azathioprine dose) increases resource utilisation of prescribed treatments and physician visits.	*Deterministic sensitivity analysis:* One-way sensitivity analysis (cost of test, test accuracy, probability of hospitalisation from adverse event). *Probabilistic sensitivity analysis:* No. *Value of information:* No.	*Base-case results:* Intervention test was dominant relative to the usual dose strategy. *Key drivers of relative cost-effectiveness:* Cost of testing.

Abbreviations: RA=rheumatoid arthritis; TPMP= thiopurine-methyltransferase.

Study: Nair et al. (2015).
Purpose of Stratified Approach: Stratify tight control methotrexate treatment decisions according to *handscan* imaging device to monitor inflammation in early RA; improve treatment effectiveness.

Study Design	Evaluation Characteristics	Data Sources	Test Characteristics	Analysis	Results
Intervention: Intensive, tight-control methotrexate informed by *handscan* imaging. *Comparator:* (i) Usual care; (ii) Intensive, tight-control methotrexate. *Population:* Patients with early RA. *Country:* The Netherlands.	*Evaluation vehicle:* Decision analytic model. *Model type:* Markov model. *Time horizon:* 2 years. *Perspective:* Societal; Health care system. *Evaluation method:* Cost-utility analysis. *Benefits:* QALYs. *Direct costs included:* Hospitalisations, rehabilitation, nursing home, aids, consultations with health care workers, alternative therapies, drug treatment, health care visits. *Indirect costs included:* Productivity loss.	*Probabilities & outcomes:* Accompanying RCT. *Health-related quality of life:* External published source. *Resource use:* External published source. *Unit cost:* Dutch costing manual; cost of testing was assumed. *Price year (currency):* Year not reported (€).	*Type of test:* Imaging. *Test accuracy evidence:* Perfect sensitivity and specificity assumed. *Timing of testing:* Two times every three months. *Consequence of testing:* None. Assumed to be equally as effective as the RCT tight-control strategy.	*Deterministic sensitivity analysis:* Cost of rheumatologist visit; change second-line therapy; test cost; test effectiveness; comparator effectiveness. *Probabilistic sensitivity analysis:* Yes. *Value of information:* No.	*Base-case results:* Intervention was dominant relative to the comparators. *Key drivers of relative cost-effectiveness:* Reduction in effectiveness of testing.

Abbreviations: QALY=quality-adjusted life year; RA=rheumatoid arthritis; RCT=randomised controlled trial

.

Study: Oh et al. (2004).

Purpose of Stratified Approach: Stratify azathioprine dose according to genetic test; avoid adverse drug reaction.

Study Design	Evaluation Characteristics	Data Sources	Test Characteristics	Analysis	Results
Intervention: Polymerase chain reaction test of TPMT activity to stratify azathioprine doses. *Comparator:* No testing – full dose of azathioprine in usual care. *Population:* Patients with moderate to severe RA or systemic lupus erythematosus. *Country:* Korea.	*Evaluation vehicle:* Decision analytic model. *Model type:* Decision tree. *Time horizon:* 12 months. *Perspective:* Societal. *Evaluation method:* Cost-effectiveness analysis. *Benefits:* Probability of not discontinuing treatment due to adverse events. *Direct costs included:* Treatment, testing, laboratory charges, hospitalisations. *Indirect costs included:* None.	*Probabilities & outcomes:* Review of the published literature. *Health-related quality of life:* Not applicable. *Resource use:* Treatment guidelines *Unit cost:* Hospitalisation costs from observed cases. *Price year (currency):* 2002 (Korean won).	*Type of test:* Genetic test. *Test accuracy evidence:* Single published source. *Timing of testing:* Before commencing azathioprine therapy. *Consequence of testing:* None reported.	*Deterministic sensitivity analysis:* One-way sensitivity analysis (adjust prevalence of decreased TPMT activity, hospitalisation cost, incidence of severe adverse events from intermediate TPMT activity, and the cost of testing). *Probabilistic sensitivity analysis:* No. *Value of information:* No.	*Base-case results:* Intervention test was dominant relative to the usual dose strategy. *Key drivers of relative cost-effectiveness:* Results were robust to one-way sensitivity analysis.

Abbreviations: RA=rheumatoid arthritis; TPMP= thiopurine-methyltransferase.

Study: Solomon et al. (2000).
Purpose of Stratified Approach: Stratify corticosteroid treatment by bone mineral density test; avoid adverse drug reaction (corticosteroid-induced osteoporosis).

Study Design	Evaluation Characteristics	Data Sources	Test Characteristics	Analysis	Results
Intervention: Test bone mineral density by dual x-ray absorptiometry scan before treating with corticosteroids. *Comparator:* (1) Treat only after fracture; (2) treat all with corticosteroids. *Population:* Postmenopausal women with RA, aged 55. *Country:* USA.	*Evaluation vehicle:* Decision analytic model. *Model type:* Markov model. *Time horizon:* Lifetime. *Perspective:* Societal. *Evaluation method:* CUA. *Benefits:* QALYs. *Direct costs included:* Treatments, fractures, nursing home. *Indirect costs included:* None.	*Probabilities & outcomes:* Review of the literature. *Health-related quality of life:* Obtained from published source. *Resource use:* Not reported. *Unit cost:* Wholesale prices for treatments, previously published sources. *Price year (currency):* 1998 ($ USA).	*Type of test:* Imaging. *Test accuracy evidence:* Perfect sensitivity and specificity assumed. *Timing of testing:* Before commencing corticosteroid treatment. *Consequence of testing:* None reported.	*Deterministic sensitivity analysis:* One-way sensitivity analysis (test cut-point threshold, treatment costs and efficacy, rate of fractures, rate of bone mineral density loss, proportion of hip fractures). *Probabilistic sensitivity analysis:* No. *Value of information:* No.	*Base-case results:* Stratified intervention had an ICER of $92,600 per QALY gained relative to treating only after a fracture. *Key drivers of relative cost-effectiveness:* Rates of fracture, cost of treatments.

Abbreviations: QALY=quality-adjusted life year; ICER=incremental cost-effectiveness ratio; RA=rheumatoid arthritis.

Study: Suter et al. (2011).

Purpose of Stratified Approach: Stratify combination DMARD therapy by including MRI in standard risk-stratification protocol; improve treatment effectiveness.

Study Design	Evaluation Characteristics	Data Sources	Test Characteristics	Analysis	Results
Intervention: Include MRI scan in standard risk stratification protocol to detect patients at-risk of radiographic progression. *Comparator:* (1) standard risk-stratification; (2) treat all with cDMARDs. *Population:* Patients with RA at high risk/low risk of developing severe erosive disease, aged 45 years. *Country:* USA.	*Evaluation vehicle:* Decision analytic model. *Model type:* Markov model. *Time horizon:* 12 months and lifetime. *Perspective:* Societal. *Evaluation method:* Cost-utility analysis. *Benefits:* QALYs. *Direct costs included:* Treatment, testing, management of RA with increasing severity. *Indirect costs included:* Productivity costs.	*Probabilities & outcomes:* Accompanying review of published sources. *Health-related quality of life:* Assumption; published values. *Resource use:* Clinical guidelines in the USA. *Unit cost:* Published Government documents, previously published sources. *Price year (currency):* 2010 ($ USD).	*Type of test:* Imaging. *Test accuracy evidence:* Multiple published sources; no method of synthesis reported. *Timing of testing:* Before commencing first-line treatment. *Consequence of testing:* False-negative result: greater probability of adverse events, QALY reduction, and increased medical costs.	*Deterministic sensitivity analysis:* One-way sensitivity analysis on all input parameters. *Probabilistic sensitivity analysis:* Yes. *Value of information:* No.	*Base-case results:* Stratified intervention had an ICER of $167,783 per QALY gained relative to standard practice risk-stratification over a lifetime. *Key drivers of relative cost-effectiveness:* Prevalence of poor-prognosis patients, comparator test accuracy, intervention test specificity.

Abbreviations: cDMARD= conventional disease-modifying antirheumatic drug; MRI=magnetic resonance imaging; QALY=quality-adjusted life year; ICER=incremental cost-effectiveness ratio; RA=rheumatoid arthritis.

A11. References

Bergquist, S., Felson, D., Prashker, M., & Freedberg, K. (1995). "The Cost-effectiveness of Liver Biopsy in Rheumatoid Arthritis Patients Treated with Methotrexate", *Arthritis & Rheumatology*, Vol. 38, 3, pp. 326-333.

Craig, D., Rice, S., Aguiar-Ibáñez, R., Glanville, J., Kleijnen, J., & Drummond, M. (2007). "Chapter 3: Guidance for Writing NHS EED Abstracts". In: Craig, D., & Rice, S. (Eds.) NHS Economic Evaluations Database Handbook. (3 ed.) pp. 13-64. York: Centre for Reviews and Dissemination.

Kim, S., Jun, J., El-Sohemy, A., & Bae, S. (2006). "Cost-effectiveness Analysis of MTHFR Polymorphism Screening by Polymerase Chain Reaction in Korean Patients with Rheumatoid Arthritis Receiving Methotrexate", *The Journal of Rheumatology*, Vol. 33, 7, pp. 1266-1274.

Konnopka, A., Conrad, K., Baerwald, C., & König, H. (2008). "Cost Effectiveness of the Determination of Autoantibodies against Cyclic Citrullinated Peptide in the Early Diagnosis of Rheumatoid Arthritis ", *Annals of the Rheumatic Diseases*, Vol. 67, 10, pp. 1399-1405.

Kowada. (2010). "Cost Effectiveness of Interferon-gamma Release Assay for Tuberculosis Screening of Rheumatoid Arthritis Patients Prior to Initiation of Tumor Necrosis Factor-α Antagonist Therapy", *Molecular Diagnosis & Therapy*, Vol. 14, 6, pp. 367-373.

Krieckaert, C., Nair, S., Nurmohamed, M., van Dongen, C., Lems, W., Lafeber, F., Bijlsma, J., Koffijberg, H., Wolbink, G., & Welsing, P. (2015). "Personalised Treatment using Serum Drug Levels of Adalimumab in Patients with Rheumatoid Arthritis: An Evaluation of Costs and Effects ", *Annals of the Rheumatic Diseases*, Vol. 74, 2, pp. 361-368.

Marra, C., Esdaile, J., & Anis, A. (2002). "Practical Pharmacogenetics: the Cost Effectiveness of Screening for Thiopurine S-methyltransferase Polymorphisms in Patients with Rheumatological Conditions Treated with Azathioprine", *The Journal of Rheumatology*, Vol. 29, 12, pp. 2507-2512.

Nair, S., Welsing, P., Jacobs, J., van Larr, J., Rensen, W., Ardine de Wit, G., Bijlsma, J., & Lafeber, F. (2015). "Economic Evaluation of a Tight-control Treatment Strategy using an Imaging Device (Handscan) for Monitoring Joint Inflammation in Early Rheumatoid Arthritis", *Clinical and Experimental Rheumatology*, Vol. 33, 6, pp. 831-838.

Oh, K., Anis, A., & Bae, S. (2004). "Pharmacoeconomic Analysis of Thiopurine Methyltransferase Polymorphism Screening by Polymerase Chain Reaction for Treatment with Azathioprine in Korea", *Rheumatology*, Vol. 43, 2, pp. 156-163.

Solomon, D., & Kuntz, K. (2000). "Should Postmenopausal Women with Rheumatoid Arthritis who are Starting Corticosteroid Treatment be Screened for Osteoporosis?: A Cost-effectiveness Analysis", *Arthritis & Rheumatology*, Vol. 43, 9, pp. 1967-1975.

Suter, L., Fraenkel, L., & Braithwaite, S. (2011). "Cost-effectiveness of Adding Magnetic Resonance Imaging to Rheumatoid Arthritis Management", *JAMA Internal Medicine*, Vol. 171, 7, pp. 657-667.

Appendix 12: Chapter Three – Complete SRQR Checklist

This appendix reports the completed twenty-one item *SRQR* checklist (Table A12.1) for the qualitative study that aimed to understand current prescribing and treatment practices for the management of patients with RA in *Chapter Three*. The checklist described the standards for reporting qualitative research clearly and transparently (O'Brien et al., 2014).

Table A12.1. *Completed SRQR checklist for the qualitative study in Chapter Three..*

Topic	No.	Item	Evidence
Title and Abstract			
Title.	1	Concise description of the nature and topic of the study. Identifying the study as qualitative or indicating the approach or data collection methods is recommended.	Not applicable for a chapter in a PhD thesis.
Abstract.	2	Summary of key elements of the study using the abstract format of the intended publication; typically includes background, purpose, methods, results, and conclusions.	Not applicable for a chapter in a PhD thesis.
Introduction			
Problem formulation	3	Description and significance of the problem/phenomenon studied; review of relevant theory and empirical work; problem statement.	Section 3.1;
Purpose of research question.	4	Purpose of the study and specific objectives or questions.	Section 3.2.
Methods			
Qualitative approach and research paradigm.	5	Qualitative approach (eg. ethnography, grounded theory, case study) and guiding theory if appropriate; Identifying the research paradigm is also recommended.	Section 3.3.3.
Researcher characteristics and reflexivity.	6	Researchers' characteristics that may influence the research, including personal attributes, qualifications/experience, relationship with the participants, and/or presuppositions.	Section 3.5; *Reflective Statement.*

Topic	No.	Item	Evidence
Context.	7	Setting/site and salient contextual factors.	Section 3.3.1; Section 3.3.2.
Sampling strategy.	8	How and why research participants were selected; criteria for deciding when no further sampling was necessary.	Section 3.3.1.
Ethical issues pertaining to human subjects.	9	Documentation of approval by an appropriate ethics review board and participant consent, or explanation for lack thereof.	Section 3.3.4.
Data collection methods.	10	Types of data collected; details of data collection procedures including (as appropriate) start and stop data collection and analysis, iterative process, modification of procedures in response to evolving study findings.	Section 3.3.2; Appendix 14.
Data collection instruments.	11	Description of instruments (eg. interview guides, questionnaires) and devices (eg. audio recorders) used for data collection.	Section 3.3.2; Appendix 14.
Units of study.	12	Number and relevant characteristics of participants included in the study; level of participation (could be reported in results).	Figure 3.1; Table 3.3.
Data processing.	13	Methods for processing data prior to and during the analysis, including transcription, data entry, data management and security, verification of data integrity, data coding, and anonymization/de-identification of excerpts.	Section 3.2.2.
Data analysis.	14	Process by which inferences, themes etc were identified and developed, including the researchers involved in data analysis; usually references a specific paradigm or approach.	Section 3.3.3.
Techniques to enhance trustworthiness.	15	Techniques to enhance trustworthiness and credibility of data analysis (eg. member checking, audit trail).	Section 3.3.3.

Topic	No.	Item	Evidence
Results/Findings			
Synthesis and interpretation.	16	Main findings (eg. interpretations, inferences, and themes); might include development of a theory of model.	Section 3.4.1; Section 3.4.2; Section 3.4.3.
Links to empirical data.	17	Evidence (eg. quotes, field notes, text excerpts) to substantiate analytic findings.	Section 3.4.2; Section 3.4.3; Appendix 15.
Discussion			
Integration with prior work, implications, transferability, and contributions to the field.	18	Short summary of main findings; explanation of how findings and conclusions connect to support, elaborate on, or challenge conclusions of earlier scholarship	Section 3.5.
Limitations.	19	Trustworthiness and limitations of findings.	Section 3.5; *Limitations.*
Other			
Conflicts of interest.	20	Potential sources of influence or perceived influence on study conduct and conclusions; how these were managed.	Not applicable.
Funding.	21	Sources of funding and other support; role of funders in data collection, interpretation, and reporting.	Not applicable for individual PhD chapter. Funders acknowledged at the start of the thesis.

Source: O'Brien et al. (2014, pp. 1247-1248).

A12. References

O'Brien, B., Harris, I., Beckman, T., Reed, D., & Cook, D. (2014). "Standards for Reporting Qualitative Research: A Synthesis of Recommendations", *Academic Medicine*, Vol. 89, 9, pp. 1245-1251.

Appendix 13: Recruitment Emails and Participant Information Sheet for Study in Chapter Three

This appendix reports the recruitment emails that were sent to the rheumatologists within the sampling frame of the qualitative study in *Chapter Three* (Section A13.1) and the participant information sheet that was attached to those emails (Section A13.2).

A13.1. Participant Recruitment Emails

There were two recruitment emails sent to the participants within the sampling frame of the study in *Chapter Three*. The first email (Figure A13.1) was sent in December 2014 to all individual rheumatologists. The second (follow-up) recruitment email was sent in March 2015 to the rheumatologists that did not respond in December 2014. The follow-up recruitment email was identical to the initial participant recruitment email. All emails were sent to each rheumatologist individually.

A13.2. Participant Information Sheet

A participant information sheet (Figure A13.2) was attached to all recruitment emails. The participant information sheet explained (i) the purpose of the research, (ii) the reason why the rheumatologist was contacted, (iii) details about their role as a participant, and (iv) information about the study design (for example, maintaining confidentiality, remuneration, and the dissemination of outputs). The participant information sheet followed the standardised template recommended by The University of Manchester's *Research Ethics Committee*.

Figure A13.1. *Participant recruitment email (December 2014 and March 2015).*

My name is Sean Gavan, and I am currently undertaking my PhD in Health Economics at the Manchester Centre for Health Economics, in The University of Manchester. I am looking for participants who are willing to be interviewed over the telephone, to help provide evidence for my forthcoming research project.

My thesis is focused on the use of a diagnostic test to help inform the prescribing process of anti-TNF treatments. For my next project, I am interested in establishing the reasons for why certain anti-TNF treatments are given to a patient in practice. In particular, I am interested in how you would treat a patient with RA after they exhibit secondary non-response to an anti-TNF. That is, a treatment that was initially beneficial for the patient, but then sees a reduction in effectiveness over time. Attached to this e-mail is a participant information sheet, which explains the aim of the research and what is expected of you as a participant.

Your participation is entirely voluntary, and anything that you say during the interview will remain fully anonymous. Ethical approval for this study has been granted by The University of Manchester Ethics Committee 2 (reference number: 14,147).

If you would like to take part, have any questions, or you would like some more information on the project, please contact me on my email address: sean.gavan@manchester.ac.uk

Kind Regards,

Sean Gavan, MSc, MSc, BA (Hons)
Manchester Centre for Health Economics
The University of Manchester
Room 4.306 Jene McFarlane Building
Oxford Road
Manchester
M13 9PL

Professor Anne Barton
Arthritis Research UK Epidemiology Unit
Centre for Musculoskeletal Research
Institute of Inflammation and Repair
The University of Manchester
Room 2.607, Stopford Building
Oxford Road
Manchester
M13 9PT

Figure A13.2. *Participant Information Sheet (page one).*

MANCHESTER 1824

The University of Manchester

MCHE MANCHESTER CENTRE FOR HEALTH ECONOMICS

A Qualitative Analysis of the Approaches Taken to Treat Patients with Rheumatoid Arthritis using Anti-TNF Therapies.

Participant Information Sheet

You are being invited to take part in a research study exploring the use of anti-TNF treatments in patients with rheumatoid arthritis. This research will contribute towards the PhD thesis of Mr Sean Gavan, at The University of Manchester. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

Who will conduct the research?
The research will be conducted by Mr Sean Gavan, who is based in the Manchester Centre for Health Economics at The University of Manchester. The research findings will contribute towards a chapter in his PhD thesis.

What is the aim of the research?
The aim of the research is to explore the current prescribing and treatment practices for the use of anti-TNF treatments in the management of patients with rheumatoid arthritis.

Why have I been chosen?
You have been chosen to take part in the research because of: (1) your association with the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate (BRAGGSS) register, and (2) your expertise and knowledge in the current prescribing and treatment practices for rheumatoid arthritis.

What would I be asked to do if I took part?
By agreeing to take part, you will be interviewed for approximately 30-45 minutes over the telephone at a time that is most convenient for you. You will be asked questions regarding how you would treat patients with severe rheumatoid arthritis, and your interpretation of the National Institute for Health and Care Excellence (NICE) guidelines for anti-TNF use in patients with rheumatoid arthritis.

What happens to the data collected?
The telephone interviews will be recorded using a digital audio recording device, and transcribed to analyse your responses. The audio recording will be deleted after being transcribed. The data collected will be stored securely for a period of 5 years at The University of Manchester.

How is confidentiality maintained?
All of the data that are collected, stored, and reported will be anonymised. Direct quotations from the interview may be used in the final reporting of results. However, it will not be possible to identify your name, the centre which you are associated with, or any other identifiable characteristics from these direct quotations. All data will be kept strictly confidential and stored in accordance with the University of Manchester's data protection regulations.

Figure A13.2. *Participant Information Sheet (page two).*

What happens if I do not want to take part or if I change my mind?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw without giving a reason and without detriment to yourself. It will not be possible to withdraw your data after it has been made anonymous.

Will I be paid for participating in the research?
You will receive no payment for taking part in this research.

What is the duration of the research?
The telephone interview will last for approximately 30 to 45 minutes. The date of the interview can be chosen by you, at a time that is most convenient for you.

Where will the research be conducted?
As this is a telephone interview, you are free to participate in any location of your choosing. It is recommended that the location you choose if quiet and free from distractions.

Will the outcomes of the research be published?
The outcomes of the research will contribute towards the PhD thesis of Sean Gavan. Findings will be submitted for publication in peer-reviewed journals, and for presentation at national conferences. All responses from your participation will remain strictly anonymous.

Who has reviewed the research project?
The University of Manchester Research Ethics Committee 2 (reference number, 14,147).

Contact for further information:

Lead Researcher Primary
Mr Sean Gavan
Manchester Centre for Health Economics
Room 4.306, Jean McFarlane Building
The University of Manchester
Oxford Road
M13 9PL
Email: sean.gavan@manchester.ac.uk

Supervisor
Prof. Katherine Payne
Manchester Centre for Health Economics
Room 4.319, Jean McFarlane Building
The University of Manchester
Oxford Road
M13 9PL
Email: katherine.payne@manchester.ac.uk

Who is funding this research?
This research is funded by a National Institute for Health Research (NIHR) Manchester Musculoskeletal Biomedical Research Unit Studentship, awarded to Sean Gavan.

What if something goes wrong?
If there are any issues regarding this research that you would prefer not to discuss with members of the research team, please contact the Research Governance and Integrity Team by either writing to 'The Research Governance and Integrity Manager, Research Office, Christie Building, The University of Manchester, Oxford Road, Manchester M13 9PL', by emailing: Research.complaints@manchester.ac.uk, or by telephoning 0161 275 7583 or 275 8093

Appendix 14: Chapter Three - Semi-structured Interview Schedule

This appendix presents the semi-structured interview schedule (Section A14.1) that was used to guide the telephone interviews with consultant rheumatologists in *Chapter Three*. The interview schedule began with an introductory statement and five preliminary questions that were asked to all rheumatologists. The remaining questions in the semi-structured interview schedule addressed the specific research objectives of *Chapter Three* and could have been asked in any order. All interviews were conducted by SG.

A14.1. Semi-structured Interview Schedule

i. Introduction Section:

"Hello, my name is Sean Gavan. I am a PhD student in the Manchester Centre for Health Economics, at The University of Manchester. Thank you for agreeing to take part in my research. I plan to talk to you for around 30 to 45 minutes, to get some information on how you would use anti-TNF treatments in practice for patients with rheumatoid arthritis. Before we begin, I must remind you that our conversation will be recorded, but any information that you give will remain fully anonymous. Your participation is voluntary, and you are free to end the session at any time without giving any reason. Feel free to ask any questions at any time, during or after the interview".

(1) *"Are you still happy to take part?"* **YES NO**

(Proceed to *Preliminary Questions Section* if YES).

---

ii. Preliminary Questions Section:

"Excellent; we will begin with just a couple of brief questions about you".

(1) "*Can you confirm your name?"*:
(2) *"What is your formal job title?";*
(3) *"Can you confirm the centre that you are based at?";*
(4) *"Approximately, how many patients with rheumatoid arthritis are treated at your centre?"*

(5) *"Can you confirm that you have had experience of prescribing anti-TNFs when treating patients with rheumatoid arthritis?".*

---

iii. Questions that Addressed the Research Objectives:

"Excellent, we can begin the interview".

Topic 1: Assessing the suitability of a patient for anti-TNF treatment:

(1) "Please explain how you would assess a patient in terms of whether they are suitable for anti-TNF treatment for RA?"

Topic 2: The decision of choosing the first anti-TNF treatment:

(1) "If a patient is determined to be suitable for anti-TNF treatment, which treatment is initially chosen for that patient?"
(2) "How and why is this decision made?"

Topic 3: The decision of choosing a treatment following the failure of an anti-TNF:

(1) "What do you understand by the terms 'secondary non-response' and 'adverse drug reactions' to anti-TNFs?"
(2) "If a patient exhibits secondary non-response or adverse drug reactions to an anti-TNF treatment, what would be the next suitable treatment strategy to try?"
(3) "If this is an anti-TNF treatment, which is most suitable for the patient?"
(4) "How and why is this decision made?"

Topic 4: The interpretation of NICE guidelines:

(1) "Are you aware of NICE guidelines? If so, which ones are you aware of?"
(2) "How is NICE guidance interpreted in your centre when prescribing anti-TNF treatments for patients with RA?"

Topic 5: How to ensure adherence to NICE guidelines:

(1) *"Are there any systems in place at your centre to ensure that NICE guidance is followed?"*

Topic 6: Beliefs about the five anti-TNF treatments for patients with RA:

(1) *"Can you rank the anti-TNF treatments approved by NICE according to an order in which you would consider them suitable for initial therapy?"*
(2) *"Can you justify this ranking by explaining why and how you chose this ordering?"*
(3) *"Can you rank the anti-TNF treatments approved by NICE according to an order in which you would consider them suitable for second-line therapy, following secondary non-response to a patient's first anti-TNF?"*
(4) *"Can you justify this ranking by explaining why and how you chose this ordering?"*

Topic 7: The effects of a test for immunogenicity in clinical practice:

(1) *"Would a test for immunogenicity to anti-TNF treatments be of any use in practice?"*
(2) *"Would the existence or results of an immunogenicity test change the initial anti-TNF treatment prescribed to patients, and the reasons for why this treatment is chosen?"*
(3) *"Would the existence or results of an immunogenicity test change the choice of second anti-TNF treatment, and the reasons for why this treatment is chosen?"*

---

iv. Closing of the Interview

"Thank you very much. That concludes the interview. Your responses have been very helpful. Thank you for your participation. If you have any further questions, feel free to contact either myself or my supervisor – our contact details can be found on the participant information sheet. Have a good day. Bye".

Appendix 15: Chapter Three - Reported Prescribing Decisions at Six Points in the Care Pathway for Rheumatoid Arthritis

This appendix reports a table (Table A15.1) of the six specific prescribing decisions that each consultant rheumatologist made along the care pathway for RA, in Section 3.4.1 of *Chapter Three*. The eleven participants were labelled alphabetically (from *A* to *K*). The six treatment decisions (illustrated in Figure 3.2) were:

Decision 1: The choice of first TNFi therapy;

Decision 2: The choice of therapy if a TNFi is unsuitable;

Decision 3: The action taken following an adverse event from a TNFi therapy;

Decision 4: The action taken following primary failure of a TNFi therapy;

Decision 5: The action taken following secondary failure of a TNFi therapy;

Decision 6: The action taken during remission, induced by a TNFi therapy.

Bold font was used to highlight the specific pharmacological treatments that each consultant rheumatologist recommended prescribing at each decision point.

Table A15.1. *Prescribing decisions reported by eleven consultant rheumatologists in England at six decision points along the care pathway for RA.*

Rheumatologist	Treatment Decision					
	Decision 1	**Decision 2**	**Decision 3**	**Decision 4**	**Decision 5**	**Decision 6**
Rheumatologist A	• Rheumatology meeting to decide most suitable TNFi based on patient characteristics; • Choose **etanercept** if prone to infection.	• **Rituximab** if patient has a serious infection risk or recent malignancy; • **Tocilizumab** if the patient can't have methotrexate.	• Change treatment to a different bDMARD (usually **rituximab**) after a serious adverse reaction; • Don't use a second TNFi.	• Change treatment to a different bDMARD (usually **rituximab**); • Don't use a second TNFi.	• Change treatment to a different bDMARD (usually **rituximab**); • Don't use a second TNFi.	• Not reported.
Rheumatologist B	• Patient can choose between **adalimumab** and **etanercept**.	• **Rituximab** if patient has pulmonary fibrosis; • **Tocilizumab** if patient can't tolerate methotrexate; • **Abatacept** if patient has infection risk.	• Change treatment to the other TNFi (**adalimumab** or **etanercept**, depending on *Decision 1*) for immediate problems, such as an injection site reaction.	• Change treatment to a different bDMARD; • Don't use a second TNFi.	• Change treatment to a different bDMARD; • Don't use a second TNFi; • **Rituximab** if the patient is seropositive; • No TNFi dose escalation.	• Reduce **methotrexate**; • Reduce the **frequency of TNFi injections**. • Patients may self-regulate their TNFi injections and compliance is unknown.

Abbreviations: TNFi=anti-tumour necrosis factor-α inhibitor; bDMARD=biologic disease-modifying anti-rheumatic drug.

Rheumatologist	Treatment Decision					
	Decision 1	**Decision 2**	**Decision 3**	**Decision 4**	**Decision 5**	**Decision 6**
Rheumatologist C	• Patient can choose between **all TNFi therapies**; • Choose **etanercept** if prone to infection.	• **Rituximab** if patient has interstitial lung disease; • **Tocilizumab** if patient has high inflammatory markers and can't receive methotrexate; • **Abatacept** or **rituximab** if patient has multiple sclerosis.	• Change to a second TNFi for injection site reactions. • **Etanercept** if patient received a monoclonal TNFi, and a non-monoclonal TNFi otherwise.	• Change treatment to a different bDMARD (**Rituximab**, **tocilizumab**, or sometimes **abatacept**); • Don't use a second TNFi;	• Change treatment to a different bDMARD; • Don't use a second TNFi; • **Rituximab** if the patient is seropositive; • **Tocilizumab** if patient has high inflammatory markers and can't receive methotrexate.	• Reduce **steroids** if applicable; • Reduce the **non-methotrexate cDMARDs** if receiving multiple cDMARDs; • Reduce/stop **methotrexate** if patient is having side-effects to methotrexate; • Plan to use **immunogenicity testing** in the future to inform reducing or stopping TNFi doses.
Rheumatologist D	• **Adalimumab**; • Patients can choose any TNFi if they have a strong preference.	• Choose a bDMARD (not a TNFi) if patient has a previous cancer, lung disease or can't receive methotrexate.	• Change to a second **TNFi** for local adverse reactions; • Change to a different bDMARD for general adverse reactions; • If seropositive, use the pathway: **rituximab, tocilizumab, abatacept**. Omit rituximab from this pathway if seronegative.	• Not reported.	• Not reported.	• Not reported.

Abbreviations: TNFi=anti-tumour necrosis factor-α inhibitor; bDMARD=biologic disease-modifying anti-rheumatic drug; cDMARD= conventional disease modifying anti-rheumatic drug.

Rheumatologist	Treatment Decision					
	Decision 1	Decision 2	Decision 3	Decision 4	Decision 5	Decision 6
Rheumatologist E	• **Certolizumab** is chosen for all patients; • No patient choice is facilitated.	• **Rituximab** if patient has chronic airways disease or infection risk. • Other bDMARDs (including rituximab) may be used if the patient has a history of tumours, melanoma or lymphomas.	• Continue TNFi if there's an injection site reaction; • Potentially consider TNFi dose-reduction; • Supplement painful injections with an anaesthetic. • Don't use a second TNFi for patients with severe, rare side-effects.	• Change treatment to a second TNFi if the patient agrees; • Participant has used **etanercept** after primary failure.	• Change treatment to a different bDMARD; • Don't use a second TNFi. • Don't escalate the dose of TNFi.	• Reduce frequency of TNFi injections.
Rheumatologist F	• Rheumatologist is encouraged to choose **certolizumab**; • Most patients begin with **etanercept, adalimumab** or **certolizumab**; • Patients can choose between a limited set of TNFi therapies.	• Choose **abatacept** or **tocilizumab**.	• Change treatment to a second TNFi for injection site reactions; • Participant explained that they would probably change to a different bDMARD (not a TNFi) if there's a serious injection.	• Not reported.	• Consider dose escalation for **infliximab**; • Potentially change treatment to a second TNFi if the patient did well on their initial TNFi; otherwise, change to a different bDMARD.	• Reduce the frequency of TNFi injections; • Patients can revert back to their original TNFi dose if they want.

Abbreviations: TNFi=anti-tumour necrosis factor-α inhibitor; bDMARD=biologic disease-modifying anti-rheumatic drug.

Rheumatologist	Treatment Decision					
	Decision 1	**Decision 2**	**Decision 3**	**Decision 4**	**Decision 5**	**Decision 6**
Rheumatologist G	• **Certolizumab** encouraged for most patients; • Choose **etanercept** for infection risk; • Choose i**nfliximab** for patients with compliance issues.	• Choose **rituximab** if the patient had a previous malignancy.	• Change treatment to a second TNFi if the patient responded well to their initial TNFi; • Change treatment to **rituximab** otherwise.	• Not reported.	• Change treatment to rituximab; • Don't escalate the dose of TNFi.	• Reduce the frequency of TNFi injections; • The patient has an input in whether the injection frequency is adjusted.
Rheumatologist H	• Patient can choose their TNFi therapy.	• Choose **rituximab** if the patient has severe disease or serology for Sjogren's syndrome.; • Choose **tocilizumab** if the patient can't tolerate methotrexate.	• Consider changing treatment to a second TNFi for injection site reactions.	• Not reported.	• Change treatment to a second TNFi; • Don't escalate the dose of TNFi.	• Not reported.

Abbreviations: TNFi=anti-tumour necrosis factor-α inhibitor; bDMARD=biologic disease-modifying anti-rheumatic drug.

Rheumatologist	Treatment Decision					
	Decision 1	Decision 2	Decision 3	Decision 4	Decision 5	Decision 6
Rheumatologist I	• Local health care commissioners have imposed the use of **etanercept**.	• Choose **rituximab** if the patient has pulmonary fibrosis or lung disease; • Choose **tocilizumab** if the patient can't tolerate methotrexate; • Choose **abatacept** if the patient has an infection risk or a family history of multiple sclerosis.	• Change treatment to either (i) a second TNFi or (ii) a different bDMARD, depending on the adverse event.	• Not reported.	• Change treatment to **rituximab**.	• Maintain full-dose TNFi therapy; • Reduce the dose of **methotrexate** if the patient has previously failed cDMARDs (as monotherapy or combination therapy).
Rheumatologist J	• Typically commence treatment with **adalimumab** or **etanercept**; • The choice of TNFi is made between the patient and a nurse.	• Most likely to choose **abatacept**; • Occasionally **rituximab** is used; • **Tocilizumab** may be used if the patient has systemically active disease.	• The treatment decision is made by the rheumatologist on a case-by-case basis.	• Not reported.	• Change treatment to a different bDMARD (not a TNFi) if the patient did not respond well to their initial TNFi; • Otherwise, change treatment to a second TNFi; • Don't escalate the dose of TNFi.	• Reduce the frequency of TNFi injections; • TNFi anti-drug antibody and drug level testing may be beneficial to inform dose-adjustments in remission.

Abbreviations: TNFi=anti-tumour necrosis factor-α inhibitor; bDMARD=biologic disease-modifying anti-rheumatic drug; cDMARD= conventional disease modifying anti-rheumatic drug.

Rheumatologist	Treatment Decision					
	Decision 1	**Decision 2**	**Decision 3**	**Decision 4**	**Decision 5**	**Decision 6**
Rheumatologist K	• Choose **certolizumab pegol** as the first-line TNFi therapy.	• Choose **rituximab** if the patient has had a recent malignancy.	• Change treatment to **rituximab** if the patient didn't respond well to their initial TNFi; • Otherwise, change treatment to a second TNFi.	• Not reported.	• Change treatment to **rituximab** for most patients; • Change treatment to a different bDMARD if the patient would like to enrol into a clinical research study.	• Reduce the frequency of TNFi injections after the patient has been in remission for twelve months.

Abbreviations: TNFi=anti-tumour necrosis factor-α inhibitor; bDMARD=biologic disease-modifying anti-rheumatic drug.

Appendix 16: Chapter Four - Previous Quantitative Studies of the Patient-level Factors that Influenced TNFi Prescribing Decisions for Rheumatoid Arthritis

This appendix describes the features and potential limitations of three previous quantitative studies that had estimated the patient-level factors that influenced TNFi prescribing decisions for patients with rheumatoid arthritis in North America (Carter et al., 2012; DeWitt et al., 2006; Zhang et al., 2013). Table A16.1 summarises the key features and potential limitations of each study's design.

Carter et al. (2012) obtained data from 1,696 patients with RA, who received treatment between 2000 and 2006, in a national commercial database of over eighty private health plans. The results of a univariate analysis concluded that infliximab-treated patients (i) were significantly older than patients treated with adalimumab or etanercept, and (ii) had higher staging of RA compared to etanercept-treated patients. However, the reliability of these results may be debated; the univariate analysis may have been confounded by unmeasured variables such as previous cDMARD use (Hudson et al., 2010). Moreover, Carter et al. (2012) did not control for any temporal effects or hospital-level heterogeneity in treatment decisions, which were identified as potentially influential factors on prescribing decisions in *Chapter Three*.

DeWitt et al. (2006) estimated the patient-level factors that influenced the prescription of etanercept and infliximab, between 2000 and 2003, in a cohort of 1,663 patients with RA in the North American *National Databank for Rheumatic Diseases*. Multivariable analyses of disease, patient-level and geographic characteristics found infliximab prescribing to be significantly associated with older age, lower education, and lower physical activity (DeWitt et al., 2006). The multivariable analysis may have addressed potential confounding in the results, however DeWitt et al. (2006) also did not control for temporal effects or hospital-level heterogeneity. In addition, DeWitt et al. (2006) excluded patients with missing data from their analysis (known as a *complete case analysis*), which may have introduced a selection bias in their results if those patients with missing data were systematically different from those patients with complete data (Rezvan et al., 2015).

Table A16.1. *Features of three previous studies of TNFi prescribing decisions.*

Author	Study Design	Variables	Results	Limitations
DeWitt et al. (2006). *Country*: USA.	*Sample size:* 1,663 patients. *Data source*: National Databank for Rheumatic Diseases *Years analysed*: 2000 to 2003 *Method*: Multivariable analysis.	*Dependent variable*: TNFi (etanercept or infliximab). *Independent variables*: Public insurance, age, sex, ethnicity, marital status, income, education, employment status, disease duration, HAQ, treatments, pain, disease severity, SF-36 physical and mental component.	*Influences*: - infliximab relative to etanercept: concomitant methotrexate; older age; public insurance.	*Limitations*: • Only included patients with full data available; • Only analysed two TNFi therapies; • No time element included in analyses; • Patients were distributed across 413 practices but the analysis did not control for hospital-level heterogeneity.
Carter et al. (2012). *Country*: USA.	*Sample size:* 1,696 patients. *Data source*: Database of over eighty private health plans. *Years analysed*: 2000 to 2006. *Method*: Univariate T-test.	*Dependent variable*: TNFi (adalimumab, etanercept or infliximab). *Independent variables*: Age, sex, comorbidities, staging of disease severity.	*Influences*: - infliximab relative to etanercept or adalimumab: older age. - infliximab relative to etanercept only: a higher staging of RA.	*Limitations*: • Only analysed three TNFi therapies; • Univariate analysis may be subject to confounding; • No time variable included in the analysis; • Did not control for hospital-level heterogeneity.
Zhang et al. (2013). *Country*: USA.	*Sample size:* 11,966 patients. *Data source*: All patients covered by Medicare. *Years analysed*: 2006 to 2009. *Method*: Logistic regression.	*Dependent variable*: TNFi (adalimumab, etanercept or infliximab). *Independent variables*: Age, sex, ethnicity, urban residence, income, previous treatments, physician preference for infusions, year.	*Influences*: - infliximab relative to subcutaneous TNFi: older age; concomitant methotrexate; physician preference for infusions. - subcutaneous TNFi relative to infliximab: low income.	*Limitations*: • Only analysed three TNFi therapies; • Did not control for hospital-level heterogeneity.

Zhang et al. (2013) used multivariable logistic regression to analyse all patients with RA covered by Medicare in North America who commenced adalimumab, etanercept, or

infliximab between 2006 and 2009. Like Carter et al. (2012), the authors found that older patients were more likely to be prescribed infliximab compared with etanercept or adalimumab. Low income patients (measured by proxy according to receipt of Government assistance for Medicare Part B premiums) were more likely to receive a subcutaneous TNFi, and physician preference for intravenous therapies was found to positively influence the likelihood of treatment with infliximab (Zhang et al., 2013). These results may be more reliable than Carter et al. (2012) and DeWitt et al. (2006) because independent variables to control for differences over time were included in the analysis. However, Zhang et al. (2013) did not control for hospital-level heterogeneity, which may have introduced omitted variable bias in the results (Wooldridge, 2010) if unobservable hospital-level factors also influenced prescribing decisions, as described by the qualitative study in *Chapter Three*.

A16. References

Carter, C., Changolkar, A., & McKenzie, R. (2012). "Adalimumab, Etanercept, and Infliximab Utilization Patterns and Drug Costs among Rheumatoid Arthritis Patients", *Journal of Medical Economics*, Vol. 15, 2, pp. 332-339.

DeWitt, E., Glick, H., Albert, D., Joffe, M., & Wolfe, F. (2006). "Medicare Coverage of Tumor Necrosis Factor Alpha Inhibitors as an Influence on Physicians' Prescribing Behavior", *Archives of Internal Medicine*, Vol. 166, 1, pp. 57-63.

Hudson, M., & Suissa, S. (2010). "Avoiding Common Pitfalls in the Analysis of Observational Studies of New Treatments for Rheumatoid Arthritis", *Arthritis Care & Research*, Vol. 62, 6, pp. 805-810.

Rezvan, P., Lee, K., & Simpson, J. (2015). "The Rise of Multiple Imputation: A Review of the Reporting and Implementation of the Method in Medical Research", *BMC Medical Research Methodology*, Vol. 15, 30, pp. 1-14.

Wooldridge, J. (2010). "Econometric Analysis of Cross Section and Panel Data". (2 ed.) London: The MIT Press.

Zhang, J., Xie, F., Delzell, E., Chen, L., Kilgore, M., Yun, H., Saag, K., Lewis, J., & Curtis, J. (2013). "Trends in the use of Biologic Agents among Rheumatoid Arthritis Patients Enrolled in the US Medicare Program", *Arthritis Care & Research*, Vol. 65, 11, pp. 1743-1751.

Appendix 17: Chapter Four - Complete STROBE Checklist

This appendix provides the completed *STROBE* checklist for the econometric study presented in *Chapter Four* (Table A17.1). The checklist ensured that the quantitative study was reported in accordance with best-practice recommendations for empirical analyses that utilised observational data (Vandenbroucke et al., 2007).

Table A17.1. *Complete STROBE checklist for quantitative analysis of observational data in Chapter Four.*

Section/Topic	Item No.	Recommendation	Evidence
Title			
Title and abstract.	1	(a) Indicate the study's design with a commonly used term in the title or abstract; (b) Provide in the abstract an informative and balanced summary of what was done and what was found.	Not applicable for a PhD thesis.
Introduction			
Background/rationale.	2	Explain the scientific background and rationale for the investigation being reported.	Section 4.1.
Objectives.	3	State specific objectives, including any pre-specified hypotheses.	Section 4.2.
Methods			
Study design.	4	Present key elements of the study design early in the paper.	Section 4.3.
Setting.	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.	Section 4.3.2.
Participants.	6	(a) Cohort study – Give the eligibility criteria, and the sources and methods of selection of participants.	Section 4.3.2.1.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers.	Section 4.3.2.2; Section 4.3.2.3; Section 4.3.3.2.

Section/Topic	Item No.	Recommendation	Evidence
Data sources/measurement.	8	For each variable of interest, give sources of data and details of methods of assessment (measurement).	Section 4.3.2; Section 4.3.2.3.
Bias.	9	Describe any efforts to address potential sources of bias.	Section 4.3.2.4; Section 4.3.3.2.
Study size.	10	Explain how the study size was arrived at.	Section 4.3.2.
Quantitative variables.	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why.	Section 4.3.2.2; Section 4.3.2.3.
Statistical methods.	12	(a) Describe all statistical methods, including those to control for confounding; (b) Describe any methods used to examine subgroups and interactions; (c) Explain how missing data were addressed; (d) Cross-sectional study – if applicable, describe analytical methods taking account of sampling strategy; (e) Describe any sensitivity analyses.	Section 4.3.2.4; Section 4.3.3.1; Section 4.3.3.2; Section 4.3.3.3; Section 4.3.3.4; Section 4.3.3.5.
Results			
Participants.	13	(a) Report the numbers of individuals at each stage of the study; (b) Give reasons for non-participation at each stage; (c) Consider use of a flow diagram.	Figure 4.2.
Descriptive data.	14	(a) Give characteristics of study participants and information on exposures and potential confounders; (b) Indicate the number of participants with missing data for each variable of interest.	Section 4.4.1; Figure 4.3; Figure 4.4; Table 4.6; Appendix 20.
Outcome data.	15	Report numbers of outcome events or summary measures.	Table 4.5.

Section/Topic	Item No.	Recommendation	Evidence
Main results.	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision. Make clear which confounders were adjusted for and why they were included; (b) Report category boundaries when continuous variables were categorised; (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.	Section 4.4.2.
Other analyses.	17	Report other analyses – eg. analyses of subgroups, interactions, and sensitivity analyses.	Section 4.3.3; Appendix 21.
Discussion			
Key results.	18	Summarise key results with reference to study objectives.	Section 4.5.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision.	Section 4.5; *Limitations*.
Interpretation	20	Give a cautious interpretation of the results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	Section 4.5.
Generalisability.	21	Discuss the generalisability of the results.	Section 4.5.
Other Information			
Funding.	22	Give the source of funding and the role of funders for the present study and, if applicable, for the original study on which the present article is based.	Not applicable for individual PhD chapter. Funders acknowledged at the start of the thesis.

Source: Vandenbroucke et al. (2007, p. 1630).

A17. References

Vandenbroucke, J., van Elm, E., Altman, D., Gøtzsche, P., Mulrow, C., Pocock, S., Poole, C., Schlesseman, J., & Egger, M. (2007). "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboation", *PLoS Medicine*, Vol. 4, 10(e297), pp. 1628-1654.

Appendix 18 – Chapter Four - List of Hospitals that Contributed Data to BRAGGSS

This appendix provides a list of the fifty-seven hospitals, across the UK, which had contributed patient-level data to the BRAGGSS cohort (see Table A18.1). These hospitals were referred to as *Contributing Centres* within BRAGGSS.

Table A18.1. *Fifty-seven BRAGGSS Contributing Centres.*

Hospital ID	Trust	Hospital
1	Cambridge University Hospitals NHS Foundation Trust	Addenbrookes Hospital
2	Mid Staffordshire General Hospitals NHS Trust	Cannock Chase Hospital
3	The Leeds Teaching Hospitals NHS Trust	Chapel Allerton Hospital
4	Derby Hospitals NHS Foundation Trust	Derbyshire Royal Infirmary
5	Doncaster And Bassetlaw Hospitals NHS Foundation Trust	Doncaster Royal Infirmary
6	Peterborough and Stamford Hospitals NHS Foundation Trust	Edith Cavell Hospital
7	The Newcastle upon Tyne Hospitals NHS Trust	Freeman Hospital
8	University Hospital of North Staffordshire NHS Trust	Haywood Hospital
9	Hereford Hospitals NHS Trust	Hereford County Hospital
10	Norfolk & Norwich University Hospital NHS Trust	Norfolk & Norwich University Hospital
11	Pennine Acute Hospitals NHS Trust	North Manchester General Hospital
12	Portsmouth Hospitals NHS Trust	Queen Alexander Hospital
13	Gateshead Health NHS Trust	Queen Elizabeth Hospital
14	Sheffield Teaching Hospitals NHS Trust	Royal Hallamshire Hospital
15	University Hospital of Morcambe Bay	Royal Lancaster Infirmary
16	Sandwell and West Birmingham Hospital NHS Trust	Sandwell General/City Hospital
17	University Hospital Birmingham NHS Foundation Trust	Selly Oak Hospital
18	St Helens and Knowsley Hospital NHS Trust	St Helens Hospital

Hospital ID	Trust	Hospital
19	South Tees Hospitals NHS Trust	James Cook University Hospital
20	County Durham and Darlington Acute Hospitals NHS Trust	University Hospital of North Durham
21	Whipps Cross University Hospital NHS Trust	Whipps Cross University Hospital
22	The West Suffolk Hosiptal NHS Trust	West Suffolk Hospital
23	Southampton University Hospital NHS Trust	Southampton General Hospital
24	Basingstoke & North Hampshire NHS Foundation Trust	Basingstoke & North Hampshire Hospital
25	Queen Mary's Sidcup NHS Trust	Queen Mary’s Sidqcup Hospital
26	Pennine Acute Hospitals NHS Trust	Royal Oldham Hospital
27	Pennine Acute Hospitals NHS Trust	Rochdale Infirmary
28	University Hospitals of Morecambe Bay NHS Trust	Furness Hospital
29	Central Manchester University Hospital NHS Foundation Trust	Manchester Royal Infirmary
30	Worcestershire Acute Hospitals NHS Trust	Worcestershire Royal Hospital
31	The Dudley Group of Hospitals NHS Foundation Trust	Russells Hall Hospital
32	Northumbria Healthcare NHS Foundation Trust	Wansbeck Hospital
33	University Hospitals of Coventry and Warwickshire NHS Trust	University Hospital Coventry
34	Wrightington, Wigan and Leigh Hospitals NHS Foundation Trust	Wrightington Hospital
35	Nottingham University Hospitals NHS Trust	Nottingham Hospital
36	Salford Royal NHS Foundation Trust	Hope Hospital
37	South Warwickshire General Hospital NHS Trust	Warwick Hospital.
38	Weston Area Health NHS Trust	Weston General Hospital.
39	The Royal Bournemouth & Christchurch Hospitals NHS Foundation Trust	Christchurch Hospital.
40	Northern Lincolnshire and Goole Hospitals NHS Foundation Trust	Diana, Princess of Wales Hospital.

Hospital ID	Trust	Hospital
41	York Hospitals NHS Foundation Trust	York District Hospital.
42	University Hospitals of Leicester NHS Trust	Leicester Royal Infirmary.
43	The Royal Wolverhampton Hospitals NHS Trust	New Cross Hospital.
44	Green Park Healthcare NHS Trust	Musgrave Park Hospital.
45	Chesterfield Royal Hospital NHS Foundation Trust	Chesterfield Royal Hospital.
46	Trafford Healthcare NHS Trust	Trafford General Hospital.
47	Harrogate and District NHS Foundation Trust	Harrogate District Hospital.
48	Royal National Hospital for Rheumatic Diseases NHS Foundation Trust	Bath Hospital.
49	Oxford Radcliffe Hospitals NHS Trust	John Radcliffe Hospital.
50	Milton Keynes Hospital NHS Foundation Trust	Milton Keynes Hospital.
51	Royal Liverpool and Broadgreen University Hospitals NHS Trust	Royal Liverpool Hospital.
52	Countess of Chester Hospital NHS Foundation Trust	Countess of Chester Hospital.
53	Royal Cornwall Hospitals NHS Trust	Royal Cornwall Hospital.
54	City Hospitals Sunderland NHS Foundation Trust	Royal Sunderland Hospital.
55	Stockport NHS Foundation Trust	Stepping Hill Hospital.
56	Trafford Healthcare NHS Trust	Trafford General Hospital.
57	Kettering General Hospital NHS Foundation Trust	Kettering General Hospital.

Appendix 19: Chapter Four - Likelihood Ratio Test to Collapse Categorical Dependent Variable

This appendix presents the rationale for collapsing the dependent variable (from five to three categories) for the quantitative study reported in *Chapter Four*.

A19.1. Introduction

The distribution of the five TNFi therapies prescribed across the sample of 894 patients with RA in *Chapter Four* is reported in Table A19.1a. The relatively few patients that received infliximab and golimumab may have led to convergence problems during statistical estimation if the dependent variable was defined by these five categories. The efficiency of a regression analysis may be improved by collapsing the categories of an unordered categorical variable that are indistinguishable with respect to the independent variables (Long et al., 2012). Collapsing a categorical dependent variable, however, should be done to maintain clinical and statistical plausibility. The proposed collapsed dependent variable ($TNFi_{Prescribed}$) comprised three mutually exclusive categories (Table A19.1b); patients that were prescribed infliximab or adalimumab were categorised as having received an *older monoclonal antibody* and patients that were prescribed certolizumab pegol or golimumab were categorised as having received a *newer monoclonal antibody*. Etanercept was categorised as the only *non-monoclonal antibody*. These three mutually exclusive categories were clinically plausible (van Vollenhoven, 2009).

Table A19.1. *Distribution of (a) five TNFi therapies prescribed across the sample and (b) the proposed three-category dependent variable ($TNFi_{Prescribed}$).*

(a) Five TNFi Therapies		(b) Proposed Three-category Dependent Variable	
TNFi	**n**	***TNFi*$_{Prescribed}$**	**n**
Etanercept	357	Non-monoclonal antibody Etanercept	357
Infliximab	34	Older Monoclonal Antibodies (infliximab and adalimumab)	373
Adalimumab	339	Newer Monoclonal Antibodies (certolizumab pegol and golimumab)	164
Certolizumab pegol	123		
Golimumab	41		
Total	894	Total	894

A19.2. Aim and Objective

The aim of this study was to determine the statistical plausibility of collapsing the five-category variable in Table A19.1a into the three-category TNFi variable in Table A19.1b.

A19.3. Method

Statistical plausibility for creating the dependent variable with three categories was informed by a likelihood ratio test (Wooldridge, 2010; Long et al., 2012). The test statistic for a categorical dependent variable was based on the log-likelihood of a full (LL^{Full}) and restricted ($LL^{Restricted}$) multinomial logistic regression. The dependent variable of this regression was the five-category TNFi variable (Table A19.1a); the independent variables of the regression were the variables that represented the patient's health status (see Section 4.3.2.3).

Written in a general case, the log-likelihood of the full model (LL^{Full}) was from a regression that included only the constant term. The log-likelihood of the restricted model ($LL^{Restricted}$), when collapsing categories *i* and *j*, was from a regression that used category *i* as the base-category and the coefficients of all other independent variables were constrained to zero except for the constant term for category *j*. The likelihood ratio test statistic (LR), defined by Equation A19.1 was chi-squared distributed with degrees of freedom (*k*) equal to the number of independent variables (Long et al. 2012).

$$LR = LL^{Full} - LL^{Restricted} \sim X_k^2 \qquad \textbf{(Equation A19.1)}$$

Under the null hypothesis, all coefficients (aside from the intercepts) for alternatives *i* and *j* were equal to zero and could be collapsed. A statistically significant test statistic was sufficient to reject the null hypothesis.

A19.4. Results

The likelihood ratio test statistics for collapsing the older monoclonal antibodies (infliximab and adalimumab) and newer monoclonal antibodies (certolizumab pegol and golimumab) into distinct categories, calculated using baseline clinical health variables, are reported in Table A19.2. The null hypothesis could not be rejected for either test statistic under conventional levels of statistical significance (in both cases, the p-value was greater

than 0.1). Therefore, the three category dependent variable reported in Table A19.1b was considered to be statistically plausible.

Table A19.2. *Likelihood ratio test for collapsing dependent variable categories.*

Collapsed Categorical Variables	$\mathbf{LR \sim X_k^2}$	P-value
Infliximab and adalimumab	11.835	0.106
Certolizumab pegol and golimumab	8.475	0.293

A19.5. Conclusion

The results of this study indicated that it was statistically plausible to collapse the dependent variable into three categories ((i) *Non-monoclonal antibody*; (ii) *Older monoclonal antibodies*; and (iii) *Newer monoclonal antibodies*). This three-category dependent variable was therefore used within the quantitative study reported in *Chapter Four*.

A19. References

Long, J., & Freese, J. (2012). "Regression Models for Categorical Dependent Variables using Stata". (3 ed.) Texas: Stata Press.

van Vollenhoven, R. (2009). “Treatment of Rheumatoid Arthritis: State of the Art 2009”, *Nature Reviews Rheumatology*, Vol. 5, 10, pp. 531-541.

Wooldridge, J. (2010). "Econometric Analysis of Cross Section and Panel Data". (2 ed.) London: The MIT Press.

Appendix 20: Chapter Four - Multiple Imputation of Missing Data

This appendix documents the methods of multiple imputation used to handle missing data in *Chapter Four*. The appendix describes the data that were missing in *Chapter Four* (Section A20.1), the three stages of handling missing data by multiple imputation (Section A20.2), multiple imputation by chained equations (Section A20.3), and the number of imputations used in *Chapter Four* (Section A20.4).

A20.1. Missing Data in Chapter Four

Missing data can be characterised as (i) *missing completely at random* (MCAR), (ii) *missing at random* (MAR), or (iii) *missing not at random* (MNAR). The definitions of these three categories are reported in Table A21.1.

Table A20.1. *Definitions to categorise missing data.*

Type of Missing Data	Acronym	Definition
Missing completely at random.	MCAR	No systematic difference between observed and missing values.
Missing at random.	MAR	Differences between the observed and missing values are explained, conditional on the observed data.
Missing not at random.	MNAR	Systematic differences between observed and missing values remain, after conditioning on observed values.

Source: Sterne et al. (2009).

Multiple imputation methods require the assumption that missing data are MAR, in order to use the *observed* data to predict values for the missing observations (see Section A20.2) (White et al., 2011; Sterne et al., 2009). The econometric study reported in *Chapter Four* used patient-level data from the *BRAGGSS* cohort. There were seven independent variables within the analysis that had missing data (reported in Table A20.2); these missing data were assumed to be MAR.

Table A20.2. *Variables with missing data in the analysis presented in Chapter Four.*

Variable	Missing (n)
Work status	309
DAS8	283
MTX	158
BMIover	156
YearswithRA	13
HAQ	12
Woman	2

A20.2. The Three Stages of Multiple Imputation

The general method of multiple imputation comprises three stages:

- **Stage one**: Generate *m* datasets, each of which have replaced the missing data for specific variables by sampling from their predicted distribution (conditional on the observed data);
- **Stage two**: Perform a separate statistical analysis (for example, a regression) on each of the *m* datasets;
- **Stage three**: Combine the *m* parameter estimates and standard errors using *Rubin's rules* (Sterne et al., 2009).

The process of sampling a different value for the missing data *m* times, in *Stage one*, appropriately represents the uncertainty in the true values of the missing observations (Sterne et al., 2009). The statistical analyses performed on each imputed dataset, in *Stage two*, will therefore likely produce different outcomes and variance-covariance matrices due to differences in the imputed missing values (White et al., 2011). The use of Rubin's rules to combining the separate analyses and produce the final parameter estimate, in *Stage three*, accounts for (i) the within-imputation uncertainty (the uncertainty of the results within one imputed dataset) and (ii) the between-imputation uncertainty (the uncertainty of the results across *m* datasets) (Rubin, 1996; White et al., 2011). The formulae (Rubin, 1996) to calculate a final parameter estimate $(\widehat{\theta})$ and variance $(VAR\hat{\theta})$ over *m* imputed datasets using Rubin's rules (which are based on the average over imputed datasets) are described in Equation A20.1 and Equation A20.2, respectively.

$$\hat{\theta} = \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_j$$ **(Equation A20.1)**

$$VAR\hat{\theta} = \left[\frac{1}{m}\sum_{j=1}^{m}VAR\hat{\theta_j}\right] + \left[\left(1+\frac{1}{m}\right) * \frac{1}{m-1}\sum_{j=1}^{m}\left(\hat{\theta_j} - \hat{\theta}\right)\right]$$ **(Equation A20.2)**

where *j* is used to index an individual dataset within *m*.

A20.3. Multiple Imputation by Chained Equations

The two predominant approaches for generating the *m* datasets in *Stage one* of multiple imputation, when data are missing from several variables, are (i) *multiple imputation by chained equations* (MICE) and (ii) *multivariate normal imputation* (Rezvan et al., 2015). The study in *Chapter Four* used MICE to impute missing data because there was some evidence to indicate its relative superiority at producing unbiased results (Romaniuk et al., 2014) and given the available commands to perform the analysis in *STATA* (Royson et al., 2011).

MICE was used to produce *m* imputed datasets by initially sampling the missing values from the observed data with replacement (Royston, 2009; Royston et al., 2011). The first variable with missing data, for example X_1, was regressed on all other variables that observed X_1 and the dependent variable (Moons et al., 2006). The type of regression depended on the properties of the variable with missing data (Romaniuk et al., 2014). For example, logistic regression was used if the variable was dichotomous (for example, *MTX*, *BMIover*, *Woman*) and multinomial logistic regression was used if the variables were unordered categorical data (for example, *Work status*). The remaining variables (*DAS28*, *YearswithRA*, *HAQ*) used ordinary least squares regression. The missing values were then imputed by repeatedly sampling from the predicted value of the regression to generate one imputed dataset. The whole process was repeated *m* times to produce *m* datasets (White et al., 2011). Variables that were not normally distributed (*HAQ*, *YearswithRA*) were transformed to a normal distribution before imputation and inverted to the original scale following imputation.

A20.4. Number of Imputations in Chapter Four

The appropriate number of imputations (*m*) was determined using the rule that *m* should be at least 100 times the *largest fraction of missing information* (FMI) test statistic (White et al., 2011). The base-case regression in *Chapter Four* had an FMI test statistic of 0.356. Therefore, sixty imputations was deemed to be sufficient for the analysis.

A20. References

Lee, K., & Carlin, J. (2010). "Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation", *American Journal of Epidemiology*, Vol. 171, 5, pp. 624-632.

Moons, K., Donders, R., Stijnen, T., & Harrell Jr, F. (2006). "Using the Outcome for Imputation of Missing Predictor Values was Preferred", *Journal of Clinical Epidemiology*, Vol. 59, 10, pp. 1092-1101.

Rezvan, P., Lee, K., & Simpson, J. (2015). "The Rise of Multiple Imputation: A Review of the Reporting and Implementation of the Method in Medical Research", *BMC Medical Research Methodology*, Vol. 15, 30, pp. 1-14.

Romaniuk, H., Patton, G., & Carlin, J. (2014). "Multiple Imputation in a Longitudinal Cohort Study: a Case Study of Sensitivity to Imputation Methods", *American Journal of Epidemiology*, Vol. 180, 9, pp. 920-932.

Royston, P. (2009). "Multiple Imputation of Missing Values: Further Update of ice, with an Emphasis on Categorical Variables", *The Stata Journal*, Vol. 9, 3, pp. 466-477.

Royston, P., & White, I. (2011). "Multiple Imputation by Chained Equations (MICE): Implementation in STATA", *Journal of Statistical Software*, Vol. 45, 4, pp. 1-20.

Rubin, D. (1996). "Multiple Imputation after 18+ Years", *Journal of the American Statistical Association*, Vol. 91, 434, pp. 473-489.

Sterne, J., White, I., Carlin, J., Spratt, M., Royston, P., Kenward, M., Wood, A., & Carpenter, J. (2009). "Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls", *British Medical Journal*, Vol. 338, b2393,

White, I., Royston, P., & Wood, A. (2011). "Multiple Imputation using Chained Equations: Issues and Guidance for Practice", *Statistics in Medicine*, Vol. 30, 4, pp. 377-399.

Appendix 21: Chapter Four - Sensitivity Analysis of the Base-case Result

This appendix presents the results of the sensitivity analysis described in Section 4.3.3.3 of *Chapter Four* when the regressions were re-estimated on a restricted sample of hospitals that had at least ten patients.

A21.1. Sensitivity Analysis: Exclude Hospitals with Fewer than Ten Observations

Table A21.1 reports the mean partial effects from the multinomial logistic regression in *Chapter Four* when hospitals with fewer than ten patient-level observations were omitted from the sample. This structural sensitivity analysis was performed to assess whether the base-case results were sensitive to the number of patients clustered within each hospital.

Regression D was still the preferred model specification (relatively higher pseudo-R^2, and relatively lower AIC and BIC test statistics. The base-case results (which identified that a patient's age, concomitant methotrexate use, and marital status had an influence on the likelihood of being prescribed etanercept, adalimumab, or infliximab) were robust to performing the analysis on the restricted sample, in terms of the partial effects' statistical significance, relative magnitude, and direction of influence. The results of this sensitivity analysis (i) confirmed the importance of including hospital-level dummy variables to control for unobservable between-hospital heterogeneity, and (ii) suggested that the base-case results were not driven by hospitals that had included fewer than ten patients to the BRAGGSS cohort (which may have been systematically *different* compared with the other hospitals that remained in the sample).

Table A21.1. *Sensitivity analysis: mean partial effects from multinomial logistic regression when hospitals with fewer than ten observations were omitted.*

	Regression A			Regression B			Regression C			Regression D		
	Non-mAb	**Older mAbs**	**Newer mAbs**	**Non-mAb**	**Older mAbs**	**Newer mAbs**	**Non-mAb**	**Older mAbs**	**Newer mAbs**	**Non-mAb**	**Older mAbs**	**Newer mAbs**
DAS28	0.0234	-0.0193	-0.0041	0.0467	-0.0459	-0.0008	0.0250	-0.0212	-0.0038	0.0477	-0.0471	-0.0006
	(0.0313)	(0.0325)	(0.0206)	(0.0391)	(0.0390)	(0.0011)	(0.0317)	(0.0333)	(0.0201)	(0.0397)	(0.0397)	(0.0011)
Woman	0.0442	0.0011	-0.0453	0.0435	-0.0406	-0.0029	0.0454	0.0055	-0.0509	0.0426	-0.0397	-0.0029
	(0.0396)	(0.0426)	(0.0357)	(0.0465)	(0.0464)	(0.0020)	(0.0426)	(0.0442)	(0.0370)	(0.0502)	(0.0500)	(0.0020)
Age/10	0.0094	-0.0198	0.0104	**0.0225***	**-0.0227***	0.0002	**0.0300***	**-0.0280***	-0.0020	**0.0416****	**-0.0415****	-0.0001
	(0.0127)	(0.0137)	(0.0119)	(0.0132)	(0.0132)	(0.0005)	(0.0154)	(0.0169)	(0.0114)	(0.0186)	(0.0186)	(0.0005)
HAQ	-0.0323	0.0062	0.0262	-0.0014	-0.0003	0.0017	-0.0221	0.0105	0.0115	0.0068	-0.0077	0.0009
	(0.0301)	(0.0333)	(0.0190)	(0.0383)	(0.0385)	(0.0011)	(0.0297)	(0.0332)	(0.0203)	(0.0372)	(0.0373)	(0.0011)
Totaldrug	0.0175	0.0105	-0.0280	-0.0024	0.0024	-0.0001	0.0164	0.0110	-0.0274	-0.0047	0.0046	0.0000
	(0.0215)	(0.0175)	(0.0199)	(0.0192)	(0.0190)	(0.0006)	(0.0213)	(0.0172)	(0.0195)	(0.0192)	(0.0190)	(0.0005)
Totalcomorb	0.0178	-0.0147	-0.0032	0.0264	-0.0264	0.0000	0.0188	-0.0129	-0.0059	0.0247	-0.0246	-0.0002
	(0.0159)	(0.0164)	(0.0095)	(0.0199)	(0.0199)	(0.0005)	(0.0168)	(0.0168)	(0.0091)	(0.0209)	(0.0209)	(0.0004)
YearswithRA	-0.0010	0.0026	-0.0016	-0.0019	0.0020	**-0.0001*****	-0.0012	**0.0030***	-0.0017	-0.0022	0.0024	**-0.0002*****
	(0.0020)	(0.0017)	(0.0014)	(0.0022)	(0.0022)	(0.0001)	(0.0021)	(0.0018)	(0.0014)	(0.0023)	(0.0023)	(0.0001)
BMIover	0.0344	-0.0172	-0.0172	0.0212	-0.0212	-0.0001	0.0317	-0.0175	-0.0142	0.0179	-0.0182	0.0003
	(0.0484)	(0.0490)	(0.0285)	(0.0602)	(0.0601)	(0.0015)	(0.0481)	(0.0500)	(0.0285)	(0.0602)	(0.0601)	(0.0014)
MTX	**-0.1079*****	0.0574	0.0506	**-0.0909***	**0.0880***	0.0029	**-0.1101*****	0.0603	**0.0497***	**-0.1012****	**0.0982***	**0.0030****
	(0.0404)	(0.0422)	(0.0294)	(0.0501)	(0.0501)	(0.0013)	(0.0424)	(0.0432)	**(**0.0292)	(0.0510)	(0.0510)	(0.0013)
Smoke							-0.0396	0.0310	0.0086	-0.0538	0.0512	**0.0026***
							(0.0463)	(0.0483)	(0.0243)	(0.0575)	(0.0576)	(0.0014)
Work1							0.0308	0.0303	-0.0611	0.0199	-0.0169	**-0.0031****
							(0.0597)	(0.0544)	(0.0334)	(0.0718)	(0.0717)	(0.0015)
Work3							-0.0259	0.0483	-0.0225	-0.0399	0.0416	-0.0017
							(0.0664)	(0.0673)	(0.0414)	(0.0852)	(0.0851)	(0.0016)
Marital1							**-0.1969*****	**0.1231****	**0.0738***	**-0.1997*****	**0.1980*****	0.0017
							(0.0666)	(0.0547)	(0.0378)	(0.0734)	(0.0729)	(0.0016)
Marital3							**-0.2111*****	0.0510	**0.1601****	**-0.1862***	**0.1827***	0.0036
							(0.0798)	(0.0842)	(0.0813)	(0.1100)	(0.1099)	(0.0036)
Homecare							0.0015	-0.0356	0.0341	0.0392	-0.0381	-0.0010
							(0.0652)	(0.0588)	(0.0360)	(0.0856)	(0.0855)	(0.0022)
Year	-0.0102	**-0.0526****	**0.0627*****	0.0012	-0.0055	**0.0043*****	-0.0108	**-0.0559****	**0.0667*****	0.0007	-0.0052	**0.0045*****
	(0.0173)	(0.0220)	(0.0181)	(0.0197)	(0.0197)	(0.0005)	(0.0183)	(0.0220)	(0.0176)	(0.0213)	(0.0212)	(0.0006)
Hospital Dummy	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Pseudo R2		0.0909			0.3077			0.1291			0.3666	
AIC		330.2344			257.7613			317.1032			237.551	
BIC		358.4035			289.0603			345.2723			268.85	

Note: Standard errors are reported in parentheses. Non-mAbs = Non-monoclonal antibody (etanercept); Older mAbs = Older monoclonal antibodies (infliximab & adalimumab); Newer mAbs = Newer monoclonal antibodies (certolizumab pegol & golimumab). *, **, *** indicates statistical significance at 10%, 5% and 1%, respectively. Partial effects of hospital dummy variables are unreported.

Appendix 22: Section 5.3 - Complete PRISMA Statement Checklist

This appendix provides the completed *PRISMA* checklist (Table A22.1) for the systematic review in Section 5.3 of *Chapter Five*. The checklist ensured that the systematic review was reported transparently and completely in accordance with the standards recommended for best-practice (Liberati et al., 2009).

Table A22.1. *Complete PRISMA Statement checklist for systematic review in Section 5.3 of Chapter Five.*

Section/Topic	Item No.	Checklist Item	Evidence
Title			
Title.	1	Identify the report as a systematic review, meta-analysis, or both.	Not applicable for PhD thesis.
Abstract			
Abstract.	2	Provide a structured summary.	Not applicable for PhD thesis.
Introduction			
Rationale.	3	Describe the rationale for the review in the context of what is already known.	Section 5.3.1.
Objectives.	4	Provide an explicit statement of questions being addressed with reference to PICOS.	Section 5.3.2.
Methods			
Protocol and registration.	5	Indicate if a review protocol exists, if and where is can be accessed, and, if applicable, provide registration number information.	No registration number exists. Review followed protocol as written in methods section.
Eligibility criteria.	6	Specify study characteristics and report characteristics used as criteria for eligibility, giving rationale.	Table 5.1; Section 5.3.3.
Information sources.	7	Describe all information sources in the search and date last searched.	Section 5.3.3; *Study Selection*.
Search.	8	Present full electronic search strategy for at least one database.	Appendix 23.
Study selection.	9	State the process for selecting studies.	Section 5.3.3; *Study Selection*.
Data collection process.	10	Describe method of data extraction from reports.	Section 5.3.3; *Data Extraction and Analysis*.
Data items.	11	List and define variables for which data were sought.	Section 5.3.3; *Data Extraction and Analysis*.

Section/Topic	Item No	Checklist Item	Evidence
Risk of bias in individual studies.	12	Describe methods used for assessing risk of bias of individual studies.	Not applicable: purpose of review was to identify published prescribing recommendations.
Summary measures.	13	State the principal summary measures.	Section 5.3.3; *Data Extraction and Analysis.*
Synthesis of results.	14	Describe the methods used for handling data and combining results of studies.	Section 5.3.3; *Data Extraction and Analysis.*
Risk of bias across studies.	15	Specify any assessment of risk of bias that may affect the cumulative evidence.	Not applicable: purpose of review was to identify published prescribing recommendations.
Additional analyses.	16	Describe methods of additional analyses.	Not applicable.
Results			
Study selection.	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusion at each state, ideally with a flow diagram.	Section 5.3.4; Figure 5.2.
Study characteristics.	18	For each study, present characteristics for which data were extracted and provide citations.	Table 5.2.
Risk of bias within studies.	19	Present data on risk of bias of each study and, if available, any outcome assessment.	Not applicable: purpose of review was to identify published prescribing recommendations.
Results of individual studies.	20	For all outcomes, present for each study, (a) a simple summary data for each intervention group and (b) effect estimates and confidence intervals.	Table 5.3; Section 5.3.4.1; Section 5.3.4.2.
Synthesis of results.	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Table 5.3; Section 5.3.4.1; Section 5.3.4.2.
Risk of bias across studies.	22	Present results of any assessment of risk of bias across studies (see item 15).	Not applicable: purpose of review was to identify published prescribing recommendations.

Section/Topic	Item No	Checklist Item	Evidence
Additional analyses.	23	Give results of additional analyses, if done.	Not applicable.
Discussion			
Summary of evidence.	24	Summarise the main findings including the strengths of evidence for each main outcome.	Section 5.3.5.
Limitations.	25	Discuss limitations at study and outcome level and at review level.	Section 5.3.5; *Limitations*.
Conclusions.	26	Provide a general interpretation of the results in the context of other evidence and implications for future research.	Section 5.3.5; *Implications for Future Research.*
Funding			
Funding.	27	Describe sources of funding for the systematic review and other support, and the role of funding for the systematic review.	Not applicable for individual PhD chapter. Funders acknowledged at the start of the thesis.

Source: Liberati et al. (2009, p. 18).

A22. References

Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ionnidis, J., Clarke, M., Devereaux, P., Kleijnen, J., & Moher, D. (2009). "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Healthcare Interventions: Explanation and Elaboration", *British Medical Journal*, Vol. 339, b2700, pp. 1-27.

Appendix 23: Search Strategies to Identify Studies with Relevance to TNFi Anti-drug Antibody and Drug Level Testing in Rheumatoid Arthritis

This appendix reports the search strategies that were used in *Chapter Five* to identify all published studies that had relevance to TNFi ADAb and drug level testing in RA, within (i) *Medline*, and (ii) *Embase*. The search strategies were used in *Section 5.3* to identify prescribing algorithms that incorporated ADAb and drug level testing, and in *Appendix 34* to perform a systematic review of test accuracy studies. The search strategies were based on an independent systematic review, conducted as part of the NICE DAP appraisal process, which identified published studies that included TNFi ADAb and drug level testing for patients with Crohn's disease (National Institute for Health and Care Excellence, 2016).

(i) Medline

1 (adalimumab or humira).mp.
2 ADA.tw.
3 (infliximab or remicade).mp.
4 IFX.tw.
5 (etanercept or enbrel).mp.
6 ETN.tw.
7 (certolizumab* or cimzia).mp.
8 CZP.tw.
9 (golimumab or simponi).mp.
10 GOL.tw.
11 ((anti-TNF* or antiTNF* or TNF*) adj2 inhibitor*).mp.
12 anti* tumo?r* necrosis* factor*.mp.
13 Tumor Necrosis Factor-alpha/ and Antibodies, Monoclonal/
14 biologic* treatment*.mp.
15 biologic* agent*.mp.
16 biologic* therap*.mp.
17 anti* drug* antibod*.tw.
18 ADAb.tw.
19 or/1-18

20 exp Enzyme-Linked Immunosorbent Assay/
21 enzyme* link* immunoassay*.mp.
22 enzyme* link* immuno* assay*.mp.
23 ELISA*.mp.
24 *Radioimmunoassay/
25 (radioimmuno* or radio immuno* or radio-immuno*).mp.
26 RIA.tw.
27 reporter* gene* assay*.mp.
28 RGA.tw.
29 semi* fluid* phase* enzyme* immuno*.mp.
30 EIA.tw.
31 ((homogenous* or homogeneous*) adj1 mobilit* shift* assay*).mp.
32 HMSA.tw.
33 or/20-32
34 ((immuno* or monitor* or pharmacokinetic* or measur* or level* or concentration*) adj3 (adalimumab or ADA or infliximab or IFX or etanercept or ETN or certolizumab* or CZP or golimumab or GOL or anti-TNF* or anti-tumo?r necrosis factor*)).mp.
35 exp Arthritis, Rheumatoid/
36 (rheumatoid* or rheumatoid arthritis).tw.
37 35 or 36
38 (((immuno* or monitor* or pharmacokinetic* or measur* or level* or concentration*) adj3 (adalimumab or ADA or infliximab or IFX or etanercept or ETN or certolizumab* or CZP or golimumab or GOL or anti-TNF* or anti-tumo?r necrosis factor*)) and (correlat* or associat* or test performance)).mp.
39 19 and 33 and 37
40 34 and 37
41 38 or 39 or 40
42 Animals/ not Humans/
43 41 not 42
44 remove duplicates from 43

(ii) Embase

1 (adalimumab or humira or ADA).tw.
2 *adalimumab/
3 (infliximab or remicade or IFX).tw.
4 *infliximab/
5 (etanercept or enbrel or ETN).tw.
6 *etanercept/
7 (certolizumab* or cimzia or CZP).tw.
8 *certolizumab pegol/
9 (golimumab or simponi or GOL).tw.
10 *golimumab/
11 ((anti-TNF* or antiTNF* or TNF*) adj2 inhibitor*).tw.
12 anti* tumo?r* necrosis* factor*.tw.
13 *tumor necrosis factor alpha inhibitor/
14 biologic* treatment*.tw.
15 biologic* agent*.tw.
16 biologic* therap*.tw.
17 anti* drug* antibod*.tw.
18 ADAb.tw.
19 *drug antibody/
20 or/1-19
21 *enzyme linked immunosorbent assay/
22 enzyme* link* immunoassay*.tw.
23 enzyme* link* immuno* assay*.tw.
24 ELISA*.tw.
25 *radioimmunoassay/
26 (radioimmuno* or radio immuno* or radio-immuno*).tw.
27 RIA.tw.
28 reporter* gene* assay*.tw.
29 RGA.tw.
30 semi* fluid* phase* enzyme* immuno*.tw.
31 EIA.tw.
32 (homogeny* adj1 mobilit* shift* assay*).tw.
33 HMSA.tw.
34 or/21-33

35 ((immuno* or monitor* or pharmacokinetic* or measure* or level* or concentration*) adj3 (adalimumab or ADA or infliximab or IFX or etanercept or ETN or certolizumab* or CZP or golimumab or GOL or Anti-TNF* or Anti-Tumo?r Necrosis Factor*)).tw.

36 *rheumatoid arthritis/

37 (rheumatoid* or rheumatoid arthritis*).tw.

38 or/36-37

39 (((immuno* or monitor* or pharmacokinetic* or measure* or level* or concentration*) adj3 (adalimumab or ADA or infliximab or IFX or etanercept or ETN or certolizumab* or CZP or golimumab or GOL or Anti-TNF* or Anti-Tumo?r Necrosis Factor*)) and (correlate* or associate* or test performance)).tw.

40 20 and 34 and 38

41 35 and 38

42 39 or 40 or 41

43 nonhuman/ not human/

44 42 not 43

45 remove duplicates from 44

A23. References

National Institute for Health and Care Excellence. (2016). "Therapeutic Monitoring of TNF-Alpha Inhibitors in Crohn's Disease (LISA-TRACKER ELISA Kits, IDKmonitor ELISA Kits, and Promonitor ELISA Kits)". *NICE Diagnostics Guidance, DG22*. London: National Institute for Health and Care Excellence.

Appendix 24: Section 5.4 - Full Derivation of Incremental Outcomes from the Algebraic Early Economic Evaluation Technique

This appendix provides the full derivation of the incremental outcomes calculated by the early economic evaluation technique in Section 5.4 of *Chapter Five*. The algebraic derivations for the three broad TNFi ADAb and drug level testing strategies to stratify treatment are reported sequentially.

A24.1. Strategy A: Testing after Loss of Response

The derivation in Section A24.1 relates to the cost and QALY profiles for *Strategy A* and *Current Practice*, illustrated in Figure 5.3.

Cost Profile

$$\text{Total Cost}_{\text{Strategy A}} = P[(t_1 - t_0)a + (t_2 - t_1)(a + b + c) + (t_3 - t_2)a]$$

$$\text{Total Cost}_{\text{Current Practice}} = P[(t_1 - t_0)a + (t_2 - t_1)(a + c) + (t_3 - t_2)a]$$

$$\text{Incremental Cost}_{\text{Strategy A}} = \boldsymbol{(t_2 - t_1)bP}$$

QALY Profile

$$\text{Total QALYs}_{\text{Strategy A}} = (t_1 - t_0)Q_1 + (t_2 - t_1)(Q_1 - X_1) + (t_3 - t_2)Q_1$$

$$\text{Total QALYs}_{\text{Current Practice}} = (t_1 - t_0)Q_1 + (t_2 - t_1)(Q_1 - X_1) + (t_3 - t_2)Q_1$$

$$\text{Incremental QALYs}_{\text{Strategy A}} = \mathbf{0}$$

Incremental Net Monetary Benefit

$$\text{Incremental Net Monetary Benefit}_{\text{Strategy A}} = \boldsymbol{-(t_2 - t_1)bP}$$

The incremental net monetary benefit of *Strategy A*, compared with current practice, was negative, which indicated that it was unlikely to be a potentially relevant comparator to include in the final decision problem.

A24.2. Strategy B: Testing During Response

The derivation in Section A24.2 relates to the cost and QALY profiles for *Strategy B* and *Current Practice*, illustrated in Figure 5.4.

Cost Profile

$$\text{Total Cost}_{\text{Strategy B}} = P[(t_1 - t_0)(a + b) + (t_3 - t_1)a] + (1 - P)[(t_3 - t_0)(a + b)]$$

$$\text{Total Cost}_{\text{Current Practice}} = P[(t_1 - t_0)a + (t_2 - t_1)(a + c) + (t_3 - t_2)a] + (1 - P)[(t_3 - t_0)a]$$

$$\begin{aligned}\text{Incremental Cost}_{\text{Strategy B}} = {} & P(t_1 - t_0)a + P(t_1 - t_0)b + P(t_3 - t_1)a + (t_3 - t_0)a + (t_3 - t_0)b \\ & - P(t_3 - t_0)a - P(t_3 - t_0)b - P(t_1 - t_0)a - P(t_2 - t_1)a - P(t_2 - t_1)c \\ & - P(t_3 - t_2)a - (t_3 - t_0)a + P(t_3 - t_0)a \\ = {} & \boldsymbol{P[(t_1 - t_0)b - (t_2 - t_1)c] + (1 - P)[(t_3 - t_0)b]}\end{aligned}$$

QALY Profile

$$\text{Total QALY}_{\text{Strategy B}} = P[(t_3 - t_0)Q_1] + (1 - P)[(t_3 - t_0)Q_1]$$

$$\begin{aligned}\text{Total QALY}_{\text{Current Practice}} = {} & P[(t_1 - t_0)Q_1 + (t_2 - t_1)(Q_1 - X_1) + (t_3 - t_2)Q_1] \\ & + (1 - P)[(t_3 - t_0)Q_1]\end{aligned}$$

$$\begin{aligned}\text{Incremental QALY}_{\text{Strategy B}} = {} & P(t_3 - t_0)Q_1 + (t_3 - t_0)Q_1 - P(t_3 - t_0)Q_1 - P(t_1 - t_0)Q_1 \\ & - P(t_2 - t_1)Q_1 + P(t_2 - t_1)X_1 - P(t_3 - t_2)Q_1 - (t_3 - t_0)Q_1 \\ & + P(t_3 - t_0)Q_1 \\ = {} & \boldsymbol{(t_2 - t_1)X_1P}\end{aligned}$$

Incremental Net Monetary Benefit

$$\text{Incremental Net Monetary Benefit}_{\text{Strategy B}} = \boldsymbol{\lambda(t_2 - t_1)X_1P - (t_1 - t_0)bP + (t_2 - t_1)cP - (t_3 - t_0)b + (t_3 - t_0)bP}$$

It was not possible to determine whether the incremental net monetary benefit of *Strategy B*, compared with *Current Practice*, was positive or negative based on the algebraic analysis alone. The sign of the incremental net monetary benefit depended on whether (i) the monetary value of the incremental QALY gains ($\lambda(t_2 - t_1)X_1P$) *and* (ii) the cost-reduction associated with avoiding the need to treat patients that had lost response ($(t_2 - t_1)cP$) were greater than the additional costs imposed by testing ($-(t_1 - t_0)bP - (t_3 - t_0)b + (t_3 - t_0)bP$).

The following section presents the first-derivative of the incremental net monetary benefit for *Strategy B* with respect to six input parameters:

Cost of Treating Loss of Response

$$\frac{\partial \text{Incremental Net Monetary Benefit}_{\text{B}}}{\partial \text{c}} = \boldsymbol{(t_2 - t_1)P} \quad > \quad \mathbf{0}$$

The positive first-derivative indicated that an increase in the cost of treating patients that had lost response (c) would increase the incremental net monetary benefit of *Strategy B*, *ceteris paribus*.

Cost of Testing Patients

$$\frac{\partial \text{Incremental Net Monetary Benefit}_{\text{B}}}{\partial \text{b}} = \boldsymbol{-(t_1 - t_0)P - (t_3 - t_0) + (t_3 - t_0)P} \quad < \quad \mathbf{0}$$

The negative first-derivative indicated that an increase in the cost of testing patients (b) would reduce the incremental net monetary benefit of *Strategy B*, *ceteris paribus*.

Cost-effectiveness Threshold

$$\frac{\partial \text{Incremental Net Monetary Benefit}_{\text{B}}}{\partial \lambda} = \boldsymbol{(t_2 - t_1)X_1P} \quad > \quad \mathbf{0}$$

The positive first-derivative indicated that an increase in the cost-effectiveness threshold (λ) would increase the incremental net monetary benefit of *Strategy B, ceteris paribus*.

Time Taken to Develop Anti-drug Antibodies

$$\frac{\partial \text{Incremental Net Monetary Benefit}_B}{\partial(t_1 - t_0)} = -bP \quad < \quad 0$$

The negative first-derivative indicated that an increase in time taken to develop ADAb ($t_1 - t_0$) would reduce the incremental net monetary benefit of *Strategy B, ceteris paribus*.

Proportion of Patients with Anti-drug Antibodies

$$\frac{\partial \text{Incremental Net Monetary Benefit}_B}{\partial \text{P}} = \lambda(t_2 - t_1)X_1 - (t_1 - t_0)b + (t_2 - t_1)c + (t_3 - t_0)b \quad > \quad 0$$

The positive first-derivative indicated that an increase in the proportion of patients that developed ADAb (P) would increase the incremental net monetary benefit of *Strategy B, ceteris paribus*.

Magnitude of QALY Loss Associated with Treatment Failure

$$\frac{\partial \text{Incremental Net Monetary Benefit}_B}{\partial X_1} = \lambda(t_2 - t_1)P \quad > \quad 0$$

The positive first-derivative indicated that an increase in the magnitude of QALY loss associated with treatment failure (X_1) would increase the incremental net monetary benefit of *Strategy B, ceteris paribus*.

A24.3. Strategy C: Test Drug Levels in Remission

The derivation in Section A24.3 relates to the cost and QALY profiles for *Strategy C* and *Current Practice*, illustrated in Figure 5.5.

Cost Profile

$$\text{Total Cost}_{\text{Strategy C}} = D[(t_1 - t_0)(a + b) + (t_2 - t_1)(a - y) + q(t_3 - t_2)a + (1 - q)(t_3 - t_2)(a - y)] + (1 - D)[(t_1 - t_0)(a + b) + (t_3 - t_1)a]$$

$$\text{Total Cost}_{\text{Current Practice}} = D[(t_3 - t_0)a] + (1 - D)[(t_3 - t_0)a]$$

$$\begin{aligned}\text{Incremental Cost}_{\text{Strategy C}} = {} & D(t_1 - t_0)a + D(t_1 - t_0)b + D(t_2 - t_1)a - D(t_2 - t_1)y + Dq(t_3 - t_2)a \\ & + D(t_3 - t_2)a - D(t_3 - t_2)y - Dq(t_3 - t_2)a + Dq(t_3 - t_2)y \\ & + (t_1 - t_0)a + (t_1 - t_0)b + (t_3 - t_1)a - D(t_1 - t_0)a - D(t_1 - t_0)b \\ & - D(t_3 - t_1)a - D(t_3 - t_0)a - (t_3 - t_0)a + D(t_3 - t_0)a\end{aligned}$$

$$= \boldsymbol{(t_1 - t_0)b - (t_2 - t_1)yD - (1 - q)(t_3 - t_2)yD}$$

QALY Profile

$$\text{Total QALY}_{\text{Strategy C}} = D[(t_1 - t_0)Q_2 + q(t_2 - t_1)(Q_2 - X_2) + (1 - q)(t_2 - t_1)Q_2 + (t_3 - t_2)Q_2] + (1 - D)[(t_3 - t_0)Q_2]$$

$$\text{Total QALY}_{\text{Current Practice}} = D[(t_3 - t_0)Q_2] + (1 - D)[(t_3 - t_0)Q_2]$$

$$\begin{aligned}\text{Incremental QALY}_{\text{Strategy C}} = {} & D(t_1 - t_0)Q_2 + Dq(t_2 - t_1)Q_2 - Dq(t_2 - t_1)X_2 + D(t_2 - t_1)Q_2 \\ & - Dq(t_2 - t_1)Q_2 + D(t_3 - t_2)Q_2 + (t_3 - t_0)Q_2 - D(t_3 - t_0)Q_2 \\ & - D(t_3 - t_0)Q_2 - (t_3 - t_0)Q_2 + D(t_3 - t_0)Q_2\end{aligned}$$

$$= \boldsymbol{- D(t_2 - t_1)X_2q}$$

$$\text{Incremental Net Monetary Benefit}_{\text{Strategy C}} = \boldsymbol{- \lambda D(t_2 - t_1)X_2q - (t_1 - t_0)b + (t_2 - t_1)yD + (1 - q)(t_3 - t_2)yD}$$

It was not possible to determine whether the incremental net monetary benefit of *Strategy C*, compared with *Current Practice*, was positive or negative based on the algebraic analysis alone. The sign of the incremental net monetary benefit depended on whether (i) the reduction in QALYs (from patients that flared), expressed in monetary units ($-\lambda D(t_2 - t_1)X_2q$), and (ii) the additional cost imposed by testing ($-(t_1 - t_0)b$), were offset by the

reduction in the cost of treatment due to using a lower-dose ($(t_2 - t_1)yD + (1 - q)(t_3 - t_2)yD$).

The following section presents the first-derivative of the incremental net monetary benefit for *Strategy C* with respect to six input parameters.

Cost of Testing

$$\frac{\partial \text{Incremental Net Monetary Benefit}_C}{\partial b} = -(t_1 - t_0) \quad < \quad 0$$

The negative first-derivative indicated that an increase in the cost of testing patients (*b*) would reduce the incremental net monetary benefit of *Strategy C*, *ceteris paribus*.

Cost-reduction of Lower-dose TNFi

$$\frac{\partial \text{Incremental Net Monetary Benefit}_C}{\partial y} = (t_2 - t_1)D + (1 - q)(t_3 - t_2)D \quad > \quad 0$$

The positive first-derivative indicated that an increase in the cost-reduction associated with using a lower-dose of TNFi (*y*) would increase the incremental net monetary benefit of *Strategy C*, *ceteris paribus*.

Cost-effectiveness Threshold

$$\frac{\partial \text{Incremental Net Monetary Benefit}_C}{\partial \lambda} = -D(t_2 - t_1)X_2 q \quad < \quad 0$$

The negative first-derivative indicated that an increase in the cost-effectiveness threshold (λ) would reduce the incremental net monetary benefit of *Strategy C*, *ceteris paribus*.

Proportion of Patients with High Drug Levels

$$\frac{\partial \text{Incremental Net Monetary Benefit}_C}{\partial D} = -\lambda(t_2 - t_1)X_2 q + (t_2 - t_1)y + (1 - q)(t_3 - t_2)y$$

It was not possible to determine whether the first-derivative of the incremental net monetary benefit, with respect to the proportion of patients with high drug levels (*D*), was

positive or negative from the algebraic analysis alone. This was because the sign of the function depended on values taken by its arguments.

Proportion of Patients that Flared from Reduced-dose TNFi

$$\frac{\partial \text{Incremental Net Monetary Benefit}_C}{\partial q} = -\lambda D(t_2 - t_1)X_2 - (t_3 - t_2)yD < 0$$

The negative first-derivative indicated that an increase in the proportion of patients that flared from reduced-dose TNFi (q) would reduce the incremental net monetary benefit of *Strategy C, ceteris paribus*.

QALY-reduction Associated with a Flare

$$\frac{\partial \text{Incremental Net Monetary Benefit}_C}{\partial X_2} = -\lambda D(t_2 - t_1)q < 0$$

The negative first-derivative indicated that an increase in magnitude of the QALY-reduction that was associated with a flare (X_2) would reduce the incremental net monetary benefit of *Strategy C, ceteris paribus*.

Appendix 25: Section 5.5 - Defining Relevant Strategies to Stratify Treatment in Current Practice

This appendix explains how the adalimumab ADAb and drug level tests were conceptually embedded within the existing care pathway for RA to ensure relevance to clinical practice in England. Sections A25.1 and A25.2 describe how relevant prescribing decisions were determined for using the tests to stratify treatment in patients that (i) responded to treatment and (ii) were in remission, respectively. These relevant treatment strategies were subsequently included in the decision problem of the economic evaluation in Section 5.5.

A25.1. Testing Adalimumab ADAb and Drug Levels in Patients that Responded

Patients with RA were expected to receive a sequence of treatments over their lifetimes due to the chronic nature of the disease (Stevenson, 2016). Early detection of ADAb and low drug levels in patients that were responding to adalimumab was assumed to be predictive of treatment failure (Garcês et al., 2013). Therefore, such patients may have responded better to a bDMARD with a different therapeutic target (not tumour necrosis factor-α).

Patients with RA in England that had lost response to a first-line TNFi typically commenced rituximab therapy as a second-line bDMARD, as demonstrated by the qualitative interviews in *Chapter Three*. Rituximab, unlike adalimumab and the other TNFi therapies, has a mechanism of action that targets B-cells (Emery, 2012). Sequential TNFi therapies, despite being recommended by some prescribing algorithms (Section 5.3), were not likely to be a relatively cost-effective use of health care resources in England (National Institute for Health and Care Excellence, 2010).

An appropriate way to embed the ADAb and drug level tests into the care pathway for RA in England, therefore, was to stratify treatment by advancing patients to rituximab therapy. Table A25.1i reports the appropriate prescribing decisions associated with each test result. If only the ADAb test was used, a positive detection of ADAb was sufficient evidence to change treatment to rituximab before adalimumab failure occurred. Patients with no detectable ADAb were recommended to remain on adalimumab therapy. An early change to rituximab was assumed to be associated with a benefit to ADAb-positive patients and a harm to ADAb-negative patient. The accuracy of ADAb testing was improved by

including drug level testing, which was assumed to reduce the harm induced by false-positive test results.

A25.2. Testing Adalimumab Drug Levels in Patients in Remission

Patients with RA that were in remission were assumed to have continued receiving full-dose adalimumab in current practice. It may have been possible, however, to reduce the dose of adalimumab in such patients that had been responding a long period of time. The dose of a TNFi can be reduced by increasing the time interval between the scheduled injections (Smolen et al., 2014).

Prescribing recommendations by *EULAR* stated that the dose of a TNFi therapy may be tapered if a patient with RA remained in persistent remission (Smolen et al., 2014). However, dose-reduction strategies may have only been appropriate for patients with high adalimumab drug levels, which was assumed to have reduced the likelihood of a subsequent flare in disease activity (Bykerk et al., 2016). The reduced-dose of adalimumab was recommended to revert back to its original dose if the patient experienced a flare in disease activity (Smolen et al., 2014). Table A25.1ii reports the appropriate prescribing decisions associated with testing adalimumab drug levels in patients during remission.

Table A25.1. *Appropriate test and prescribing decisions for current practice in England.*

Test Outcome	Prescribing Decision
(i) Test Patients during Response to Adalimumab	
ADAb Test Only	
ADAb-positive.	Prescribe rituximab & methotrexate.
ADAb-negative.	Continue adalimumab & methotrexate.
ADAb & Drug Level Test	
ADAb-positive & drug level low.	Prescribe rituximab & methotrexate.
ADAb-positive & drug level high; ADAb-negative & drug level high; ADAb-negative and drug level low.	Continue adalimumab & methotrexate.
(ii) Test Adalimumab Drug Levels of Patients in Remission	
Drug level high.	Half-dose of adalimumab & maintain methotrexate.
Drug level low; Drug level normal.	Continue adalimumab & methotrexate.

A25. References

Bykerk, V., Bingham, C., Choy, E., Lin, D., Alten, R., Christensen, R., Furst, D., Hewlett, S., Leong, A., March, L., Woodworth, T., Boire, G., Haraoui, B., Hitchon, C., Jamal, S., Keystone, E., Pope, J., Tin, D., Thorn, J., & Bartlett, S. (2016). "Identifying Flares in Rheumatoid Arthritis: Reliability and Construct Validation of the OMERACT RA Flare Core Domain Set", *RMD Open*, Vol. 2, e000225, pp. 1-10.

Emery, P. (2012). "Optimizing Outcomes in Patients with Rheumatoid Arthritis and an Inadequate Response to Anti-TNF Treatment", *Rheumatology*, Vol. 51, s5, pp. v22-v30.

Garcês, S., Demengeot, J., & Benito-Garcia, E. (2013). "The Immunogenicity of Anti-TNF Therapy in Immune-Mediated Inflammatory Diseases: A Systematic Review of the Literature with a Meta-Analysis", *Annals of the Rheumatic Diseases*, Vol. 73, 1, pp. 1947-1955.

National Institute for Health and Care Excellence. (2010). "Adalimumab, Etanercept, Infliximab, Rituximab and Abatacept for the Treatment of Rheumatoid Arthritis after the Failure of a TNF Inhibitor". NICE Technology Appraisal Guidance, TA195. London: National Institute for Health and Care Excellence.

Stevenson, M., Archer, R., Tosh, J., Simpson, E., Everson-Hock, E., Stevens, J., Hernandez-Alava, M., Paisley, S., Dickinson, K., Scott, D., Young, A., & Wailoo, A. (2016). "Adalimumab, Etanercept, Infliximab, Certolizumab Pegol, Golimumab, Tocilizumab, and Abatacept for the Treatment of Rheumatoid Arthritis not Previously Treated with Disease-Modifying Antirheumatic Drugs and After Failure of Conventional Disease-Modifying Antirheumatic Drugs Only: Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 20, 35, pp. 1-610.

Smolen, J., Landewé, R., Breedveld, F., Buch, M., Burmester, G., Dougados, M., Emery, P., Gaujoux-Viala, C., Gossec, L., Nam, J., Ramiro, S., Winthrop, K., de Wit, M., Aletaha, D., Betteridge, N., Biklsma, J., Boers, M., Buttgereit, F., Combe, B., Cutolo, M., Damjanov, N., Hazes, J., Kouloumas, M., Kvien, T., Mariette, X., Pavelka, K., van Riel, P., Rubbert-Roth, A., Scholte-Voshaar, M., Scott, D., Sokka-Isler, T., Wong, J., & van der Heijde, D. (2014). "EULAR Recommendations for the Management of Rheumatoid Arthritis with Synthetic and Biological Disease-modifying Antirheumatic Drugs: 2013 Update", *Annals of the Rheumatic Diseases*, Vol. 73, 3, pp. 492-509.

Appendix 26: Section 5.7 - Checklist to Inform the Choice of Model Type

Table A26.1. presents the checklist, reported by Brennan et al. (2006, p. 1304), regarding the choice of model type for an economic evaluation with respect to fifteen issues. An example was provided in the final column of the table if an issue was considered to be relevant for adalimumab ADAb and drug level testing.

Table A26.1. *Checklist to inform model selection by Brennan et al. (2006), applied to thesis case study.*

Issue†	Relevant	Choice of model†	Example
I.1: *Does the decision maker require knowledge of variability to inform the decision?*	Yes	Need for stochastic output.	Needed to know variability in estimates of expected cost and health outcomes.
I.2: *Is the decision maker uncertain about which sub-groups are relevant and likely to change his/her mind?*	Yes	Individual level models are more flexible to further covariates or changed assumptions.	Unknown which population should have used ADAb and drug level testing *a priori*.
I.3: *Is a probabilistic sensitivity analysis required?*	Yes	Need for PSA should not drive model structure.	NICE required a probabilistic sensitivity analysis.
I.4: *Do individual risk factors affect outcome in a non-linear fashion?*	Yes	Need to subdivide states or consider individual level model if the number of states is large.	Increased risk of death with more severe RA. Hospitalisations may have increased with disease severity.
I.5: *Do covariates have multiple effects, which cause interaction?*	Yes	Individual level modelling likely to be necessary.	Development of adalimumab ADAb and low drug levels reduced the time to secondary non-response of adalimumab.
I.6: *Are times in states non-Markovian?*	Yes	Need to use 'fixes' in Markovian models or use non-Markovian models.	Development of adalimumab ADAb and low drug levels reduced the time to secondary non-response of adalimumab.
I.7: *Is the dimensionality too great for a cohort approach?*	Yes	Individual level modelling likely to be necessary.	Patient characteristics may have affected the probability of death and the time to treatment failure. Rate of disease progression have depended on treatment type.
I.8: *Do states 'recycle'?*	Yes	Decision tree approach is probably not appropriate.	Treatments were prescribed in sequence in which patients responded and then lost response.

Issue†	Relevant	Choice of model†	Example
I.9: *Is phasing or timing of events decisions important?*	Yes.	Possible to have different branches in the decision tree but Markov model or simulation may be necessary.	The time to failure of adalimumab was dependent on whether the patient had previously developed ADAb and low drug levels.
I.10: *Is there interaction directly between patients?*	No.	Models with interactions.	Not applicable.
I.11: *Is there interaction due to constrained resources?*	No.	Models with interaction.	Not applicable.
I.12: *Could many events occur in one unit of time?*	Yes.	Need for small time intervals or continuous time models.	In a particular period of time, a patient's disease may have progressed, they may have lost response to treatment, they may have developed ADAb, or they may have been tested for ADAb and drug levels.
I.13. *Are interactions occurring in small populations?*	No.	Need to consider individual level modelling.	Not applicable.
I.14: *Are there delays in response due to the resource constraints which affect cost or health outcome?*	No.	Need for stochastic output and interaction.	Not applicable.
I.15: *Is there non-linearity in system performance when inherent variability occurs?*	No.	Discrete event simulation useful.	Not applicable.

Note: †=As reported in Brennan et al. (2006, p. 1304).

A26. References

Brennan, A., Chick, S., & Davies, R. (2006). "A Taxonomy of Model Structures for Economic Evaluation of Health Technologies", *Health Economics*, Vol. 15, 12, pp. 1295-1310.

Appendix 27: Section 5.8 - Systematic Review of Individual-level Model-based Economic Evaluations in Rheumatoid Arthritis

This appendix presents a systematic review of model-based economic evaluations in RA that have estimated expected outcomes by simulating the histories of individual patients over time. The primary purpose of this systematic review was to identify the modelling assumptions that could be used to inform the design of the DES in this thesis. This appendix is reported as a standalone study, with the following subsections: an introduction (Section A27.1), aim and objectives (Section A27.2), method (Section A27.3), results (Section A27.4), and a summary of key findings (Section A27.5).

A27.1. Introduction

A decision analytic model necessarily requires simplifying assumptions to be made in order to simulate the progression of a patient's disease over time. These assumptions can be represented by the different choices that decision-analysts have made regarding (i) the input parameters included in a model, and (ii) how those parameters were assumed to be related to each other (Tappenden et al., 2014). Published model-based economic evaluations (within the same disease area) have been recommended as a potentially useful source of evidence to help inform the development of a *de novo* decision analytic model, by identifying and critically appraising the assumptions within these models (Tappenden et al., 2014).

Previous reviews of model-based economic evaluations in RA have described, in general terms, the differences in structural assumptions that have been made by different authors (Bansback et al., 2005b; Drummond et al., 2005; Bansback et al., 2008; Barton, 2011; Madan et al., 2011; Tosh et al., 2011a; Tsao et al., 2012; Scholz et al., 2014; Tosh et al., 2014; Ganz et al., 2015; Madan et al., 2015). Section 5.7 of this thesis concluded that a DES was the most appropriate type of decision analytic model to estimate the relative cost-effectiveness of adalimumab ADAb and drug level testing to stratify treatment for patients with RA. A principal feature of using a DES for an economic evaluation is that patients are simulated through the structure of the model individually (Caro et al., 2016). Individual-level decision analytic models may also be referred to as *individual sampling models* or *microsimulations* (Barton et al., 2004a; Brennan et al., 2006; Davis et al., 2014). In the context of this thesis, it was therefore potentially useful to identify the specific assumptions

that have been made in published individual-level decision analytic models for RA, in order to inform the development of the *de novo* decision analytic model in *Chapter Five*.

A27.2. Aim and Objectives

The aim of this study was to identify and summarise the key assumptions used in published decision analytic models for RA that have estimated expected outcomes by simulating patients individually. There were three objectives to meet this aim:

Objective 1: Identify all model-based economic evaluations for RA that have simulated patients individually through the structure of the model;

Objective 2: Identify the key assumptions that were made during the design of these individual-level decision analytic models.

Objective 3: Summarise the key assumptions to inform the design of the *de novo* DES model in this thesis.

A27.3. Method

A systematic review of economic evaluations in RA that simulated patients individually through the structure of a decision analytic model was performed according to PRISMA reporting standards (Liberati et al., 2009). This study built on the systematic review of published model-based economic evaluations of stratified medicine in RA that was reported in *Chapter Two*.

Study Selection

The study inclusion criteria (reported in Table A27.1) was designed to identify all model-based economic evaluations for RA that had performed an individual patient simulation. A study was included if it used a decision analytic model that simulated patients with RA individually (and not as a cohort) for any pharmacological therapy.

Following the approach of *Chapter Two*, *Medline*, *Embase*, *Web of Science*, and the NHS *EED and HTA* databases were searched electronically between January 1990 and January 2014. A subsequent electronic search was conducted to update the review by searching *Medline* and the NHS *EED and HTA* databases until December 2016. The search strategies to identify the published model-based economic evaluations were identical to that of *Chapter Two* (reported in *Appendix 10*).

The title and abstract of all publications identified by the search strategies were screened by SG against the inclusion criteria in Table A27.1. Four researchers at *the Manchester*

Centre for Health Economics, The University of Manchester, were allocated an equal proportion of titles and abstracts to independently second-screen. Abstracts were not excluded at the screening stage if there were disagreements between SG and the independent reviewers. Studies that remained after the screening stage were read in full by SG to determine whether a full model-based economic evaluation was performed that simulated patients with RA individually through the structure of the model.

Table A27.1. *Systematic review inclusion criteria: model-based economic evaluations in RA that simulated individual patients.*

Study Characteristic	Inclusion Criteria
Population	Adults (>16 years) with RA.
Intervention	Any pharmacological therapy.
Comparator	Any comparator.
Outcome	Expected costs and expected patient benefits per intervention strategy.
Study Design	Full economic evaluation (cost-benefit analysis, cost-effectiveness analysis, cost-utility analysis), comparing at least two treatment strategies, using a decision analytic model that simulated the histories of individual patients over time (DES, microsimulation, or individual sampling model).
Language	English; full-text publication.

Data Extraction and Analysis

The principal motivation for this study was to identify the specific assumptions that had been made within published individual-level model-based economic evaluations for RA. Therefore, the following eight key assumptions were extracted from all included studies: (i) how the sample of patients was simulated; (ii) how treatment response was classified; (iii) how time-to-treatment failure was determined; (iv) how disease was assumed to progress over time; (v) how RA-specific mortality has handled; (vi). how direct health care costs were included; (vii) how QALYs were estimated (if relevant); and (viii) how a PSA was conducted. All data extraction was performed by SG. The key assumptions were summarised by a narrative synthesis of the extracted data.

A27.4 Results

A flow diagram of the studies included in the systematic review is reported in Figure A27.1. Twenty-nine published model-based economic evaluations in RA, that had simulated patients individually, were identified by the search strategy. Following the approach of Scholz et al. (2014), Table A27.2 categorised these twenty-nine economic evaluations by the general structure of their decision analytic model. The models were categorised into eight 'families' that represented incremental developments to their general structures over time.

Figure A27.1. *Flow diagram of included studies.*

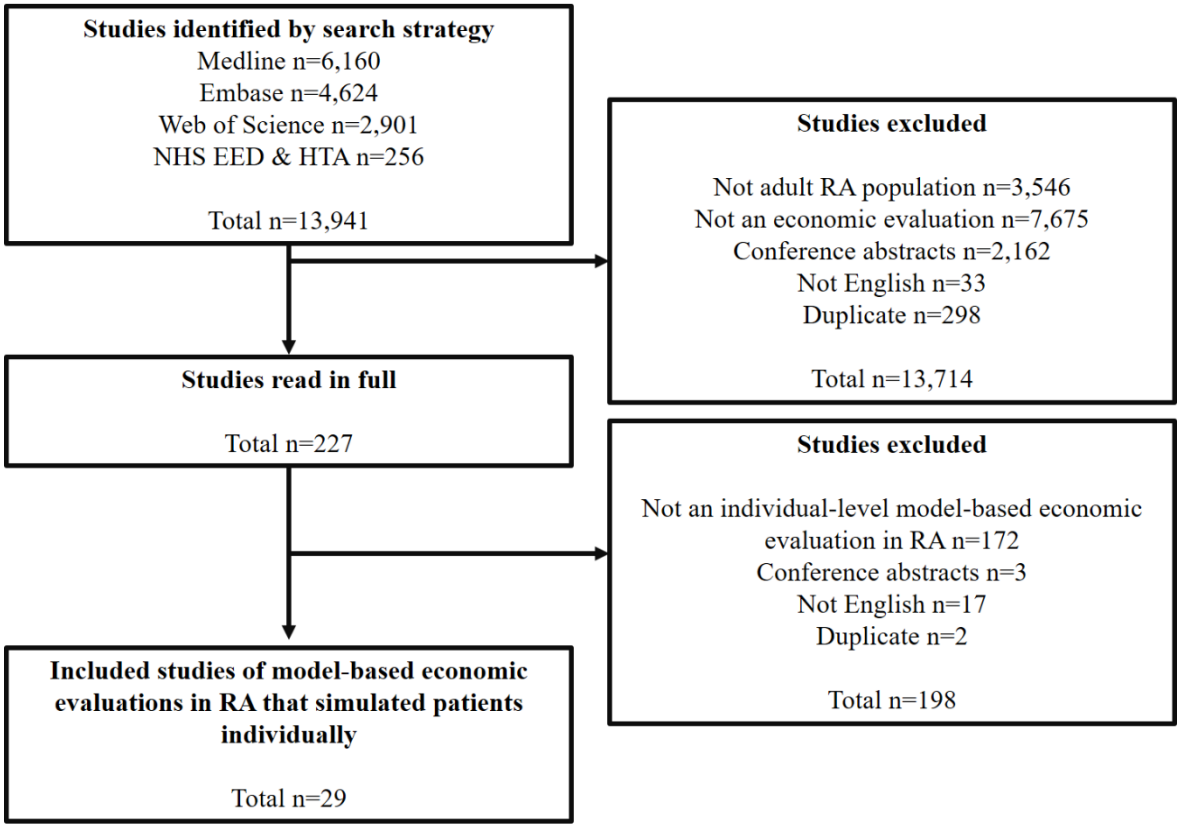


Table A27.2. *Twenty-nine individual-level model-based economic evaluations in RA, categorised by their general structure.*

(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
Brennan et al. (2004)	Kielhorn et al. (2008)	Jobanputra et al. (2002)	Vera-Llonch et al. (2008a)	Kobelt et al. (2009)	Stephens et al. (2015)	Tran-Duy et al. (2014)	Wu et al. (2015)
Bansback et al. (2005a)	Merkesdal et al. (2010)	Barton et al. (2004b)	Vera-Llonch et al. (2008b)	Lindgren et al. (2009)			
Brennan et al. (2007)	Hallinen et al. (2010)	Clark et al. (2004)	Yuan et al. (2010)				
Wailoo et al. (2008)	Diamantopoulos et al. (2012)	Chen et al. (2006)					
Finckh et al. (2009)	Soini et al. (2012)	Malottki et al. (2011)					
Davies et al. (2009)	Diamantopoulos et al. (2014)						
Tosh et al. (2011b)	Athanasakis et al. (2015)						
Stevenson et al. (2016)	Carlson et al. (2015)						

Each model simulated individual patients over their lifetime through a sequence of treatments relevant to the respective decision problem and clinical setting. The appropriate treatment sequences were informed by clinical guidelines, patient registry data, or expert input from rheumatologists. The majority of models followed a common sequence of phases:

- Phase One: An individual patient may have had an initial response to the first treatment in the sequence. Treatment was changed to the next in the sequence if the patient did not respond;
- Phase Two: The patient's disease activity may have changed immediately as an initial response to treatment and/or over time for the duration of treatment;
- Phase Three: The patient may have eventually stopped responding to treatment, at which point they returned to the start of the model and received the next treatment in the sequence. Loss of response to a treatment may have had a negative impact on disease activity;
- Phase Four: The patient could die at any time. The next patient was selected to proceed through the model if the current patient died. The whole process was repeated until all patients had been simulated through the model.

Over half of the models simulated patients individually through time by using time cycles of either six months (n=14) (Brennan et al., 2004; Bansback et al., 2005; Finckh et al., 2009; Davies et al., 2009; Tosh et al., 2011; Kielhorn et al., 2008; Merkesdal et al., 2010; Hallinen et al., 2010; Soini et al., 2012; Diamantopoulos et al., 2012; Athanasakis et al., 2015; Carlson et al., 2015; Diamantopoulos et al., 2014; Stephens et al., 2015) or three months (n=3) (Vera-Llonch et al., 2008a; Vera-Llonch et al., 2008b; Yuan et al., 2010). The twelve remaining models advanced time by estimating time-to-event values from survival curves.

Model Population

The patient population simulated within sixteen of the models were representative of the baseline characteristics observed within pivotal drug trials (Stephens et al., 2015; Athanasakis et al., 2015; Carlson et al., 2015; Vera-Llonch et al., 2008a; Vera-Llonch et al., 2008b; Yuan et al., 2010; Brennan et al., 2004; Davies et al., 2009; Wu et al., 2015; Bansback et al., 2005a; Kielhorn et al., 2008; Merkesdal et al., 2010; Lindgren et al., 2009; Soini et al., 2012; Diamantopoulos et al., 2012). The fourteen remaining studies had representative populations of a more-general distribution of RA patients, based on data

from incidence cohorts or national registries. Four models, in particular, had a population representative of patients with RA in the UK by utilising patient characteristics derived from the *British Society for Rheumatology Biologics Registry* (Diamantopoulos et al., 2014; Brennan et al., 2007; Malottki et al., 2011; Stevenson et al, 2016).

The majority of studies allowed patient-level characteristics to vary between individual patients, however seven studies simulated identical patients through their model structure (Bansback et al., 2005a; Kobelt et al., 2009; Lindgren et al., 2009; Diamantopoulos et al., 2012; Diamantopoulos et al., 2014; Athanasakis et al., 2015; Wu et al., 2015). The majority of models described a patient's baseline characteristics according to at least three variables: age (apart from Davies et al. (2009) and Stephens et al. (2015)), sex (apart from Bansback et al. (2005a) and Davies et al. (2009)), and HAQ score (apart from Jobanputra et al. (2002) and Stephens et al. (2015)). Other characteristics to describe patients included disease duration (n=7), previous cDMARDs (n=5), DAS28 (n=3), and socioeconomic characteristics such as education and income (n=2).

Initial Treatment Response

Patients were assumed to have an initial short-term response after receiving treatment for six months in nineteen models. The majority of these models (n=15) categorised a six-month treatment response according to ACR response, which was reported by most trials of biologic therapies for RA. Two models included treatment response as a EULAR response, obtained from accompanying registry data (Brennan et al., 2007) and a review of published trials (Stevenson et al., 2016). Tran-Duy et al. (2014) included treatment response as a direct change in DAS28 score only.

The initial six-month response to therapy was estimated predominantly by indirect treatment comparison, using multiple sources of evidence, where a common comparator existed between trials (n=13) (Diamantopoulos et al., 2014; Athanasakis et al., 2015; Bansback et al., 2005a; Diamantopoulos et al., 2012; Stevenson et al, 2016; Wailoo et al., 2008; Finckk et al., 2009; Kielhorn et al., 2008; Merkesdal et al., 2010; Soini et al., 2012; Tosh et al., 2011; Carlson et al., 2015; Davies et al., 2009). Alternatively, Brennan et al. (2004), Hallinen et al. (2010), Wu et al. (2015), and Stephens et al. (2015) estimated a six-month treatment response using evidence from a single trial.

Treatment Failure

There were two different methods to determine the time for remaining on a treatment until its withdrawal due to failure. The first approach applied a probability of treatment failure

within each time cycle of the model (n=13). The second approach directly assigned a time of treatment failure to each individual patient (n=16).

Different methods were used to directly assign the time-to-treatment failure for individual patients. A fixed time-to-treatment failure was assigned to each patient in three models (Hallinen et al., 2010; Merkesdal et al., 2010; Kielhorn et al., 2008). The thirteen remaining models sampled individual time-to-event values from statistical survival distributions. The most common distribution used to sample an individual patient's time-to-treatment failure was the Weibull distribution (n=10) (Malottki et al., 2011; Chen et al., 2006; Brennan et al., 2007; Wu et al., 2015; Kobelt et al., 2009; Lindgren et al., 2009; Wailoo et al., 2008; Barton et al., 2004; Clark et al., 2004; Jobanputra et al., 2002). However, multivariate normal distributions (Finckh et al., 2009), gamma distributions (Stevenson et al., 2016), exponential distributions (Tran-Duy et al., 2014), and log-normal distributions (Diamantopoulos et al., 2014) were also used.

Disease Progression

Most models determined RA disease progression via changes in patients' HAQ scores over time. Five models, which were developed for specific NICE Technology Appraisals, used *legitimate* HAQ scores, expressed only in multiples of 0.125 (Barton et al., 2004b; Clark et al., 2004; Chen et al., 2006; Malottki et al., 2011; Stevenson et al., 2016). There were three common clinical events within the models that were assumed to prompt disease progression:

(i) Initial treatment response;
(ii) Treatment withdrawal;
(iii) Long-term progression while receiving treatment.

Initial Treatment Response

Twenty-six models represented the benefit of treatment response by an immediate reduction in the a patient's HAQ score. This HAQ reduction was conditional on the response criteria in fifteen models; for example, the greatest HAQ reduction was for a good EULAR response (Finckh et al., 2009; Diamantopoulos et al., 2012; Diamantopoulos et al., 2014; Wu et al., 2015; Athanasakis et al., 2015; Stevenson et al., 2016; Hallinen et al., 2010; Kielhorn et al., 2008; Merkesdal et al., 2010; Soni et al., 2012; Carlson et al., 2015; Davies et al., 2009; Bansback et al., 2005a; Wailoo et al., 2008; Tosh et al., 2011), or the treatment received in eleven models; for example, bDMARDs provided a greater HAQ reduction than cDMARDs (Kobelt et al., 2009; Lindgren et al., 2009; Jobanputra et al.,

2002; Barton et al., 2004b; Clark et al., 2004; Brennan et al., 2004; Chen et al., 2006; Malottki et al., 2011; Vera-Llonch et al., 2008a; Vera-Llonch et al., 2008b; Yuan et al., 2010).

Treatment Withdrawal

The worsening of a patient's disease, associated with treatment withdrawal, was represented in most models (n=22) by an increase in the HAQ score. Upon treatment failure, fifteen models assumed a complete reversal of the initial HAQ reduction (known as a *perfect rebound*), (Chen et al., 2006; Malottki et al., 2011; Diamantopoulos et al., 2012; Diamantopoulos et al., 2014; Athanasakis et al., 2015; Wu et al., 2015; Stevenson et al., 2016; Kielhorn et al., 2008; Merkesdal et al., 2010; Soni et al., 2012; Davies et al., 2009; Brennan et al., 2004; Bansback et al., 2005a; Wailoo et al., 2008; Tosh et al., 2011); three models assumed that HAQ returned to the value before treatment initiation (Carlson et al., 2015; Kobelt et al., 2009; Lindgren et al., 2009); and three models assumed that the HAQ score increased to the value observed in the comparator arm (Vera-Llonch et al., 2008a; Vera-Llonch et al., 2008b; Yuan et al., 2010).

Long-term Progression

RA is characterised by a worsening of disease over time, most frequently represented in the models by gradual increases in the HAQ score according to the type of treatment (n=16; HAQ worsened more quickly on cDMARD than bDMARD therapy) (Brennan et al., 2004; Wailoo et al., 2008; Kielhorn et al., 2008; Merkesdal et al., 2010; Hallinen et al., 2010; Vera-Llonch et al., 2008a; Vera-Llonch et al., 2008b; Yuan et al., 2010; Brennan et al., 2007; Malottki et al., 2011; Diamantopoulos et al., 2012; Diamantopoulos et al., 2014; Athanasakis et al., 2015; Wu et al., 2015; Soini et al., 2012; Carlson et al., 2015), or initial response to therapy (n=7; HAQ worsened more quickly for a poorer response to treatment) (Bansback et al., 2005a; Tosh et al., 2011; Davies et al., 2009; Finckh et al., 2009; Kobelt et al., 2009; Lindgren et al., 2009; Stevenson et al., 2016). Three models specifically incorporated the impact of radiographic progression on HAQ progression over time (Brennan et al., 2004; Brennan et al., 2007; Finckh et al., 2009).

Table A27.3 presents the annual rate of HAQ progression in eleven models, by treatment type, for moderate to severe patients with RA. A range of annual HAQ progression rates were assumed for cDMARD (between 0.034 and 0.065 per year) and TNFi (between 0 and 0.034 per year) therapies. The lower the annual rate of HAQ progression for a treatment, the more favourable its estimated relative cost-effectiveness is likely to be, *ceteris paribus*.

Table A27.3. *Assumed rate of annual HAQ progression for patients with RA on cDMARD and bDMARD therapies.*

cDMARD		bDMARD		
Author (Year)	**Annual HAQ Progression**	**Author (Year)**	**Annual HAQ Progression**	**Specific Biologic**
Brennan et al. (2004); Bansback et al. (2005a); Hallinen et al. (2010); Kielhorn et al. (2008); Merkesdal et al. (2010).	0.034 per year.	Brennan et al. (2007); Diamantopoulos et al. (2012); Diamantopoulos et al. (2014); Malottki et al. (2011); Soini et al. (2012).	0 per year.	TNFi, rituximab, abatacept
Brennen et al. (2007).	0.042 per year.	Brennan et al. (2004); Vera-Llonch et al. (2008a); Vera-Llonch et al. (2008b); Yuan et al. (2010).	0.015 per year.	Etanercept, abatacept
Malottki et al. (2011); Soini et al. (2012); Diamantopoulos et al. (2014); Athanasakis et al. (2015).	0.045 per year.	Bansback et al. (2005a); Hallinen et al. (2010); Kielhorn et al. (2008); Merkesdal et al. (2010).	0.034 per year.	TNFi, rituximab
Vera-Llonch et al. (2008a); Vera-Llonch et al. (2008b); Yuan et al. (2010).	0.065 per year.	Diamantopoulos et al. (2010).	-0.037 per year.	Tocilizumab
		Soini et al. (2012).	-0.032 per year.	Tocilizumab

Mortality

Most studies adjusted standardised age/sex mortality rates to account for the increased mortality risk associated with RA in one of three ways: (i) direct adjustment for elevated mortality due to RA (n=7) (Wailoo et al., 2008; Merkesdal et al., 2010; Bansback et al., 2005a; Kobelt et al., 2009; Lundgren et al., 2009; Brennan et al., 2004; Jobanputra et al.,

2002); (ii) HAQ-dependent adjustment to reflect higher mortality with greater disease severity (n=15), (Hallinen et al., 2010; Barton et al., 2004b; Athanasakis et al., 2015; Carlson et al., 2015; Vera-Llonch et al., 2008a; Vera-Llonch et al., 2008b; Stevenson et al., 2016; Yuan et al., 2010; Kielhorn et al., 2008; Clark et al., 2004; Chen et al., 2006; Malottki et al., 2011; Soini et al., 2012; Wu et al., 2015; Davies et al., 2009); or (iii) mortality adjustment by type of treatment (n=1) (Finckh et al., 2009).

Relative risks (between 1.3 and 1.975) were applied to population mortality rates when adjusting for RA-specific mortality only. Four approaches were used when adjusting mortality by HAQ, by applying: (i) a relative risk of 1.33 per HAQ unit (n=10), (ii) a relative risk of 2.73 per HAQ unit (n=1), (iii) an odds ratio of 1.33 per HAQ unit (n=3), or (iv) a different hazard ratio depending on baseline HAQ score (n=1). One study applied a relative risk reduction in mortality of 0.65 when the patient received a biologic therapy (Finckh et al., 2009).

Direct Medical Costs

All models included the cost of treatments, monitoring and, where relevant, the cost of treatment administration. Patients with RA may incur additional direct medical costs over their lifetime for hospitalisations and joint replacements. These additional direct medical costs were most commonly included by HAQ-dependency, by assigning an annual cost to mutually exclusive HAQ intervals (n=12) (Hallinen et al., 2010; Athanasakis et al., 2015; Carlson et al., 2015; Vera-Llonch et al., 2008a; Vera-Llonch et al., 2008b; Stevenson et al., 2016; Tosh et al., 2011; Diamantopoulos et al., 2012; Diamantopoulos et al., 2014; Merkesdal et al., 2010; Kielhorn et al., 2008; Soini et al., 2012), or by using a monotonic HAQ-based cost algorithm (n=6) (Malottki et al., 2011; Wu et al., 2015; Bansback et al., 2005a; Kobelt et al., 2009; Davies et al., 2009; Brennan et al., 2004).

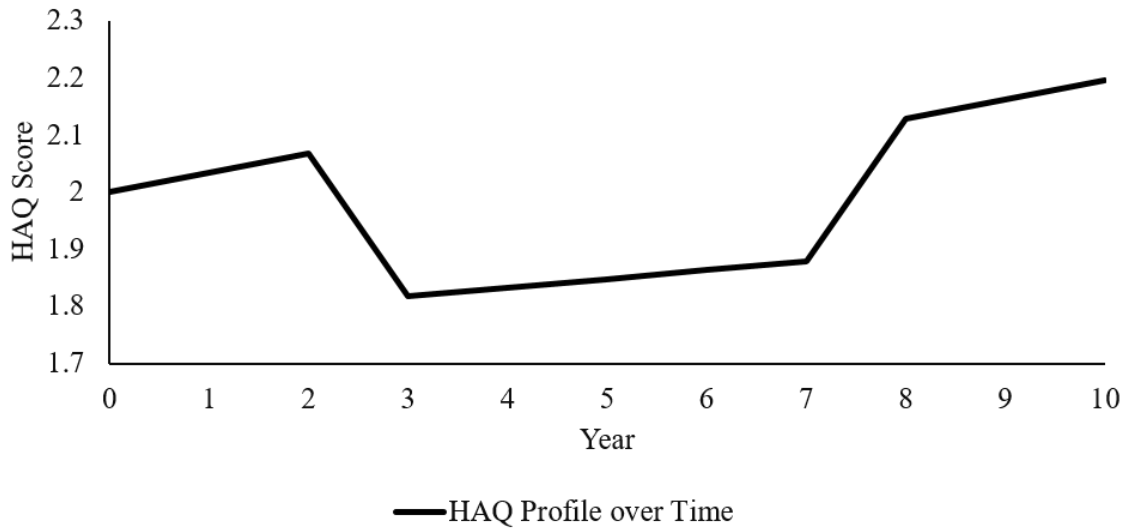
Six models used multivariable regression-based algorithms that included patient characteristics to calculate direct medical costs (Wailoo et al., 2008; Stephens et al., 2015; Tran-Duy et al., 2014; Finckh et al., 2009; Lindgren et al., 2009; Brennan et al., 2007). One model included an additional clinical state (for joint replacement) which, if experienced, increased the patient's direct medical cost (Barton et al., 2004b).

QALYs

All models used QALYs as the outcome measure of health benefit. Four models calculated QALYs according to HAQ intervals (Vera-Llonch et al., 2008a; Vera-Llonch et al., 2008b; Wailoo et al., 2008; Yuan et al., 2010), whereas the remaining models estimated QALYs

from a regression-based algorithm. Five algorithms (in ten models) calculated QALYs according to current HAQ score only (reported in Table A27.4). Figure A27.2 illustrates a hypothetical HAQ profile of a patient with RA over ten years.

Figure A27.2. *10-Year HAQ profile of a hypothetical patient.*



The patient in Figure A27.2 commenced cDMARD therapy with a HAQ score of 2 in year zero. HAQ was assumed to progress by 0.034 per year (lowest value for cDMARDs in Table A28.3), and biologic therapy was initiated in year three, which was assumed to have immediately reduced their HAQ score by 0.25. Biologic therapy remained effective for four years, and HAQ progressed by 0.015 per year (lowest non-zero value for bDMARDs in Table A28.3). A perfect rebound of the HAQ score (by 0.25) was assumed upon biologic failure in year eight, and HAQ was assumed to progress by 0.034 per year for the remainder of the time horizon. The total undiscounted QALYs calculated for this HAQ profile, using the five HAQ-based algorithms identified within the individual economic evaluations, are reported in Table A27.4. Different methods to estimate QALYs from HAQ led to substantial differences in total QALYs estimated for this particular individual patient.

Probabilistic Sensitivity Analysis

The majority of models (n=24) reported performing a probabilistic sensitivity analysis, the details of which are presented in Table A27.5 (seven studies did not report the number of PSA samples or individuals simulated). There was variation between the studies in the number of PSA simulations and samples generated, which were most likely due to differences in the computational burden between models.

Table A27.4. *Five HAQ-based algorithms to calculate QALYs and 10-year total undiscounted QALYs gained.*

Author (Year)	QALY Algorithm	Total Undiscounted 10-Year QALYs per Patient†
Jobanputra et al. (2002).	0.2 x HAQ	4.37 QALYs.
Barton et al. (2004b); Clark et al. (2004); Chen et al. (2006).	0.862 – (0.327 x HAQ)	2.34 QALYs.
Malottki et al. (2011).	0.804 – (0.203 x HAQ) – (0.045 x HAQ^2)	2.45 QALYs.
Davies et al. (2009).	0.76 – (0.28 x HAQ)	2.25 QALYs.
Soini et al. (2012); Diamantopoulos et al. (2010); Wu et al. (2015); Carlson et al. (2015).	0.82 – (0.11 x HAQ) – (0.07 x HAQ^2)	3.57 QALYs.

Note: †Total QALYs were calculated by applying the QALY algorithm to the HAQ profile in Figure A27.2.

Table A27.5. *The reported number of PSA parameter samples and individuals simulated in seventeen models.*

Author	PSA Samples	PSA Individuals
Brennan et al. (2007)	100.	50.
Stevenson et al. (2016)	100.	1,000.
Vera-Llonch et al. (2008a)	100.	1,000.
Vera-Llonch et al. (2008b)	100.	1,000.
Yuan et al. (2010)	100.	1,000.
Stephens et al. (2015)	250.	1,000.
Kobelt et al. (2009)	1,000.	Not reported.
Lindgren et al. (2009)	1,000.	Not reported.
Diamantopoulos et al. (2014)	1,000.	Not reported.
Wu et al. (2015)	1,000.	Not reported.
Tosh et al. (2011b)	1,000.	100.
Kielhorn et al. (2008)	1,000.	Not reported.
Soini et al. (2012)	1,000.	3,000.
Tran-Duy et al. (2014)	1,000.	100,000.
Malottki et al. (2011)	2,000.	5,000.
Carlson et al. (2015)	2,000.	10,000.
Athanasakis et al. (2015)	10,000.	Not reported.

A27.5 Summary of Key Findings

This systematic review identified the specific assumptions that had been made within twenty-nine published model-based economic evaluations for RA that had simulated patients individually. Key structural assumptions were made within each model regarding the sample population, initial response and failure of treatment, disease progression, mortality, direct medical costs, and the calculation of lifetime QALYs. The key assumptions revealed within each model could subsequently be used, through a design-oriented conceptual model, to inform the development of a *de novo* individual-level decision analytic model for RA (Tappenden et al., 2014).

It is possible that the different modelling assumptions were due to differences in the underling decision problems and settings between models, which themselves may not be applicable to the routine management of RA in England. For example, assessing treatment response in terms of an ACR response would not be used in clinical practice in England (National Institute for Health and Care Excellence, 2016). This potential limitation was mitigated by focusing on identifying the *types* of assumptions that were necessary to implement an individual-level decision-analytic model in RA in general, rather than identifying which specific assumptions were most applicable to the decision problem in Section 5.5.

The different structural assumptions between the published individual-level decision analytic models for RA represented different modelling choices made by decision analysts. Structural assumptions are unavoidable when designing a *de novo* decision analytic model for any economic evaluation, and any simplification should be transparent and justified. The findings of this systematic review were therefore able to provide justification for the specific assumptions that were made when designing the DES model in *Chapter Five* of this thesis.

A27. References

Athanasakis, K., Tarantilis, F., Tsalapati, K., Konstantopoulou, T., Vritzali, E., & Kyriopoulos, J. (2015). "Cost-utility Analysis of Tocilizumab Monotherapy in First Line Versus Standard of Care for the Treatment of Rheumatoid Arthritis in Greece", *Rheumatology International*, Vol. 35, 9, pp. 1489-1495.

Bansback, N., Ara, R., Karnon, J., & Anis, A. (2008). "Economic Evaluations in Rheumatoid Arthritis: A Critical Review of Measures Used to Define Health States", *Pharmacoeconomics*, Vol. 26, 5, pp. 395-408.

Bansback, N., Brennan, A., & Ghatnekar, O. (2005a). "Cost Effectiveness of Adalimumab in the Treatment of Patients with Moderate to Severe Rheumatoid Arthritis in Sweden", *Annals of the Rheumatic Diseases*, Vol. 64, 7, pp. 995-1002.

Bansback, N., Regier, D., Ara, R., Brennan, A., Shojania, K., Esdaile, J., Anis, A., & Marra, C. (2005b). "An Overview of Economic Evaluations for Drugs Used in Rheumatoid Arthritis: Focus on Tumour Necrosis Factor-α Antagonists", *Drugs*, Vol. 65, 4, pp. 473-496.

Barton, P. (2011). "Development of the Birmingham Rheumatoid Arthritis Model: Past, Present and Future Plans", *Rheumatology*, Vol. 50, S4, pp. iv32-iv38.

Barton, P., Bryan, S., & Robinson, S. (2004a). "Modelling in the Economic Evaluation of Health Care: Selecting the Appropriate Approach", *Journal of Health Services Research & Policy*, Vol. 9, 2, pp. 110-118.

Barton, P., Jobanputra, P., Wilson, J., Bryan, S., & Burls, A. (2004b). "The use of Modelling to Evaluate New Drugs for Patients with a Chronic Condition: the Case of Antibodies against Tumour Necrosis Factor in Rheumatoid Arthritis", *Health Technology Assessment*, Vol. 8, 11, pp. 1-91.

Brennan, A., Bansback, N., Nixon, R., Madan, J., Harrison, M., Watson, K., & Symmons, D. (2007). "Modelling the Cost Effectiveness of TNF-alpha Antagonists in the Management of Rheumatoid Arthritis: Results from the British Society for Rheumatology Biologics Registry", *Rheumatology*, Vol. 46, 8, pp. 1345-1354.

Brennan, A., Bansback, N., Reynolds, A., & Conway, P. (2004). "Modelling the Cost-effectiveness of Etanercept in Adults with Rheumatoid Arthritis in the UK", *Rheumatology*, Vol. 43, 1, pp. 62-72.

Brennan, A., Chick, S., & Davies, R. (2006). "A Taxonomy of Model Structures for Economic Evaluation of Health Technologies", *Health Economics*, Vol. 15, 12, pp. 1295-1310.

Carlson, J., Ogale, S., Dejonckheere, F., & Sullivan, S. (2015). "Economic Evaluation of Tocilizumab Monotherapy Compared to Adalimumab Monotherapy in the Treatment of Severe Active Rheumatoid Arthritis", *Value in Health*, Vol. 18, 2, pp. 173-179.

Caro, J., Möller, J., Karnon, J., Stahl, J., & Ishak, J. (2016). "Discrete Event Simulation for Health Technology Assessment". (1 ed.) Boca Raton: CRC Press.

Chen, Y., Jobanputra, P., Barton, P., Jowett, S., Bryan, S., Clark, W., Fry-Smith, A., & Burls, A. (2006). "A Systematic Review of the Effectiveness of Adalimumab, Etanercept and Infliximab for the Treatment of Rheumatoid Arthritis in Adults and an Economic Evaluation of their Cost-effectiveness.", *Health Technology Assessment*, Vol. 10, 42, pp. 1-229.

Clark, W., Jobanputra, P., Barton, P., & Burls, A. (2004). "The Clinical and Cost-effectiveness of Anakinra for the Treatment of Rheumatoid Arthritis in Adults: a Systematic Review and Economic Analysis", *Health Technology Assessment*, Vol. 8, 18, pp. 1-105.

Davies, A., Cifaldi, M., Segurado, O., & Weisman, M. (2009). "Cost-effectiveness of Sequential Therapy with Tumor Necrosis Factor Antagonists in Early Rheumatoid Arthritis", *The Journal of Rheumatology*, Vol. 36, 1, pp. 16-26.

Davis, S., Stevenson, M., Tappenden, P., & Wailoo, A. (2014). "Cost-effectiveness Modelling using Patient-level Simulation". NICE DSU Technical Support Document 15, Sheffield: Decision Support Unit.

Diamantopoulos, A., Benucci, M., Capri, S., Berger, W., Wintfeld, N., Giuliani, G., & Ricciardi, W. (2012). "Economic Evaluation of Tocilizumab Combination in the Treatment of Moderate-to-severe Rheumatoid Arthritis in Italy", *Journal of Medical Economics*, Vol. 15, 3, pp. 576-585.

Diamantopoulos, A., Finckh, A., Huizinga, T., Sungher, D., Sawyer, L., Neto, D., & F., D. (2014). "Tocilizumab in the Treatment of Rheumatoid Arthritis: A Cost-effectiveness Analysis in the UK", *Pharmacoeconomics*, Vol. 32, 8, pp. 775-787.

Drummond, M., Barbieri, M., & Wong, J. (2005). "Analytic Choices in Economic Models of Treatments for Rheumatoid Arthritis: What Makes a Difference?", *Medical Decision Making*, Vol. 25, 5, pp. 520-533.

Finckh, A., Bansback, N., Marra, C., Anis, A., Michaud, K., Lubin, S., White, M., Sizto, S., & Liang, M. (2009). "Treatment of Very Early Rheumatoid Arthritis with Symptomatic Therapy, Disease-modifying Antirheumatic Drugs, or Biologic Agents: A Cost-effectiveness Analysis", *Annals of Internal Medicine*, Vol. 151, 9, pp. 612-622.

Ganz, M., Hansen, B., Valencia, X., & Strandberg-Larsen, M. (2015). "Key Data Elements for use in Cost-utility Modeling of Biological Treatments for Rheumatoid Arthritis", *Journal of Medical Economics*, Vol. 18, 5, pp. 366-375.

Hallinen, T., Soini, E., Eklund, K., & Puolakka, K. (2010). "Cost–utility of Different Treatment Strategies after the Failure of Tumour Necrosis Factor Inhibitor in Rheumatoid Arthritis in the Finnish Setting", *Rheumatology*, Vol. 49, 4, pp. 767-777.

Jobanputra, P., Barton, P., Bryan, S., & Burls, A. (2002). "The Effectiveness of Infliximab and Etanercept for the Treatment of Rheumatoid Arthritis: a Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 6, 21, pp. 1-110.

Kielhorn, A. P., D., Diamantopoulos, A., & Lewis, G. (2008). "UK Cost-utility Analysis of Rituximab in Patients with Rheumatoid Arthritis that Failed to Respond Adequately to a Biologic Disease-modifying Antirheumatic Drug", *Current Medical Research and Opinion*, Vol. 24, 9, pp. 2639-2650.

Kobelt, G., Lindgren, P., & Geborek, P. (2009). "Costs and Outcomes for Patients with Rheumatoid Arthritis Treated with Biological Drugs in Sweden: a Model Based on Registry Data", *Scandinavian Journal of Rheumatology*, Vol. 38, 6, pp. 409-418.

Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ionnidis, J., Clarke, M., Devereaux, P., Kleijnen, J., & Moher, D. (2009). "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Healthcare Interventions: Explanation and Elaboration", *British Medical Journal*, Vol. 339, b2700, pp. 1-27.

Lindgren, P., Geborek, P., & Kobelt, G. (2009). "Modeling the Cost-effectiveness of Treatment of Rheumatoid Arthritis with Rituximab using Registry Data from Southern Sweden", *International Journal of Technology Assessment in Health Care*, Vol. 25, 2, pp. 181-189.

Madan, J., Ades, A., & Welton, N. (2011). "An Overview of Models used in Economic Analyses of Biologic Therapies for Arthritis - from Current Diversity to Future Consensus", *Rheumatology*, Vol. 50, S4, pp. iv10-iv18.

Madan, J., Ades, T., Barton, P., Bojke, L., Choy, E., Helliwell, P., Jobanputra, P., Stein, K., Stevens, A., Tosh, J., Verstappen, S., & Wailoo, A. (2015). "Consensus Decision Models for Biologics in Rheumatoid and Psoriatic Arthritis: Recommendations of a Multidisciplinary Working Party", *Rheumatology and Therapy*, Vol. 2, 2, pp. 113-125.

Malottki, K., Barton, P., Tsourapas, A., Uthman, A., Lui, Z., Routh, K., Cannock, M., Jobanputra, P., Moore, D., Fry-Smith, A., & Chen, Y. (2011). "Adalimumab, Etanercept, Infliximab, Rituximab and Abatacept for the Treatment of Rheumatoid Arthritis after the Failure of a Tumour Necrosis Factor Inhibitor: a Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 15, 14, pp. 1-278.

Merkesdal, S., Kirchhoff, T., Wolka, D., Ladinek, G., Kielhorn, A., & Rubbert-Roth, A. (2010). "Cost-effectiveness Analysis of Rituximab Treatment in Patients in Germany with Rheumatoid Arthritis after Etanercept-failure", *The European Journal of Health Economics*, Vol. 11, 1, pp. 95-104.

National Institute for Health and Care Excellence. (2016a). "Adalimumab, Etanercept, Infliximab, Certolizumab Pegol, Golimumab, Tocilizumab, and Abatacept for Rheumatoid Arthritis not Previously Treated with DMARDs or after Conventional DMARDs only have Failed". NICE Technology Appraisal Guidance, TA375. London: National Institute for Health and Care Excellence

Scholz, S., & Mittendorf, T. (2014). "Modeling Rheumatoid Arthritis using Different Techniques - a Review of Model Construction and Results", *Health Economics Review*, Vol. 4, 18, pp. 1-16.

Soini, E., Hallinen, T., Puolakka, K., Vihervaara, V., & Kauppi, M. (2012). "Cost-effectiveness of Adalimumab, Etanercept, and Tocilizumab as First-line Treatments for Moderate-to-severe Rheumatoid Arthritis", *Journal of Medical Economics*, Vol. 15, 2, pp. 340-351.

Stephens, S., Botteman, M., Cifaldi, M., & van Hout, B. (2015). "Modelling the Cost-effectiveness of Combination Therapy for Early, Rapidly Progressing Rheumatoid Arthritis by Simulating the Reversible and Irreversible Effects of the Disease", *BMJ Open*, Vol. 5, 6, e006560, pp. 1-10.

Stevenson, M., Archer, R., Tosh, J., Simpson, E., Everson-Hock, E., Stevens, J., Hernandez-Alava, M., Paisley, S., Dickinson, K., Scott, D., Young, A., & Wailoo, A. (2016). "Adalimumab, Etanercept, Infliximab, Certolizumab Pegol, Golimumab, Tocilizumab, and Abatacept for the Treatment of Rheumatoid Arthritis not Previously Treated with Disease-Modifying Antirheumatic Drugs and After Failure of Conventional Disease-Modifying Antirheumatic Drugs Only: Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 20, 35, pp. 1-610.

Tappenden, P. (2014). "Problem Structuring for Health Economic Model Development". In: Culyer, A. (Ed.) Encyclopaedia of Health Economics Vol. 3. (1 ed.) pp. 168-179. Oxford: Elsevier.

Tosh, J., Brennan, A., Wailoo, A., & Bansback, N. (2011a). "The Sheffield Rheumatoid Arthritis Health Economic Model", *Rheumatology*, Vol. 50, S4, pp. iv26-iv31.

Tosh, J., Stevenson, M., & Akehurst, R. (2014). "Health Economic Modelling of Treatment Sequences for Rheumatoid Arthritis: A Systematic Review", *Current Rheumatology Reports*, Vol. 16, 447, pp. 1-19.

Tosh, J., Wailoo, A., Scott, D., & Deighton, C. (2011b). "Cost-effectiveness of Combination Nonbiologic Disease-modifying Antirheumatic Drug Strategies in Patients with Early Rheumatoid Arthritis", *The Journal of Rheumatology*, Vol. 38, 8, pp. 1593-1600.

Tran-Duy, A., Boonen, A., Kievit, W., van Riel, P., van de Laar, M., & Severens, J. (2014). "Modelling Outcomes of Complex Treatment Strategies Following a Clinical Guideline for Treatment Decisions in Patients with Rheumatoid Arthritis", *Pharmacoeconomics*, Vol. 32, 10, pp. 1015-1028.

Tsao, N., Bansback, N., Shojania, K., & Marra, C. (2012). "The Issue of Comparators in Economic Evaluations of Biologic Response Modifiers in Rheumatoid Arthritis", *Best Practice & Research Clinical Rheumatology*, Vol. 26, 5, pp. 659-676.

Vera-Llonch, M., Massarotti, E., Wolfe, F., Shadick, N., Westhovens, R., Sofrygin, O., Maclean, R., Li, T., & Oster, G. (2008a). "Cost-effectiveness of Abatacept in Patients with Moderately to Severely Active Rheumatoid Arthritis and Inadequate Response to Tumor Necrosis Factor-alpha Antagonists.", *The Journal of Rheumatology*, Vol. 35, 9, pp. 1745-1753.

Vera-Llonch, M., Massarotti, E., Wolfe, F., Shadick, N., Westhovens, R., Sofrygin, O., Maclean, R., Yuan, Y., & Oster, G. (2008b). "Cost-effectiveness of Abatacept in Patients with Moderately to Severely Active Rheumatoid Arthritis and Inadequate Response to Methotrexate", *Rheumatology*, Vol. 47, 4, pp. 535-541.

Wailoo, A., Bansback, N., Brennan, A., Michaud, K., Nixon, R., & Wolfe, F. (2008). "Biologic Drugs for Rheumatoid Arthritis in the Medicare Program: A Cost-effectiveness Analysis", *Arthritis & Rheumatology*, Vol. 58, 4, pp. 939-946.

Wu, B., Song, Y., Leng, L., Bucala, R., & Lu, L. (2015). "Treatment of Moderate Rheumatoid Arthritis with Different Strategies in a Health Resource-limited Setting: A Cost-Effectiveness Analysis in the Era of Biosimilars", *Clinical and Experimental Rheumatology*, Vol. 33, 1, pp. 20-26.

Yuan, Y., Trivedi, D., Maclean, R., & Rosenblatt, L. (2010). "Indirect Cost-effectiveness Analyses of Abatacept and Rituximab in Patients with Moderate-to-severe Rheumatoid Arthritis in the United States", *Journal of Medical Economics*, Vol. 13, 1, pp. 33-41.

Appendix 28: Chapter Six – Complete CHEERS Checklist

This appendix reports the completed twenty-four item *CHEERS* checklist (Table A28.1) for the early model-based economic evaluation in *Chapter Six*. The checklist ensured that the economic evaluation was reported transparently and consistently with the standards recommended for best-practice (Husereau et al. 2013).

Table A28.1. *Completed CHEERS checklist for the early economic evaluation in Chapter Six.*

Section/Item	Item No.	Recommendation	Evidence
Title and Abstract			
Title.	1	Identify the study as an economic evaluation, or use more specific terms such as "cost-effectiveness analysis" and describe the interventions compared.	Title of *Chapter Six*.
Abstract.	2	Provide a structured summary of objectives, perspectives, setting, methods (including study design and inputs), results (including base-case and uncertainty analyses), and conclusions.	Not applicable for PhD thesis.
Introduction			
Background and objectives.	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	Section 6.2.
Methods			
Target population and subgroups.	4	Describe characteristics of the base-case population and subgroups analysed including why they were chosen.	Section 6.3.2.
Setting and location.	5	State the relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Section 6.3.1.4.
Study perspective.	6	Describe the perspective of the study and relate this to the costs being evaluated.	Section 6.3.1.1.

Section/Item	Item No.	Recommendation	Evidence
Comparators.	7	Describe the interventions or strategies being compared and state why they were chosen.	Section 6.3.1.4; Table 6.1.
Time horizon.	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Section 6.3.1.2.
Discount rate.	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Section 6.3.3.2; Section 6.3.3.4.
Choice of health outcomes.	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Section 6.3.3.2.
Measurement of effectiveness.	11	*Synthesis-based estimates*: Describe fully the methods used for the identification of included studies and synthesis of clinical effectiveness data.	Section 6.3.3.1.
Measurement and valuation of preference-based outcomes.	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	Section 6.3.3.2.
Estimating resources and costs.	13	*Model-based economic evaluation*: Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each item in terms of its unit cost. Describe any adjustments made to approximate opportunity costs.	Section 6.3.3.3; Section 6.3.3.4.
Currency, price date, and conversion.	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	Section 6.3.3.4.
Choice of model.	15	Describe and give reasons for the specific type of decision-analytic model used. Providing a figure to show the model structure is strongly recommended.	Section 5.7; Figure 5.9.

Section/Item	Item No.	Recommendation	Evidence
Assumptions.	16	Describe all structural or other assumptions underpinning the decision-analytic model.	Section 6.3.1.5; Section 6.3.3.
Analytic methods.	17	Describe all analytic methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments to a model; and methods for handling population heterogeneity and uncertainty.	Section 6.3.3; Section 6.3.4; Section 6.3.5.
Results			
Study parameters.	18	Report the values, ranges, references, and if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty were appropriate. Providing a table to show the input values is strongly recommended.	Section 6.3.3; Table 6.4; Table 6.6; Appendix 37.
Incremental costs and outcomes.	19	For each intervention, report mean values for the main categories of estimates costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Section 6.4.1; Table 6.7; Table 6.8; Table 6.9.
Characterising uncertainty.	20	*Model-based economic evaluation*: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Section 6.4.2; Appendix 38; Section 6.4.3; Section 6.4.4.
Characterising heterogeneity.	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Not applicable.

Section/Item	Item No.	Recommendation	Evidence
Discussion			
Study findings, limitations, generalisability, and current knowledge.	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Section 6.5.
Other			
Source of funding.	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other nonmonetary sources of support.	Not applicable for individual PhD chapter. Funders acknowledged at the start of the thesis.
Conflicts of interest.	24	Describe any conflicts of interest among study contributors in accordance with journal policy. In the absence of journal policy, we recommend authors comply with International Committee of Medical Journal Editors' recommendations.	Not applicable.

Source: Husereau et al. (2013, pp. 235-236).

A28. References

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A., Mauskopf, J., & Loder, E. (2013). "Consolidated Health Economic Evaluation Reporting Standards (CHEERS) - Explanation and Elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force", *Value in Health*, Vol. 16, 2, pp. 231-250.

Appendix 29: Background to Parametric Survival Analysis

This appendix provides a background to the general method of parametric survival analysis. Parametric survival analysis was performed in *Chapter Six* to estimate the values of time-to-event parameters within the DES model. The appendix is structured by an introduction to survival analysis in economic evaluation (Section A29.1); the types of parametric survival model (Section A29.2); how to select a survival model (Section A29.3); and how to perform a survival analysis in the absence of individual patient-level data (Section A29.4).

A29.1. Introduction to Survival Analysis

Survival analysis is concerned with the estimating the time between two events, such as the time from a patient entering a particular study until their death (Latimer, 2011). The events of interest for this thesis were:

(i) The time to developing ADAb against adalimumab;
(ii) The time to treatment failure;
(iii) The time to a patient's death.

The principal challenge with observed time-to-event data is that, over the duration of a particular study, it is unlikely that all patients will have experienced the event of interest (Altman et al., 1998). Time-to-event data are routinely censored for two reasons: (i) some patients may drop-out of the study resulting in loss to follow-up and (ii) secondly, some patients may not have experienced the event by the conclusion of the study's data collection period (Latimer, 2011). However, decisions regarding the cost-effectiveness of a health technology must be made irrespective of the quality of time-to-event evidence to support those decisions (Davies et al., 2013). Consequently, survival analysis methods can be used to extrapolate analyses beyond the duration observed within a single study (Latimer, 2011). Within a DES, survival analysis is required to define a distribution from which event times can be simulated (by using random numbers) for each individual patient (Ishak et al., 2013).

A29.2. Parametric Survival Models

Parametric survival analysis characterises observed time-to-event data by using a mathematical model that makes an assumption about how the risk of an event changes over time (Ishak et al., 2013). A *hazard function* is the event rate at time *t*, conditional on survival up to time *t*. A *probability density function (F(t))* is the probability that survival time is less than time *t*. A *survivor function (1-F(t))* is the probability that the survival time is at least equal to time *t* (Latimer, 2011).

There are different parametric survival models that can be fit to time-to-event data. The following sections describe the properties of five survival models (exponential, Weibull, Gompertz, log-logistic, and log-normal) that have been recommended for use used when estimating input parameter values for a decision analytic model (Latimer, 2013).

Exponential Distribution

Hazard Function: $h(t) = \alpha \; for \; 0 \leq t < \infty$

Survivor Function: $S(t) = exp^{-\alpha t}$

The exponential distribution is defined by one parameter (α: the rate) and the hazard function is constant over time for all values of *t* (Latimer, 2011).

Weibull Distribution

Hazard Function: $h(t) = \beta\gamma t^{\gamma-1} \; for \; 0 \leq t < \infty$

Survivor Function: $S(t) = \exp(-\beta t^{\gamma})$

The Weibull distribution is defined by two parameters (β: the scale; γ: the shape). The hazard can increase (γ>1) or decrease (γ<1) monotonically over time. The Weibull distribution collapses to an exponential distribution when γ=1 (Latimer, 2011).

Gompertz Distribution

Hazard Function: $h(t) = \beta exp^{\gamma t} \; for \; 0 \leq t < \infty$

Survivor Function: $S(t) = exp\left[\frac{\beta}{\gamma}(1 - exp^{\gamma t})\right]$

The Gompertz distribution is defined by two parameters (β: the scale; γ: the shape). The hazard can increase (γ>0) or decrease (γ<0) monotonically over time (Latimer, 2011).

Log-logistic Distribution

Hazard Function: $h(t) = \frac{exp^{\theta}\delta t^{\delta-1}}{1+exp^{\theta}t^{\delta}}\ for\ 0 \leq t < \infty, \delta > 0$

Survivor Function: $S(t) = \left[1 + exp^{\theta}t^{\delta}\right]^{-1}$

The log-logistic distribution is an accelerated failure time model defined by two parameters (δ and θ). The hazard can decrease with time (δ≤1) or have a single mode (δ>1) where the hazard initially increases and then decreases over time (Latimer, 2011).

Log-normal Distribution

Hazard function: $h(t) = \frac{f(t)}{S(t)}\ for\ 0 \leq t < \infty, f(t) =$ probability density function of t.

Survivor Function: $S(t) = 1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right), \Phi = standard\ normal\ distribution.$

The log-normal distribution has two parameters (μ, σ) with a hazard that initially increases, and then later decreases, with time (Latimer, 2011).

A29.3. Selecting a Parametric Survival Model

Selection of the most appropriate parametric survival model requires a consideration of internal and external validity (Latimer, 2013). Internal validity can be assessed statistically using the AIC and BIC criteria (defined in Section 4.3.3.5.2 and Section 4.3.3.5.3, respectively) to see how the estimated curve fits the observed time-to-event data. The curve with the lowest AIC and BIC values fits the data best (Akaike, 1974; Schwarz, 1978). External validity can be assessed by considering the clinical plausibility of the extrapolation beyond the observed study length. Visual inspection of the extrapolated survival curve can be used to rationalise whether the gradient is clinically plausible, supported by independent clinical evidence or input from experts (Ishak et al., 2013; Connock et al., 2011).

A29.4. Survival Analysis without Individual-level Patient Data

In the absence of individual-level patient data, parametric survival models can be fit using count-time data from published secondary sources. Count-time data report, for each time period in the analysis, the number of patients that experienced the event of interest or were censored. If published time-to-event data are only displayed graphically, software can be

used (such as *DigitizeIt*) (Bormann et al., 2016) to reconstruct the count-time data by plotting the co-ordinates of the graph electronically (Guyot et al., 2012; Diaby et al., 2014; Ishak et al., 2013).

A29. References

Akaike, H. (1974). "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, Vol. 19, 6, pp. 716-723.

Altman, D., & Bland, M. (1998). "Time to Event (Survival) Data", *British Medical Journal*, Vol. 317, 7156, pp. 468-469.

Bormann, I. (2016). "DigitizeIt: Digitizer Software". Version 2.2., Braunschweig: DigitizeIt.

Connock, M., Hyde, C., & Moore, D. (2011). "Cautions Regarding the Fitting and Interpretation of Survival Curves: Examples from NICE Single Technology Appraisals of Drugs for Cancer", *Pharmacoeconomics*, Vol. 29, 10, pp. 827-837.

Davies, C., Briggs, A., Lorgelly, P., Garellick, G., & Malchau, H. (2013). "The "Hazards" of Extrapolating Survival Curves", *Medical Decision Making*, Vol. 33, 3, pp. 369-380.

Diaby, V., Adunlin, G., & Montero, A. (2014). "Survival Modeling for the Estimation of Transition Probabilities in Model-based Economic Evaluations in the Absence of Individual Patient Data: A Tutorial", *Pharmacoeconomics*, Vol. 32, 2, pp. 101-108.

Guyot, P., Ades, A., Ouwens, M., & Welton, N. (2012). "Enhanced Secondary Analysis of Survival Data: Reconstructing the Data from Published Kaplan-Meier Survival Curves", *BMC Medical Research Methodology*, Vol. 12, 9, pp. 1-13.

Ishak, K., Krief, N., Benedict, A., Muszbek, N. (2013). "Overview of Parametric Survival Analysis for Health-economic Applications", *Pharmacoeconomics*, Vol. 31, 8, pp. 663-675.

Latimer, N. (2011). "Survival Analysis for Economic Evaluations alongside Clinical Trials – Extrapolation with Patient-level Data", *NICE DSU Technical Support Document*, No. 14. Sheffield: Decision Support Unit.

Latimer, N. (2013). "Survival Analysis for Economic Evaluations alongside Clinical Trials – Extrapolation with Patient-level Data: Inconsistencies, Limitations, and a Practical Guide", *Medicine Decision Making*, Vol. 33, 6, pp. 743-754.

Schwarz, G. (1978). "Estimating the Dimension of a Model", *The Annals of Statistics*, Vol. 6, 2, pp. 461-464.

Appendix 30: Chapter Six - Application of Coyle et al. (2010)'s Hierarchy of Data Sources for Health Economic Analyses

This appendix provides justification for the sources of evidence that were used to populate the decision analytic model in *Chapter Six* by applying the hierarchy of data sources for health economic analyses by Coyle et al. (2010). The appendix first presents the hierarchy of data sources (Section A30.1) and then applies the hierarchy to all input parameters of the DES in *Chapter Six* (Section A30.2).

A30.1. The Hierarchy of Data Sources for Health Economic Analyses

The hierarchy of data sources for health economic analyses, described by Coyle et al. (2010), explained that (i) different study designs may be more appropriate sources of evidence for different types of input parameter, and (ii) the relevance of a specific data source is as important as its quality. The hierarchy graded the standards of evidence for a model's input parameters on a scale of one (*high quality*) to six (*low quality*). The following five tables outline the hierarchy of evidence for five specific types of input parameter: *Clinical effect size* (Table A30.1); *Baseline clinical data* (Table A30.2); *Resource use* (Table A30.3); *Unit costs* (Table A30.4); and *Utilities* (Table A30.5).

Table A301. *Hierarchy of evidence for clinical effect sizes.*

Rank	Data Components
1+	Meta-analysis of RCTs with direct comparison between comparator therapies, measuring final outcomes.
1	Single RCT with direct comparison between comparator therapies, measuring final outcomes.
2+	Meta-analysis of RCTs with direct comparison between comparator strategies, measuring surrogate outcomes; Meta-analysis of placebo-controlled RCTs with similar trial populations, measuring final outcomes for each individual therapy.
2	Single RCT with direct comparison between comparator therapies, measuring surrogate outcomes; Single placebo-controlled RCTs with similar trial populations, measuring final outcomes for each individual therapy.
3+	Meta-analysis of placebo-controlled RCTs with similar trial populations, measuring surrogate outcomes.
3	Single placebo-controlled RCTs with similar trial populations, measuring surrogate outcomes for each individual therapy.
4	Case-control or cohort study.
5	Non-analytic studies, such as case reports.
6	Expert opinion.

Source: Coyle et al. (2010, p. 107); Abbreviation: RCT=randomised-controlled trial.

Table A30.2. *Hierarchy of evidence for baseline clinical data.*

Rank	Data Components
1	Case series or analysis of reliable administrative databases specifically conducted for the study covering patients solely from the jurisdiction of interest.
2	Recent case series or analysis of reliable administrative databases covering patients solely from the jurisdiction of interest.
3	Recent case series or analysis of reliable administrative databases covering patients from another jurisdiction.
4	Old case series or analysis of reliable administrate databases; Estimates from RCTs.
5	Unsourced estimates from a previously published economic analysis.
6	Expert opinion

Source: Coyle et al. (2010, p. 107); Abbreviation: RCT=randomised-controlled trial.

Table A30.3. *Hierarchy of evidence for resource use.*

Rank	Data Components
1	Prospective data collection or analysis of reliable administrative data from same jurisdiction for specific study.
2	Recently published results of prospective data collection or recent analyses of reliable administrative data from same jurisdiction of interest.
3	Unsourced data from a previous economic evaluation from the same jurisdiction of interest.
4	Recently published results of prospective data collection or recent analysis of reliable administrative from a different jurisdiction.
5	Unsourced data from a previous economic evaluation from a different jurisdiction.
6	Expert opinion.

Source: Coyle et al. (2010, p. 107).

Table A30.4. *Hierarchy of evidence for unit costs.*

Rank	Data Components
1	Cost calculations based on reliable databases or data sources conducted for the specific study, from the same jurisdiction of interest.
2	Recently published cost calculations based on reliable databases or data sources from the same jurisdiction of interest.
3	Unsourced data from a previous economic evaluation, from the same jurisdiction of interest.
4	Recently published cost calculations based on reliable databases or data sources, from a different jurisdiction.
5	Unsourced data from a previous economic evaluation, from a different jurisdiction.
6	Expert opinion.

Source: Coyle et al. (2010, p. 107).

Table A30.5. *Hierarchy of evidence for utilities.*

Rank	Data Components
1	Direct utility assessment for the specific study from a sample: (a) of the general population, (b) with knowledge of the disease of interest, (c) of patients with the disease of interest; Indirect utility assessment from a specific study from a patient sample with the disease of interest: using a tool validated for the patient population.
2	Indirect utility assessment from a patient sample with the disease of interest: using a tool not validated for the patient population.
3	Direct utility assessment from a previous study from a sample: (a) of the general population, (b) with knowledge of the disease of interest, (c) of patients with the disease of interest; Indirect utility assessment from a previous study from a patient sample with the disease of interest: using a tool validated for the patient population;
4	Unsourced utility data from a previous study, with an unknown elicitation method.
5	Patient preference values obtained from a visual analogue scale.
6	Expert opinion.

Source: Coyle et al. (2010, p. 108).

A30.2 Application of the Hierarchy to the Input Parameters in Chapter Six

Table A30.6 describes each input parameter of the decision analytic model in *Chapter Six* in terms of (i) its source of evidence, (ii) the quality standard of this evidence according to the hierarchy by Coyle et al. (2010), and (iii) an explanation for this grade.

Table A30.6. *Application of the hierarchy of evidence to the parameters of the decision analytic model in Chapter Six.*

Parameter	Section in Thesis	Source of Evidence	Grade of Quality	Explanation
Clinical Effect Sizes – Table A31.1				
EULAR response to all treatments	6.3.3.1.3.	Network meta-analysis of RCT evidence by Stevenson et al. (2016).	2+	No RCT existed that directly compared all therapies. Stevenson et al. (2016) was most relevant source of evidence because (i) the final outcomes (EULAR response) were measured for each individual treatment (ii) in the relevant trial population (bDMARD-naïve patients with RA having failed to respond to methotrexate).
HAQ multipliers for monitoring test.	6.3.3.1.12.	Assumption.	6	No data were available to estimate the likely HAQ multiplier associated with a pre-emptive change in treatment due to treatment stratification. Extensive sensitivity analyses were therefore performed on these values.
Baseline Clinical Data – Table A31.2.				
All-cause mortality for men and women.	6.3.3.1.1.	Office for National Statistics (2015) national life tables for England.	2	These data were reliable administrative data that covered patients solely from the jurisdiction of interest.
RA-specific mortality adjustment.	6.3.3.1.2.	Analysis of *National Data Back for Rheumatic Diseases* cohort by Michaud et al. (2012).	3	These data were from reliable administrative databases that covered patients with RA from a different jurisdiction (North America). The source was deemed to be suitable because it was used in the generation of evidence for the *NICE Technology Appraisal 375* for RA by Stevenson et al. (2016).
Time to treatment failure.	6.3.3.1.5.	Meta-analysis of biologic therapy discontinuation studies that used data from registry or health care databases by Souto et al. (2016).	3	These data were from reliable administrative databases that covered patients with RA from different jurisdictions. There was a trade-off between jurisdiction-specific estimates for England (with a smaller sample size) and jurisdiction non-specific estimates (with a larger sample size). It was therefore assumed that the estimated time to failure of a biologic therapy could be generalised across jurisdictions.

Parameter	Section in Thesis	Source of Evidence	Grade of Quality	Explanation
Time to developing adalimumab ADAb.	6.3.3.1.7.	Cohort study by Bartelds et al. (2011) of patients with RA, followed over three years, to identify the development of adalimumab ADAb.	3	These data were a recent cohort study, identified by a systematic review, which recorded the development of ADAb against adalimumab over three years. The data used patients from a different jurisdiction (The Netherlands) but was deemed to be appropriate because of the assumption that TNFi immunogenicity was not heterogeneous across different jurisdictions.
Consequence of developing adalimumab ADAb.	6.3.3.1.8.	Systematic review and meta-analysis by Garcês et al. (2013).	3	These data were sourced from a systematic review that synthesised all available evidence from patients with RA across different jurisdictions.
Time to testing.	6.3.1.4.	Defined by the decision problem.	1	The evidence for time-to-testing was sourced from an extensive model conceptualisation process in *Chapter Five*. The time-to-testing was varied according to different intervention strategies in the cost-effectiveness analysis.
Annual HAQ progression.	6.3.3.1.6.	Evidence from previous NICE Technology Appraisals for RA by Stevenson et al. (2016) and Malottki et al. (2011).	2	The sources of evidence, from which the annual rate of HAQ progression was based, were used in previous NICE Technology Appraisals that were relevant to patients with RA in the jurisdiction of interest.
HAQ reduction following treatment response.	6.3.3.1.4.	Evidence provided by Stevenson et al. (2016) that used a sample of patients with RA in England enrolled to the *BSRBR* register.	2	The source of evidence was a recent analysis of patient-level data that were solely from the jurisdiction of interest.

Parameter	Section in Thesis	Source of Evidence	Grade of Quality	Explanation
Probability of low adalimumab drug levels in remission.	6.3.3.1.9.	Systematic review and meta-analysis of studies that investigated the risk of flare in patients with RA following a de-escalation of a TNFi therapy whilst in remission or low disease activity by Kuijper et al. (2015).	3	The source of evidence was a recent meta-analysis of all evidence (from different jurisdictions) on the proportion of patients with RA that flared following a de-escalation of a TNFi therapy. It was assumed that the proportion of patients that flared corresponded to the proportion of patients with low adalimumab drug levels in remission.
HAQ increase due to flare.	6.3.3.1.10.	Single study that expressed flare in disease activity in terms of a change in HAQ by Markusse et al. (2015).	3	This source provided evidence for patients with RA from a recent study (the BeSt study) in a different jurisdiction (The Netherlands).
Duration of flare.	6.3.3.1.0.	Cohort study reported by Bykerk et al. (2013).	3	This source provided evidence for patients with RA from a recent study (the BRASS Registry) in a different jurisdiction (North America).
ADAb test accuracy.	6.3.3.1.11.	Jani et al. (2016a) performed a ROC analysis for the accuracy of measuring adalimumab ADAb by ELISA.	2	This recent source was good quality evidence because it had evaluated test performance using a sample of patients with RA, treated with adalimumab, within the jurisdiction of interest.
Drug level (full-dose) test accuracy.	6.3.3.1.11.	Appendix 34 reports a systematic review and hierarchical meta-analysis of test accuracy studies.	3	This source of evidence synthesised all available test accuracy data, which included patients from different jurisdictions.
Drug level (half dose) test accuracy.	6.3.3.1.11.	Chen et al. (2016) performed a ROC analysis for the accuracy of measuring adalimumab drug levels by ELISA to predict whether patients would maintain response.	3	This source of evidence evaluated test performance in a sample of patients with RA from another jurisdiction (Taiwan).

Parameter	Section in Thesis	Source of Evidence	Grade of Quality	Explanation
Resource use - Table A31.3.				
Treatment dosing schedules.	6.3.3.3.1.	Recommended dosing schedule according to the *British National Formulary* (2016).	1	High-quality source of evidence because treatments are administered in routine practice in England according to the *British National Formulary*.
Treatment administration.	6.3.3.3.1.	Evidence provided by Stevenson et al. (2016) for the *NICE Technology Appraisal 375*.	3	Study assumed that ten percent of subcutaneous injections were performed by a nurse, to generate evidence that was consistent with previous NICE technology appraisals for RA.
Testing.	6.3.3.3.3.	A microcosting study of testing by Jani et al. (2016b), which accompanied this thesis, was reported in Appendix 35.	1	This source was high-quality evidence because it collected prospective data specifically for this economic evaluation within the same jurisdiction.
Hospitalisations.	6.3.3.3.2.	Based on patient-level data within the *Norfolk Arthritis Register*, produced for submission to a *NICE Single Technology Appraisal* for rituximab.	3	This source of evidence used data that had been used previously in published economic evaluations for RA. The evidence was produced using patient-level data from the same jurisdiction as the study.
Unit Costs - Table A31.4.				
Treatments.	6.3.3.4.1.	*British National Formulary* (2016).	1	High-quality evidence because costs were calculated based on reliable data sources from the same jurisdiction of interest.
Treatment administration.	6.3.3.4.1.	Evidence provided by Stevenson et al. (2016) for the *NICE Technology Appraisal 375*	3	Unit cost of nurse-based subcutaneous injections and intravenous infusions were consistent with the values chosen for previous NICE technology appraisals for RA.

Parameter	Section in Thesis	Source of Evidence	Grade of Quality	Explanation
Testing.	6.3.3.4.3.	A microcosting study of testing by Jani et al. (2016b), which accompanied this thesis, was reported in Appendix 35.	1	Cost calculations were based on reliable databases within the same jurisdiction of the economic evaluation.
Hospitalisations.	6.3.3.4.2.	*NHS National Schedule of Reference Costs* (Department of Health, 2016).	1	Cost calculations were based on reliable databases within the same jurisdiction of the economic evaluation.
Utilities - Table A31.5.				
QALYs.	6.3.3.2.	Estimated by a mapping algorithm (from HAQ to EQ-5D) used by Malottki et al. (2011) when generating evidence for the *NICE Technology Appraisal 195*.	1	This source of evidence was high-quality because it indirectly assessed EQ-5D using (i) a sample a patients with the disease of interest, (ii) and a tool validated for the patient population (HAQ).

A30. References

Bartelds, G., Krieckaert, C., Nurmohamed, M., van Schouwenburg, P., Lems, W., Twisk, J., Dijkmans, B., Aarden, L., & Wolbink, G. (2011). "Development of Antidrug Antibodies Against Adalimumab and Association with Disease Activity and Treatment Failure During Long-term Follow-up", *Journal of the American Medical Association*, Vol. 305, 14, pp. 1460-1468.

British National Formulary. (2016). "British National Formulary: Accessed via MedicinesComplete". London: Pharmaceutical Press.

Bykerk, V., Shadick, N., Frits, M., Bingham, C., Jeffery, I., Iannaccone, C., Weinblatt, M., & Solomon, D. (2014). "Flares in Rheumatoid Arthritis: Frequency and Management. A Report from the BRASS Registry", *The Journal of Rheumatology*, Vol. 41, 2, pp. 227-234.

Chen, D., Chen, Y., Hsieh, T., Hung, W., Hsieh, C., Chen, H., Tang, K., & Lan, J. (2016). "Drug Trough Levels Predict Therapeutic Responses to Dose Reduction of Adalimumab for Rheumatoid Arthritis Patients During 24 Weeks of Follow-up", *Rheumatology*, Vol. 55, 1, pp. 143-148.

Coyle, D., Lee, K., & Cooper, N. (2010). “Chapter 9: Use of Evidence in Decision Models”. In: Shemilt, I., Mugford, M., Vale, L., Marsh, K., & Donaldson, C. (Eds). *Evidence-based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice*. pp. 106-113. Oxford: Wiley-Blackwell.

Department of Health. (2016). "NHS Reference Costs: 2015 to 2016". London: Department of Health.

Garcês, S., Demengeot, J., & Benito-Garcia, E. (2013). "The Immunogenicity of Anti-TNF Therapy in Immune-Mediated Inflammatory Diseases: A Systematic Review of the Literature with a Meta-Analysis", *Annals of the Rheumatic Diseases*, Vol. 73, 1, pp. 1947-1955.

Jani, M., Gavan, S., Dixon, W., Harrison, B., Moran, A., Barton, A., & Payne, K. (2016b). "A Microcosting Study of Immunogenicity and TNFi Drug Level Tests for Therapeutic Monitoring in Clinical Practice", Rheumatology, Vol. 55, 12, pp. 2131-2137.

Jani, M., Isaacs, J., Morgan, A., Wilson, A., Plant, D., Hyrich, K., Chinoy, H., & Barton, A. (2016a). "Detection of Anti-drug Antibodies using a Bridging ELISA Compared with Radioimmunoassay in Adalimumab-treated Rheumatoid Arthritis Patients with Random Drug Levels", *Rheumatology*, Vol. 55, 11, pp. 2050-2055.

Kuijper, T., Lamers-Karnebeek, F., Jacobs, J., Hazes, J., & Luime, J. (2015). "Flare Rate in Patients with Rheumatoid Arthritis in Low Disease Activity or Remission When Tapering or Stopping Synthetic or Biologic DMARD: A Systematic Review", *The Journal of Rheumatology*, Vol. 42, 11, pp. 2012-2022.

Malottki, K., Barton, P., Tsourapas, A., Uthman, A., Lui, Z., Routh, K., Cannock, M., Jobanputra, P., Moore, D., Fry-Smith, A., & Chen, Y. (2011). "Adalimumab, Etanercept, Infliximab, Rituximab and Abatacept for the Treatment of Rheumatoid Arthritis after the Failure of a Tumour Necrosis Factor Inhibitor: a Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 15, 14, pp. 1-278.

Markusse, I., Dirven, L., Gerards, A., van Groenendal, J., Ronday, H., Kerstens, P., Lems, W., Huizinga, T., & Allaart, C. (2015). "Disease Flares in Rheumatoid Arthritis are Associated with Joint Damage Progression and Disability: 10-year Results from the BeSt Study", *Arthritis Research & Therapy*, Vol. 17, 232, pp. 1-9.

Michaud, K., Vera-Llonch, M., & Oster, G. (2012). "Mortality Risk by Functional Status and Health-Related Quality of Life in Patients with Rheumatoid Arthritis", *The Journal of Rheumatology*, Vol. 39, 1, pp. 54-59.

Office for National Statistics. (2015). "National Life Tables: England (2012-2014)". Newport: Office for National Statistics.

Souto, A., Maneiro, J., & Gómez-Reino, J. (2016). "Rate of Discontinuation and Drug Survival of Biologic Therapies in Rheumatoid Arthritis: A Systematic Review and Meta-Analysis of Drug Registries and Health Care Databases", *Rheumatology*, Vol. 55, 3, pp. 523-534.

Stevenson, M., Archer, R., Tosh, J., Simpson, E., Everson-Hock, E., Stevens, J., Hernandez-Alava, M., Paisley, S., Dickinson, K., Scott, D., Young, A., & Wailoo, A. (2016). "Adalimumab, Etanercept, Infliximab, Certolizumab Pegol, Golimumab, Tocilizumab, and Abatacept for the Treatment of Rheumatoid Arthritis not Previously Treated with Disease-Modifying Antirheumatic Drugs and After Failure of Conventional Disease-Modifying Antirheumatic Drugs Only: Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 20, 35, pp. 1-610.

Appendix 31 – Chapter Six - Estimating Time to Death

The decision analytic model in *Chapter Six* required an estimate of each patient's time to death. National life table data provide a representative empirical distribution of mortality by age and sex. By fitting a parametric survival curve (see Appendix 29) to these observed data, the model could draw random numbers to simulate each patient's age of death. This appendix reports how parametric survival analysis was performed to estimate a time-to-event distribution for all-cause mortality in England.

A31.1. Method

The most recent national life table data (for the years 2012-2014) for England were obtained from the *Office for National Statistics* (2015). These life table data provided, for a representative 100,000 individuals, the number of individuals expected to die annually (by age) from all causes.

The five parametric survival curves described in *Appendix 29* (exponential, Weibull, Gompertz, log-logistic, log-normal) were fit to these all-cause mortality data. The parametric survival curve used to simulate time to death in the decision analytic model was chosen according to the lowest AIC and BIC statistics (Akaike, 1974; Schwarz, 1978), and by visual inspection to ensure biologic plausibility of the estimated survival curve. Parametric survival analysis was estimated separately on life table data for men and women to account for differences in all-cause mortality by sex. All analyses were performed in *STATA Version 13* (StataCorp, 2013).

A31.2. Results

The Kaplan-Meier curves for all-cause mortality are plotted separately for men and women in Figure A31.1 and Figure A31.2, respectively.

Figure A31.1. *Kaplan-Meier national life table data for the United Kingdom: men.*

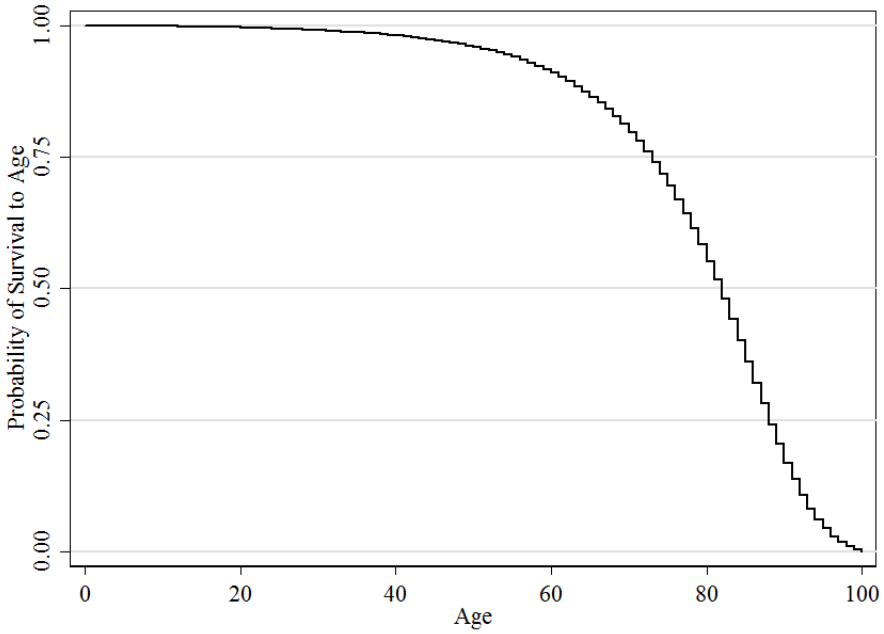
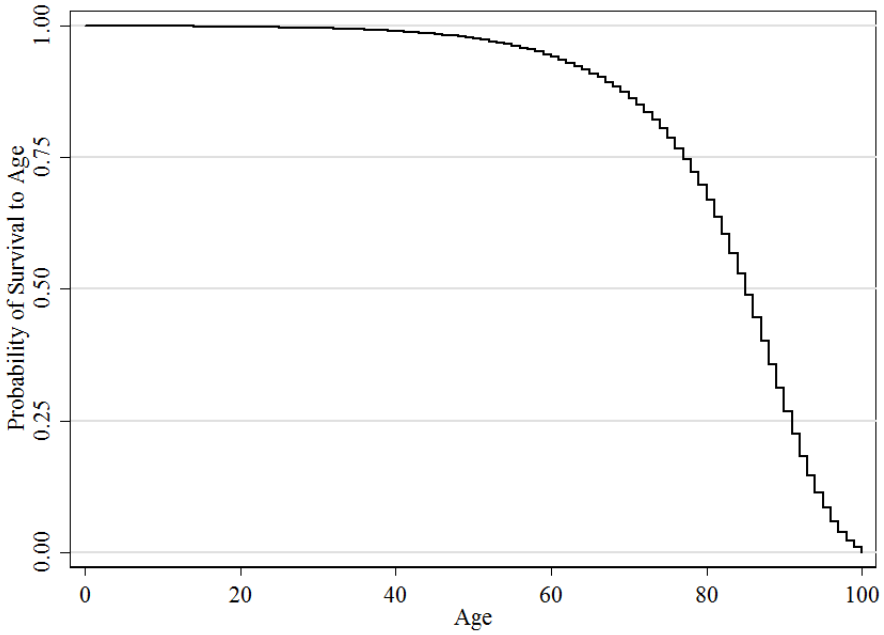


Figure A31.2. *Kaplan-Meier national life table data for the United Kingdom: women.*



The lifetable distribution for men was further to the left than for women, indicating a greater risk of all-cause mortality for men.

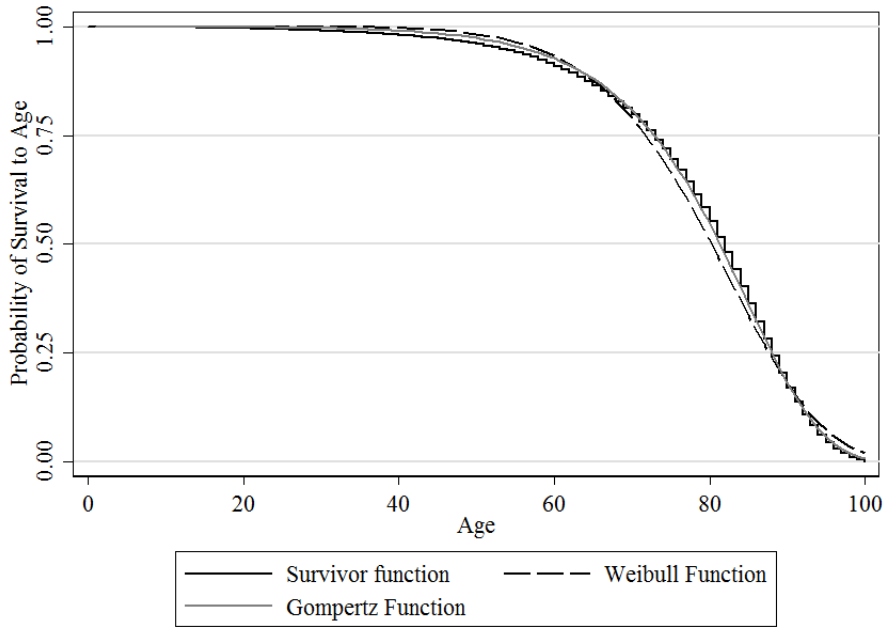
The parameter estimates from estimating the parametric survival curves using male life table data are reported in Table A31.1.

Table A31.1. *Parametric survival analysis on national life table data for the United Kingdom: Men.*

	Parametric Survival Distribution				
Parameter	**Exponential**	**Weibull**	**Gompertz**	**Log-normal**	**Log-logistic**
Rate	0.012653				
Shape		7.886955	0.103924	0.240344	0.099351
Scale		6.62E-16	1.54E-05		4.382697
Location				4.348525	
AIC	201954.5	-81196.6	**-95104.1**	-1329.86	-50264.2
BIC	201964	-81177.6	**-95085.1**	-1310.85	-50245.2

The Gompertz survival curves had the lowest AIC and BIC test statistics, indicating that it fit the observed data best. The Weibull survival function was the second-best fit to the observed data. The estimated Gompertz and Weibull survival functions have been overlaid on the Kaplan-Meier survival curve for men in Figure A31.3 (other functions were omitting from the figure to ease interpretation).

Figure A31.3. *Estimated Gompertz and Weibull survival functions with Kaplan-Meier national life table data: men.*



The Gompertz curve (grey line) was closer to the observed survival data over all time periods compared to the Weibull curve (dashed line). Therefore, the decision analytic model used a Gompertz survival curve to estimate the time to death for male patients, with a shape parameter of 0.103924 and a scale parameter of 0.0000154.

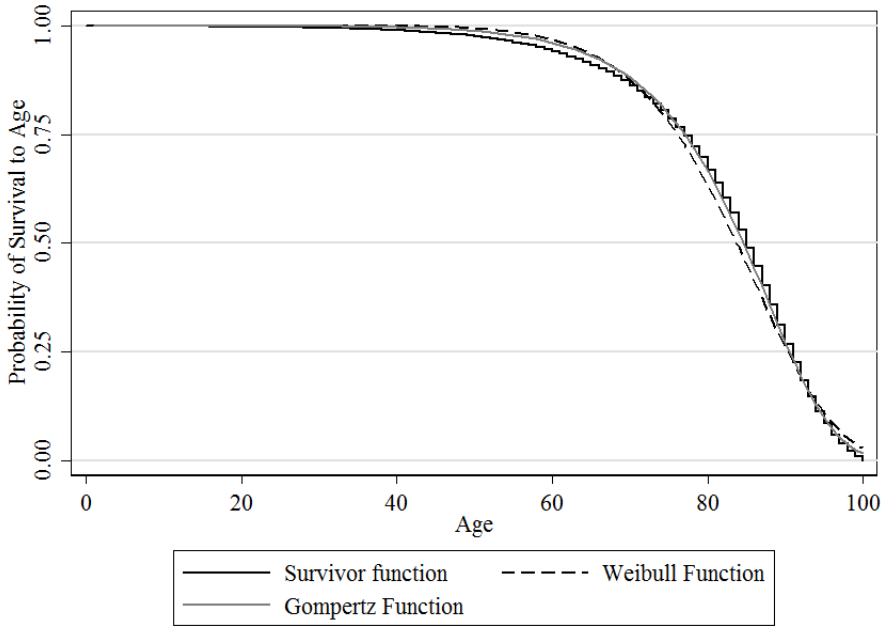
The parameter estimates from estimating the parametric survival curves using the national life table data for women are reported in Table A31.2.

Table A31.2. *Parametric survival analysis on national life table data for the United Kingdom: Women.*

	Parametric Survival Distribution				
Parameter	**Exponential**	**Weibull**	**Gompertz**	**Log-normal**	**Log-logistic**
Rate	0.012124				
Shape		9.210187	0.1162922	0.211123	4.425642
Scale		1.35E-18	4.307E-06		0.085314
Location				4.396185	
AIC	198857	-110419.6	**-123841.5**	-26677.1	-79522
BIC	198866.5	-110400.7	**-123822.5**	-26658.1	-79503

Similar to the mortality data for men, the Gompertz survival distribution fit the all-cause mortality data for women best (lowest AIC and BIC) followed by the Weibull survival distribution. The Gompertz and Weibull survival curves were overlaid on the Kaplan-Meier survival curve for women in Figure A31.4; the Gompertz curve (grey line) was also closer to the observed survival data over time compared with the Weibull curve (dashed line). Therefore, the decision analytic model used a Gompertz survival function to estimate the time to death for women patients, with a shape parameter of 0.1162922 and a scale parameter of 0.000004307.

Figure A31.4. *Estimated Gompertz and Weibull survival functions with Kaplan-Meier national life table data: women.*



A31. References

Akaike, H. (1974). "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, Vol. 19, 6, pp. 716-723.

Office for National Statistics. (2015). “National Life Tables: England (2012-2014)”. Newport: Office for National Statistics.

Schwarz, G. (1978). "Estimating the Dimension of a Model", *The Annals of Statistics*, Vol. 6, 2, pp. 461-464.

StataCorp. (2013). "Stata Statistical Software". Release 13, College Station, TX: StataCorp LP.

Appendix 32 – Chapter Six - Estimating Time to Treatment Failure

The decision analytic model in *Chapter Six* required an estimate of the time that each patient would lose response to each bDMARD treatment. These time-to-event values were estimated by randomly sampling from a parametric survival curve that was fit to secondary data. This appendix describes how these parametric survival curves were estimated and selected.

A32.1. Method

The times to failing all bDMARD therapies (adalimumab, rituximab, and tocilizumab) were assumed to be sampled from the same parametric survival curve, due to limited data availability, following the modelling approach assumed by Stevenson et al. (2016) for the previous *NICE Technology Appraisal 375* of bDMARD therapies for RA.

Souto et al. (2016) reported a systematic review and meta-analysis of studies that estimated the discontinuation of bDMARD therapies in patients with RA using data from registry or health care databases. The annual percentage of patients that discontinued any TNFi therapy over four years since starting treatment, estimated by the meta-analysis, is reported in Table A32.1.

Table A32.1. *Annual percentage of patients that discontinued TNFi therapy: meta-analysis estimate from Souto et al. (2016).*

Time (Years)	Discontinue any TNFi (%)	95% Confidence Interval
1	27%	23% - 32%
2	37%	35% - 40%
3	44%	40% - 49%
4	52%	46% - 57%

Source: Souto et al. (2016, p.525).

The meta-analysis demonstrated that approximately half of patients with RA had lost response to their therapy by four years from first commencing a TNFi. A parametric survival analysis was required to extrapolate beyond the four-year follow-up period to estimate the time-to-treatment failure for the remaining patients that were still responding to therapy.

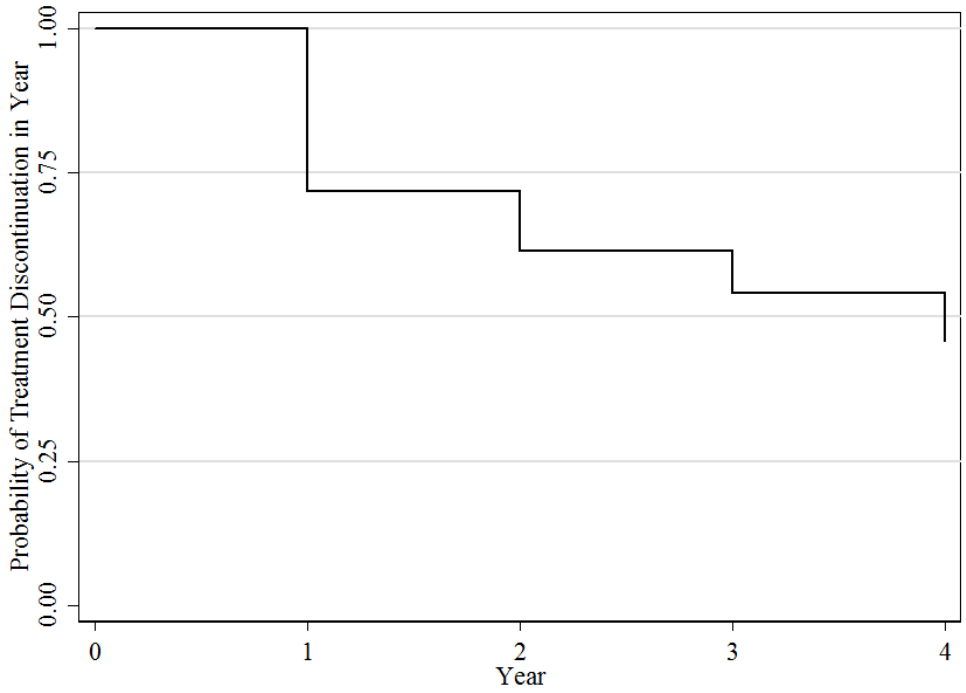
Count-time Kaplan-Meier curves were constructed for 100,000 hypothetical patients over a duration of four years by applying the annual percentage of patients that discontinued

therapy reported in Table A32.1. The five parametric survival curves reported in *Appendix 29* (exponential, Weibull, Gompertz, log-normal, and log-logistic) were estimated using the observed data on treatment discontinuation. The most appropriate parametric form was selected for the decision analytic model according to the lowest AIC and BIC values (Akaike, 1974; Schwarz, 1978), and according to clinical plausibility by visual inspection of the extrapolated distribution. All survival analyses were performed in *STATA Version 13* (Statacorp, 2013).

A32.2. Results

The Kaplan-Meier curve of treatment failure for the first four years after commencing therapy is illustrated in Figure A32.1.

Figure A32.1. *Kaplan-Meier bDMARD treatment failure in RA.*



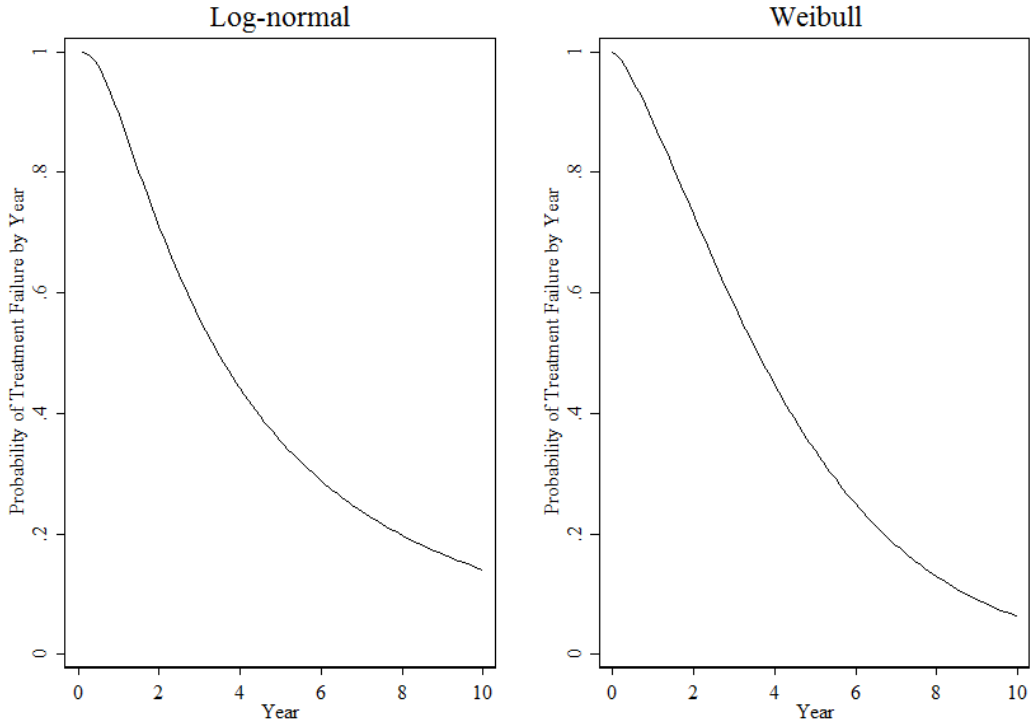
The parameters that were estimated to define the parametric survival curves that were fit to these data, and the summary statistics to assess the fit of each curve to the data, are reported in Table A32.2.

Table A32.2. *Parametric survival analysis on data for bDMARD failure in RA.*

	Parametric Survival Distribution				
Parameter	**Exponential**	**Weibull**	**Gompertz**	**Log-normal**	**Log-logistic**
Rate	0.188406	N/A	N/A	N/A	N/A
Shape	N/A	1.351142	0.132325	0.985526	0.61154
Scale	N/A	4.708286	0.147838	N/A	1.236764
Location	N/A	N/A	N/A	1.240371	N/A
AIC	226170.1	220932.9	224925.3	**212247.4**	217461.2
BIC	226179.6	220951.9	224944.2	**212266.3**	217480.2

The log-normal survival curve fit the observed treatment failure data best and had the lowest values AIC and BIC test statistics. However, the most common way to model treatment failure in previously published individual-level simulation models for RA was to use a Weibull survival curve (see *Appendix 27; Treatment Failure*). Figure A32.2 plots the estimated log-normal and Weibull survival curves, extrapolated over ten years, to enable a comparison between the two functional forms.

Figure A32.2. *Plot of estimated log-normal and Weibull survival curves, extrapolated over ten years.*



The log-normal and Weibull curves followed a similar trajectory up to year four, for the time period where the meta-analysis data were available. However, after year four, the two curves diverged in their extrapolation of the likelihood of treatment failure. The Weibull extrapolation was more conservative than the log-normal curve because it estimated that treatment failure was more likely to occur earlier than estimated by the log-normal curve. The estimated Weibull survival curve therefore appeared to have greater clinical plausibility, irrespective of the statistical fit of the distribution. The base case analysis of

the decision analytic model simulated the time to bDMARD treatment failure from a Weibull survival curve with a shape parameter of 1.351142 and a scale parameter of 4.708286. The impact of this structural assumption was explored in a sensitivity analysis that replaced the Weibull curve in the model with the estimated log-normal curve.

A32. References

Akaike, H. (1974). "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, Vol. 19, 6, pp. 716-723.

Souto, A., Maneiro, J., & Gómez-Reino, J. (2016). "Rate of Discontinuation and Drug Survival of Biologic Therapies in Rheumatoid Arthritis: A Systematic Review and Meta-analysis of Drug Registries and Health Care Databases", *Rheumatology*, Vol. 55, 3, pp. 523-534.

Schwarz, G. (1978). "Estimating the Dimension of a Model", *The Annals of Statistics*, Vol. 6, 2, pp. 461-464.

StataCorp. (2013). "Stata Statistical Software". Release 13, College Station, TX: StataCorp LP.

Stevenson, M., Archer, R., Tosh, J., Simpson, E., Everson-Hock, E., Stevens, J., Hernandez-Alava, M., Paisley, D., Dickinson, K., Scott, D., Young, A., & Wailoo, A. (2016). "Adalimumab, Etanercept, Infliximab, Certolizumab Pegol, Golimumab, Tocilizumab and Abatacept for the Treatment of Rheumatoid Arthritis not Previously Treated with Disease-modifying Antirheumatic Drugs and after the Failure of Conventional Disease-modifying Antirheumatic Drugs only: Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 20, 35, pp. 1-610.

Appendix 33: Chapter Six – Estimating the Time to Developing Adalimumab Anti-drug Antibodies

The decision analytic model in *Chapter Six* required an estimate for each patient's time to developing ADAb against adalimumab that was simulated from a parametric survival curve. However, no individual-level data were available to perform a patient-level survival analysis. Therefore, the parameters of the survival curve were estimated by using evidence identified by a systematic review of the published clinical literature. This appendix reports how these parameters were estimated for the parametric survival curve used in the decision analytic model, from which an individual's time to developing adalimumab ADAb were simulated.

A33.1. Method

The method section of this appendix describes how (i) published studies were identified that included evidence of the time to developing ADAb against adalimumab in patients with RA, and (ii) how parametric survival analyses were performed with these data.

Systematic Review

A systematic review was conducted to identify all published studies that met the inclusion criteria in Table A33.1 that explicitly reported the time taken for patients with RA to develop ADAb against adalimumab.

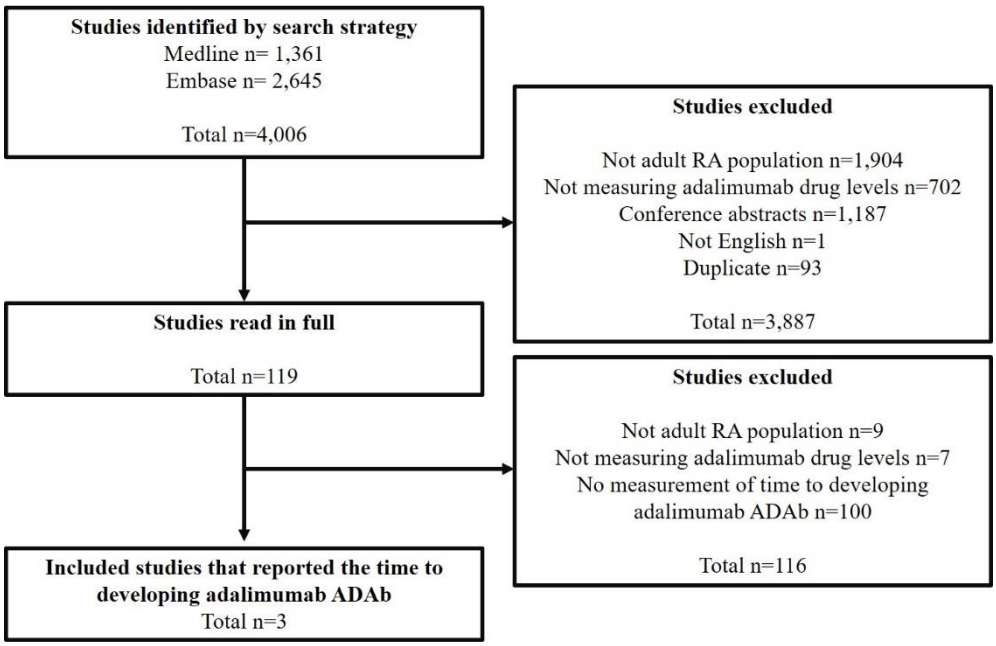
Table A33.1. *Inclusion criteria for identifying temporal studies assessing the development of adalimumab ADAb.*

Study Feature	Inclusion Criteria
Population	Adults with RA, receiving 40mg adalimumab every two weeks.
Intervention	Any test to detect adalimumab ADAb.
Comparator	None.
Outcome	Time to developing ADAb.
Study design	Any study with a longitudinal design.
Language	English.

An electronic search strategy (*Appendix 23*) was used to identify the titles and abstracts of the published clinical studies. *Medline* and *Embase* were searched electronically from the

date of inception until August 2016. All abstracts were screened by SG and second-screened independently by six members of the *Manchester Centre for Health Economics, The University of Manchester*. Studies were not excluded if there was a disagreement between SG and the second-screening process. Abstracts that remained after screening were read in full by SG and evaluated against the inclusion criteria. A flow diagram of included studies is illustrated in Figure A33.1.

Figure A33.1. *Flow diagram of included studies that assessed the time to developing adalimumab ADAb.*



The search strategy identified 4,006 abstracts. Three studies were identified that reported a time to developing adalimumab ADAb across a sample of patients with RA (Bartelds et al., 2011; Krickaert et al., 2012; van Schouwenburg et al., 2013). Krickaert et al. (2012) reported a time-to-developing ADAb, stratified by the dose of concomitant methotrexate, and was of limited relevance to this thesis because all patients were assumed to be prescribed the maximum dose of methotrexate in the decision analytic model. van Schouwenburg et al. (2013) reported the percentage of patients that developed adalimumab ADAb over time, but failed to report the number of patients censored from the analysis at each time point. Bartelds et al. (2011) reported the time to developing adalimumab ADAb across a starting cohort of 272 patients with RA at nine time intervals over three years, and reported the number of patients that remained in the study at each time point. Therefore, the economic evaluation in *Chapter Six* used the data from Bartelds et al. (2011) to estimate the time to developing adalimumab ADAb by accounting for sample attrition.

Data Extraction

The number of patients that were censored at each time point in Bartelds et al. (2011) was reported in the study. However, the number of patients that developed adalimumab ADAb were reported graphically. The computer software *DigitizeIt* (Bormann, 2016) was used to extract data points from the graph to estimate the number of corresponding patients that developed ADAb at each follow-up period, as recommended by Guyot et al. (2012). Table A33.2 reports the number of patients with RA in Bartelds et al. (2011) that developed adalimumab ADAb over three years**.**

Table A33.2. *Estimated number of patients to develop adalimumab ADAb over three years, reported in Bartelds et al. (2011).*

Week	Patients that Remained in study (n)	Patients that were Censored (n)	Patients that develop ADAb (n)
0	272	0	0
4	261	11	22
16	247	14	22
28	228	19	7
40	201	27	3
52	192	9	3
78	175	17	9
104	156	19	5
130	137	19	3
156	118	19	2

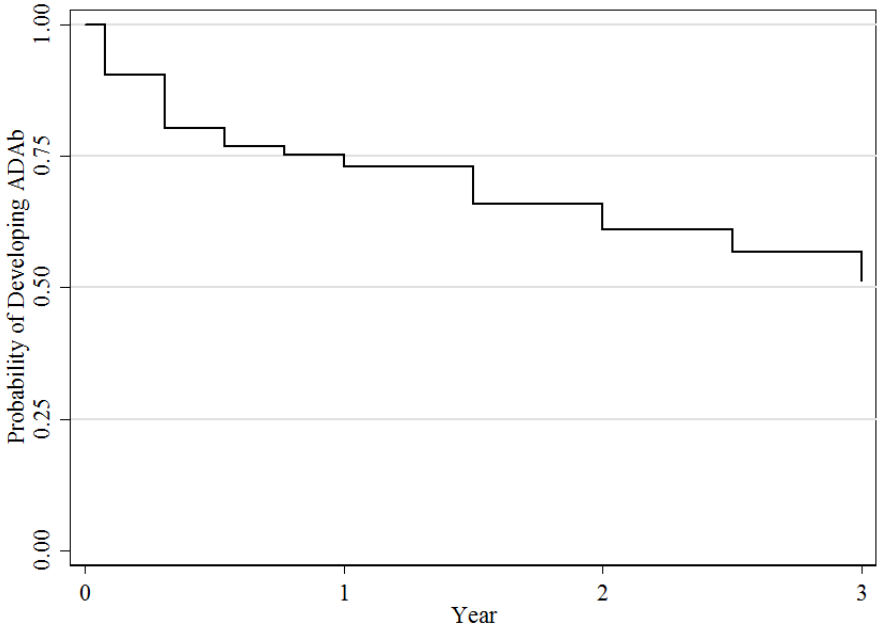
Parametric Survival Analysis

The five parametric survival curves reported in *Appendix 29* (exponential, Weibull, Gompertz, log-normal, and log-logistic) were estimated using the data from Bartelds et al. (2011) in Table A33.2. The most appropriate parametric form was selected for the decision analytic model according to the lowest AIC and BIC values (Akaike, 1974; Schwarz, 1978), and according to clinical plausibility by visual inspection of the extrapolated distribution. All survival analyses were performed in *STATA Version 13* (StataCorp, 2013).

A33.2. Results

The Kaplan-Meier curve that illustrated the time-to-event data for developing adalimumab ADAb is reported in Figure A33.2.

Figure A33.2. *Kaplan-Meier curve for the time to developing adalimumab ADAb based on Bartelds et al. (2011).*



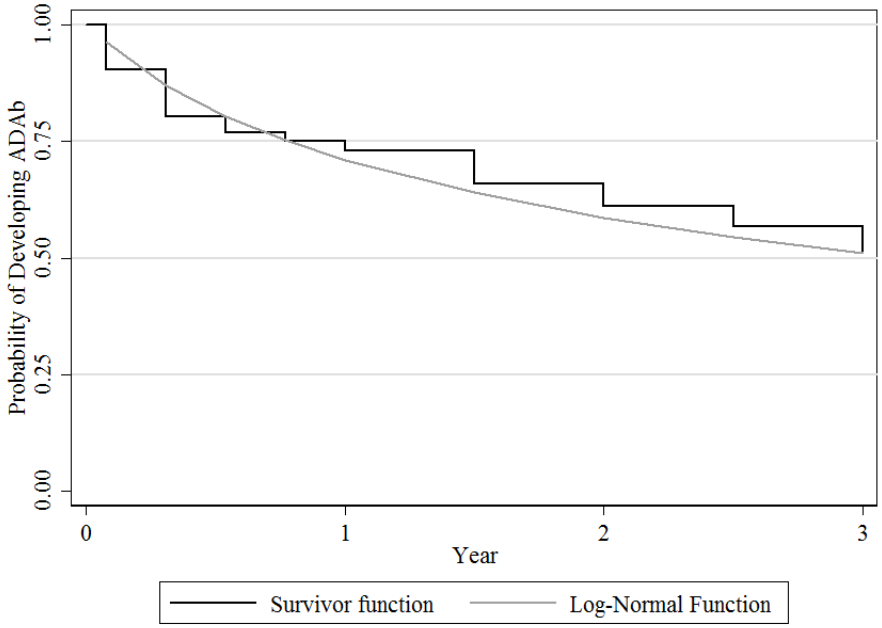
The results from fitting the five parametric survival curves to the data from Bartelds et al. (2011), to extrapolate beyond the three-year follow-up, is reported in Table A33.2. The log-normal distribution had the lowest AIC and BIC values and therefore fit the data best.

Table A33.2. *Parametric survival analysis using data from Bartelds et al. (2011) on time to developing adalimumab ADAb.*

	Parametric Survival Distribution				
Parameters	**Exponential**	**Weibull**	**Gompertz**	**Log-normal**	**Log-logistic**
Rate	0.283646				
Shape		0.744583	-0.52128	0.728429	1.185381
Scale		0.317068	0.439401		1.09911
Location				1.146684	
AIC	494.2614	486.1691	486.7302	**478.8197**	484.3277
BIC	497.6995	493.0453	493.6063	**485.6958**	491.2038

Figure A33.3 plots the estimated log-normal survival curve on the observed Kaplan-Meier curve to illustrate the fit. The estimated time to developing ADAb appeared to be conservative after one year of receiving treatment because the log-normal function lay below the observed data.

Figure A33.3. *Estimated log-normal survival function with Kaplan-Meier data for time to ADAb development.*



The decision analytic model therefore simulated each patient's time to developing adalimumab ADAb from a log-normal survival curve, defined by a location parameter of 1.146684 and a scale parameter of 0.7284289.

A33. References

Akaike, H. (1974). "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, Vol. 19, 6, pp. 716-723.

Bartelds, G., Krieckaert, C., Nurmohamed, M., van Schouwenburg, P., Lems, W., Twisk, J., Dijkmans, B., Aarden, L., & Wolbink., G. (2011). "Development of Antidrug Antibodies Against Adalimumab and Association with Disease Activity and Treatment Failure During Long-term Follow-up", *Journal of the American Medical Association*, Vol. 305, 14, pp. 1460-1468.

Bormann, I. (2016). "DigitizeIt: Digitizer Software". Version 2.2., Braunschweig: DigitizeIt.

Guyot, P., Ades, A., Ouwens, M., & Welton, N. (2012). "Enhanced Secondary Analysis of Survival Data: Reconstructing the Data from Published Kaplan-Meier Survival Curves", *BMC Medical Research Methodology*, Vol. 12, 9, pp. 1-13.

Krieckaert, C., Nurmohamed, M., Wolbink, G. (2012). "Methotrexate Reduces Immunogenicity in Adalimumab Treated Rheumatoid Arthritis Patients in a Dose Dependent Manner", *Annals of the Rheumatic Diseases*, Vol. 71, 11, pp. 1914-1915.

Schwarz, G. (1978). "Estimating the Dimension of a Model", *The Annals of Statistics*, Vol. 6, 2, pp. 461-464.

StataCorp. (2013). "Stata Statistical Software". Release 13, College Station, TX: StataCorp LP.

van Schouwenburg, P., Krieckaert, C., Rispens, T., Aarden, L., Wolbink, G., & Wouters, D. (2013). “Long-term Measurement of Anti-adalimumab using pH-shift-anti-idiotype Antigen Binding Test Shows Predictive Value and Transient Antibody Formation”, *Annals of the Rheumatic Diseases*, Vol. 71, 10, pp. 1680-1686.

Appendix 34: Chapter Six - Systematic Review of Test Accuracy Studies for Measuring Adalimumab Anti-drug Antibodies and Drug Levels by ELISA

The decision analytic model in *Chapter Six* required evidence for the accuracy of using the ELISA tests to measure adalimumab ADAb and drug levels in three scenarios, as described in Section 6.3.3.1.11:

(i) The accuracy of measuring adalimumab ADAb by ELISA, compared with measurement by radioimmunoassay;
(ii) The accuracy of measuring adalimumab drug levels by ELISA to identify patients that would maintain response to full-dose adalimumab;
(iii) The accuracy of measuring adalimumab drug levels by ELISA to identify patients that would maintain response to reduced-dose adalimumab.

However, the accuracy of measuring adalimumab ADAb and drug levels by ELISA was unknown, *a priori*. NICE require evidence of test accuracy, used to inform the values of input parameters in a model-based economic evaluation, to be identified by a systematic review (National Institute for Health and Care Excellence, 2011).

Test accuracy studies compare the performance of an index test with a reference standard when detecting a target condition (Macaskill et al., 2010). An *index test* is the test of interest with unknown accuracy. The *target condition* is the health condition that is being detected. A *reference standard* is the existing best method to identify the target condition. Test accuracy studies comprise a receiver operating characteristic (ROC) curve analysis to estimate the sensitivity and specificity of an index test, relative to a reference standard, at a particular test cut-point (Macaskill et al., 2010).

A34.1. Aim and Objectives

The aim of this study was to identify evidence to inform the accuracy of using adalimumab ADAb and drug level testing by ELISA in patients with RA. There were two objectives to meet this aim:

Objective 1: Identify all studies that had estimated the accuracy of testing adalimumab (i) ADAb by ELISA, compared with a radioimmunoassay, in patients with RA; (ii) drug levels by ELISA to identify patients with RA that would maintain response to full-dose adalimumab; and (iii) drug levels by ELISA to identify patients with RA that would maintain response to reduced-dose adalimumab.

Objective 2: To synthesise the evidence from multiple test accuracy studies to estimate input parameter values for the decision analytic model regarding test sensitivity and specificity.

A34.2. Method

This study conducted a systematic review and hierarchical meta-analysis of test accuracy studies for measuring adalimumab ADAb and drug levels by ELISA in patients with RA. The method section describes the systematic review (Section A34.2.1), quality assessment (Section A34.2.2), and the hierarchical meta-analysis (Section A34.2.3).

A34.2.1. Systematic Review

The study inclusion criteria, to identify all published test accuracy studies relevant for the economic evaluation, is reported in Table A34.1. The search strategy reported in *Appendix 23* was used to identify the titles and abstracts of published test accuracy studies. *Medline* and *Embase* were searched electronically from the date of inception until August 2016. All abstracts were screened by SG and second-screened by six researchers at the *Manchester Centre for Health Economics, The University of Manchester*. Abstracts were not excluded if there were disagreements at the screening stage. Abstracts that remained after screening were read in full by SG and evaluated against the inclusion criteria in Table A34.1.

The following data were extracted from each study included in the review by SG: (i) study characteristics (country, ELISA test used, test cut-point, definition of the target condition, sample size, treatment regime, test sensitivity and specificity) and (ii) sample summary statistics (baseline mean DAS28, age, proportion of women, proportion receiving concomitant methotrexate). If more than two studies were identified for a particular testing scenario, to synthesise the published evidence, the proportion of patients classified as true-positive, false-positive, true-negative, and false-negative by the ELISA were also extracted. Patient classifications were obtained directly from the 2x2 table of test accuracy

reported in each manuscript. If a 2x2 table was not reported, the values were calculated by SG according to the number of patients with the target condition and the reported ELISA sensitivity and specificity. If calculation was not possible, the authors of the manuscript were contacted by SG to obtain the true-positive, false-positive, true-negative, and false-negative values of the test.

Table A34.1. *Inclusion criteria for systematic review of drug level test accuracy studies*.

Study feature	Inclusion criteria
Population	Adults with rheumatoid arthritis receiving 40mg adalimumab every two weeks.
Intervention (Index test)	(i) ELISA to measure adalimumab ADAb; (ii) ELISA to measure adalimumab drug levels for detection of treatment response with full-dose adalimumab; (iii) ELISA to measure adalimumab drug levels for detection of treatment response with reduced-dose adalimumab.
Comparator (Reference standard)	(i) Radioimmunoassay; (ii) Observed treatment response; (iii) Observed treatment response.
Outcome (Target condition)	(i) Detection of adalimumab ADAb; (ii) Treatment response; (iii) Treatment response.
Study design	ROC analysis of test accuracy in a peer-reviewed publication (exclude conference abstracts).
Language	English.

Abbreviations: ADAb=Anti-drug Antibody; ELISA=Enzyme-linked Immunosorbent Assay; ROC=Receiver Operating Characteristic.

A34.2.2. Quality Assessment

The quality of each study in the review was assessed according to the QUADAS-2 (Whiting et al., 2011), as recommended by the NICE DAP (National Institute for Health and Care Excellence, 2011). The QUADAS-2 is a checklist to assess the quality of diagnostic accuracy studies across four domains (patient selection, the index test, the reference standard, and the flow of patients through the study) (Whiting et al., 2011). The results of the quality assessment were presented in graphical form if multiple studies were identified, independently for risk of bias and concern regarding study applicability.

A34.2.3. Hierarchical Meta-analysis

The appropriate methods for a meta-analysis of test accuracy studies are different to the conventional methods for a meta-analysis of relative treatment effects (Macaskill et al., 2010). Heterogeneity is often present between test accuracy studies and may arise from a number of sources, including the patient population and the protocol for using the index

test and reference standard (Dinnes et al., 2005; Reitsma et al., 2005). The cut-point, used to define whether the result of an index test is *positive*, may vary between test accuracy studies when not defined in advance (Macaskill et al., 2010). The higher the cut-point, the lower the sensitivity and the higher the specificity of the index test (Dinnes et al., 2005). Different cut-points between different studies will induce a negative correlation between the study-specific estimates of sensitivity and specificity, indicating that they should not be treated as independent values in a meta-analysis (Dinnes et al., 2005). Consequently, hierarchical methods for meta-analysis have demonstrated greater accuracy for synthesising test performance data, compared with conventional meta-analysis methods (Harbord et al., 2008), and are recommended by the *Cochrane Collaboration* (Macaskill et al., 2010) and used in evidence generation for the appraisal of health technologies by the NICE DAP (National Institute for Health and Care Excellence, 2011).

Random effects models can provide an estimate of the average test accuracy, controlling for systematic heterogeneity between individual studies (Macaskill et al., 2010; Dinnes et al., 2005). Moreover, by jointly synthesising the evidence of a test's sensitivity and specificity, the covariance between these parameters can be estimated to enable parameter correlation in a PSA of a decision analytic model (Novielli et al., 2010; see Appendix 4). The two hierarchical meta-analysis methods reported in the literature are the bivariate meta-analysis (Reitsma et al., 2005) and the hierarchical summary ROC (HSROC) (Rutter et al., 2001). Both methods are mathematically equivalent in the absence of study-level covariates but are parameterised differently (Harbord et al., 2007). This study estimated a bivariate meta-analysis of test accuracy studies, which directly modelled the test's sensitivity and specificity, and the correlation between them (Reitsma et al., 2005). The bivariate meta-analysis method was chosen because of the ability to perform the analysis in *STATA Version 13* (StataCorp, 2013) using the method of Harbord et al. (2009).

Hierarchical models of test accuracy studies are estimated in two levels; the first level performs an analysis on the study-level 2x2 tables and the second level uses random effects to control for between-study heterogeneity in test accuracy (Macaskill et al., 2010). The bivariate meta-analysis jointly estimated Equation A34.1 by maximum likelihood.

$$\begin{bmatrix}\mu_{Ai}\\ \mu_{Bi}\end{bmatrix} \sim Normal\left[\begin{matrix}\mu_A\\ \mu_A,\end{matrix}\begin{bmatrix}\sigma_A^2 & \sigma_{AB}\\ \sigma_{AB} & \sigma_B^2\end{bmatrix}\right] \qquad \textbf{(Equation A34.1)}$$

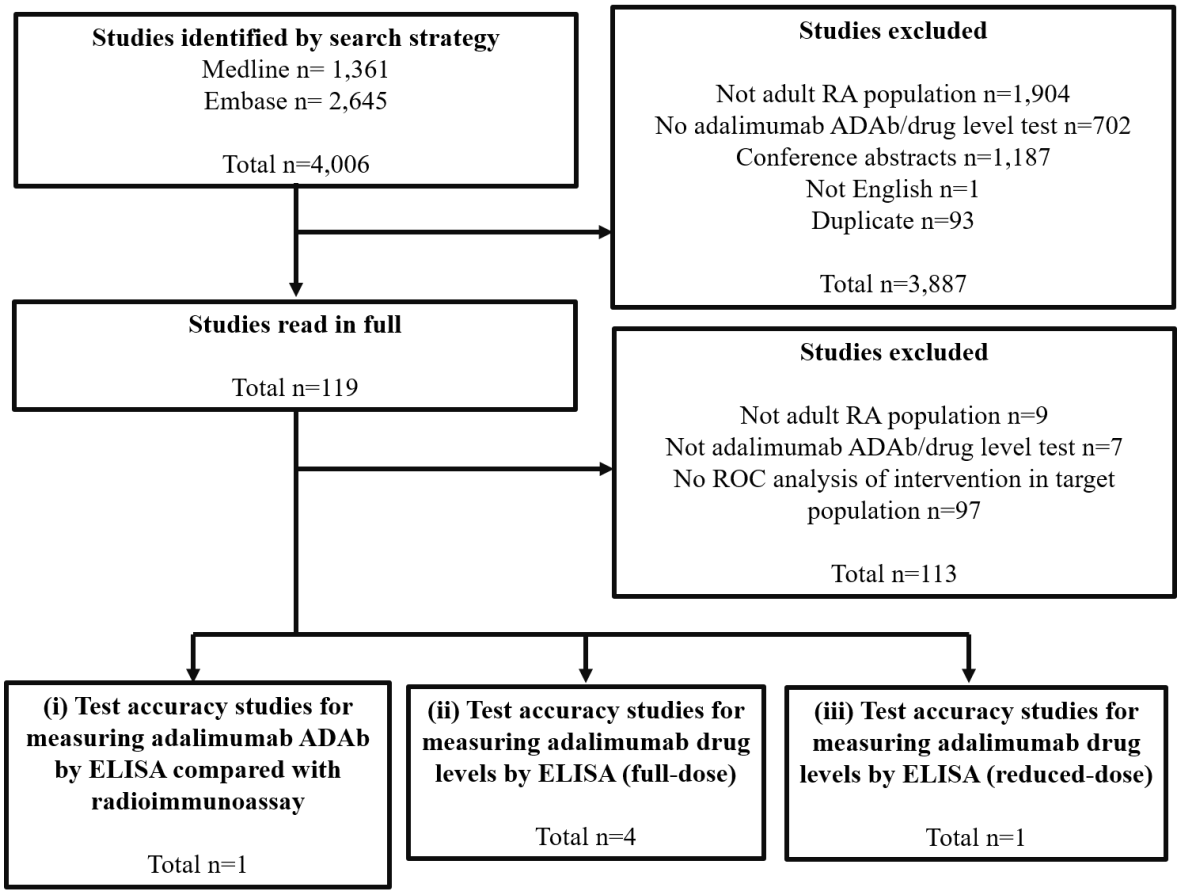
Where μ_A is the logit-transformed sensitivity with variance σ_A^2, and μ_B is the logit-transformed specificity with variance σ_B^2. The subscript *i* denotes *study i*. Therefore, the

model estimated five parameters: μ_A, μ_B, σ_A^2, σ_B^2 and $\rho_{AB} = (\frac{\sigma_{AB}}{\sigma_A \sigma_B})$. A test's average sensitivity and specificity values were derived by transforming back from their logit specifications $\left[\text{if } y=\text{logit}(x), \text{ then } x=\left(\frac{e^y}{e^y+1}\right)\right]$ (Macaskill et al., 2010).

A34.3. Results

A flow diagram of included studies is illustrated in Figure A34.1. The search strategy identified 4,006 abstracts, 119 of which were read in full. There was (i) one study that estimated the accuracy of measuring adalimumab ADAb by ELISA compared with radioimmunoassay (Jani et al., 2016); (ii) four studies that reported a ROC analysis to detect treatment response by measuring full-dose adalimumab drug levels (Chen et al., 2015; Pouw et al., 2015; Rosas et al., 2014; Jani et al., 2015); and (iii) one study that reported a ROC analysis to detect treatment response by measuring reduced-dose adalimumab drug levels (Chen et al., 2016). The results are presented for each of these testing scenarios in turn.

Figure A34.1. *Flow diagram of included test accuracy studies*.



A34.3.1. Accuracy of Measuring Adalimumab ADAb by ELISA

The systematic review identified one study that assessed the concordance between measuring adalimumab ADAb by ELISA and by radioimmunoassay (the reference standard) (Jani et al., 2016). It was therefore not possible to synthesise multiple sources of evidence using a bivariate meta-analysis.

Jani et al. (2016) measured adalimumab ADAb using the commercial *Promonitor* bridging-ELISA (see Table 1.8) in 159 serum samples of patients with RA in England that met the NICE eligibility criteria for TNFi therapy (DAS28 score ≥ 5.1; had failed two previous attempts of cDMARDs including methotrexate). Following the commercial manufacturer's instructions, the cut-point to define adalimumab ADAb positivity was 12 AU/mL for the radioimmunoassay and 3.5 AU/mL for the ELISA. The ELISA test detected adalimumab ADAb, compared with the radioimmunoassay, with a sensitivity of 32.2% (95% CI: 20.6% to 45.6%) and a specificity of 98% (95% CI: 93% to 99%) (Jani et al., 2016).

Table A34.2 reports the QUADAS-2 checklist for Jani et al. (2016), which assessed the study's risk of bias and applicability to the research question.

Table A34.2. *QUADAS-2 checklist for Jani et al. (2016).*

Domain	Risk of Bias	Domain	Concern Regarding Applicability
1. Patient selection.	Low.	1. Patient selection.	Low.
2. Index test.	Low.	2. Index test.	Low.
3. Reference standard.	Low.	3. Reference standard.	Low.
4. Flow and timing.	Low.		

The study reported by Jani et al. (2016) had a low risk of bias according to the QUADAS-2 checklist (patients were enrolled by a prospective cohort study; index test cut-points were pre-specified; the reference standard was likely to correctly classify the target condition; and all patients were included in the analysis). There was also low concern regarding the study's applicability to the research question (the patient population were identical to the population in the cost-effectiveness analysis in *Chapter Six*; the index test was the test being evaluated in *Chapter Six*).

A34.3.2. Accuracy of Measuring Adalimumab Drug Levels (Full Dose) by ELISA

The systematic review identified four studies that estimated the accuracy of measuring drug levels by ELISA to identify patients that would maintain response to full-dose

adalimumab (Chen et al., 2015; Rosas et al., 2014; Jani et al. 2015; Pouw et al., 2015). It was therefore possible to synthesise these sources of evidence using a bivariate meta-analysis. The study design and sample characteristics of these four studies are reported in Table A34.3. The 2x2 table extracted from each study is reported in Table A34.4.

Chen et al. (2015) and Pouw et al. (2015) estimated the optimal cut-point of trough adalimumab levels to determine a good EULAR response by a ROC analysis. Jani et al. (2015) estimated the optimal cut-point of adalimumab drug levels to determine any EULAR response (good or moderate). Rosas et al. (2014) performed a ROC analysis to estimate the adalimumab drug level cut-point that distinguishes patients with low disease activity from those with moderate or higher disease activity.

In all studies, patients with RA received 40mg adalimumab every two weeks, consistent with the target population of the economic evaluation in *Chapter Six*. However, there was heterogeneity across the study designs and patient samples. Notably, mean disease severity measured by DAS28 at study entry ranged between 2.7 and 6.08, and the proportion of patients that received concomitant methotrexate ranged between 51% and 89%. Similarly to other meta-analyses of test accuracy, the cut-point varied between the four studies (Macaskill et al., 2010).

All studies used ELISA-based assays to measure drug levels. Three studies used the commercial index test that was the focus of the economic evaluation in *Chapter Six* (described in Table 1.8) (Chen et al., 2015; Rosas et al., 2014; Jani et al. 2015). Pouw et al. (2015) used an in-house ELISA assay and, to facilitate evidence synthesis, it was assumed that the ability of this test to measure therapeutic drug levels was equivalent to the ability of the commercial assays. There were differences in how response was defined between studies: two studies predicted a *good EULAR response* (Chen et al., 2015; Pouw et al., 2015), one study predicted a *good or moderate EULAR response* (Jani et al., 2015), and one study predicted *low disease activity* (Rosas et al., 2014). Values for the 2x2 tables (Table A35.4) were reported fully by one study (Rosas et al., 2014), were calculated according to sensitivity and specificity values in two studies (Chen et al., 2015; Pouw et al., 2015), and were obtained by correspondence with the author for one study (Jani et al., 2015).

The quality of the four studies included in the review, determined by the QUADAS-2 checklist, is presented graphically by risk of bias (Figure A34.2) and concern of applicability to the research question (Figure A34.3).

Table A34.3. *Study design and sample characteristics of four test accuracy studies included in review.*

	Study Design			Sample Characteristics				
Study Country	**Assay**	**Response Outcome**	**Drug level cut-point**	**Sample size (n)**	**Mean age (SD)**	**% women**	**Baseline DAS28 (SD)**	**% with methotrexate**
Chen et al. (2015) Taiwan	Sandwich ELISA (*Promonitor*)	Good EULAR response at 12 months Vs. moderate/no EULAR response	1.046 μg/mL	36	52.9 (15)	89%	6.08 (0.81)	89%
Pouw et al. (2015) The Netherlands	ELISA	Good EULAR response at 28 weeks Vs. moderate/no EULAR response	5 μg/mL	221	54 (12)	80%	5.3 (1.1)	77%
Rosas et al. (2014) Spain	Sandwich ELISA (*Promonitor*)	Low disease activity Vs. moderate/high disease activity	4.3 mg/L	70	63 (12)	79%	2.7 (1.1)	63%
Jani et al. (2015) England	Sandwich ELISA (*Promonitor*)	Good/moderate EULAR response at 12 months Vs. No EULAR response	5 μg/mL	118	56.8 (11)	76%	5.8 (0.9)	51%

Abbreviations: ELISA=Enzyme-linked immunosorbent assay; EULAR=European League Against Rheumatism.

Table A34.4. *2x2 tables extracted from the four test accuracy studies included in the review.*

	Chen et al. (2015)			
	Normal drug level ≥ 1.046 µg/mL	**Low drug level** < 1.046 µg/mL	**Total (n)**	
Good EULAR response	20	0	20	**Sensitivity** 100%
Moderate or no EULAR response	0	16	16	**Specificity** 100%

	Pouw et al. (2015)			
	Normal drug level ≥ 5 µg/mL	**Low drug level** < 5 µg/mL	**Total (n)**	
Good EULAR response	79	8	87	**Sensitivity** 91%
Moderate or no EULAR response	76	58	134	**Specificity** 43%

	Rosas et al. (2014)			
	Normal drug level ≥ 4.3 mg/L	**Low drug level** < 4.3 mg/L	**Total** (n)	
Low disease activity	44	6	50	**Sensitivity** 88%
Moderate or high disease activity	8	12	20	**Specificity** 60%

	Jani et al. (2015)			
	Normal drug level ≥ 5 µg/mL	**Low drug level** < 5 µg/mL	**Total (n)**	
Good or moderate EULAR response	98	4	102	**Sensitivity** 96%
No EULAR response	10	6	16	**Specificity** 38%

Abbreviations: EULAR=European League Against Rheumatism.

Figure A34.2. *QUADAS-2 checklist: risk of bias assessment.*

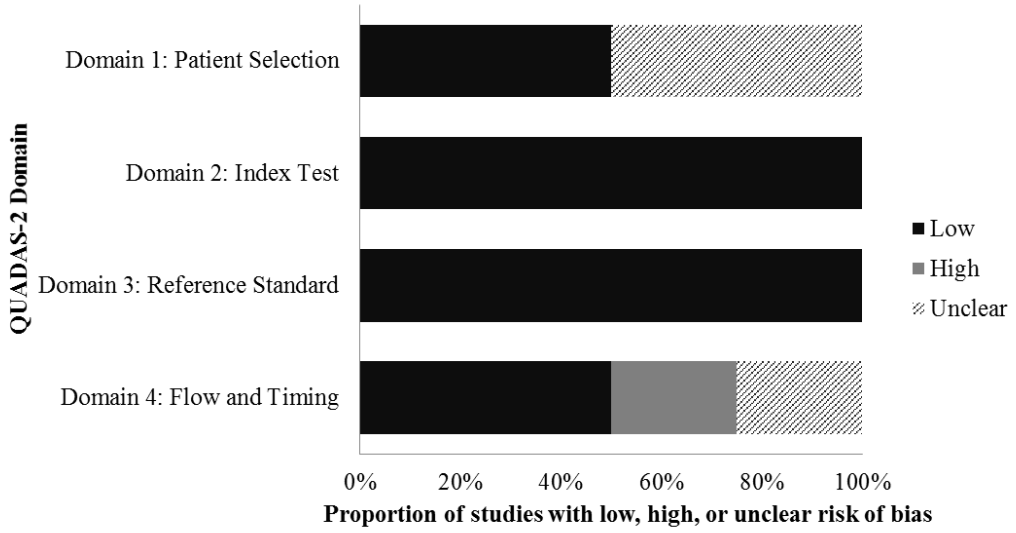
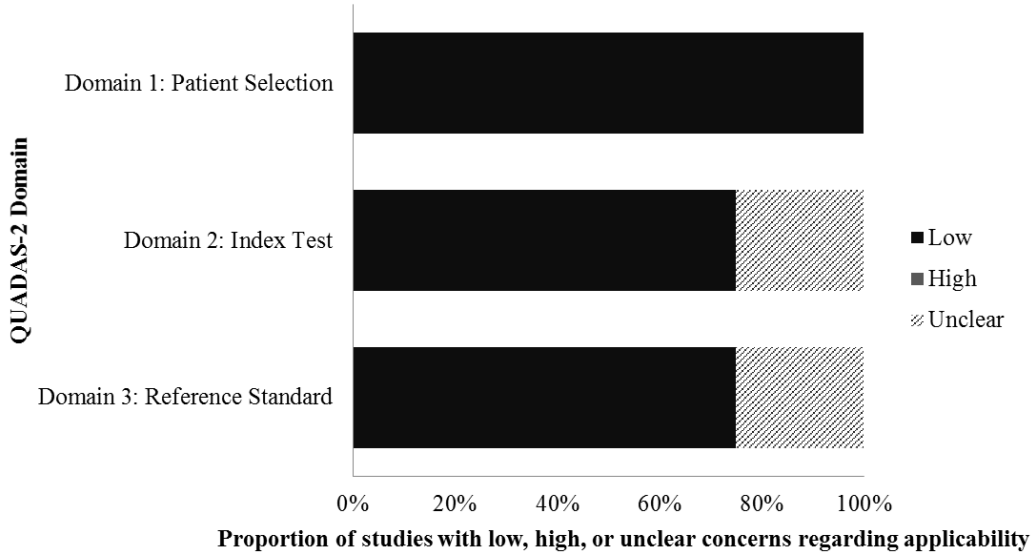


Figure A34.3. *QUADAS-2 checklist: concerns regarding applicability assessment.*



The four studies, as reported, had a low risk of bias according to the QUADAS-2 checklist (Figure A34.2). All patients received the index and reference standard tests, and the reference standard (direct observation of treatment response) was objective. The methods of patient selection were unclear in two studies (Chen et al., 2015; Rosas et al., 2014) and one study did not did not explain why the final results of the ROC analysis excluded a minority of patients (Rosas et al., 2014).

The four studies, as reported, also had low concern regarding applicability to the final research question according to the QUADAS-2 checklist (Figure A34.3). All included patients, index tests, and reference standards were relevant to the final economic evaluation in *Chapter Six*.

A bivariate meta-analysis was performed to synthesise the results of the four test accuracy studies reported in Table A34.4. The results of the analyses estimated an average test sensitivity of 95% (95% CI: 0.85-0.98) and specificity of 68% (95% CI: 0.28-0.92). The covariance between sensitivity and specificity was estimated to be 0.255.

A35.3.3. Accuracy of Measuring Adalimumab Drug Levels (Reduced Dose) by ELISA

The systematic review identified one study that estimated the accuracy of measuring adalimumab drug levels by ELISA to identify patients that would maintain response to reduced-dose adalimumab (Chen et al., 2016). It was therefore not possible to synthesise multiple sources of evidence using a bivariate meta-analysis.

Chen et al. (2016) measured adalimumab drug levels using the commercial *Promonitor* sandwich ELISA (see Table 1.8) in twenty-five patients with RA in Taiwan that were in remission and had been prescribed a half-dose of adalimumab (40mg every month plus methotrexate). The reference standard was whether those patients maintained response at twenty-four weeks. Patients had a mean age of 57.1 years (SD: 14.7), 91.3% were women, 91.3% were receiving concomitant methotrexate, and the mean DAS28 score was 2.21 (SD: 0.13). Chen et al. (2016) used a ROC analysis to estimate the optimal adalimumab drug level cut-off that was predictive of maintaining treatment response. The optimal drug level cut-point was estimated to be 6.4ug/mL, which predicted persistent remission with a sensitivity of 100% and a specificity of 93.4%. No standard error or confidence interval was reported within the published manuscript.

Table A34.5 reports the QUADAS-2 checklist for Chen et al. (2016), which assessed the study's risk of bias and applicability to the research question.

Table A34.5. *QUADAS-2 checklist for Chen et al. (2016).*

Domain	Risk of Bias	Domain	Concern Regarding Applicability
1. Patient selection.	Unclear.	1. Patient selection.	Low
2. Index test.	Low	2. Index test.	Low.
3. Reference standard.	Low.	3. Reference standard.	Low.
4. Flow and timing.	Low.		

The study, as reported by Chen et al. (2016), generally had a low risk of bias (all patients were included in the analysis; the reference standard was likely to correctly classify the target condition). However, it was unclear whether the patients in the study were sampled consecutively or randomly. The study was also deemed to be applicable to the economic

evaluation in *Chapter Six* (similar patient population; relevant index test and reference standard).

A34.4. Discussion

This study performed a systematic review and meta-analysis of test accuracy studies that assessed (i) the accuracy of measuring adalimumab ADAb by ELISA, compared with measurement by radioimmunoassay, and the accuracy of measuring adalimumab drug levels by ELISA to identify patients that would maintain response to (ii) full-dose adalimumab and (iii) reduced-dose adalimumab. The results suggested that using a bridging ELISA to detect adalimumab ADAb (Section A34.3.1) was (i) relatively accurate at detecting the absence of adalimumab ADAb correctly (specificity = 98%) but (ii) relatively poor at detecting the presence of adalimumab ADAb correctly (sensitivity = 32.2%). The measurement of high adalimumab drug levels by a sandwich ELISA, in patients that were receiving full-dose therapy (Section A34.3.2), was estimated to be highly predictive of treatment response (sensitivity = 95%) but low drug levels were not necessarily predictive of no response (specificity = 68%). The drug level cut-point estimated by Chen et al. (2016) in Section A34.3.3 was found to be accurate at predicting whether patients maintained response to reduced-dose adalimumab in remission (Sensitivity = 100%; Specificity = 93.4%).

Hierarchical meta-analyses of test accuracy studies have previously been undertaken during technology appraisals for the NICE DAP in England. For example, the DAP technology appraisal of TNFi ADAb and drug level monitoring for Crohn's disease included a hierarchical meta-analysis of four test accuracy studies that predicted response to treatment by measuring adalimumab drug levels (National Institute for Health and Care Excellence, 2015, Appendix 12.3 of the DAP report). The analysis found that low drug levels (the definition of a positive test) were predictive of no response (the target condition) with a sensitivity of 68% (95% CI: 59% to 76%) and specificity of 79% (95% CI: 64% to 88%) (National Institute for Health and Care Excellence, 2015, Appendix 12.3). These results were similar to the results of the present study (high adalimumab drug were found levels predict response); the point-estimate probability of no response given low drug levels was identical between the two studies (68% sensitivity in the NICE DAP assessment, and 68% specificity in the present study) (National Institute for Health and Care Excellence, 2015).

The application of hierarchical meta-analysis methods for test accuracy studies is variable within the literature, and many studies have performed a conventional meta-analysis to

synthesise evidence of a test's performance (Novielli et al., 2010; Ochodo et al., 2013). An advantage of this study was that the use of a bivariate meta-analysis accounted for the between-study heterogeneity and correlation between ELISA sensitivity and specificity. This correlation was subsequently incorporated during the PSA of the decision analytic model (described in *Appendix 37*). One potential limitation of the study reported in this appendix was that there were too few studies identified by the systematic review to explore between-study heterogeneity by meta-regression, despite observed differences in the proportion of patients receiving concomitant methotrexate, disease severity, and the drug level cut-point. One study also used a different ELISA test to measure drug levels, which may have had a different predictive ability than the assay used in the three remaining studies (Pouw et al., 2015). However, the QUADAS-2 assessment of bias and risk of study applicability were generally low, indicating that the four studies that were synthesised in the bivariate meta-analysis were of good quality and relevant to the economic evaluation in *Chapter Six*.

The values of the input parameters that related to test accuracy in the decision analytic model (Section 6.3.3.1.11) were chosen based on the results of this systematic review. A sensitivity analysis of these input parameters was performed (Section 6.3.5.2) to assess how the relative cost-effectiveness of treatment stratification varied according to the accuracy of testing.

A34.5. References

Chen, D., Chen, Y., Hsieh, T., Hung, W., Hsieh, C., Chen, H., Tang, K., & Lan, J. (2016). "Drug Trough Levels Predict Therapeutic Responses to Dose Reduction of Adalimumab for Rheumatoid Arthritis Patients During 24 Weeks of Follow-up", *Rheumatology*, Vol. 55, 1, pp. 143-148.

Chen, D., Chen, Y., Tsai, W., Tseng, J., Chen, Y., Hsieh, C., Hung, W., Lan, J. (2015). "Significant Associations of Antidrug Antibody Levels with Serum Drug Trough Levels and Therapeutic Response of Adalimumab and Etanercept Treatment in Rheumatoid Arthritis", *Annals of the Rheumatic Diseases*, Vol. 74, 3, pp. 1-9.

Dinnes, J., Deeks, J., Kirby, J., Roderick, P. (2005). "A Methodological Review of how Heterogeneity has been Examined in Systematic Reviews of Diagnostic Test Accuracy", *Health Technology Assessment*, Vol. 8, 12, pp. 1-113.

Harbord, R., Deeks, J., Egger, M., Whiting, P., & Sterne, J. (2007). "A Unification of Models for Meta-analysis of Diagnostic Accuracy Studies", *Biostatistics*, Vol. 8, 2, pp. 239-251.

Harbord, R., & Whiting, P. (2009). "Metandi: Meta-analysis of Diagnostic Accuracy using Hierarchical Logistic Regression", *The Stata Journal*, Vol. 9, 2, pp. 211-229.

Harbord, R., Whiting, P., Sterne, J., Egger, M., Deeks, J., Shang, A., & Bachmann, L. (2008). “An Empirical Comparison of Methods for Meta-analysis of Diagnostic Accuracy Showed Hierarchical Models are Necessary”, *Journal of Clinical Epidemiology*, Vol. 61, 11, pp. 1095-1103.

Jani, M., Chinoy, H., Warren, R., Griffiths, C., Plant, D., Fu, B., Morgan, A., Wilson, A., Isaacs, J., Hyrich, K., & Barton, A. (2015). “Clinical Utility of Random Anti-tumor Necrosis Factor Drug-level Testing and Measurement of Antidrug Antibodies on the Long-term Treatment Response in Rheumatoid Arthritis”, *Arthritis & Rheumatology*, Vol. 67, 8, pp. 2011-2019.

Jani, M., Isaacs, J., Morgan, A., Wilson, A., Plant, D., Hyrich, K., Chinoy, H., & Barton, A. (2016). "Detection of Anti-drug Antibodies using a Bridging ELISA Compared with Radioimmunoassay in Adalimumab-treated Rheumatoid Arthritis Patients with Random Drug Levels", *Rheumatology*, Vol. 55, 11, pp. 2050-2055.

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. (2010). “Chapter 10: Analysing and Presenting Results”. In: Deeks, J., Bossuyt, P., & Gatsonis, C. (Eds). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. (1 Ed). London: The Cochrane Collaboration.

National Institute for Health and Care Excellence. (2011). *Diagnostics Assessment Programme Manual*. Manchester: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence. (2015). *Diagnostics Assessment Programme: Crohn’s Disease: Tests for Therapeutic Monitoring of TNF Inhibitors (LISA-TRACKER ELISA Kits, TNF-α Blocker ELISA Kits, and Promonitor ELISA Kits – Evaluation Report*. Manchester: National Institute for Health and Care Excellence.

Novielli, N., Cooper, N., Abrams, K., & Sutton, A. (2010). “How is Evidence on Test Performance Synthesized for Economic Decision Models of Diagnostic Tests? A Systematic Appraisal of Health Technology Assessments in the UK Since 1997”, *Value in Health*, Vol. 13, 8, pp. 952-957.

Ochodo, E., Reitsma, J., Bossuyt, P., & Leeflang, M. (2013). “Survey Revealed a Lack of Clarity about Recommended Methods for Meta-analysis of Diagnostic Accuracy Studies”, *Journal of Clinical Epidemiology*, Vol. 66, 11, pp. 1281-1288.

Pouw, M., Krieckaert, C., Nurmohamed, M., van der Kleij, D. Aarden, L., Rispens, T., & Wolbink, G. (2015). “Key Findings Towards Optimising Adalimumab Treatment: the Concentration-effect Curve”, *Annals of the Rheumatic Diseases*, Vol. 74, 3, pp. 513-518.

Reitsma, J., Glas, A., Rutjes, A., Scholten, R., Bossuyt, P., & Zwinderman, A. (2005). “Bivariate Analysis of Sensitivity and Specificity Produces Informative Summary Measures in Diagnostic Reviews”, *Journal of Clinical Epidemiology*, Vol. 58, 10, pp. 982-990.

Rosas, J., Llinares-Tello, F., de la Torre, I., Santos-Ramírez, C., Senabre-Gallego, J., Valor, L., Barber-Vallés, X., Hernández-Flórez, D., Santos-Soler, G., Salas-Heredia, E., & Carreño, L. (2014). “Clinical Relevance of Monitoring Serum Levels of Adalimumab in Patients with Rheumatoid Arthritis is Daily Practice”, *Clinical and Experimental Rheumatology*, Vol. 32, 6, pp. 942-948.

Rutter, C., & Gatsonis, C. (2001). “A Hierarchical Regression Approach to Meta-analysis of Diagnostic Test Accuracy Evaluations”, *Statistics in Medicine*, Vol. 20, 19, pp. 2865-2884.

StataCorp. (2013). "Stata Statistical Software". Release 13, College Station, TX: StataCorp LP.

Whiting, P., Rutjes, A., Westwood, M., Mallett, S., Deeks, J., Reitsma, J., Leeflang, M., Sterne, J., & Bossuyt, P. (2011). “QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies”, *Annals of Internal Medicine*, Vol. 155, 8, pp. 529-536.

Appendix 35: Published Microcosting Study of TNFi Anti-drug Antibody and Drug Level Testing

This appendix presents a published microcosting study that estimated the direct health care costs associated with using the ELISA tests to measure TNFi ADAb and drug levels in routine rheumatology practice in England. The evidence from this study was a necessary input (the cost of testing) to the *de novo* decision analytic model in *Chapter Six*. The study was published in *Rheumatology* in December 2016.

The microcosting study was conducted independently from this PhD thesis with Dr. Meghna Jani, and contributed towards the PhD thesis of Dr. Jani. SG contributed to the study design and made substantive comments on the final manuscript. All data collection and analysis was performed by MJ.

The appropriate citation for the study is:

- Jani, M., **Gavan, S**., Dixon, W., Harrison, B., Moran, A., Barton, A., & Payne, K. (2016). "A Microcosting Study of Immunogenicity and TNFi Drug Level Tests for Therapeutic Monitoring in Clinical Practice", *Rheumatology*, Vol. 55, 12, pp. 2131-2137.

RHEUMATOLOGY

Rheumatology 2016;55:2131–2137
doi:10.1093/rheumatology/kew292
Advance Access publication 29 August 2016

Concise report

A microcosting study of immunogenicity and tumour necrosis factor alpha inhibitor drug level tests for therapeutic drug monitoring in clinical practice

Meghna Jani[1,2], Sean Gavan[3,4], Hector Chinoy[1,4,5], William G. Dixon[1,2,5], Beverley Harrison[5], Andrew Moran[6], Anne Barton[1,4,7] and Katherine Payne[3]

Abstract

Objectives. To identify and quantify resource required and associated costs for implementing TNF-α inhibitor (TNFi) drug level and anti-drug antibody (ADAb) tests in UK rheumatology practice.

Methods. A microcosting study, assuming the UK National Health Service perspective, identified the direct medical costs associated with providing TNFi drug level and ADAb testing in clinical practice. Resource use and costs per patient were identified via four stages: identification of a patient pathway with resource implications; estimation of the resources required; identification of the cost per unit of resource (2015 prices); and calculation of the total costs per patient. Univariate and multiway sensitivity analyses were performed using the variation in resource use and unit costs.

Results. Total costs for TNFi drug level and concurrent ADAb testing, assessed using ELISAs on trough serum levels, were £152.52/patient (range: £147.68–159.24) if 40 patient samples were tested simultaneously. For the base-case analysis, the pre-testing phase incurred the highest costs, which included booking an additional appointment to acquire trough blood samples. The additional appointment was the key driver of costs per patient (67% of the total cost), and labour accounted for 10% and consumables 23% of the total costs. Performing ELISAs once per patient (rather than in duplicate) reduced the total costs to £133.78/patient.

Conclusion. This microcosting study is the first assessing the cost of TNFi drug level and ADAb testing. The results could be used in subsequent cost-effectiveness analyses of TNFi pharmacological tests to target treatments and inform future policy recommendations.

Key words: microcosting, immunogenicity, TNFi drug levels, opportunity costs, health economics

Rheumatology key messages

- Microcosting analysis enabled quantification of resource use and costs required to implement TNF inhibitor pharmacological monitoring in practice.
- The cost of £152.52/patient for TNF inhibitor pharmacological monitoring (base case analysis) was comparable to other novel diagnostics.
- The additional appointment for trough level TNF inhibitor pharmacological monitoring was the key driver of costs per patient.

[1]Centre for Musculoskeletal Research, [2]Arthritis Research UK Centre for Epidemiology, Institute of Inflammation and Repair, [3]Manchester Centre for Health Economics, Institute of Population Health, Manchester Academic Health Science Centre, University of Manchester, [4]National Institute of Health Research Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust Manchester Academic Health Science Centre, Manchester, [5]Department of Rheumatology, Salford Royal NHS Foundation Trust, Salford, [6]Department of Immunology and [7]Kellgren Centre for Rheumatology, Manchester Royal Infirmary, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK

Submitted 29 November 2015; revised version accepted 29 June 2016

Correspondence to: Katherine Payne, Manchester Centre for Health Economics, The University of Manchester, Oxford Road, 4th floor, Jean McFarlane Building, Manchester M13 9PL, UK. E-mail: katherine.payne@manchester.ac.uk

Introduction

TNF-α inhibitors (TNFi) have transformed the treatment of several chronic inflammatory diseases. Given their effectiveness in the most severely affected patients, the use of biologics in rheumatology continues to increase, but is associated with significant expenditure (£10 000/year/patient). TNFi agents such as adalimumab, etanercept and infliximab are currently represented within the top five highest medicinal expenditures in England [1], with an estimated cost to the National Health Service (NHS) of ∼£160 million annually for RA [2]. A targeted approach using robust predictive biomarkers of response in TNFi-treated patients may add value to the clinical decision-making process by potentially informing the selection of which TNFi drug to use first in specific patients, the appropriate biologic sequence and whether to continue the drug in patients established on therapy. However, there remain considerable gaps in the evidence base supporting the introduction of a targeted approach into clinics [3]. In the era of finite budgets, robust economic evidence is required in order to ensure that the alternative uses for funds are considered in any decision, and decision-making groups must be aware of other funding pressures and service developments that will otherwise be forgone (opportunity costs) [4].

An important mechanism for treatment failure of certain TNFi agents is immunogenicity involving the formation of anti-drug antibodies (ADAb) and low drug levels [5, 6]. While the presence of ADAbs and low TNFi drug levels, detected soon after treatment initiation, have been shown to predict subsequent treatment response [7], tests quantifying levels are not currently available in rheumatology clinical practice in the UK NHS. Such testing needs to be both effective in improving outcomes and a cost-effective use of the healthcare budget before it can be recommended for implementation into the clinic. To date, a description of the types and quantity of resources needed to provide the test is not available in the published literature. Identifying the resources required will facilitate the calculation of the costs of implementing these tests in a UK clinical setting if the introduction of such testing is shown to be clinically useful.

Microcosting is a method that allows robust assessment of the types and quantities of resources and associated costs of health interventions consumed [8]. It is particularly useful for estimating the costs of new interventions and for interventions with large variability across providers, thereby potentially providing a key input for undertaking subsequent economic evaluations. The aim of this study was to identify and quantify the resource use and associated costs required for introducing drug level and ADAb testing to assess response to TNFi drugs in routine practice in the UK setting.

Methods

A microcosting study assumed the NHS (service provider) perspective for identifying the resource use and cost per patient of providing TNFi drug level and ADAb testing (the test). Costs of providing the test were determined from the point of a patient established on treatment (for ⩾3 months) presenting to clinic, to the results being fed back to the clinician to inform a treatment decision. Direct medical costs associated with providing the test were identified; indirect non-medical costs (such as absence from work) were not consistent with the study perspective and beyond the scope of the paper. Ethical approval was not required. This study was essentially an audit of practice in North West England. (Regional guidelines for biologics in RA [9] allow use of these tests in rheumatology practice if clinicians have access.) The four study stages are now described.

Stage 1: identifying the testing pathway

The test is not routinely available in UK rheumatology practice, and it was necessary to define an explicit pathway for a patient being offered testing with input from six experts from North West England (four rheumatology consultants and two clinical/laboratory staff) (Fig. 1A and B). The pathway eventually encompassed three phases: pre-testing, analysis of samples, and treatment decision (Fig. 1A). This study assumed that the test is reliant on identifying a pre-defined drug trough level, requiring an additional outpatient appointment, rather than using random sampling, which mirrors the current availability of the technology available in the UK (bridging ELISAs) to measure ADAbs.

Stage 2: use of resources

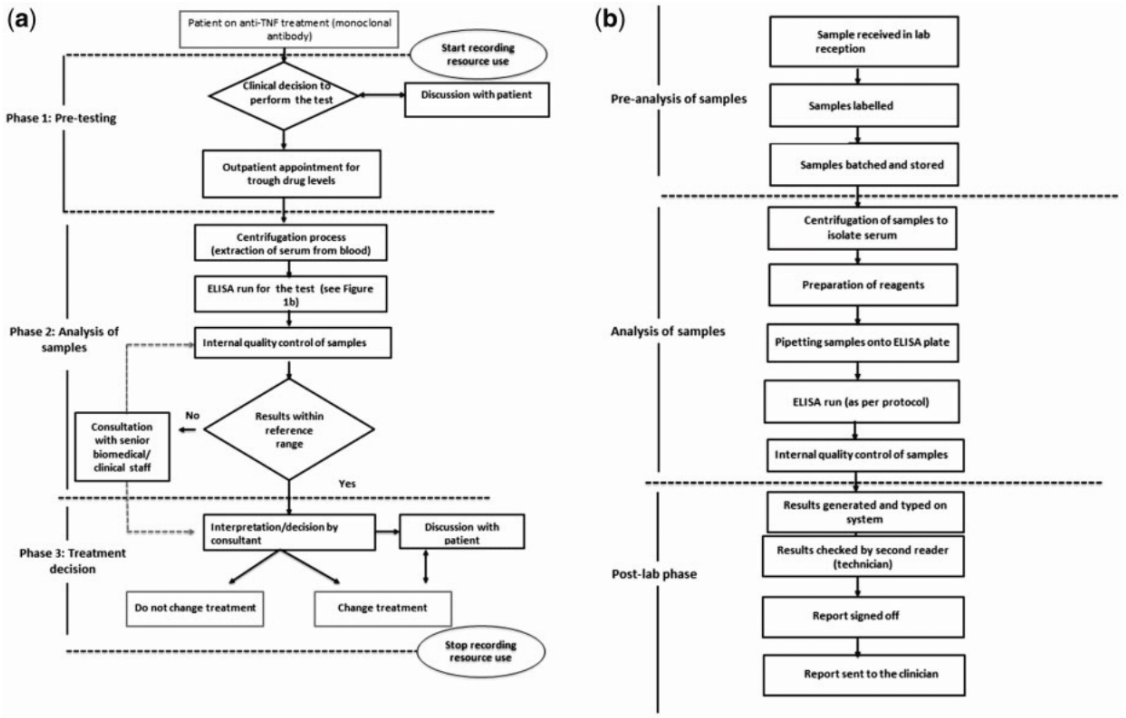
The use of direct medical resources for the pathway was estimated using structured face-to-face interviews and elicitation with six experts. The expert elicitation process is described in more detail in supplementary Table S1, available at *Rheumatology* Online. Direct non-participant observation of staff was undertaken in a hospital setting to generate an estimate of the time taken for selected procedures. The Central Manchester Foundation NHS Trust Immunology Department was asked to name resources required relating to laboratory staff time. The level of each resource use was estimated per patient for each phase and per batch of 40 samples for the laboratory processes (Fig. 1A, phase 2).

Stage 3: identifying unit costs

It was assumed that most hospital laboratories would have the necessary room requirements and stock standard equipment required to perform ELISAs, and the following items of resource use were therefore excluded: equipment costs of centrifuge systems; ELISA readers; pipettes; personal protective equipment; phlebotomy equipment costs; overhead; and capital costs.

A variety of approaches were taken to identify unit costs (price year 2015) for each type of resource use (Table 1). The unit cost of a rheumatology blood-monitoring appointment was obtained from operational managers for rheumatology directorates of two hospitals (Central Manchester University Hospitals and Salford Royal NHS Foundation Trust). Published estimates of unit costs for

FIG. 1 Pathway for immunogenicity and drug level testing (the test)



(**A**) Overview of the pathway from the clinical decision to perform the test. (**B**) Summary of laboratory processes.

labour time were not available for all types of staff by the Personal Social Services Research Unit [10]. Therefore hospital-based health care staff time was valued using relevant labour unit costs from the national pay system for the NHS (Agenda for change—pay rates 2015–16) [11] and the British Medical Association pay scale for medical staff in England (2015–16) [12]. Salary scales per annum were converted to a per-minute rate by dividing the number of workable minutes per year, as described previously [13] (see supplementary Table S2, available at *Rheumatology* Online).

Stage 4: data analysis

The base-case analysis calculated the total cost of the test by multiplying unit costs with the identified items and quantities of resource for each phase of the pathway (see Fig. 1A). Multiway sensitivity analyses were conducted by varying the following parameters using lower and upper ranges of estimated resource use: lowest time taken to perform tasks using the lowest pay grade (best case scenario) and highest amount of time taken to perform procedures using the highest pay grade (worst case scenario). Three one-way sensitivity analyses and one two-way sensitivity analysis were used to understand the impact of varying pre-defined assumptions made when calculating the cost of the test.

Results

Table 1 summarizes the items and quantity of resource use and unit costs for each of the three phases of the pathway (Fig. 1A).

Base-case analysis

The total cost for performing the test was £152.52/patient for the base-case analysis. The most expensive element of the pathway was the cost of the additional appointment to conduct blood sampling for drug trough levels. Therefore the pre-testing phase incurred the highest costs due to the additional appointment to perform trough blood sampling (total costs: £105.50/patient). The total cost for processing 40 samples during laboratory phase (phase 2, analysis of samples) was £749.34 [£18.73 (cost in phase 2 divided by 40) × 2 (for both tests) = £37.47/patient to simultaneously perform the test]. The final treatment decision cost was £9.55/patient. The additional trough level appointment accounted for 67% of the total cost, and labour and consumables accounted for 10% and 23% of the total costs, respectively.

Sensitivity analyses

The multiway sensitivity analysis varied the estimated and directly observed time and pay grade for each phase (see Table 1). Using the lowest values, the estimated best-case

Table 1 Resource use and costs of implementing drug level and immunogenicity testing per patient in a hospital setting

Type of resource use	Staff member	Mean volume of resource use, minutes	Range of resource use, minutes	Source to obtain resource use	Unit costs for base-case analysis 2015, £[a]	Range of unit costs 2015, £	Source to obtain unit costs [reference]	Total costs (£) (minimum to maximum range)
Phase 1: pre-testing								
Outpatient appointment for discussion about need for test	Consultant rheumatologist	3.6 min	2–5 min	Expert estimation	£39.24/h	£33.42–45.06/h	BMA pay scales [12]	£2.35 (£1.11–3.76)
Clerical staff (to book the appointment and send out a letter to patient)	Clerical assistant	8.2 min	6–11 min	Direct observation	£8.43/h	£7.72–9.12/h	NHS pay scale, band 5 [11]	£1.15 (£0.77–1.67)
Appointment for trough blood levels	Phlebotomist/ clinical support worker	Resource use incorporated in unit cost	NA	NA	£102 per appointment	NA	DOH national tariff [14]	£102 (NA)
Phase 2: analysis of samples								
Receipt and labelling of samples—central specimen reception[b]	Medical lab assistant	15 min	NA	Expert estimation	£8.86/h	£7.74–9.98/h	NHS pay scale band 5 [11]	£2.22 (£1.94–2.50)
Data entry of patient information to lab system[b]	Medical lab assistant	15 min	NA					£2.22 (£1.94–2.50)
Sample preparation—extraction of serum from blood[b]	Medical lab assistant	15 min	NA					£2.22 (£1.94–2.50)
Transport, receipt and storage of sample—immunology lab[b]	Medical lab assistant	15 min	NA					£2.22 (£1.94– 2.50)
Preparation of reagents (wash solution, setting up assay, conjugate)[b]	Biomedical scientist	15 min	NA	Expert estimation	£12.79/h	£11.12–14.45/h	NHS pay scale band 5 [11]	£3.20 (£2.78–3.61)
ELISA kit for ADAbs or drug levels[b]	NA	NA	NA	NA	£700.00 per ELISA	NA	UK commercial price from Grifols[c]	£700.00
Pipette tips for ELISAs[b]	NA	NA	NA	NA	£6.00 per ELISA	NA	University of Manchester Centre for Musculoskeletal Research laboratory	£6.00
Semi-deep well plates for ELISAs[b]	NA	NA	NA	NA	£2.20 per ELISA	NA	Laboratory costs	£2.20
Troughs for ELISAs[b]	NA	NA	NA	NA	£1 per ELISA	NA	Laboratory costs	£1.00
Retrieval of patient/IQC samples from storage[b]	Biomedical scientist	10 min	NA	Expert estimation	£12.79/h	£11.12–14.45/h	NHS pay scale- band 5 [11]	£2.13 (£1.85 –2.40)
Checking and sorting samples to match worklist[b]	Biomedical scientist	10 min	NA					£2.13 (£1.85–2.40)
Pipetting samples onto ELISA plate[b]	Biomedical scientist	20 min	NA					£4.26 (£3.71– 4.81)
Pipetting calibrators, IQC samples and incubation of samples[b]	Biomedical scientist	10 min	NA					£2.13 (£1.85 –2.40)
Washing ELISA plate and addition of conjugate[b]	Biomedical scientist	10 min	NA					£2.13 (£1.85 –2.40)
Washing ELISA plate and addition of substrate[b]	Biomedical scientist	10 min	NA					£2.13 (£1.85–2.40)
Addition of stop solution[b]	Biomedical scientist	5 min	NA					£1.06 (£0.93 –1.20)
ELISA plate reading and printing of results[b]	Biomedical scientist	10 min	NA					£2.13 (£1.85–2.40)
Technical validation involving review of Internal quality control[b]	Biomedical scientist	5 min	NA					£1.06 (£0.93–1.20)
Results transcribed to worksheet[b]	Biomedical scientist	5 min	NA					£1.06 (£0.93–1.20)
Data entry of results to patient record in lab system[b]	Biomedical scientist	10 min	NA					£2.13 (£1.85 – 2.40)
Transcribed results/data entry reviewed by a second independent biomedical scientist[b]	Biomedical scientist	5 min	NA					£1.06 (£0.93 – 1.20)
Clinical authorisation using reference range/delta check failure results[b]	Senior Biomedical scientist or Consultant immunologist	5 min	NA	Expert estimation	£30.48/h	£15.9–45.06/h	NHS pay scale band 7 assumed [11]; BMA pay scale [12]	£2.54 (£1.33 –3.76)
Hardcopy report sent to clinician[b]	Clerical assistant	15 min	NA	Expert estimation	£8.43/h	£7.74 –9.12/h	NHS pay scale [11]	£2.11 (£1.93 –2.28)
Phase 3: Treatment decision								
Interpretation of results by rheumatologist	Consultant Rheumatologist	6 min	4–10	Expert estimation	£39.24/h	£33.42–45.06/h	NHS pay scale [11]	£3.92 (£2.23 –7.51)
Discussion with patient (phone call)		5.3 min	5–6					£3.47 (£2.79–3.92)
Letter with results and decision		3.3 min	3–4					£2.16 (£1.61–2.62)
Total costs (best case to worst case scenario)[d]								£152.52 (£147.68–159.24)

[a]Mid-point of salary grade used to calculate base-case sample. [b]Resource use estimated per batch (40 samples). [c]Previously known as Progenika Biopharma. [d]Base-case (multiway sensitivity analyses were conducted by varying the following parameters using pre-defined lower and upper ranges of estimated resource use: lowest time taken to perform tasks using the lowest pay grade (best case scenario) and highest amount of time taken to perform procedures using the highest pay grade (worst case scenario). BMA: British Medical Association; DOH: department of health; ELISA: enzyme-linked immunosorbent assay; IQC: internal quality control; NA: not applicable; NHS: National Health Service.

scenario was £147.68/patient/test. Using the highest values, the worst-case scenario estimated a cost of £159.24.

Three one-way sensitivity analyses were performed (see Sensitivity analysis in the supplementary data, available at *Rheumatology* Online). Performing the tests singly and not in duplicate may reduce test accuracy, but lowered the total cost to £133.78/patient. If the patient was due to take their TNFi on the day following their rheumatology appointment, an additional trough level appointment was not required, lowering the test cost to £50.52. If there were 50 samples to be processed by the laboratory, a new batch would need to be started, increasing the resource use in phase 2 and the total cost to £173.79/patient.

One two-way sensitivity analysis examined the impact of using various pay grades. For costs attributed to consultant time (base-case), varying the pay scale to the lower grade using the mean volume of resource use (Table 1) changed the total costs to £145.26/patient. The variation in grade included a specialty trainee in rheumatology at £38 588.50/annum (mid-point of paygrade, supplementary Table S2, available at *Rheumatology* Online), a consultant rheumatologist (Table 1, phases 1 and 3) and a senior clinical biochemist (mid-point of paygrade £35 891/annum, supplementary Table S2, available at *Rheumatology* Online) instead of a consultant immunologist (Table 1, phase 2).

Discussion

This microcosting study has identified the potential direct medical costs associated with TNFi pharmacological testing from a service provider's perspective in the UK. Since these tests for TNFi-treated patients are not routinely performed in UK clinical practice, a testing pathway was developed to allow a detailed estimation of the quantities of resources required in order to calculate a total cost. The developed pathway provides a framework for reporting resource use, presenting unit costs and allowing decision-makers from various jurisdictions to use their country-specific data if required.

There is accumulating evidence that monoclonal TNFi drug levels and ADAb levels correlate with future response to the drugs [7, 15]. If the testing strategy is to translate to clinical practice, a number of points will need to be addressed. First, the test must be shown to be useful in changing clinical decision-making; second, robust evidence must confirm that the change in practice will result in better outcomes for patients; finally the test intervention should be a cost-effective use of health care budgets. The current work is the first step in informing the last requirement. To date, the costs associated with providing TNFi drug level/ADAb testing are not known because no national tariff exists for diagnostic tests. Emerging numbers of microcosting studies in other areas have enabled rigorous comparison of health interventions in order to inform efficient resource allocation [16]. A recent NICE diagnostic assessment committee evaluating test performance of ELISA kits for ADAbs and TNFi levels in Crohn's disease was not able to draw definitive conclusions about the relative cost-effectiveness of the test compared with current practice because of insufficient evidence to inform the analysis. Early analysis suggested that the test may save the NHS money, but would also result in some loss of health in the population tested. The high degree of uncertainty in the economic analysis, particularly around the impact of the test on quality-adjusted life-years meant that the committee concluded that further research was required before the test could be recommended for use in clinical practice [17].

Our microcosting analysis identified a unit cost of £152.52/patient, making this biomarker test for guiding decisions regarding future treatment with TNFi comparable with that of other novel diagnostics and theranostics [18]. A robust economic evaluation that identifies the incremental costs and health benefits (quality-adjusted life-years) of using the test for targeting TNFi treatments compared with current prescribing practice in RA is required to determine whether this targeted approach is a cost-effective use of health care budgets.

The overall cost of testing per patient in the UK was influenced most by the cost of an additional appointment for obtaining trough levels. When the cost of trough levels was excluded, the cost per patient reduced to £50.52. To deal with batching and capacity, the base-case analysis assumed batching of samples from 40 patients/ELISA. However, uneven sample numbers would require a new batch with changes in marginal costs (cost of doing one more test) impacting on consumables, staff resources and time. If results are to be fed back in sufficient time for referring clinicians to make treatment decisions, it is unlikely that samples from 40 patients would be available for testing unless test sites were restricted to regional or national laboratories. When processing 50 rather than 40 samples, the cost per sample rose to £173.79 because each ELISA kit only allows for 40 samples to be analysed at a time.

This analysis made several assumptions in order to estimate the total cost. The base-case assumed a tertiary level setting in the north-west of England; however, it is acknowledged that the cost of a trough level appointment may vary elsewhere in the UK, thus influencing the total costs/sample. We assumed a concurrent testing strategy for all samples, in which tests for TNFi drug levels and ADAbs were performed at the same time, rather than reflex testing, which may be an alternative to reduce costs. Reflex testing would involve testing the TNFi drug levels first and only testing for ADAbs if the drug was undetectable. Direct non-medical costs such as patient out-of pocket expenses for trough level testing were not included, which may have wider societal implications [19]. Capital/overhead costs were not included in the analysis. While some hospital laboratories have automated ELISA systems or multiplex platforms, this was not assumed and was deemed unlikely to significantly lower resource estimates, following consultation with the hospital laboratory team. Furthermore, while numerous ELISA kits are commercially available for ADAb and TNFi drug

level testing, unit costs were based on those ELISA kits frequently used in the literature [20].

In conclusion, using a microcosting approach, we have explicitly identified and quantified the types and quantities of resources required in order to provide TNFi drug and ADAbs level testing in an NHS clinical setting and found that the costs were comparable with those of other tests already available. The identified cost of the test will be of use for future cost-effectiveness analysis of TNFi pharmacological testing. The results of the study will also help inform potential resource implications per patient for hospital trusts considering incorporating pharmacological monitoring into clinical practice.

Acknowledgements

M.J. is currently supported by a National Institute for Health Research (NIHR) clinical lectureship and was a Medical Research Council (MRC) Clinical Training Fellow supported by the North West England Medical Research Council Fellowship Scheme in Clinical Pharmacology and Therapeutics, which is funded by the MRC (grant number G1000417/94909), ICON, GlaxoSmithKline, AstraZeneca and the Medical Evaluation Unit. We thank Arthritis Research UK for their support (grant reference 20385 and 20380) and the NIHR Manchester Musculoskeletal Biomedical Research Unit. This report includes independent research supported by the National Institute for Health Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. W.G.D. was supported by an MRC Clinician Scientist Fellowship (G092272). We also acknowledge support from MATURA, a stratified medicines initiative jointly funded by MRC and Arthritis Research UK (MR/K015346/1) and Innovate UK.

Funding: No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this manuscript.

Disclosure statement: A.B. has received honoraria or grants from Pfizer and Eli-Lilly in the past 2 years. H.C. has received honoraria, lecture fees and/or research grants from Abbvie, Janssen, Pfizer, UCB, Roche, Celgene, outside the submitted work. M.J. has received honoraria/travel support from Pfizer, Abbvie and UCB. S.G. is supported by a National Institute for Health Research Manchester Musculoskeletal Biomedical Research Unit PhD studentship. All other authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

1 Prescribing & Medicines Team, Health and Social Care Information Centre. Prescribing Costs in Hospitals and the Community, England 2014–15. November 2015.

2 National Audit Office. Services for People with Rheumatoid Arthritis. July 2009. https://www.nao.org.uk/wp-content/uploads/2009/07/0809823.pdf (28 June 2016, date last accessed).

3 Gavan S, Harrison M, Iglesias C *et al*. Economics of stratified medicine in rheumatoid arthritis. Curr Rheumatol Rep 2014;16:468.

4 Faculty of Public Health. Priority Setting in the NHS. Position Statement by Faculty of Public Health, 2014. http://www.fph.org.uk/uploads/Policy_reports (28 June 2016, date last accessed).

5 Garcês S, Demengeot J, Benito-Garcia E. The immunogenicity of anti-TNF therapy in immune-mediated inflammatory diseases: a systematic review of the literature with a meta-analysis. Ann Rheum Dis 2013;72:1947–55.

6 Jani M, Isaacs JD, Morgan AW *et al*. High frequency of anti-drug antibodies and association of random drug levels with efficacy in certolizumab pegol-treated patients with rheumatoid arthritis: results from the BRAGGSS cohort. Ann Rheum Dis 2016 May 31. pii: annrheumdis-2015-208849.

7 Jani M, Chinoy H, Warren RB *et al*. Clinical utility of random anti–tumor necrosis factor drug–level testing and measurement of antidrug antibodies on the long-term treatment response in rheumatoid arthritis. Arthritis Rheumatol 2015;67:2011–9.

8 Xu X, Grossetta Nardini HK, Ruger JP. Micro-costing studies in the health and medical literature: protocol for a systematic review. Syst Rev 2014;3:47.

9 Jani M, Reid V, Parker B, *et al*. Harmonised Biologics Pathway for Rheumatoid Arthritis. Greater Manchester Medicines Management/MAHSC Guidelines. http://gmmmg.nhs.uk/docs/guidance/GMMMG RA Pathway 22 april_2015.pdf (28 June 2016, date last accessed).

10 Personal Social Services User Research Unit (PSSURU). Unit Costs of Health and Social Care 2014. http://www.pssru.ac.uk/project-pages/unit-costs/2014/ (28 June 2016, date last accessed).

11 Health Careers, NHS. NHS Careers, Agenda for Change – Pay Rates from 1 April 2015. http://www.nhscareers.nhs.uk/working-in-the-nhs/pay-and-benefits/agenda-for-change-pay-rates/ (1 September 2015, date last accessed).

12 British Medical Association. Pay Scales for 2015–2016. 2015. http://bma.org.uk/practical-support-at-work/pay-fees-allowances/pay-scales (28 June 2016, date last accessed).

13 Hanly P, Céilleachair AÓ, Skally M *et al*. Direct costs of radiotherapy for rectal cancer: a microcosting study. BMC Health Serv Res 2015;15:184.

14 GOV.UK. NHS National Tariff Payment System 2015/16: draft price-setting models - Publications - GOV.UK (July 2014). https://www.gov.uk/government/publications/nhs-national-tariff-payment-system-201516-draft-price-setting-models (31 August 2015, date last accessed).

15 Bartelds G, Krieckaert C. Development of antidrug antibodies against adalimumab and association with disease activity and treatment failure during long-term follow-up. JAMA 2011;305:1460–8.

16 Tan SS, Oppe M, Zoet-Nugteren SK *et al*. A microcosting study of diagnostic tests for the detection of coronary

artery disease in the Netherlands. Eur J Radiol 2009;72:98–103.

17 NICE diagnostic guidance [DG22]. Therapeutic Monitoring of TNF-alpha Inhibitors in Crohn's Disease (LISA-TRACKER ELISA kits, IDKmonitor ELISA kits, and Promonitor ELISA kits). London: NICE, 2016. https://www.nice.org.uk/guidance/dg22 (28 June 2016, date last accessed).

18 Parker D, Belaud-Rotureau M-A. Micro-cost analysis of ALK rearrangement testing by FISH to determine eligibility for crizotinib therapy in NSCLC: implications for cost effectiveness of testing and treatment. Clin Med Insights Oncol 2014;8:145–52.

19 Robinson S, Roberts T, Barton P *et al*. Healthcare and patient costs of a proactive chlamydia screening programme: the Chlamydia Screening Studies project. Sex Transm Infect 2007;83:276–81.

20 Chen D-Y, Chen Y-M, Tsai W-C *et al*. Significant associations of antidrug antibody levels with serum drug trough levels and therapeutic response of adalimumab and etanercept treatment in rheumatoid arthritis. Ann Rheum Dis 2015;74:e16.

Appendix 36: Chapter Six - Internal Validity of the Decision Analytic Model

This appendix presents the results of the four procedures, described in Section 6.3.4.2, that were used to assess the internal validity of the decision analytic model in *Chapter Six*. Each procedure is described in turn: (i) Set all cost parameters equal to zero (Section A36.1), (ii) Set all QALY parameters equal to zero (Section A36.2), (iii) Increase the discount rate (Section A36.3), and (iv) Sample 100,000 random values from the survival curves (Section A36.4).

A36.1 Set all Cost Parameters Equal to Zero

The input parameters that accounted for the unit costs of testing, treatments, and hospitalisation (Section 6.3.3.4) were all set to equal zero. The model was then run for *Current Practice* and the most complex intervention strategy (*Strategy 3*). The model subsequently estimated the total expected costs for each strategy to be zero. Therefore, the logic written within the model's code to calculate total expected costs was internally valid.

A36.2. Set all QALY Parameters Equal to Zero

The three input parameters that were used in the QALY mapping algorithm within the model (Section 6.3.3.2) were set to equal zero. The model was then run for *Current Practice* and the most complex intervention strategy (*Strategy 3*). The model subsequently estimated the total expected QALYs for each strategy to be zero. Therefore, the logic written within the model's code to calculate total expected QALYs was internally valid.

A36.3. Increase the Discount Rate

The rate at which future costs and QALYs were discounted was adjusted to 0%, 3.5%, and 5%. The model was run for current practice and the most complex intervention strategy (*Strategy 3*) for each of the discount rates. The model subsequently estimated expected outcomes such that, the higher the discount rate, the lower the present value of expected costs and QALYs. Therefore, the logic written within the model's code to discount future costs and QALYs was internally valid.

A36.4. Sample from Survival Distributions

Four survival curves were used within the model that simulated each patient's time to (i) death, for men, (ii) death, for women, (iii) bDMARD treatment failure, and (iv) developing ADAb against adalimumab. 100,000 values were sampled randomly from these survival curves to appraise their internal validity and clinical plausibility. Table A36.1 reports the mean values from these random samples. A histogram of the 100,000 random samples from each survival curve was produced; each survival curve is now discussed.

Table A36.1. *Mean time-to-event values from 100,000 random samples drawn from the survival curves in the decision analytic model.*

Parameter Estimated by Survival Curve	Mean (Years)
Time to death; women	82.78 years.
Time to death; men	79.27 years.
Time to bDMARD failure	4.32 years.
Time to developing ADAb	4.10 years.

Time to Death

Figure A36.1 and Figure A36.2 illustrate the histograms of the 100,000 random samples from the Gompertz survival curves that described time-to-death for men and women, respectively. The histograms demonstrated a clinically plausible distribution of ages at which individuals were expected to die from all-causes, in alignment with the life table data provided by the *Office for National Statistics* (2016). The mean age of death was greater for women than men (82 years compared with 79 years) which was also consistent with the national life table data for England.

Time to bDMARD Treatment Failure

Figure A36.3 illustrates the histogram of the 100,000 random samples from the Weibull survival curve that described the time-to-treatment failure of a bDMARD therapy. The histogram demonstrated a clinically plausible distribution of the time (in years) at which a patient would have maintained response to a bDMARD therapy. The mean time to bDMARD treatment failure was 4.32 years, consistent with the meta-analysis performed by Souto et al. (2016) which estimated that approximately fifty percent of patients with RA would lose response to their TNFi therapy after four years.

Time to Developing ADAb against Adalimumab

Figure A36.4 illustrates the histogram of 100,000 random samples from the log-normal survival curve that described the time-to-developing ADAb against adalimumab. The histogram demonstrated a clinically plausible distribution of the times-to-developing ADAb against adalimumab; the evidence within the clinical literature suggested that patients who had developed ADAb against a TNFi therapy had done so within the years that immediately followed treatment initiation (Bartelds et al., 2011).

Figure A36.1. *Histogram of 100,000 random samples from the Gompertz survival curve that characterised all-cause mortality for men.*

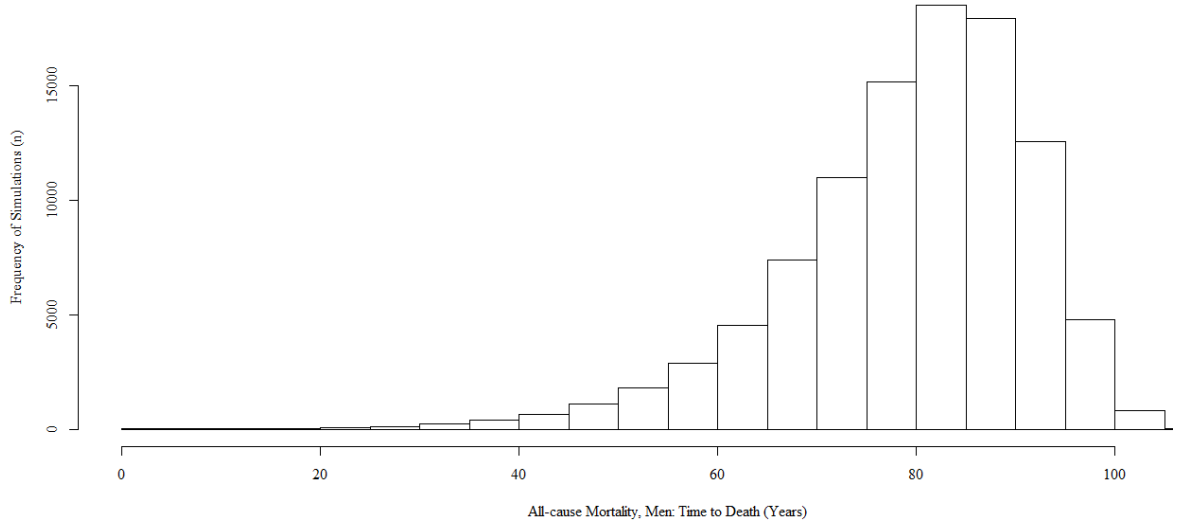


Figure A36.2. *Histogram of 100,000 random samples from the Gompertz survival curve that characterised all-cause mortality for women.*

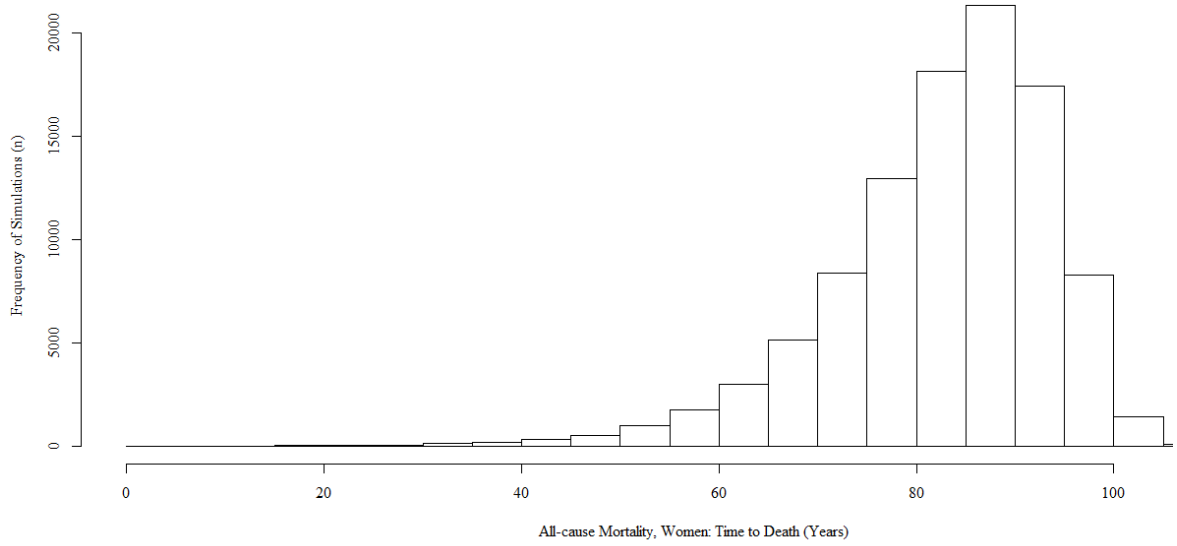


Figure A36.3. *Histogram of 100,000 random samples from the Weibull survival curve that characterised bDMARD treatment failure.*

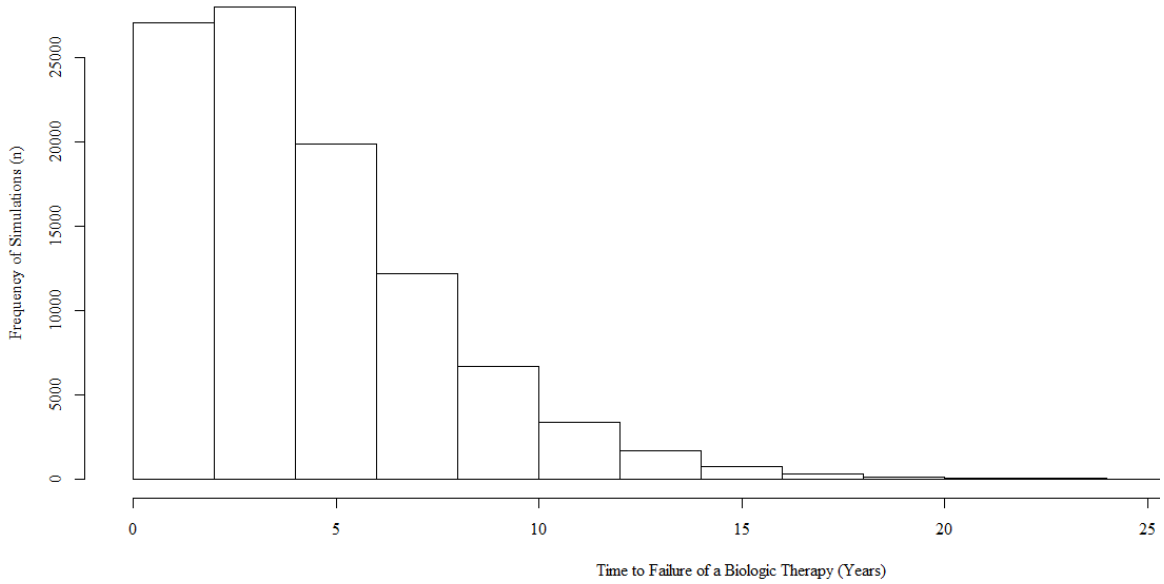
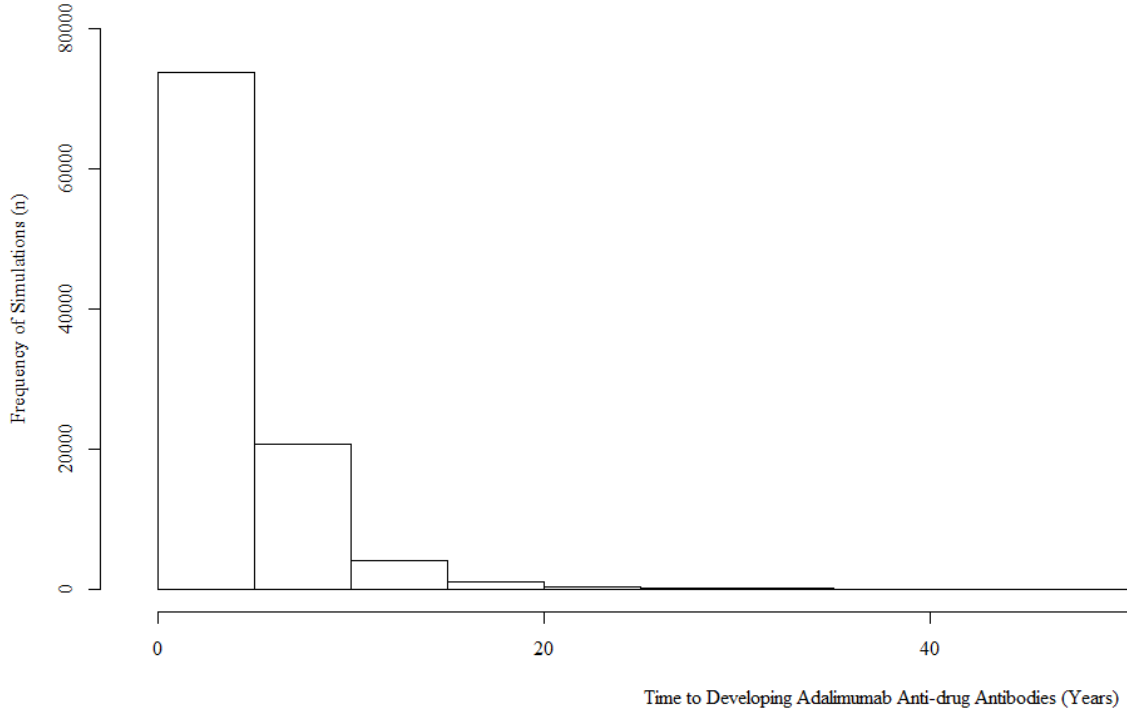


Figure A36.4. *Histogram of 100,000 random samples from the log-normal survival curve that characterised the time to developing ADAb against adalimumab.*



A36. References

Bartelds, G., Krieckaert, C., Nurmohamed, M., van Schouwenburg, P., Lems, W., Twisk, J., Dijkmans, B., Aarden, L., & Wolbink, G. (2011). "Development of Antidrug Antibodies Against Adalimumab and Association with Disease Activity and Treatment Failure During Long-term Follow-up", *Journal of the American Medical Association*, Vol. 305, 14, pp. 1460-1468.

Office for National Statistics. (2015). "National Life Tables: England (2012-2014)". Newport: Office for National Statistics.

Souto, A., Maneiro, J., & Gómez-Reino, J. (2016). "Rate of Discontinuation and Drug Survival of Biologic Therapies in Rheumatoid Arthritis: A Systematic Review and Meta-Analysis of Drug Registries and Health Care Databases", *Rheumatology*, Vol. 55, 3, pp. 523-534.

Appendix 37: Chapter Six - Distributions that Characterised the Input Parameters

It was necessary to characterise the input parameters of the decision analytic model in *Chapter Six* as probability distributions in order to perform a PSA (Briggs et al., 2006). This appendix describes how these distributions were chosen, with reference to the distributions described in *Appendix 4* (Beta; Dirichlet; Lognormal; Gamma; Normal; Multivariate normal). The distributions are described for clinical parameters (Section A37.1), resource use (Section A37.2), and QALYs (Section A37.3). Unit costs were not assigned probability distributions because they were assumed to be fixed (and not subject to parameter uncertainty) (Briggs et al., 2006).

A37.1. Clinical Parameters

Section A37.1 explains the distributions that characterised each input parameter in Table 6.4.

Time to Death

National life table data from the Office for National Statistics (2015) were used to estimate a patient's time to death. Gompertz survival curves were fit to these life table data for men and women, which were defined by two parameters (the shape and rate). The variance-covariance matrices were known for the shape and rate parameters based on the regression-output from the survival analysis (reported in *Appendix 31*). A Cholesky decomposition of the variance-covariance matrices was performed (described in *Appendix 4*), which enabled the sampled shape and rate parameters to be correlated, by characterising their probability distribution as a multivariate normal distribution. The parameters that defined this multivariate normal distribution are reported in Table 6.4.

Mortality Adjustment for Rheumatoid Arthritis

The mortality adjustment for patients with RA was described by the hazard ratios estimated by Michaud et al. (2012). Lognormal distributions were used to characterise uncertainty in the hazard ratios reported by Michaud et al. (2012) by following the procedure described in *Appendix 4*.

EULAR Response to Treatment

A EULAR response to treatment was a multinomial outcome (there were three possible outcomes: a *good* response, a *moderate* response, or no response). Stevenson et al. (2016) provided evidence from a network meta-analysis on the probabilities associated with each EULAR response for each treatment. A three-parameter Dirichlet distribution was used to characterise the uncertainty in the EULAR response probabilities for the PSA by using a series of conditional beta distributions, described in *Appendix 4*.

HAQ Improvement following a EULAR Response

The HAQ improvement following a EULAR response was always a negative value (a HAQ improvement meant that the HAQ score *reduced*). The potential value of a reduction in HAQ was therefore bound between zero and negative infinity. A gamma distribution was used to characterise the uncertainty in the HAQ reduction, based on the mean and standard errors estimated by Stevenson et al. (2016), according to the method of moments (described in *Appendix 4*).

Time to Treatment Failure

The time to treatment failure was estimated by survival analysis (*Appendix 32*), similarly to estimating the time to death. A Cholesky decomposition of the variance-covariance matrix from the survival analysis was therefore performed, to induce correlation between the shape and scale parameters of the Weibull survival curve, to enable those parameters to be sampled from a multivariate normal distribution.

Time to Developing ADAb against Adalimumab

A Cholesky decomposition of the variance-covariance matrix from the survival analysis that estimated the time to developing ADAb against adalimumab (*Appendix 33*) was performed, to induce correlation between the parameters that described the log-normal survival distribution. These parameters were sampled from a multivariate normal distribution in the PSA, the values for which are defined in Table 6.4.

Consequence of Developing ADAb against Adalimumab

Garcês et al. (2013) provided evidence for a relative risk parameter that was used to adjust the time to adalimumab treatment failure in the model. This relative risk parameter was sampled from a lognormal distribution in the PSA, using the technique described in *Appendix 4*.

Proportion of Patients with Low Adalimumab Drug Levels in Remission

The annual probability of a disease flare, one year after TNFi dose reduction, was estimated using data from a systematic review and meta-analysis by Kuijper et al. (2015), reported in Section 6.3.3.1.9. This probability was characterised by a beta distribution in the PSA, by the method of moments (see *Appendix 4*), based on the mean and confidence interval reported by Kuijper et al. (2015).

Consequence of a Flare in Remission

A gamma distribution (bound between zero and positive infinity) was used to characterise the uncertainty associated with a change in HAQ following a flare in disease activity. Using the evidence reported by Markusse et al. (2015), a legitimate HAQ increase of 0.125 units was expected following a flare. However, Markusse et al. (2015) did not report a standard error for this estimate. Therefore, following the approach of Briggs et al. (2006), it was assumed that the standard error was equal to the mean (0.125), and the method of moments was used to characterise a gamma distribution (see Appendix 4) with the parameter values reported in Table 6.4.

Accuracy of Adalimumab ADAb Test

The accuracy of adalimumab ADAb testing were derived from the ROC analysis by Jani et al. (2016a). Beta distributions (see *Appendix 4*) were used to characterise the uncertainty in the sensitivity and specificity of testing, by using the mean and confidence interval reported by Jani et al. (2016a). No correlation was induced between test sensitivity and specificity because their co-variance was not reported within the published study.

Accuracy of Adalimumab Drug Level Testing (Full Dose)

Appendix 34 reported a hierarchical meta-analysis of test accuracy studies for adalimumab drug level testing. This study estimated the covariance between the synthesised estimates of sensitivity and specificity. Therefore, a Cholesky decomposition of the variance-covariance matrix between test sensitivity and specificity was performed, to induce correlation between their sampled values in the PSA. The parameter values that defined the multivariate normal distribution are reported in Table 6.4.

Accuracy of Adalimumab Drug Level Test (Half Dose)

The accuracy of drug level testing when the dose of adalimumab was halved was estimated by Chen et al. (2016). Chen et al. (2016) reported the mean sensitivity and specificity of testing, but did not report an associated standard error. Using the data reported by Chen et al. (2016) it was possible to calculate the number of patients that had a true-positive, false-positive, true-negative, and false-negative test result. A beta distribution was fit to these data, separately for sensitivity and specificity, such that the number of *true* test outcomes defined the "number of events" (α) parameter of the distribution (see Appendix 4).

Consequence of Treatment Decisions following Routine Monitoring

Routine monitoring of adalimumab by ADAb and drug level testing may have led to a pre-emptive change of treatment to rituximab (Section 6.3.3.1.12). The effect of this treatment change was represented by a multiplier on the patient's HAQ rebound (Table 6.5). However, there was no evidence to support the value of these HAQ multipliers because the tests were relatively early in their product lifecycle. Therefore, uninformative uniform distributions were used to characterise the uncertainty in the HAQ multipliers during the PSA. The uniform distributions sampled any value between (and inclusive of) zero and one with equal probability (Briggs et al., 2006).

A37.2. Resource use

Section A37.2 explains how parameter uncertainty was characterised in the model's input parameters for resource use.

Treatments

All treatments were assumed to be administered at the dose recommended in the *British National Formulary* (2016). The resources required to perform an intravenous infusion, and the proportion of subcutaneous injections performed by a nurse, were also assumed to be fixed. Therefore, all health care resources that were specifically related to treatments were not subject to parameter uncertainty in the PSA.

Testing

The resources required to incorporate testing in routine practice were quantified by the supplementary microcosting study (reported in *Appendix 35*) by Jani et al. (2016b). These resources were assumed to be necessary to perform the ADAb and drug level tests and were therefore assumed to be fixed in the PSA in *Chapter Six*.

Hospitalisations

The mean days hospitalised per year were estimated for six categories of HAQ score by using evidence submitted to a previous *NICE Single Technology Appraisal* for RA (Section 6.3.3.3.2). The distribution of health care resources is often positively-skewed and can therefore be characterised by a gamma distribution (bound between zero and positive infinity) (see *Appendix 4*; Briggs et al., 2006). However, there were no standard errors reported with the mean days hospitalised. Therefore, following the approach of Briggs et al. (2006), it was assumed that for each category of HAQ score, the mean days hospitalised was equal to the standard error. A gamma distribution was subsequently used in the PSA, with the parameter values reported in Table 6.4, to characterise parameter uncertainty in the mean days hospitalised per year.

A38.3. Quality-adjusted Life Years

The value of a patient's EQ-5D score was estimated from their prevailing HAQ score by using the mapping algorithm reported by Malottki et al. (2011) in Section 6.3.3.2. Malottki et al. (2011) reported the mean and confidence intervals for each of the three parameters within the mapping algorithm. The PSA therefore sampled these mapping algorithm parameter values from three independent normal distributions.

A37. References

Briggs, A., Claxton, K., & Sculpher, M. (2006). *Decision Modelling for Health Economic Evaluation*. (1 Ed.). Oxford: Oxford University Press.

British National Formulary. (2016). "British National Formulary: Accessed via MedicinesComplete". [Online] Available at: http://www.medicinescomplete.com/mc/bnf/current/ [Accessed 8 April 2016].

Chen, D., Chen, Y., Hsieh, T., Hung, W., Hsieh, C., Chen, H., Tang, K., & Lan, J. (2016). "Drug Trough Levels Predict Therapeutic Responses to Dose Reduction of Adalimumab for Rheumatoid Arthritis Patients During 24 Weeks of Follow-up", *Rheumatology*, Vol. 55, 1, pp. 143-148.

Garcês, S., Demengeot, J., & Benito-Garcia, E. (2013). "The Immunogenicity of Anti-TNF Therapy in Immune-Mediated Inflammatory Diseases: A Systematic Review of the Literature with a Meta-Analysis", *Annals of the Rheumatic Diseases*, Vol. 73, 1, pp. 1947-1955.

Jani, M., Gavan, S., Dixon, W., Harrison, B., Moran, A., Barton, A., & Payne, K. (2016b). "A Microcosting Study of Immunogenicity and TNFi Drug Level Tests for Therapeutic Monitoring in Clinical Practice", *Rheumatology*, Vol. 55, 12, pp. 2131-2137.

Jani, M., Isaacs, J., Morgan, A., Wilson, A., Plant, D., Hyrich, K., Chinoy, H., & Barton, A. (2016a). "Detection of Anti-drug Antibodies using a Bridging ELISA Compared with Radioimmunoassay in Adalimumab-treated Rheumatoid Arthritis Patients with Random Drug Levels", *Rheumatology*, Vol. 55, 11, pp. 2050-2055.

Kuijper, T., Lamers-Karnebeek, F., Jacobs, J., Hazes, J., & Luime, J. (2015). "Flare Rate in Patients with Rheumatoid Arthritis in Low Disease Activity or Remission When Tapering or Stopping Synthetic or Biologic DMARD: A Systematic Review", *The Journal of Rheumatology*, Vol. 42, 11, pp. 2012-2022.

Malottki, K., Barton, P., Tsourapas, A., Uthman, A., Lui, Z., Routh, K., Cannock, M., Jobanputra, P., Moore, D., Fry-Smith, A., & Chen, Y. (2011). "Adalimumab, Etanercept, Infliximab, Rituximab and Abatacept for the Treatment of Rheumatoid Arthritis after the Failure of a Tumour Necrosis Factor Inhibitor: a Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 15, 14, pp. 1-278.

Markusse, I., Dirven, L., Gerards, A., van Groenendal, J., Ronday, H., Kerstens, P., Lems, W., Huizinga, T., & Allaart, C. (2015). "Disease Flares in Rheumatoid Arthritis are Associated with Joint Damage Progression and Disability: 10-year Results from the BeSt Study", *Arthritis Research & Therapy*, Vol. 17, 232, pp. 1-9.

Michaud, K., Vera-Llonch, M., & Oster, G. (2012). "Mortality Risk by Functional Status and Health-Related Quality of Life in Patients with Rheumatoid Arthritis", *The Journal of Rheumatology*, Vol. 39, 1, pp. 54-59.

Office for National Statistics. (2015). "National Life Tables: England (2012-2014)". Newport: Office for National Statistics.

Stevenson, M., Archer, R., Tosh, J., Simpson, E., Everson-Hock, E., Stevens, J., Hernandez-Alava, M., Paisley, S., Dickinson, K., Scott, D., Young, A., & Wailoo, A. (2016). "Adalimumab, Etanercept, Infliximab, Certolizumab Pegol, Golimumab, Tocilizumab, and Abatacept for the Treatment of Rheumatoid Arthritis not Previously Treated with Disease-Modifying Antirheumatic Drugs and After Failure of Conventional Disease-Modifying Antirheumatic Drugs Only: Systematic Review and Economic Evaluation", *Health Technology Assessment*, Vol. 20, 35, pp. 1-610.

Appendix 38: Chapter Six - Deterministic Sensitivity Analysis of Base-case Result

This appendix reports the results of all deterministic sensitivity analyses of the base-case result derived from the *de novo* model-based economic evaluation in *Chapter Six*. The table that corresponds to each sensitivity analysis is as follows:

Test Characteristics

Table A38.1: Reduce cost of all tests to £20;

Table A38.2: Assume no benefit from a true-positive monitoring test result;

Table A38.3: Assume no harm from a false-positive monitoring test result;

Table A38.4: Assume no benefit from a true-positive monitoring test result *and* no harm from a false-positive monitoring test result;

Table A38.5: Assume that all tests are perfectly accurate.

Disease Characteristics

Table A38.6: Assume a lower probability of having low adalimumab drug levels in remission;

Table A38.7: Assume a higher probability of having low adalimumab drug levels in remission;

Table A38.8: Assume a lower relative risk of losing response to adalimumab after developing ADAb;

Table A38.9: Assume a higher relative risk of losing response to adalimumab after developing ADAb;

Table A38.10: Assume no HAQ progression whilst receiving methotrexate therapy only;

Treatment Characteristics

Table A38.11: Assume that the cost of adalimumab is one third lower to represent the use of biosimilar adalimumab.

Structural Assumptions

Table A38.12: Assume a discount rate of 0% for costs and QALYs;

Table A38.13: Assume a discount rate of 6% for costs and 1.5% for QALYs;

Table A38.14: Assume that time to treatment failure was sampled from a log-normal survival curve;

Table A38.15: Assume that QALYs were estimated from HAQ by using the mapping algorithm by Adams et al. (2011);

Table A38.16: Assume that QALYs were estimated from HAQ by using the mapping algorithm by Barton et al. (2004).

Table A38.1. *Deterministic sensitivity analysis: test cost of £20.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	110,802	2.739957	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	111,131	2.761210	328.99	0.021253	15,480	96.07	308.60	733.6
Strategy 1	113,882	2.839540	2,750.60	0.078330	35,116	-1,184	-400.70	1,165.90

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.2. *Deterministic sensitivity analysis: no benefit from true-positive monitoring test result.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 1	114,989.01	2.697814	Dominated by Strategy 11		N/A	N/A	N/A	N/A
Strategy 3	111,679.41	2.726842	Dominated by Strategy 11		N/A	N/A	N/A	N/A
Strategy 11	110,802.42	2.739957	N/A	N/A	N/A	N/A	N/A	N/A

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.3. *Deterministic sensitivity analysis: no harm from false-positive monitoring test result.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	110,802.42	2.739957	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	111,580.49	2.764169	778.06	0.024212	32,135.30	-293.82	-51.70	432.54
Strategy 1	114,597.03	2.844088	3,016.54	0.079919	37,745.16	-1,418.17	-618.98	979.39

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.4. *Deterministic sensitivity analysis: no harm from false-positive monitoring test result and no benefit from true-positive monitoring test result.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 1	114,974.76	2.702361	Dominated by Strategy 11		N/A	N/A	N/A	N/A
Strategy 3	111,670.20	2.729801	Dominated by Strategy 11		N/A	N/A	N/A	N/A
Strategy 11	110,802.42	2.739957	N/A	N/A	N/A	N/A	N/A	N/A

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.5. *Deterministic sensitivity analysis: perfect accuracy of all tests.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	110,802.42	2.739957	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	111,636.16	2.805580	833.74	0.065624	12,704.89	478.73	1,134.97	2,447.44
Strategy 1	114,401.34	2.908273	2,765.18	0.102692	26,926.83	-711.33	315.59	2,369.44

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.6. *Deterministic sensitivity analysis: lower probability of low drug levels in remission.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	110,599.95	2.739994	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	111,435.38	2.761507	835.44	0.021513	38,833.92	-405.18	-190.04	240.22
Strategy 1	114,611.27	2.839540	3,175.89	0.078033	40,699.37	-1,615.23	-834.90	725.76

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.7. *Deterministic sensitivity analysis: higher probability of low drug levels in remission.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	111,065.38	2.739907	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	111,908.25	2.761982	842.87	0.022074	38,183.35	-401.39	-180.64	260.85
Strategy 1	114,611.27	2.839540	2,703.03	0.077559	34,851.41	-1,151.85	-376.27	1,174.90

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.8. *Deterministic sensitivity analysis: lower relative risk of ADAb development.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	110,506.67	2.749627	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	111,303.78	2.773032	797.11	0.023404	34,057.89	-329.02	-94.97	373.12
Strategy 1	114,572.62	2.845931	3,268.84	0.072899	44,840.58	-1,810.86	-1,081.87	376.12

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.9. *Deterministic sensitivity analysis: higher relative risk of ADAb development.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	110,586.19	2.749962	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	111,480.92	2.770236	894.73	0.020274	44,131.75	-489.25	-286.51	118.97
Strategy 1	114,593.94	2.845209	3,113.01	0.074973	41,522.05	-1,613.56	-863.84	635.61

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.10. *Deterministic sensitivity analysis: no HAQ progression on cDMARDs.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	108,822.15	3.456676	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	109,596.50	3.483455	774.35	0.026779	28,916.25	-238.77	29.02	564.60
Strategy 1	112,600.48	3.574900	3,003.98	0.091445	32,850.04	-1,175.07	-260.62	1,568.28

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.11. *Deterministic sensitivity analysis: biosimilar adalimumab.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	105,598.54	2.739957	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 1	108,959.86	2.839540	3,361.32	0.099584	33,753.82	-1,369.65	-373.82	1,617.85
Strategy 3†	106,441.67	2.761210	ED†	ED†	N/A	N/A	N/A	N/A

Note: †=*Strategy 3* was extendedly dominated by 75% of *Strategy 11* and 25% of *Strategy 1*; λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.12. *Deterministic sensitivity analysis: Discount rate – 0% for costs and QALYs.*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	166,845.71	3.601181	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	167,599.07	3.634202	753.37	0.033021	22,814.69	-92.94	237.27	897.69
Strategy 1	170,945.11	3.766379	3,346.04	0.132178	25,314.68	-702.48	619.30	3,262.85

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.13. *Deterministic sensitivity analysis: Discount rate – 6% for costs and 1.5% for QALYs.*

						Incremental Net Monetary Benefit (£)		
Comparator Strategy	**Mean Cost (£)**	**Mean QALY**	**Incremental Cost (£)**	**Incremental QALY**	**ICER (£ per QALY)**	**λ = £20,000 per QALY**	**λ = £30,000 per QALY**	**λ = £50,000 per QALY**
Strategy 11	87,811.23	3.190146	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	88,594.81	3.217159	783.57	0.027013	29,007.62	-243.32	26.81	567.06
Strategy 1	91,377.43	3.321408	2,782.62	0.104249	26,692.01	-697.64	344.86	2,429.84

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.14. *Deterministic sensitivity analysis: Log-normal time to bDMARD treatment failure.*

						Incremental Net Monetary Benefit (£)		
Comparator Strategy	**Mean Cost (£)**	**Mean QALY**	**Incremental Cost (£)**	**Incremental QALY**	**ICER (£ per QALY)**	**λ = £20,000 per QALY**	**λ = £30,000 per QALY**	**λ = £50,000 per QALY**
Strategy 11	109,754.80	2.841893	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	110,694.97	2.861321	940.17	0.019428	48,393.41	-551.62	-357.34	31.21
Strategy 1	114,171.00	2.932965	3,476.03	0.071644	48,517.92	-2,043.14	-1,326.70	106.18

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.15. *Deterministic sensitivity analysis: QALY mapping algorithm by Adams et al. (2011).*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	110,802.42	5.383739	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	111,589.70	5.393951	787.27	0.010211	77,097.86	-583.05	-480.93	-276.71
Strategy 1	114,611.27	5.431183	3,021.58	0.037232	81,154.93	-2,276.93	-1,904.61	-1,159.97

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

Table A38.16. *Deterministic sensitivity analysis: QALY mapping algorithm by Barton et al. (2004).*

Comparator Strategy	Mean Cost (£)	Mean QALY	Incremental Cost (£)	Incremental QALY	ICER (£ per QALY)	Incremental Net Monetary Benefit (£)		
						λ = £20,000 per QALY	λ = £30,000 per QALY	λ = £50,000 per QALY
Strategy 11	110,802.42	2.815876	N/A	N/A	N/A	N/A	N/A	N/A
Strategy 3	111,589.70	2.835752	787.27	0.019876	39,609.91	-389.76	-191.00	206.51
Strategy 1	114,611.27	2.908222	3,021.58	0.072470	41,694.28	-1,572.18	-847.48	601.91

Note: λ = cost-effectiveness threshold; QALY = quality-adjusted life-year.

A38. References

Adams, R., Craig, B., Veale, D., Bresnihan, B., Walsh, C., FitzGerald, O., & Barry, M. (2011). "The Impact of a Revised EQ-5D Population Scoring on Preference-based Utility Scores in an Inflammatory Arthritis Cohort", *Value in Health*, Vol. 14, 6, pp. 921-927.

Barton, P., Jobanputra, P., Wilson, J., Bryan, S., & Burls, A. (2004). "The use of Modelling to Evaluate New Drugs for Patients with a Chronic Condition: the Case of Antibodies against Tumour Necrosis Factor in Rheumatoid Arthritis", *Health Technology Assessment*, Vol. 8, 11, pp. 1-91.