Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

COST-EFFECTIVENESS OF ANTI-TUMOUR NECROSIS FACTOR DRUGS FOR THE TREATMENT OF RHEUMATOID ARTHRITIS IN IRELAND

ROISIN C. ADAMS

MPharm, MSc.

A thesis submitted for the degree of Doctor in Philosophy

National Centre for Pharmacoeconomics

&

Department of Pharmacology and Therapeutics

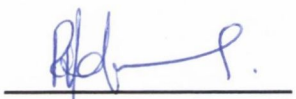
University of Dublin, Trinity College

2012

Declaration

I certify that none of the work in this thesis has been submitted previously for any degree or diploma at this, or any other, university, and that the work described in this thesis, except where duly acknowledged, is my own.

I hereby agree that the Library of Trinity College Dublin may lend or copy this thesis upon request.



Roisin C. Adams

Summary

Rheumatoid Arthritis (RA) is the most common form of inflammatory arthritis. It is a chronic condition that places a substantial burden on patients and their carers, imposing a negative effect on quality of life (QOL), including physical, psychological, and social functioning and is associated with premature mortality. Treatment pathways include disease modifying agents (DMARDs) followed by anti-TNF therapy if poor or loss of response is evident. In 2009, expenditure for anti-TNF drugs reached approximately €100 million. All of the anti-TNF agents are used and reimbursed either under the community drugs schemes, or by hospitals. These include adalimumab, certolizumab, etanercept, golimumab, and infliximab.

This thesis examines the cost-effectiveness of anti-TNF agents for treatment of RA in the Irish healthcare setting. Two of these agents have undergone pharmacoeconomic assessment and were initially refused reimbursement due to uncertainty associated with their cost-effectiveness. In order to assess the cost effectiveness of these agents in the Irish setting an economic evaluation was performed.

The population chosen was those patients with established RA, who have had an inadequate response to methotrexate. The Birmingham Rheumatoid Arthritis Model (BRAM) was adapted and used for the assessment. The BRAM inputs were derived from a combination of sources including randomised controlled data, international observational data, and local data. The analyses for these inputs were divided into three main sections; QOL, effectiveness and costs.

QOL instruments used in inflammatory arthritis trials can include either a generic measure such as the EQ-5D or the SF-36 or a disease specific instrument such as the HAQ, or both. Data used in this study was derived from an Irish biological population of RA patients which collected EQ-5D, SF-6D, HAQ, and DAS 28. Preference based utility indicated that the EQ-5D produced twice the QALY gain compared to the SF-6D. A mapping exercise was performed from both the HAQ and the DAS 28 to the SF-6D and the EQ-5D. A further analysis was performed which attempted to correct the anomaly associated with the high number of worse than death

states shown in the EQ-5D utilities. The utility values were rescored using this adjusted scoring system.

A mixed treatment comparison was performed to combine data from the RCTs and calculate a relative effect for HAQ improvement; categorical outcomes were also assessed - ACR 20 and the ACR 50. Observational data was combined to estimate the length of time patients typically remain on these agents and a meta-analysis combined data for short term discontinuations. Irish cost data was used for the model. The perspective was that of the payer over a lifetime time horizon. The comparator was leflunomide. The basecase used the mapping coefficients for the revised EQ-5D scoring. The basecase results indicate that the anti-TNF agents are not cost-effective at a willingness to pay threshold of less than €100,000/ QALY. The methods used to calculate utility gain and the assumptions around long term HAQ improvement have most effect on the results of the model.

These results present an aggregate estimate of the cost-effectiveness of anti-TNF therapy in RA patients in Ireland, which is the relevant question for the payer. It may be of benefit to explore further whether these agents are cost-effective in particular subgroups. Additional strategies to reduce the drug costs associated with these agents may also be of benefit.

Table of Contents **Page**

Acknowledgements

I would like to thank my supervisor, Professor Michael Barry, for his support and advice throughout this thesis. His complete support of our role as students and researchers provided a great PhD experience.

Professor Oliver FitzGerald's interest in the area of pharmacoeconomics set the seed for this project and without the complete access to real life data in the rheumatology unit in St. Vincents University hospital which he provided, this project would not have grown and developed as it did.

The research presented here could not have been completed without the collaboration of Dr. Pelham Barton. Dr. Barton provided the BRAM model and facilitated the adaptation of the model. Professor Benjamin Craig collaborated on the revised scoring for the EQ-5D. I am very grateful for the assistance of both Dr. Barton and Professor Craig.

Dr. Lesley Tilson has very often been my first port of call for queries which she almost always seemed to have an answer for. Lesley encouraged ideas, provided clear guidance, and was enthusiastic about all aspects of the research.

Throughout my work on this PhD I have worked closely with Susanne Schmitz. Susi performed the technical analysis associated with the mixed treatment comparison work. Working with a fellow PhD student in this way was both exciting and rewarding. I am immensely grateful to Susi, for her support.

I started this PhD around the time that Professor Cathal Walsh joined the team here and was incredibly lucky to be able to avail of his expertise. Cathal has a unique ability to translate the often mind-boggling pages of squiggles into a simple one line message. There were many occasions when he walked into a despairing office and turned it into a box of creative ideas. I am immensely grateful for his guidance and support.

The learning roller coaster I have been on since joining the NCPE has been one I have enjoyed with a number of really great colleagues; Cara, Aisling and Emer - thank you! My journey here started with Laura. We have shared lots in the last three and half years and in all Laura has been a great support and I have learnt an immense amount through discussion of ideas with her.

Thanks to Cliona, Regina and Tom for their help with proof reading and formatting crises. Last and certainly not least I thank my family. They have encouraged me through everything I do. I am very, very grateful for that.

List of Tables **Page**

List of Figures

Page

List of Appendices

List of Abbreviations

ACR	American College of Rheumatology
ACR20	20% improvement in the ACR criteria
ACR50	50% improvement in the ACR criteria
ACR-N	American College of Rheumatology N
AIMS	Arthritis Impact Score
ALT	Alanine Transaminase
AS	Ankylosing Spondylitis
AST	Aspartate Transaminase
AUC	Area Under the Curve
BRAM	Birmingham Rheumatoid Arthritis Model
BSR	British Society of Rheumatology
BSRBR	British Society of Rheumatology Biologics Register
BTD	Better than Death
CADTH	Canadian Agency for Drugs and Technology in Health
CASPAR	Classification criteria for Psoriatic Arthritis
CBA	Cost benefit Analysis
CCP	Cyclic Citrullinated Peptide
CE	Cost-effectiveness
CEA	Cost-effectiveness Analysis
CEAC	Cost-effectiveness Acceptability Curve
CEAF	Cost-effectiveness Acceptability frontier
CMA	Cost minimisation Analysis
CNS	Clinical Nurse Specialist
CPU	Corporate Pharmaceutical Unit
CRP	C Reactive Protein
CSV	Comma Separated Value
CUA	Cost Utility Analysis
DAS	Disease Activity Score
DAS28	Disease Activity Score (28 joint count)
DMARD	Disease Modifying Anti-Rheumatic Drug
DOH&C	Department of Health and Children
DPS	Drugs Payment Scheme

DVT	Deep Vein Thrombosis
EMA	European Medicines Agency
EQ-5D	Euro QOL
ESR	Erythrocyte Sedimentation Rate
EULAR	European League Against Rheumatism
EVPI	Expected Value of Perfect Information
FDA	Food and Drugs Agency
FIT	Faecal Immunochemical Tests
FSIG	Flexible Sigmoidoscopy
GI	Gastrointestinal
GMS	General Medical Service
HAQ	Health Assessment Questionnaire
HAQDI	Health Assessment Questionnaire Disease Index
HIQA	Health and Information and Quality Authority
HIV	Human Immunodeficiency Virus
HPV	Human Papillovirus
HR	Hazard ratio
HRQOL	Health Related Quality of Life
HSE	Health Services Executive
HTA	Health Technology Assessment
HTDS	High Technology Drugs Scheme
HUI	Health Utilities Index
IA	Inflammatory Arthritis
ICER	Incremental Cost-effectiveness Ratio
IgG	Immunoglobulin
IL-1	Interleukin 1
IL-1α	Interleukin 1α
IL-6	Interleukin 6
IPHA	Irish Pharmaceutical Healthcare Association
IQR	Interquartile Range
ISR	Irish Society of Rheumatology
JIA	Juvenile Idiopathic Arthritis
KHAQ	Korean Health Assessment Questionnaire

LTI	Long Term Illness
LYG	Life Year Gained
MAUT	Multi-Attribute Utility Theory
MCID	Minimum Clinical Important Difference
MCMC	Markov Chain Monte Carlo
MHAQ	Modified Health Assessment Questionnaire
MID	Minimal Important Difference
MRI	Magnetic Resonance Imaging
MTA	Meta-Analysis
MTC	Mixed Treatment Comparison
MTX	Methotrexate
MVH	Measurement and Valuation of Health
NCPE	National Centre for Pharmacoeconomics
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NIH	National Institute for Health
NMA	Network Meta- Analysis
NMB	Net Monetary Benefit
NNT	Number needed to treat
OMERACT	Outcome Measures in Rheumatology
OR	Odds Ratio
OWSA	One Way Sensitivity Analysis
PAS	Patient Access Scheme
PBAC	Pharmaceutical Benefits and Advisory Committee
PCRS	Primary Care Reimbursement Service
PRO	Patient Reported Outcome
PROMIS	Patient Reported Outcome Information System
PSA	Probabilistic Sensitivity Analysis
PsA	Psoriatic Arthritis
QALY	Quality Adjusted Life Year
QOL	Quality of Life
RA	Rheumatoid Arthritis
RAQOL	Rheumatoid Arthritis Quality of Life Measure
RCT	Randomised Control Trial

RD	Risk Difference
RF	Rheumatoid Factor
RR	Relative Risk
RUM	Random Utility Model
SD	Standard Deviation
SE	Standard Error
SEM	Standard Error of the Mean
SENS	Simple Erosion Narrowing Score
SF-36	Short Form 36
SF-6D	Short Form 6 Dimension
SG	Standard Gamble
SJC	Swollen Joint Count
SMC	Scottish Medicines Consortium
SMR	Standardised Mortality Ratio
SPC	Summary of Product Characteristics
SPSS	Statistical Package for the Social Sciences
SSZ	Sulphasalazine
STA	Single Technology Appraisal
TA	Technology Appraisal
TJC	Tender Joint Count
TNFα	Tumour Necrosis Factor alpha
TTO	Time Trade Off
VAS	Visual Analogue Scale
VOI	Value of Information
WTA	Willingness to Accept
WTD	Worse Than Death
WTP	Willingness to Pay

Chapter 1

Chapter 1.0 Health Technology Assessment in Ireland

Expenditure on healthcare in Ireland has increased dramatically from €5.7 billion in 2000 to €15.4 billion in 2009.[1] Fifteen per cent of this figure was spent on drugs in 2009 (€2.24 billion). Not surprisingly, the issue of value for money arises and a number of reports have been commissioned to investigate this. In 2003, the Commission on Financial Management and Control Systems in the Health Service (Brennan report) recommended that a review be undertaken to ensure such expenditure provides good value for money. [1] Recently, reports have focused on specific areas for achieving the best value for money. In 2009, a report titled "Economies in Drug Usage in the Irish Healthcare Setting", examined potential ways to make savings within the drugs budget. The report recommended a number of initiatives including generic drug prescribing, price reductions for generic drugs, disinvestment for drugs without a sound evidence base, and review of the community drugs schemes and in particular the high technology drugs scheme (HTDS). [2]

In September 2006, a formal requirement for evidence of cost–effectiveness in reimbursement decisions for certain medicines was introduced under an agreement between the Irish Pharmaceutical Healthcare Association (IPHA) and the Health Services Executive (HSE). This agreement was renewed in 2010. [3] The agreement stipulated that high-cost products and those with a significant budget impact may be subjected to formal pharmacoeconomic assessment.

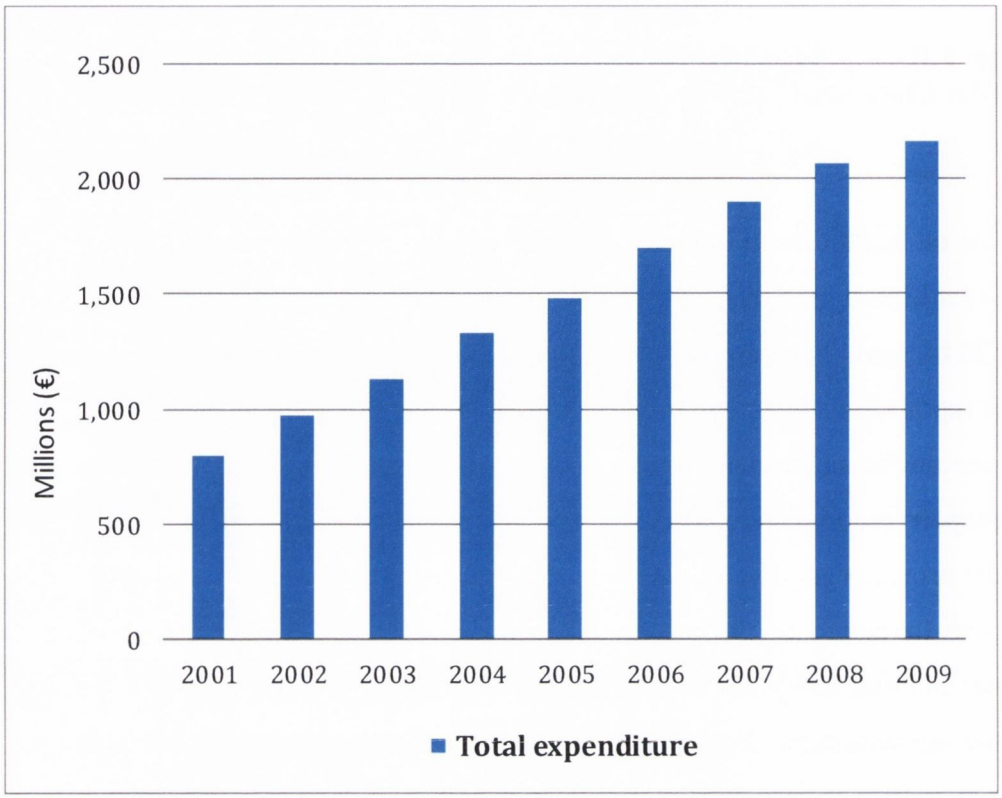


Figure 1: Medicine expenditure trends over a ten year period under the Primary Community Drugs Schemes in Ireland (GMS, LTI, DPS, and HTDS).

The drugs budget is an area which is frequently targeted for making savings in health expenditure. The challenge is to ensure maximum health benefit from the cost incurred using a systematic evidence based approach. Evidence based decision making for the payers is now more important than ever in order to achieve the best value for money in healthcare.

1.1 Health Technology Assessment

New technologies, including, but not limited to, pharmaceuticals are increasingly being subjected to Health Technology Assessment (HTA). There has been a drive to extract maximum efficiency so as to maintain the volume and quality of patient services at a time of very tight financial constraints. Introduction of new health technologies continues to accelerate.

HTA was first conceptualised in 1976. [4] The following general definition of technology assessment was used: ''a comprehensive form of policy research that

examines the short- and long-term social consequences of the application or use of technology". [4] Since the late 1990s HTA has been increasingly used as a research and evaluation approach to support health care policy. Those making healthcare coverage decisions rely on HTA for crucial technical information. Australia, Canada, and many European countries now use various forms of HTA in decision making regarding the reimbursement of drugs and other health technologies. [5]

1.1.1 Pharmacoeconomic assessment

Pharmacoeconomics has been defined as "the corporative analysis of costs and benefits of one technology over another". It identifies, measures, and compares the costs (i.e. resources consumed) and consequences (i.e. clinical, economic, and humanistic) of pharmaceutical products and services. [6]

1.1.1.2 Pharmacoeconomic assessment in Ireland

The National Centre for Pharmacoeconomics (NCPE) was established in 1998. The centre is funded by the Irish Department of Health and Children (DoH&C). The primary roles of the centre are to perform economic evaluations of pharmaceutical products and to promote cost-effective prescribing. The NCPE conducts all the pharmacoeconomic assessments for the HSE and pharmacoeconomic evaluations to inform public health policy (e.g. universal infant pneumococcal vaccination, cervical cancer vaccination and colorectal cancer screening) and prescribing in primary care (e.g. statins for primary and secondary prevention of coronary heart disease).[7-9]

A statutory HTA agency was established in Ireland in 2007, under the Health Act 2007, as part of the Health Information and Quality Authority (HIQA). This agency is responsible for ensuring that the resources in our health services are used in a way that ensures the best outcome for the patient or service user and for setting standards for conducting HTA in line with national and international practice. Their statutory remit is to assess the clinical and cost-effectiveness of the medicines, devices, diagnostics, and health promotion used across our health system.

The HSE-IPHA agreement

This agreement was drawn up between IPHA, which represents the pharmaceutical industry, and the HSE. Under the IPHA-HSE agreement a pharmacoeconomic assessment can be requested on any medicines deemed to be high cost or have a significant budget impact. The agreement links the price of medicines in Ireland to nine EU states including Belgium, Denmark, France, Germany, The Netherlands, Spain, the UK, Finland and Austria. Following receipt of market authorization a new product will be reimbursed within 60 days of the reimbursement application. However, the HSE reserves the right to assess the cost-effectiveness of new and existing technologies that may be high cost or have a significant budget impact. Where such a review is requested the 60-day rule will not apply. Where a new medicine is subjected to pharmacoeconomic assessment the reimbursement decision will be notified within 90 days of receipt of the reimbursement application. Products will be reimbursed within 40 days of a positive decision. Should reimbursement be refused an appeal may be made to an expert committee whose final decision will be made within a further 90 days. The agreement will be reviewed again in 2012.

Further developments in the pharmacoeconomic evaluation of medicines have since taken place and more recently pharmacoeconomic evaluation of medicines in Ireland include: a rapid assessment of all new medicines, a dynamic cost–effectiveness threshold, an increasing importance of budget-impact assessment, conditional reimbursement and cost–effectiveness evaluation of existing medicines that are currently reimbursed. [10] The rapid assessment process is completed within two weeks of receipt of the rapid review assessment form from the company in question. [11] Following this, a decision is made on whether the drug requires a full pharmacoeconomic assessment or not. If the drug is selected for full PE assessment, the company are notified. A dossier is then prepared by the company according to the NCPE guidelines for submission of economic evaluations and the 90 day process begins. [12] Once all information and clarifications are received by the review group, a critical assessment is completed and recommendations are made to the Corporate Pharmaceutical Unit (CPU) in the HSE. A decision is then made on reimbursement by the HSE.

1.1.1.3 Pharmaceutical pricing in Ireland

In Ireland, expenditure on medicines in the community has increased from €300 million in 1998 to €2.24 billion in 2009. The HSE has examined all aspects of the drugs supply chain in an attempt to obtain value for money. The 2006 agreement between the HSE and IPHA resulted in a 35% reduction in the price of patent-expired medicines with estimated savings of €248 million over a 4 year period. The agreement was extended to 2012 and a further "one-off" 40% price reduction for those off-patent products was implemented in February 2010. [3] Reductions in wholesaler margins and pharmacy reimbursement were introduced to provide savings of approximately €130 million per annum. Patient co-payment under the Drugs Payment Scheme (DPS) increased to €120 per month and a new co-payment for medical card holders has been introduced (50 cent per item up to a maximum of €10 per month). Since September 2009, all new pharmaceutical products are considered for pharmacoeconomic assessment. Generic substitution and reference pricing are to be introduced in 2012.

Patient Access Schemes

Patient access schemes (PAS) have been adopted in a small number of reimbursement cases. These are schemes, often proposed by pharmaceutical companies, and agreed between the payer and the company. The schemes are usually put in place to improve the cost-effectiveness of a drug.

PASs can be broadly divided into two main categories; financial schemes where the price remains high but discounts and rebates are offered or outcomes-based schemes where a price is agreed based on the achievement of outcomes and can be reviewed in light of the outcomes. There have been very few cases in Ireland and these cases have not been made publically available. In the UK, the Department of Health approved a PAS scheme for certolizumab pegol (Cimzia®) in RA, whereby the first 12 weeks of the drug is made available free of charge. [13] It is unclear whether further PAS will be introduced in Ireland.

1.1.1.4 Pharmaceutical Reimbursement in Ireland

There are a number of drug reimbursement schemes available in Ireland. Changes were made to the drug reimbursement schemes in March 1999 to replace existing schemes such as the Drugs Cost Subsidization and Drug Refund schemes. The main schemes now are the GMS, DPS, LTI, and the HTDS.

General Medical Service Scheme

This is the largest reimbursement scheme. The medical card scheme is available for all patients eligible for a full medical card. It has been in place since 1970 primarily for those with low income. In 2001 the scheme was extended to provide medical cards to all persons 70 years and over without a means test; however this was reversed in 2009 and persons 70 years and over are now subjected to a means test. The number of eligible patients for the medical card scheme has increased steadily from 2005 to 2010 (Figure 2). Twenty three per cent of all medical cards issued are in respect of people over 70 years of age.

Drugs Payment Scheme

The DP scheme is a co-payment scheme aimed at those who don't have a medical card and normally have to pay the full cost of their medication or for those who have a GP Visit Card. A maximum fee of €120 per month is paid by a family. The HSE reimburses the pharmacy the cost of the drug and a 20% mark-up and dispensing fee. There is a list of reimbursable items available on the DP scheme available from the HSE. The number of eligible patients for the scheme increased from 2005 to 2009, from which point the numbers have decreased (Figure 2). The decline may be due to fiscal constraints of the population due to the recession.

Long Term Illness Scheme

This scheme was established to enable patients suffering from certain chronic conditions (15 specified) to be eligible for free medicines for the management of these conditions. The scheme is not means tested. In 2008, 120,407 people were eligible for

the LTI, representing 2.84% of the population (Figure 2). Pharmacies are reimbursed the price of the drug + 20% mark-up + a dispensing fee.

High Technology Drugs Scheme

The HTDS was introduced in 1996 in order to facilitate the supply of high cost medicines through community pharmacies. Previous to this the cost of the drugs was reimbursed to pharmacies plus a further 50% which incurred a large cost to the payer. The cost of the medicines is now paid directly to the wholesaler and a patient care fee of €62.03 per month is paid to the dispensing pharmacy. [14]

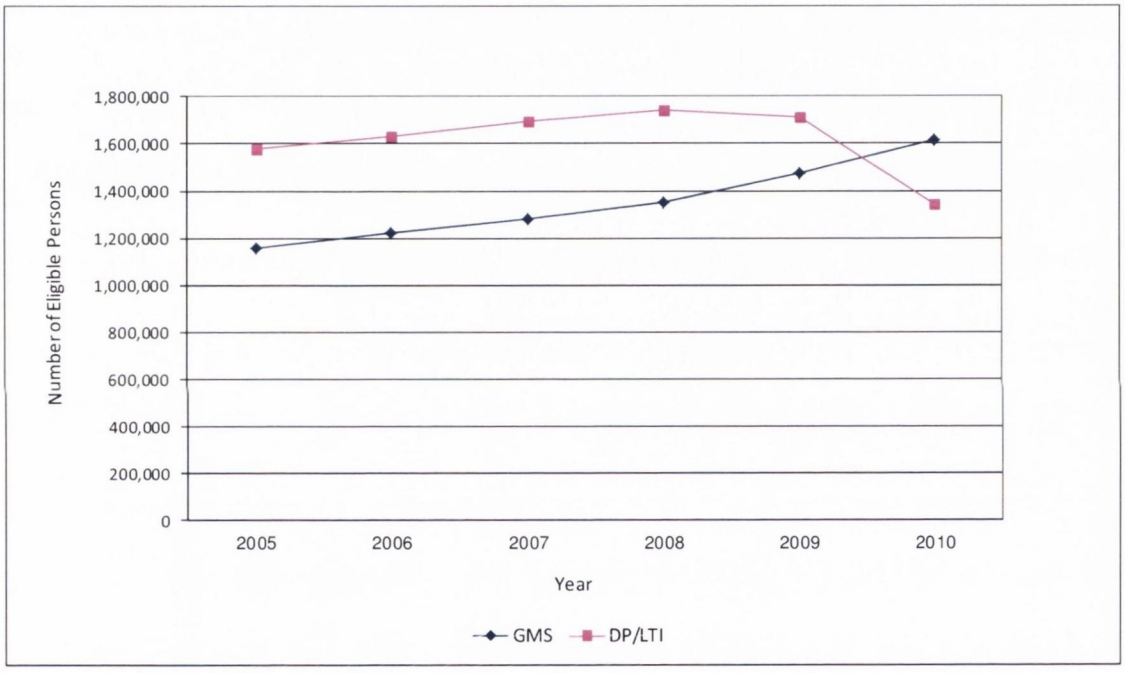


Figure 2: Numbers of eligible patients on the community drugs scheme 2005 - 2010 (GMS, DPS and LTI)

In 2009 the total ingredient cost of the scheme was €279,613,031, which represented 12.5% of the overall expenditure (Figure 3). The expenditure under this scheme has shown an exponential growth over the last ten years and has the greatest growth in rate of expenditure of all the community drugs schemes. The number of eligible patients for the HTDS is much lower than the other schemes at approximately 55,000 patients. (Figure 4) Therefore the average current expenditure (based on 2009 data) for a patient on the HTDS is €6,010 per annum compared with €941 on the GMS + LTI schemes.

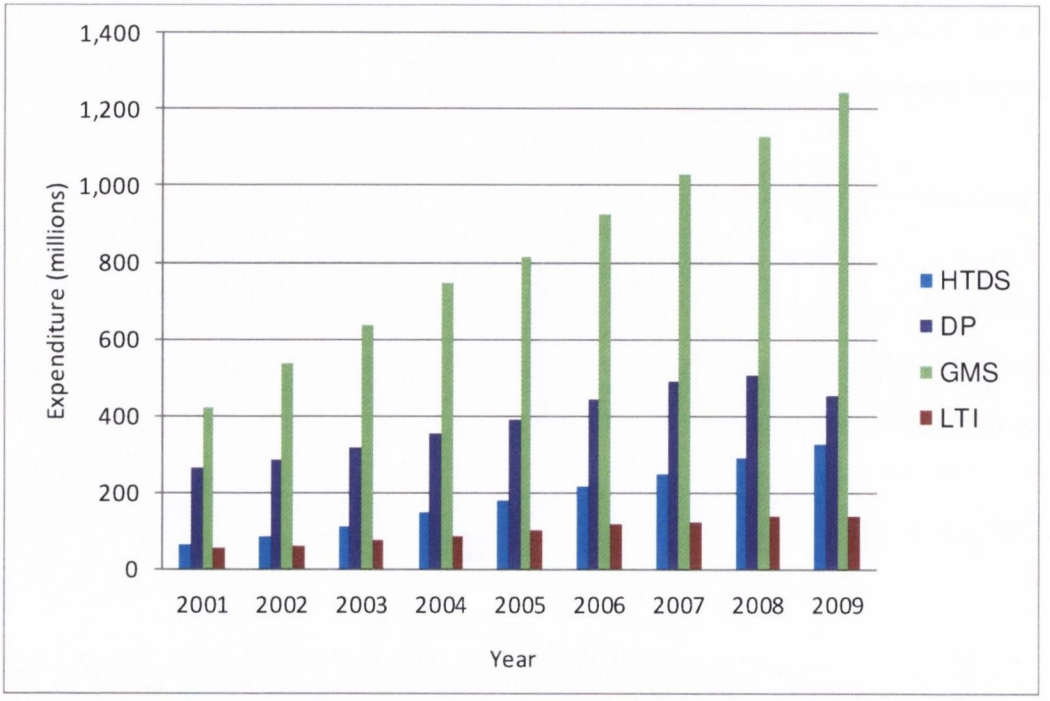


Figure 3: Expenditure (euros) on Community Drugs Scheme (HTDS, DP, GMS, LTI) 2001-2009.

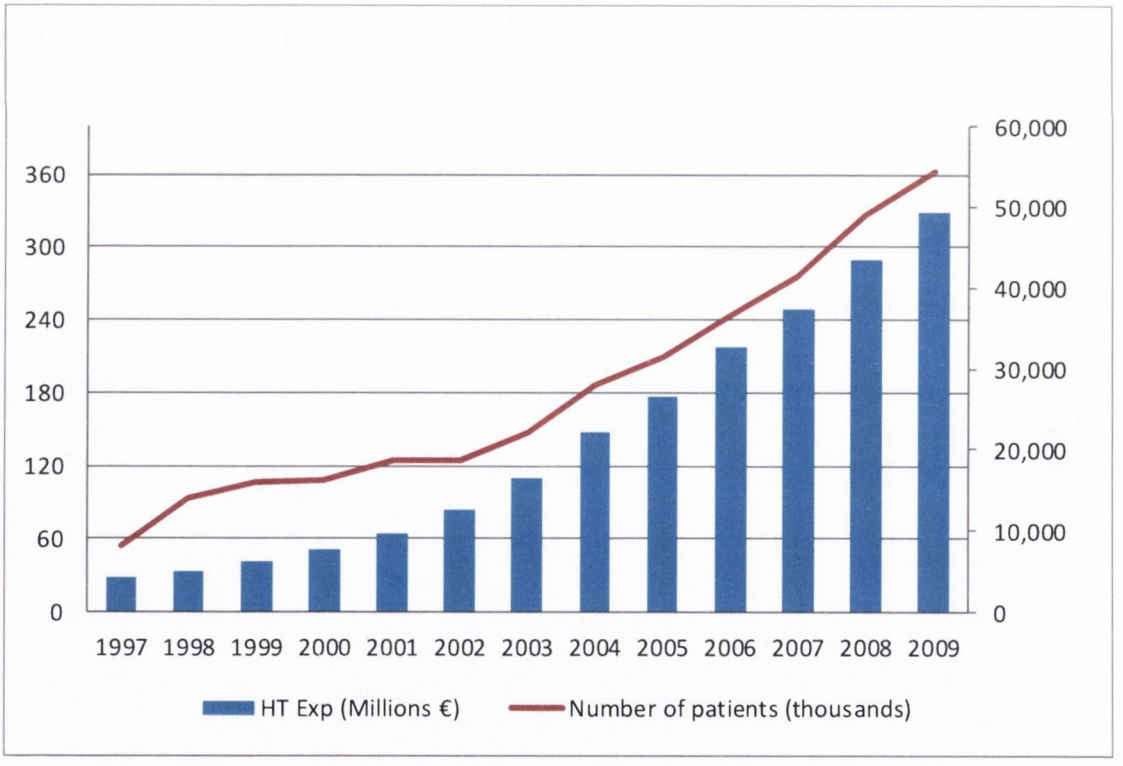


Figure 4: Medicine expenditure trend over a ten year period on the high technology drugs scheme.

In 2009 the top two drugs on the scheme represented 30% (€84,315,316) of the overall expenditure on the scheme; these drugs were adalimumab (Humira®) and

etanercept (Enbrel®). Both of these drugs are used primarily for inflammatory arthritis but also have indications in gastroenterology (Crohn's disease and ulcerative colitis) and dermatology (psoriasis). Because of the high cost of these drugs, their use in RA in Ireland was identified as an area to examine from a pharmacoeconomic viewpoint. This was the basis of the work carried out in this thesis.

1.2 Economic Evaluation

Economic evaluation is now an integral component of Irish healthcare and one of the guiding tools for deciding whether a technology should be reimbursed. The systematic process involves examining all consequences and costs of an intervention and applying a formula to arrive at an estimate of cost-effectiveness.

Economic evaluation is an umbrella term which incorporates a range of techniques. The decision on which technique or method to use, will depend on the question required to be answered. There are four main methods employed when assessing an intervention. These are cost-effective analysis, cost utility analysis, cost minimisation analysis and cost benefit analysis.

1.2.1 Types of economic analysis

Cost Utility Analysis

Cost Utility Analysis (CUA) incorporates an extra dimension into the cost and consequences analysis; quality of life (QOL). QOL is combined with mortality to produce a quality adjusted life year gained (QALY). Quality of life can be assessed using different QOL questionnaires from which we can derive a weighting or a utility. The results derived from CUA are typically expressed as cost per QALY. CUA is the most common form of economic evaluation carried out in Ireland. This is the preferred approach and recommended in the national HTA guidelines. [15]

Cost-effectiveness Analysis

Cost-effectiveness Analysis (CEA) measures costs and consequences or outcomes in natural units, e.g. life years gained (LYG) or cases of DVT detected. The outcomes

chosen are those which are relevant to the decision maker and are common to the interventions compared. A CEA usually involves populating an economic model with the associated costs and consequences (or outcomes) for each intervention. CEA is one of the most common assessments performed after CUA; however, in Ireland, it is only acceptable if a clear justification is provided for using this method over CUA. [15]

Cost Minimisation Analysis

Cost Minimisation Analysis (CMA) can be done where there is little or no difference in the consequences of the two interventions, and the only difference lies with costs. CMA is rarely appropriate for new technologies due to the uncertainty that generally surround both costs and consequences.[16] The rare circumstances it may be appropriate might include almost identical technologies from the same pharmacological class (e.g. in the case of biosimilars). CADTH recently assessed both golimumab and certolizumab (under a common drug review) for RA as CMA submissions, on the grounds that anti-TNF agents display similar efficacy. [17, 18] However in the review of efficacy for certolizumab, the review group commented that there were considerable limitations to the clinical trial data. [17] CMA has been suggested as an option for economic evaluation in RA due to some evidence from the literature proposing that anti-TNF agents are broadly similar; however in assessing the evidence there is considerable uncertainty around estimates of efficacy.

Cost Benefit Analysis

Cost benefit Analysis (CBA) is perhaps the least used form of assessment within healthcare. In this type of assessment, the value of the intervention is defined in monetary terms. For healthcare, this can often be a difficult analysis to undertake as it requires all outcomes to be expressed in monetary terms. Examples of CBA may be cost of inpatient vs. outpatient rheumatology services where the benefits may be less time off work for patients or return to employment.

The choice of analysis will depend on the disease area to be studied, the intervention to be examined and the question the decision maker would like answered. In the case of RA, the most common type of analysis is CUA and the outcome is therefore the QALY.[19]

1.2.2 Decision Analytical Models

The use of pharmacoeconomic decision analytical models has increased greatly in recent years.[7, 9, 20, 21] They are a key tool for those conducting or appraising technologies by providing an explicit framework to address the decision.[22] They can combine evidence from a range of sources which allows extrapolation of costs and outcomes over time.[23] Decision analytical models are recommended for submissions of economic evaluations in many countries national guidelines, including Ireland. [15, 24-26] Often a copy of a decision model is also requested to be included for the submissions.

A Consensus Statement on Decision Analytic Modelling in the Economic Evaluation of Health Technologies outlined the properties of good decision analytical models [23]. The model should be:

- Customised for the purposes for which it is to be used,
- Useful in answering the questions which are asked,
- Easily communicated.

A framework for a pharmacoeconomic evaluation is essentially composed of two main components; the decision analytical model and the evidence pertaining to the intervention(s). The evidence may be sourced from a number of areas including natural history, epidemiology, efficacy and effectiveness, costs and other outcomes such as QOL. This evidence may also be subjected to analytical modelling, as is the case in evidence synthesis for efficacy and effectiveness, in order to combine different sources of data for use in the model.

1.2.3 Decision Rules, ICERs and Thresholds

When decision analytical models are used to assess interventions, an estimate of cost-effectiveness is derived, which is most commonly the Incremental Cost-Effectiveness Ratio (ICER).

The ICER is represented by the following equation:

$$\text{ICER} = \frac{\text{Cost A} - \text{Cost B}}{\text{Effect A} - \text{Effect B}}$$

Where A and B could be a drug or alternative technology.

Within a constrained budget, the healthcare payer must choose the intervention that will provide most benefit for the least acceptable cost. In deciding which healthcare intervention to choose it can be useful to compare ICERs across these interventions. This has been done in the form of league tables, where each intervention is ranked according to their ICER. The decision maker may choose to use their budget to implement all interventions with the lowest ICERs. However league tables do present problems with regard to equity and an alternative strategy such as opportunity cost may be more reasonable by freeing up healthcare resources to pay for the more cost-effective option. [27]

A further approach to decide between ICERs may be to define a maximum cut-off value for the ICER. In this case a line is drawn through the CE plane above or below which the intervention may or may not be cost-effective. This is represented by the orange and black lines in Figure 6.

In an environment of scarcity, the payer is faced with a decision on whether the technology is cost-effective. In order to make such a decision it follows that there must be some form of threshold, beyond which the technology would not be classed as good value for money. The concept of the threshold was first introduced in 1973.[28] It was recognised as a level which costs and effects of an intervention must achieve to be acceptable within a healthcare system. A threshold may be adopted by a healthcare organisation and be made available in advance of any assessments and is therefore explicit. Alternatively, an implicit approach may be adopted. In this case the thresholds may not be officially stated but may be inferred by previous decisions made. The latter practice is perhaps more common and is the practice in Ireland. [15]

There are a number of challenges associated with using set thresholds for decision making; health care budgets are rarely fixed and therefore what may be an affordable threshold now may not be the case in the future. This has been demonstrated recently in the Irish situation, where an implicit threshold of €45,000 is now rarely deemed as cost-effective and a threshold of €20,000 is employed more often. [29] There is no formal threshold in Ireland under which a technology will definitely be reimbursed. In the past, most technologies with an ICER under €45,000 were reimbursed. This was in line with that used in the UK .[30] Recent evaluations (since 2009) have included cost–effectiveness at a €20,000/QALY threshold level, reflecting the decision makers' interest in how threshold level influences the cost–effectiveness of new technologies. [31] The probability of cost-effectiveness is currently presented to the decision maker in Ireland under a willingness to pay of both €20,000 and €45,000 per QALY; however it is drugs under the lower threshold of €20,000 which are most likely to be considered cost-effective.

A technology may be classed as cost-effective but the budget impact of introducing an intervention could be considerable. In 2008, the cervical cancer vaccine was assessed in the Irish healthcare setting and was deemed to be very cost-effective with an ICER of €17,385/LYG. [7] However the budget impact of vaccinating 12 year old girls was estimated at €9.7 million and therefore deemed unaffordable by the payer at that time.[10]

1.2.4 Uncertainty in economic evaluations

The provision of timely, accurate HTAs facilitate the decision making process, by providing all relevant information that may impact on the decision made. With any assessment, however, there is residual uncertainty as to what the optimal decision is. Uncertainty arises for two distinct reasons. The first of these is uncertainty about model parameters such as efficacy; this is due to the fact that we only have limited data with which to estimate these. This uncertainty can be explored using methods such as probabilistic analysis or probabilistic sensitivity analysis (PSA), one way sensitivity analysis (OWSA), cost-effectiveness acceptability curves (CEACs) and value of information calculations (such as Expected Value of Perfect Information (EVPI)).

One Way Sensitivity Analysis

One way sensitivity analysis involves varying the parameters in the model across a range (often +/- 20%) and examining how this changes the results of the model. OWSA is useful in determining which parameters have most impact on the results of the model e.g. drug costs, efficacy, survival rates. The results of a OWSA are often presented on a tornado diagram (figure 5). As can be seen in figure 5, variation in the discount rate has most impact on the cost-effectiveness estimates. The basecase ICER is usually represented as a line in the middle of the plot and extensions to the left or right indicate an increased or decreased cost-effectiveness estimate respectively.

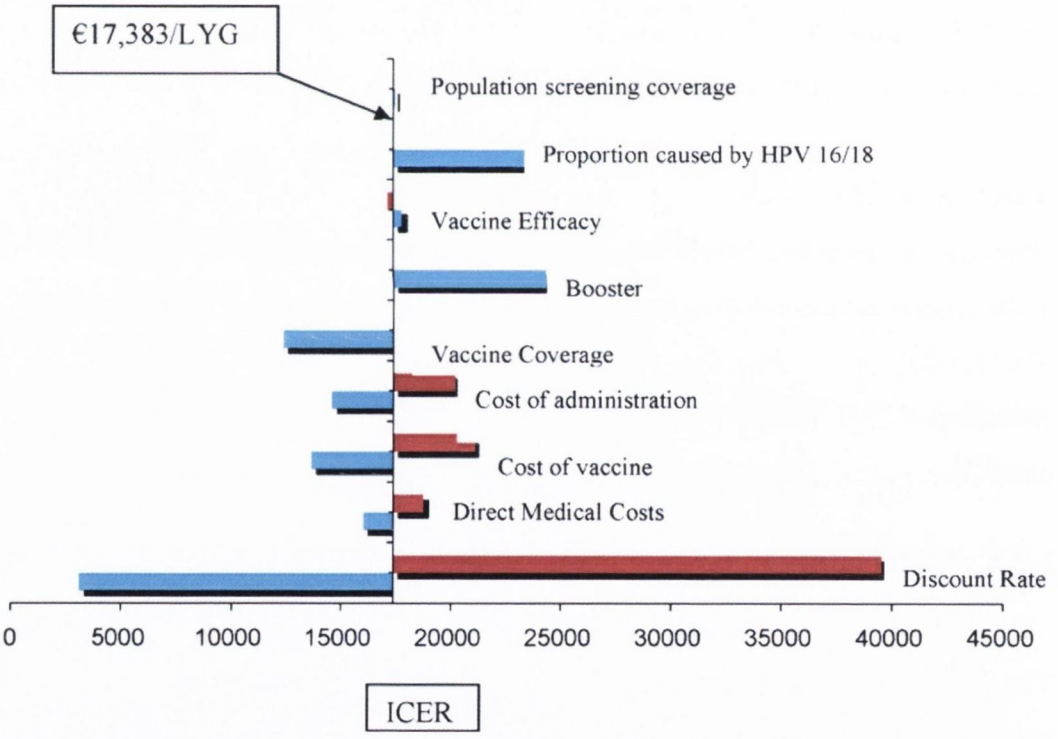


Figure 5: Tornado plot demonstrating the parameters which increase or decrease the basecase ICER.
Reproduced with permission from Usher et al.[7]

Probabilistic Analysis

A probabilistic analysis allows exploration of all parameter uncertainty simultaneously. This is done by assigning distributions to each parameter input. These distributions are then sampled at random (usually using Monte Carlo

simulation) and the result of the model using each particular simulation is recorded. The process is repeated many thousands of times (often 10,000 but may be more or less). [32] From this the proportion of times (probability) that each treatment alternative is cost-effective can be estimated. The results of the PSA are usually presented on a scatter plot (with each point on the scatter representing a simulation) and via a cost-effectiveness acceptability curve. These proportions are usually presented in pharmacoeconomic evaluations as percentages. [33]

The other sources of uncertainty arise because of decisions made during the formulation and implementation of the assessment. These can include for example the choice of comparators, the economic model (such as Markov model or discrete event simulation) or as explored in this work the different choices of statistical model for the evidence synthesis. These types of uncertainties can be referred to collectively as structural uncertainties. [32, 34]

In order for a decision maker to hold all of the 'tools' necessary for making a decision, it is vital that they are aware of the uncertainty associated with the estimates presented. The presentation of this uncertainty in a clear, transparent, understandable manner is a fundamental part of any economic evaluation. The Irish HTA guidelines recommend graphical presentation of uncertainty for more complex models, using simulation methods. [15] The choice of graphics may depend on the type of analysis undertaken but may include the following:

- cost-effectiveness plane to present the incremental costs and effects of two (or more) comparator technologies
- tornado diagrams to display the results of subgroup effects and one-way sensitivity analysis (figure 5)
- scatter plots to present ICER results from probabilistic sensitivity analysis of two comparator technologies on the cost-effectiveness plane
- cost-effectiveness acceptability curve to present the probability that a technology is more cost-effective than its comparator.

- in a study comparing more that two technologies, the graphic should present the probability that a technology is the most cost-effective as a function of the threshold willingness to pay for one additional unit of benefit.

1.2.5 Presentation of Results of Economic Evaluations

Communicating the message of pharmacoeconomic evaluations could be identified as the most important part of the process. Studies have shown that decision makers knowledge about formal methodology is limited and find the concepts behind QALYs difficult to understand.[35, 36]

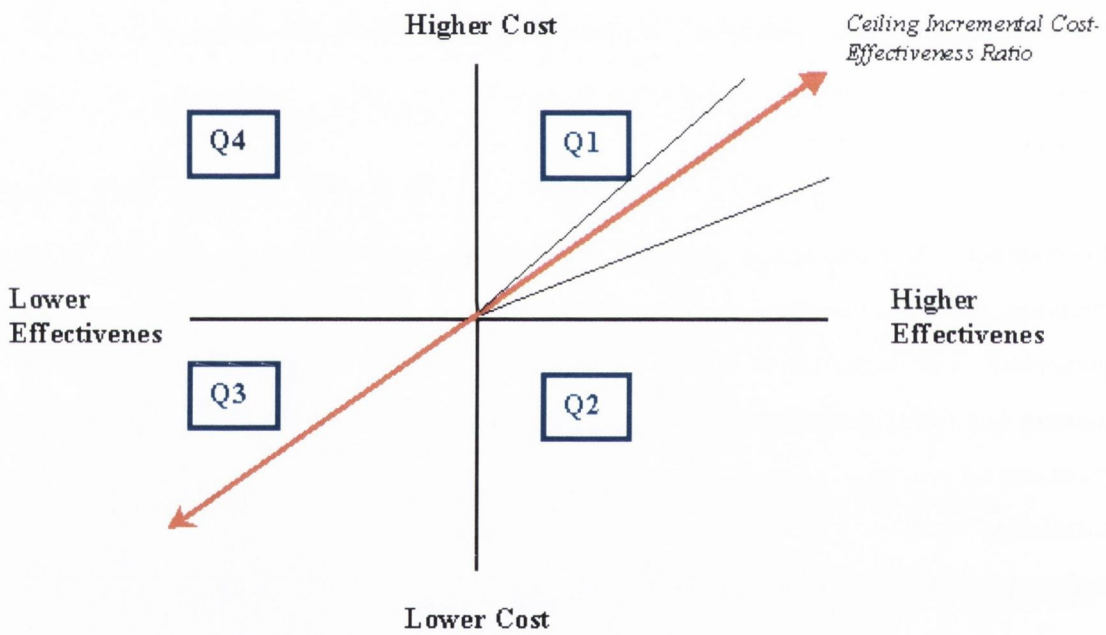


Figure 6: The Cost-effectiveness Plane.
The orange line through the origin represents the threshold for decision makers – this may increase or decrease (black lines)
Reproduced with permission from Drummond et al. Methods for the Economic Evaluation of Health Care Programmes 3rd Edition.)

1.2.5.1 Cost-effectiveness Plane

The cost-effectiveness (CE) plane is a graphical presentation of cost-effectiveness (figure 6). It was first introduced in 1990 and is used as a key method of illustrating

the results of an economic analysis. [37] It represents the incremental costs and consequences of using an alternative intervention. The comparator or the standard to which the alternative treatment is compared is represented by the origin. For example in RA, the origin may be represented by methotrexate (standard therapy) and the ICERs of anti-TNF therapy tends to fall within quadrant 1 (Q1). [38-40] The x and y axis represent the incremental cost and consequences, respectively. Quadrant 2 (Q2) is an ideal situation where an intervention is more effective but less costly. Quadrant 3 (Q3) presents an interesting choice; because the threshold passes through the origin, a drug may appear to be cost-effective and a decision then needs to be made whether a payer would be willing to accept less benefit for a lower cost. This situation arose recently in an Irish pharmacoeconomic evaluation for golimumab in RA. [41]
The majority of new medicines tend to fall within Q1 - more effective and more costly. Medicines falling within Q4 are rarely reimbursed.
In the case of multiple treatments it is possible to calculate the mean costs and effects plot them on the plane and draw a line through the non-dominated options (i.e. the cost-effective options). This line then forms the efficiency frontier. [42]

The ICER and the CE plane do not in themselves represent the uncertainty associated with the estimate and therefore using these values alone may lead to the wrong decision being made. [43] For this reason the cost-effectiveness acceptability curve (CEAC) and the cost-effectiveness acceptability frontier (CEAF) were developed.

1.2.5.2 Cost-Effectiveness Acceptability Curve

The CEAC provides a graphical representation of the probability of cost-effectiveness at different thresholds. In order to produce a CEAC, a full PSA must be carried out. The probability of CE is calculated by determining the expected net monetary benefit (NMB), for each PSA iteration.

$$NMB = \lambda \times E - C$$

λ = threshold (WTP), E = Effect (QALY), C = Cost

The option with the highest NMB is then chosen. The probability is calculated from the proportion of iterations where this is the case. An example of a CEAC is given in Figure 7 below which is reproduced from the health technology assessment of a population-based colorectal screening programme in Ireland.[44] The probability of cost-effectiveness of each of these options, flexible sigmoidoscopy (FSIG) and faecal immunochemical test (FIT) at different ages and no screening is shown. At a WTP of €1,000/QALY FIT at 55-64 years (orange curve) shows the highest probability of being cost-effective. If the decision maker is willing to pay more (€4,000/QALY), then the FIT in the 55-74 age groups would be most cost-effective (red curve).

Figure 7: Cost-effectiveness Acceptability Curves for Colorectal Cancer Screening.
FSIG at 60 years, FIT for ages 55-64 and FIT and ages 55-74. (Taken from Health Technology Assessment of a population-based colorectal cancer screening programme, HIQA, March 2009)

While the CEAC gives an indication of uncertainty, it does not show all uncertainty and for this reason it has been criticised. [45]

Limitations of the CEAC

The CEAC is generated on the assumption that the WTP equals the willingness to accept (WTA). In other words cost-effectiveness outcomes in the southwest (Q3) and northeast (Q1) of the CE plane are treated equally. However there has been some evidence that this is not always the case. [46] Severens *et al.* argue that when the cost-effectiveness pairs lie in Q3, the difference between the WTP and WTA, is likely to influence the CEAC. [47]

Groot Koerkamp *et al.* described further limitations of the CEAC. [45] The authors of this paper state that CEACs cannot distinguish different joint distributions, which restricts the ability to synthesise evidence from other sources. They also argue that the CEAC does not allow for integration of risk attitude, may mislead policymakers, and is unhelpful as regards value of future research. A counterargument was given by Fenwick and Briggs. [48] They accept that the CEAC is not sensitive to changes in the joint distributions; however this is the result of the cost-effectiveness analysis itself rather than the CEAC which assumes all points in Q1 and Q3 to be equivalent in cost-effectiveness terms. While not displaying all uncertainty, the CEAC draws attention to the uncertainty associated with the decision and moves away from classical thinking on confidence intervals and statistical significance.

The CEAC presents the probability of cost-effectiveness of a treatment over a range of values of maximum acceptable ICERs (WTP/ thresholds). However due to some limitations with the presentation of results, as highlighted, it may be appropriate to present further analysis to compliment it; the cost-effectiveness acceptability frontier (CEAF) provides additional information to support that of the CEAC. [48]

1.2.5.3 Cost-effectiveness Acceptability Frontier

While the CEAC may demonstrate the option with the highest probability of cost-effectiveness at any threshold, this may not be the optimal option to choose. [49] The CEAF, in contrast, plots the probability that the optimal option is cost-effective at different WTP thresholds. In order to calculate this, it is necessary to establish which

option is most cost-effective at each threshold. The mean cost and effect is calculated for each iteration from the PSA (as above) but in this case the optimal option is calculated at different threshold values. The probability of the optimal option being cost-effective is then plotted (y-axis) against the threshold values (x-axis). In essence the CEAF provides a graphical presentation of the probability of making an error. [33] An example of this is given in Figure 8. The CEAF is shown as the light blue curve. The CEAF demonstrates here that the optional choice (FIT in 55-74 yrs) is also the most cost-effective option at a WTP of €10,000/QALY (red curve). The CEAF, in identifying the probability of error is intrinsically linked to the expected value of information.

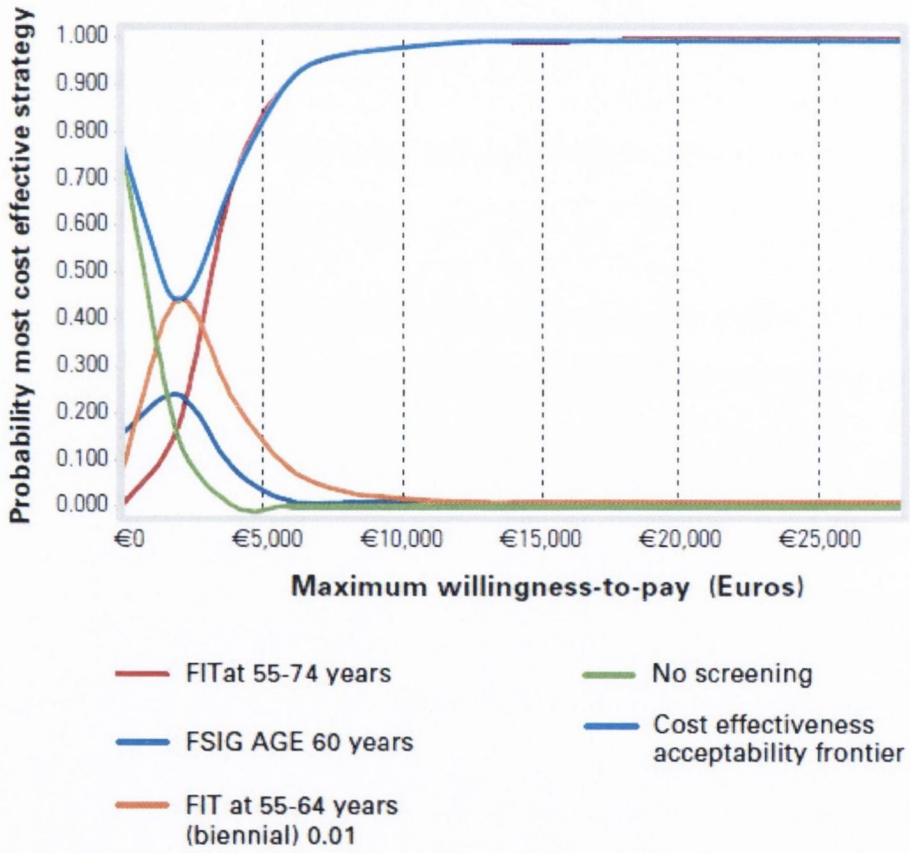


Figure 8: Cost-Effectiveness Acceptability Curve and Cost-Effectiveness Acceptability Frontier

FSIG at 60 years, FIT for ages 55-64 and FIT and ages 55-74. (Taken from Health Technology Assessment of a population-based colorectal cancer screening programme, HIQA, March 2009)

1.2.5.4 Expected Value of Perfect Information

The EVPI or Value of Information (VOI) is a method used to calculate a cost associated with making the wrong decision. The EVPI gives the upper bound estimate of the value of further research in order to eliminate uncertainty around a decision. [42] In order to calculate the EVPI, one must first calculate the probability of the wrong decision being made (from the CEAF) and also the consequences of this wrong decision.[42] These consequences are associated with foregoing benefit, due to a wrong decision. [22] If further research costs more than the calculated EVPI, then the research is unlikely to be efficient.

The EVPI, while providing additional insight into uncertainty surrounding the decision made, does not specifically indicate where the research should be done; on costs or effects or both. In order to do this further analysis (and more computationally complex) would need to be done such as expected value of perfect parameter information.[32] The value of EVPPI is pinpointing exactly where the research needs to be done, for example in relation to cost or utility.

1.3 Conclusion

HTA is recognised as a vital tool to decision making around health expenditure. In Ireland HTA results of projects such as HPV vaccination to prevent cervical cancer and population based screening for colorectal cancer have been instrumental in guiding decision makers toward the most effective and cost-effective choice. Single technology appraisals completed by the NCPE have provided the decision maker with the information to allow decisions to be made around pricing as well as advising against technologies thought to be ineffective. While the majority of assessments in Ireland are around new drugs seeking reimbursement, the issue of disinvestment must also be considered by the payer. In order to continue to allow for investment in cost-effective strategies, obsolete and inefficient strategies must be examined.

1.4 Overall Aim and Objectives of Thesis

This thesis examines the cost-effectiveness of anti-TNF agents for the treatment of rheumatoid arthritis in the Irish healthcare setting. These agents, while highly effective, incur a large cost to the healthcare payer. All of the anti-TNF agents are used and reimbursed either under the community drugs schemes or by hospitals. Only two of these agents have undergone pharmacoeconomic assessment and were initially refused reimbursement due to uncertainty associated with their cost-effectiveness. The thesis focuses on two main parts; the estimation of data inputs for the economic model and the pharmacoeconomic modelling of these agents in RA. The overall aim and objectives of the thesis are outlined below.

1.4.1 Primary Aim:

To examine the cost-effectiveness of anti-TNF agents for RA in Ireland.

1.4.2 Objectives:

- To provide an overview of health technology assessment in Ireland.
- To review health technology assessments on anti-TNF drugs in other jurisdictions.
- To provide an overview of outcomes assessment in rheumatoid arthritis.
- To measure the quality of life preferences for an Irish rheumatoid arthritis cohort pre and post initiation with anti-TNF therapy using different quality of life measures.
- To examine the differences in utility (quality of life) of these patients using different measures.
- To explore the background methodology to utility measurement using the data collected.
- To examine the efficacy of the anti-TNF agents for the treatment of rheumatoid arthritis and estimate the relative efficacy of these agents.
- To examine the rates of discontinuations for these agents in both the short term and the long term.
- To estimate the costs of treatment associated with anti-TNF agents and the comparator.

- To calculate the incremental cost effectiveness ratio for these individual agents in the Irish healthcare setting, using a pharmacoeconomic model, with the data inputs calculated (utility, relative efficacy and costs).
- To provide recommendations for the decision maker based on the results of the economic evaluation in Ireland.

The chapters of this thesis can be categorised under three key sections; background, inputs or data analysis and pharmacoeconomic modelling. Chapter one, two and three are background and review chapters. Chapter four and five contain analyses on utilities and evidence synthesis. Chapter six is a final analysis chapter which uses results from chapter four and five and background information from chapters two and three. Chapter seven discusses the implications of the results and provides recommendations for future research.

Chapter 2

Chapter 2.0 Rheumatoid Arthritis

2.1 Rheumatoid Arthritis

RA is the most common form of inflammatory arthritis. It is a chronic condition that places a substantial burden on patients and their carers, imposing a negative effect on QOL, including physical, psychological and social functioning and is associated with premature mortality.[50]

The incidence of RA is higher in women than in men, at a ratio of 2:1. While this suggests that reproductive and hormonal factors play a role in the occurrence of the disease, this relationship has not been confirmed. There appears to have been a fall in the incidence of RA over previous decades. Two theories have been postulated to explain this; the protective effect of the oral contraceptive pill and a birth cohort effect where women from a certain time period carry a higher risk of RA than women from later generations. [51]

The age of onset of RA peaks in the fifth decade of life, but some studies do suggest a later onset of disease. [52-54]

2.1.1 Classification of RA

The American College of Rheumatology originally classified RA in the 1980s. [55] However these criteria are no longer as relevant mainly due to the success of current treatments (two of the seven criteria, nodules and erosions are generally not present at early diagnosis). The EUropean League Against aRthritis (EULAR) developed classification criteria which devised a score based on 28 joints, an inflammatory marker and a global health assessment score. [56] This classification system has been used both to select patients for clinical trials and to monitor a patient's response to treatment. A joint initiative between ACR and EULAR re-examined the classification

criteria for RA with a focus on early disease and these new classification criteria have been published. [57] Improved classification systems for RA have strengthened the studies on epidemiological evidence which were difficult to compare prior to this robust classification criteria.

2.1.2 Epidemiology of RA

The prevalence data for RA in Ireland is poor. One older study suggests it to be 0.5% but the study was carried out in a specific geographical area and then extrapolated to represent the whole country. [58] Arthritis Ireland estimate that 40,000 people in Ireland have RA[59] (approximately 1% prevalence). However this figure is based on UK and US data and not on one particular study. On discussion with consultant rheumatologists in Ireland, it is estimated that the current prevalence is closer to 1% and this apparent increase in prevalence has been attributed to increased early diagnosis and treatment.

Pathogenesis of RA

RA is a disease leading to inflammation; primarily joint inflammation. There are a series of inflammatory cascades which may be triggered by adaptive immunity and eventually lead to the joint destructive behaviour that is seen in many patients.[60] Figure 9 presents the current understanding of inflammation in the joints in RA. [61]

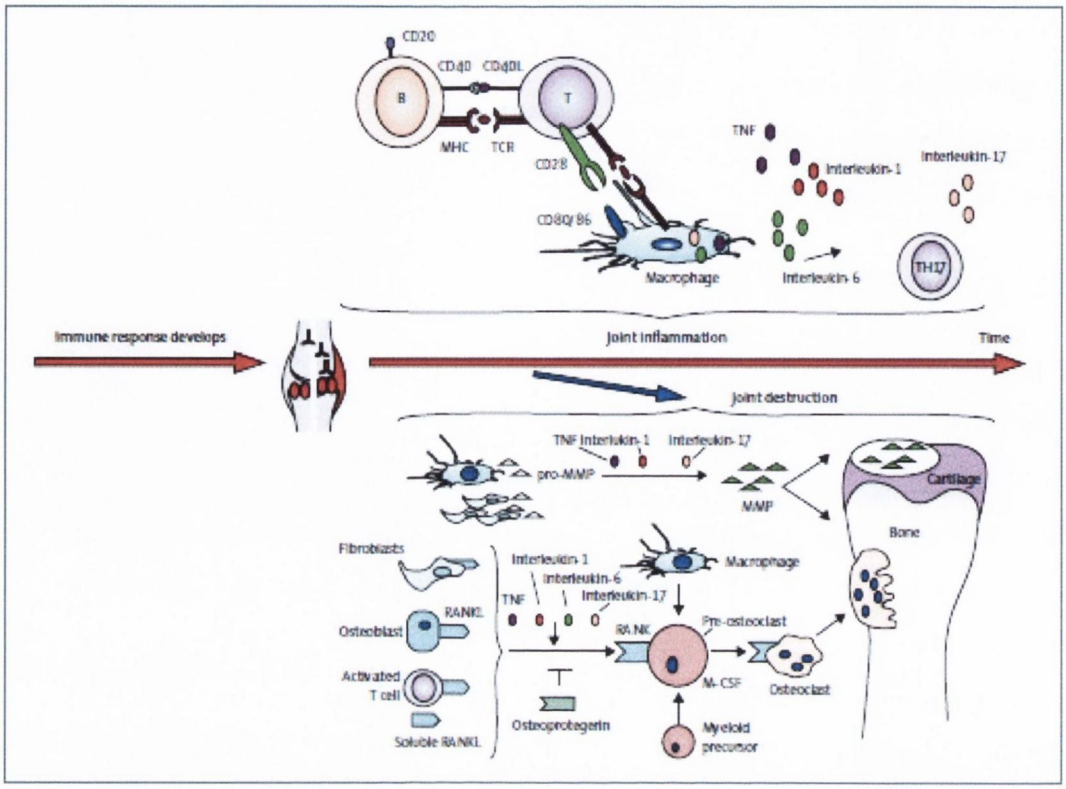


Figure 9: **Immunological pathways in the arthritic joint**
The diagram shows the involvement of different cytokines in inflammation and destruction. The upper part shows inflammation and lower part joint destruction which is the result of inflammation[61]

Synovial inflammation is characterised by many different interacting immune cells. Macrophages activated by signals produce proinflammatory cytokines such as tumour necrosis factor (TNF), interleukin 1 (IL-1) and interleukin 6 (IL-6). The identification of these cytokines has led to the development of specifically targeted treatments. Anti-TNF α (e.g. infliximab, adalimumab, etanercept, golimumab, certolizumab) and IL-1(e.g. anakinra) inhibitors were the first targeted agents to be developed. B-cell (e.g. rituximab), T-cell (e.g. abatacept) and IL-6 (e.g.tocilizumab) therapies have since been licensed for use in RA.

The discovery of the cyclic citrullinated peptide (CCP) has provided a further link to the question to why some patients have more aggressive disease than others. The detection of anti-CCP antibodies may allow the clinical rheumatologist to better predict the diagnosis and prognosis of individual patients with RA. Several observations have indicated that anti-CCP positive early RA patients may develop a

more erosive disease than those without anti-CCP. [62] Whether this or other serologic tests will allow more rational therapeutic decision-making and hence influence the long-term outcome of the disease has yet to be determined.

Risk Factors

The influence of genetic susceptibility factors has a major effect on the pathway and outcome of RA. [61] Twin studies have estimated the relative contribution of genetic factors to RA to be approximately 50% with the remaining 50% to be in part due to environment and in part to chance. [63] There is strong evidence to show that cigarette smoking is a risk factor for patients. [64, 65] Other environmental exposures are silica dust, mineral oils and other airway exposures. [66, 67] There is also evidence to suggest that moderate alcohol consumption can actually reduce risk for RA. [68] It is clear from the design of these studies, where both genes and environment and immunity were accounted for, that the genetic environmental interaction is an important one (Figure 10). Further research into this area will perhaps in time provide understanding into which specific immune reaction contributes to the CCP positive form (most aggressive) of RA.

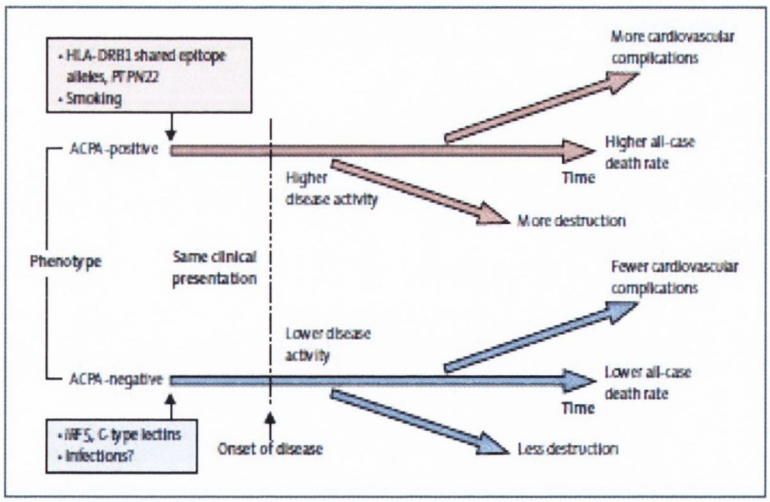


Figure 10: Differences in risk factors, immune events, and disease course between two major subsets of rheumatoid arthritis.

Poor socioeconomic status has been presented as a risk factor for RA and may indicate a worse prognosis. [69] Several infectious agents have been suggested as being risk factors for RA including Epstein-Barr virus, rubella virus, parvovirus and

others. However the role of infectious agents in the occurrence of RA remains unclear.

The manifestations of RA are many and have a significant impact on a patient's quality of life and morbidity.

2.1.3 Mortality and RA

Despite some studies indicating that RA increases mortality,[70] other studies have not supported this theory. [71] [72] A review article gathered evidence from a wide range of studies and compared mortality up to 1990 and beyond 1990 (Table 1). [70] The Standardised Mortality Ratio (SMR) for studies before 1990 and after 1990 showed increased mortality for patients with RA. The article also considered the reasons why some studies indicate no increased mortality and others did not. The authors highlighted that the studies showing no increased mortality were mainly conducted in newly diagnosed patients who were treated aggressively (inception cohort). Possible explanations for this included a change in the natural history of the disease or the disease was becoming less severe or that the early management was changing the course of an otherwise aggressive disease. Ward examined these hypotheses by performing a meta-analysis of 18 studies.[73] The SMR in inception cohorts (<2 years disease) was 1.2 and in established disease (non-inception cohorts) was 1.9 In this case the authors concluded that the differences were due to study design. Further studies have shown an increased mortality in early arthritis. [74, 75] In choosing which estimate to use, it is reasonable to choose an estimate that includes both those who were treated at early diagnosis and those for established disease since this is likely to be the pattern seen in the overall population.

An American study analysed data on survival over 40 years and found that the SMR was 1.27 (95% CI 1.13–1.41). [76] The SMR was higher for women than men (1.41 and 1.08 respectively). The mean SMR over all 15 studies was 1.7. (Table 1)

In the NICE economic evaluations of anti-TNF agents for RA, a SMR of 1.33 per HAQ was applied. [77, 78] This was calculated based on studies up to 1994. An

increased SMR of 2.73 was included in the range for the sensitivity analysis. The increased SMR was taken from Sokka *et al.* [79]

Table 1. Large mortality studies in RA over the last 15 years (studies of >300 subjects). (Naz *et al.* [70])

First author (country) and ref. no.	Year	Size of series	Mean duration of follow-up (years)	SMR	Category
Jacobsson (Pima Indians, USA)[80]	1993	2979	24	1.28	Community
Wolfe (USA/Canada)[81]	1994	3501	9–35	2.26	Clinic-based
Myllykangas (Finland)[82]	1995	1186	5	1.37	Clinic-based
Wallberg-Jonsson (Sweden)[83]	1997	606	15	1.57	Clinic-based
Symmons (UK)[84]	1998	448	22	2.70	Clinic-based
Gabriel (USA)[85]	1999	450	40	1.38	Clinic-based
Wolfe (USA)[86]	1999	1865	25	1.94	Clinic-based
Kroot (Netherlands)[72]	2000	622	10		Clinic-based
Chehata (UK)[87]	2001	309	14	1.65	Clinic-based
Bjornadal (Sweden)[88]	2002	46,917	31	2.03	Clinic-based
Goodson (UK)[75]	2002	1236	7	1.13M	Primary care
				1.01F	
Gabriel (USA)[76]*	2003	609	39	1.27	Clinic-based
Pincus (USA)[89]	2004	1378	10	1.60	Clinic-based
Goodson (UK)[75]	2005	1010	15	1.45M	Clinic-based
				1.84F	
Young (UK)[90]	2006	1429	18	1.27	Clinic-based

F, females; M, males; SMR, standardised mortality ratios.

* Inflammatory polyarthritis including a subgroup with RA.

2.2 Treatment of Rheumatoid arthritis

One of the major advances in halting disease progression has been the employment of disease modifying anti-rheumatic drug (DMARD) therapy early in the disease. A further key to management of the disease is tight control of inflammation. The development of valid and responsive methods to measure disease activity, functional status, and joint damage has been fundamental in monitoring the control of the disease. A brief explanation is given of the primary outcome measures in this chapter. This is expanded in chapter 3.

The American College of Rheumatology Response Criteria (ACR)

The ACR criteria specifically measure the change or improvement in disease activity of active drug against placebo. The ACR 20 and ACR 50 indicate improvements of at least 20% or 50% respectively, on a number of measurements determined by the American College of Rheumatology.

EULAR response criteria

To be classified as responders, patients should have a significant change in the Disease Activity Score (DAS) (>1.2 for a good response and ≤1.2 and >0.6 for a moderate response) and also low current disease activity. Three categories are defined: good, moderate, and non-responders.

2.2.1 Non-biological DMARDs

The Irish Society of Rheumatology (ISR) produced guidelines for the treatment of RA in Ireland and these treatment pathways closely follow those of the UK where first line treatment for RA is non-biological DMARD therapy. [91] Methotrexate (MTX) is the current first line standard of care in the treatment of RA in Ireland. Table 2 includes the non-biological DMARDs included in this economic evaluation.

Table 2. Non Biological DMARDs included in this economic evaluation

Name of Drug	Mode of action	Licence Indication	Dose	Reimbursement	Mode of Administration
Methotrexate	Anti-metabolite	RA, Psoriasis	7.5-20mg weekly	GMS, DPS	Oral and subcutaneous injection
Leflunomide	Immuno-modulatory	RA, PsA,	10-20mg od	GMS, DPS	Oral
Azathioprine	Anti-metabolite	RA	1-mg/Kg/day	GMS, DPS	Oral
Gold (Sodium Aurothiomalate)	Unknown	RA, JIA	50mg 4-6 weekly	GMS, DPS	Intramuscular injection
Ciclosporin	Immuno-suppressant	RA, Psoriasis	3-5mg/kg daily	GMS, DPS	Oral

JIA - Juvenile Idiopathic Arthritis

2.2.1.1 Review of methotrexate clinical trials in RA

Although MTX is established as one of the most widespread DMARDs used for the treatment of RA, there is considerable variation in the way in which it is used and prescribed by rheumatologists. [92] A number of RCTs have been carried out examining how it is most effectively used and at what dose. The MASCOT (n=687) study compared combination therapy with sulphasalazine (SSZ) and MTX with MTX or SSZ alone using DAS as the primary outcome measure. [93] The study concluded that combination therapy was more effective than monotherapy but there was little significant difference between SSZ and MTX. The COBRA (n=155) study compared a step-down regime of prednisolone, MTX and SSZ to SSZ alone in early arthritis. [94] The study concluded that combination therapy may be more effective than SSZ monotherapy. The TICORA (n=111) single blinded study compared tight control of rheumatic drugs to routine practice. [95] Disease activity, radiographic disease progression, physical function, and quality of life were all more favourable in the tight control group than in the routine care group, at no additional financial costs. The FIN RACo study (n=195) was a randomised open-parallel group trial comparing combination therapy with monotherapy in early RA. [96] The study concluded that combination therapy as a tight control strategy in patients with early RA aiming for

remission seems to be more efficacious than monotherapy. The BeST study (n=508) is a randomised open clinical trial comparing sequential monotherapy, step-up combination therapy, and initial combination therapy with either high-dose prednisone or infliximab in early RA. [97] The study concluded after 2 years that there was no difference between the arms and that highly effective treatment can be achieved by tight control of therapy.

Summary of evidence related to methotrexate.

MTX is the anchor drug used first-line and in 2008 a multinational group reviewed all the recommendations and guidelines and produced 10 points on the use of methotrexate in RA which include screening, dosing, monitoring, toxicity and use around pregnancy. [92]

2.2.1.2 Dose and route of administration of methotrexate

The recommendation is that the oral route should be first choice and escalated from a dose of 10-15mg to a dose of 20-30mg once weekly. In the case of adverse events or lack of response the parenteral route may be tried. [92]

2.2.1.3 Methotrexate monotherapy vs. combination DMARD therapy

In light of the toxicity/efficacy balance, DMARD naïve patients should be initiated on MTX monotherapy. If step-up therapy is to be used, MTX should still be used as the anchor drug. [92] A recent Cochrane review of MTX versus combination non-biological treatment found no advantage in using combination therapy over monotherapy.[98]

2.2.1.4 Toxicity of methotrexate

Patients should be screened for susceptibility to adverse effects of MTX. Patients identified as more susceptible are those with higher than normal alcohol intake, liver

function abnormalities, pulmonary abnormalities, and co-morbidities such as hepatitis or HIV.

Visser *et al.* pooled data from 2009 patients who were on MTX for 3.3 years (mean) and found that the cumulative incidence of abnormal liver function tests (ALT/AST) was 48.9% above the normal range. [92] However it is noted that liver enzymes can be raised transiently in RA and that MTX induced fibrosis is rare.

Gastro-intestinal (GI) upset is increased with combination therapy (and in particular in combination with SSZ and leflunomide). [92] Concomitant prescription of folic acid 5mg reduces both GI and hepatic toxicity without reducing efficacy.[92]

2.2.1.5 Discontinuation rates of methotrexate

Long Term use

Patients on MTX appear to remain on the drug longer than those on alternative DMARDs such as gold, hydroxychloroquine, or SSZ. Maetzel *et al.* examined the termination rates for MTX, hydroxychloroquine, SSZ and gold therapy. There were 110 studies included, with information up to 72 months for MTX, SSZ and gold and up to 24 months for hydroxychloroquine. The study used extensive search and inclusion criteria.

The median survival time for MTX was longer than that for parenteral gold or SSZ. SSZ withdrawal for lack of efficacy was higher than that for parenteral gold or methotrexate. SSZ and parenteral gold withdrawals because of toxicity were higher than that for MTX.

2.2.1.6 Efficacy of leflunomide

Leflunomide is a disease modifying agent and is classed pharmacologically as an isoxazole. It differs in its mode of action to methotrexate and therefore lack of response to MTX does not indicate a similar outcome for leflunomide. [99]

The efficacy of leflunomide against placebo has been demonstrated in RCTs. [99, 100] Radiographic data, physical function and QOL scores all demonstrated improvement over placebo. In a trial comparing leflunomide to placebo and leflunomide to MTX (n=482), the primary outcome measure at 52 weeks was the ACR 20 and secondary outcomes included ACR 50, ACR 70, HAQ scores, QOL and radiographic improvement. [100] Exclusion criteria included previous MTX therapy and liver impairment. Forty one per cent of patients in the leflunomide group achieved an ACR 20 response compared to 19% in the placebo group and 35% in the MTX group. The mean change in HAQ score was -0.3 in the leflunomide group, +0.1 in the placebo group and -0.2 in the MTX group. The mean dose in the MTX group was considerably lower than usual treatment doses (7.5mg/week) and it was not clear what the mean dose overall was (including dose increase). A further trial (n=358) compared the efficacy and safety of leflunomide with placebo and with sulphasalazine. [99] The primary outcomes measures at 24 weeks were tender and swollen joint counts and physician and patients overall assessment. Leflunomide and SSZ were significantly more effective than placebo. Only the inflammatory marker, erythrocyte sedimentation rate (ESR), showed a significant difference between leflunomide and SSZ. A RCT examined treatment of leflunomide, in patients on stable doses of MTX with active disease (n=263) vs placebo with MTX. [101] The primary outcome measure was ACR 20 and secondary outcomes were HAQ and QOL (SF-36) at 24 weeks. Forty six per cent of the leflunomide group and 19.5% of the placebo group achieved an ACR 20 response. ACR 50 response rates were 26.2% vs. 6% in the leflunomide and placebo groups respectively. ACR 70 response rates were 10% and 2.3% for the leflunomide and placebo groups respectively. The mean HAQ change in the leflunomide group was -0.4 and -0.1 in the placebo group.

2.2.1.7 Toxicity of leflunomide

The most common adverse effects associated with leflunomide include GI upset, rash, and elevated liver function tests. Liver damage is classed as a rare adverse effect. [102] In a study examining the safety and pharmacokinetics of leflunomide, the most common adverse effects were respiratory infection, alopecia, diarrhoea, increased cough and rhinitis. [103]

2.2.1.8 Discontinuation rates of leflunomide

Experience with a national cohort in the United States (n=3,325) 33 months after first approval was published. [104] The overall discontinuation rates of leflunomide in 3 major trials was reported as 20-30%. [104] Discontinuation rates in an observational cohort in the US found that 52% of patients discontinued leflunomide within 1 year; 35% of these were due to inefficacy and 17% due to adverse effects. [105] Adverse events and lack of efficacy were also the two main reasons for discontinuation in the trials. [104] A 2008 systematic review compared the efficacy and harms of RA drugs. [106] This review concluded that there was no difference in the discontinuation rates between leflunomide and methotrexate or sulphasalazine.

2.2.1.9 Leflunomide usage in Ireland.

Leflunomide has been available in Ireland since 1999 which is approximately the same time as the anti-TNF drugs were beginning to enter the market. Examining the use of leflunomide in recent years indicates that total expenditure on leflunomide in 2009 reached approximately €550,000 (GMS and DPS data, PCRS) for approximately 950 patients. This expenditure increased slightly in both 2010 and 2011. It is thought that the uptake of leflunomide for the treatment of RA was possibly affected by the newer anti-TNF agents launch onto the market. In addition the reimbursement of the newer agents was not restricted and therefore there was no impetus to use this cheaper agent which may have demonstrated greater toxicity in certain patients.

2.2.1.9 Other non-biological DMARDs

Other non-biological DMARDs used include SSZ, azathioprine, gold, ciclosporin, and low dose steroids. The practice with regard to these agents varies considerably in Ireland and may depend on factors such as geographical area, availability of alternative therapies such as biologics, physician and patient preference and disease. There is an evidence base for these agents but it is not as strong as that of MTX. [107-109]

2.2.2 Biological DMARDs

Tumour necrosis factor alpha antagonists (anti-TNF-α) are the first of the biologic treatment groups used in RA. Anakinra is an interleukin 1 (IL-1) antagonist but is used much less frequently than the anti-TNF agents and this is thought to be related to poorer efficacy data and daily subcutaneous injections. A recent review of these agents in Australia by the Prescribing Benefits Advisory Board (PBAC), identified anakinra, as significantly less effective than the anti-TNF agents and the review recommended that reimbursement be withdrawn.[110] A number of other biologic groups are now also licensed; B cell antagonist (rituximab), T cell modulator (abatacept), and the interleukin 6 (IL-6) antagonist (tocilizumab). These are all administered in the hospital setting in Ireland and therefore fall under a different reimbursement process than the other agents. These are reimbursed by the individual hospitals under local budgets. A full review of the clinical efficacy and effectiveness of the anti-TNF agents is given in chapter 5. The biological drugs to be included in this evaluation are shown in Table 3.

Table 3. Biological drugs included within this economic evaluation

Name of Drug	Mode of action	Licence Indication	Dose	Reimbursement	Mode of Administration
Adalimumab (Humira)	Anti-TNF- α	RA, PsA, AS, Crohns, Psoriasis	40mg every 2 weeks	High technology drugs scheme	Subcutaneous Injection
Certolizumab (Cimzia®)	Pegylated Anti-TNF- α	RA	400mg at weeks 0,2,4 and 200mg every 2 weeks thereafter	High technology drugs scheme (with PAS)	Subcutaneous injection
Etanercept (Enbrel®)	Anti-TNF- α	RA, PsA, AS, Psoriasis, JIA	50mg weekly Or 25mg twice weekly	High technology drugs scheme	Subcutaneous injection
Golimumab (Simponi®)	Anti-TNF- α	RA	50mg once monthly	High technology drugs scheme	Subcutaneous injection
Infliximab (Remicade®)	Anti-TNF- α	RA, PsA, AS, Crohns, Psoriasis, Ulcerative colitis	3-5mg/kg at week 0,2,6, and 8 weekly thereafter	Hospital funding	Intravenous infusion over 2 hours

JIA - Juvenile Idiopathic Arthritis, AS Ankylosing Spondylitis, RA Rheumatoid arthritis, PsA Psooriatic arthritis PAS Patient Access Scheme

2.3 Summary of existing economic evaluations in RA

A review of the literature was completed for existing economic evaluations examining the efficacy, costs and cost-effectiveness of anti-TNF therapy in established RA. Articles on the cost-effectiveness of drugs for RA after the failure of one or more DMARDs were identified. In addition, the NHS Economic Evaluation Database (NHS EED), Cochrane Library database, TUFTs CEA database and the websites of NICE, SMC, and CADTH. The methods used for this are outlined in Appendix 1.

Cost-effectiveness evaluations of anti-TNF agents by HTA agencies

2.3.1 Economic evaluations by NICE

To date NICE have commissioned 19 technology appraisals for biologicals in musculoskeletal disease. Guidance has been issued on six of these for RA, three of which include cost-effectiveness of anti-TNF agents as the primary question (Table 4).

Table 4. Summary of economic evaluations completed to date on anti-TNF agents in RA by HTA agencies.

Technology appraisal	Interventions Included	Comparator	Appraisal Type	Guidance / Recommendations	Year completed
NICE					
TA 195	Adalimumab, abatacept, etanercept, infliximab, rituximab	Conventional DMARD sequence beginning with leflunomide	MTA	Rituximab is most cost-effective following inadequate response to DMARDs including at least 1 anti-TNF agent.	**2010**
TA 186	Certolizumab	Adalimumab and etanercept (Monotherapy scenario) Adalimumab, etanercept, infliximab, rituximab and tocilizumab + methotrexate (Combination therapy scenario)	STA	Certolizumab pegol is cost-effective when a PAS* is in place	**2010**
TA 130	Adalimumab, etanercept, infliximab	DMARD sequence	MTA	Agents are cost-effective if patients have DAS >5.1 and have had an inadequate response to 2 or more DMARDs. Treatment should not be continued if inadequate response at 6 months	**2007**
TA 225	Golimumab	TNF-inhibitors + Methotrexate	STA	Golimumab is cost-effective if used as described in TA 130 and TA 195 in combination with a PAS**	**N/A**

Technology appraisal	Interventions Included	Comparator	Appraisal Type	Guidance / Recommendations	Year completed
	Abatacept, adalimumab, certolizumab, etanercept, golimumab, infliximab, rituximab	N/A	Therapeutic Review Panel (TRP)	In adult patients with rheumatoid arthritis with an inadequate response on optimal doses of DMARDs, one of the following biologics: abatacept, adalimumab, etanercept, golimumab, or infliximab could be used in combination with methotrexate or other DMARDs. Anakinra and certolizumab were not recommended. Rituximab is restricted for patients who have failed an anti-TNF agent.	**2010**
CADTH					
Issue 85	Etanercept, infliximab	N/A	Clinical Review	Agents are not cost-effective at a lower than C$100,000 willingness to pay threshold	**2007**
S0174	Golimumab	N/A	Common Drug Review: Cost minimisation analysis	Golimumab + methotrexate be listed in a similar manner to other anti-TNFs in RA. Dosing should be restricted to a maximum of 50 mg/month. Response to golimumab should be assessed after 14 to 16 weeks of treatment and therapy be continued only if there is a clinical response.	**2010**

Technology appraisal	Interventions Included	Comparator	Appraisal Type	Guidance / Recommendations	Year completed
S0175	Certolizumab	N/A	Common Drug Review: Cost minimisation analysis	Was not listed for reimbursement	**2010**
SMC					
81/03	Adalimumab	N/A	STA	Accepted for restricted use within NHS Scotland for the treatment of rheumatoid arthritis by specialist physicians in accordance with guidance by the British Society for Rheumatology.	**2003**
305/06	Etanercept	N/A	STA abbreviated submission	Accepted for use within NHS Scotland	**2006**
590/09	Certolizumab	Anti-TNF therapy (Infliximab, adalimumab and etanercept) + methotrexate	STA	Accepted following resubmission under the same indications as for NICE submission in combination with PAS only.	**2009**

**Patient Access Scheme: manufacturer provides the initial 12 week supply of drug free of charge.*

*** Patient Access Scheme: manufacturer supplies golimumab 100mg dose for the same price as the 50mg dose.*

TA 130

Technology assessment report no. 130 examined the clinical and cost-effectiveness of TNF inhibitors (adalimumab, infliximab and etanercept) for the treatment of adult RA patients (both early disease and established >3yrs). [77] The Birmingham Rheumatoid Arthritis Model (BRAM) Markov model was used. [111] The model is a simulation model, with a lifetime horizon, which considered improvements in QOL and mortality, but assumed no effect of the TNF inhibitors on the need for joint replacement. The incremental cost per QALY for therapy with MTX was £24,000 for etanercept, £30,000 for adalimumab and £38,000 for infliximab. For monotherapy, the ICERs were higher with both adalimumab and etanercept in the region of £50,000 per QALY. When the effectiveness values for early RA were used for TNF inhibitors in third place, the results for the three TNF inhibitors were broadly similar. They were sensitive to assumptions around HAQ progression while on treatment, and to assumptions around effectiveness and long-term survival on conventional DMARDs. When the effectiveness values for late RA were used instead, the results were considerably less favourable (ICERs between £50,000 and £140,000). The key assumption, influencing the result of the model was in relation to HAQ progression. Assuming no HAQ progression while on anti-TNF therapy reduced the ICERs by approximately 50% and assuming a slow HAQ progression increased the ICERs by 50%.

Along with company submissions, an independent report was carried out by the University of Sheffield using data from the British Society for Rheumatology Biologics Registry as part of the NICE review[20]. The lifetime cost of anti-TNF therapy was £57,919 per 5.15 QALYs versus £20,706/ 3.59 QALYs for conventional DMARDs. The incremental cost/QALY was £23,882/QALY. The probability of cost-effectiveness at a WTP of £20,000 was 11% and at £30,000 was 84%. A number of sensitivity analyses were carried out; the impact of using SF-6D over EQ-5D to estimate utility gain was explored. The cost per QALY gained using the SF-6D was £48,206 with 0% probability of cost-effectiveness at either WTP threshold. The impact of disability progression (HAQ score) whilst on DMARDs (in the absence of anti-TNF therapy) was also explored. The cost per QALY gained in this case was

£18,537 and had a probability of being cost-effective at WTP £20,000 of 68% and at £30,000 of 100%.

There were a number of limitations with this analysis. Anti-TNF therapy is viewed as a group rather than individually. There may be differences between the agents, which was not possible to see with the data that was available. The data from BSRBR was limited in providing sufficient information on doses; a variable that was shown to influence the result in sensitivity analysis. Assumptions in relation to long term disability on DMARD therapy were based on a paper by Scott *et al.* [112] which showed considerable heterogeneity between the studies. Data on the long-term progression of HAQ, whilst on anti-TNF therapy, was not available for this study but was recommended by the researchers of the study.

TA 195 (Sequential anti-TNF therapy)

NICE examined the anti-TNF agents used sequentially in 2009 (TA 195). [113] This was a multiple treatment assessment of adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of RA after the failure of a TNF inhibitor. In this assessment rituximab was the most cost-effective drug following the failure of one other anti-TNF inhibitor. The comparators used were other anti-TNF's and DMARD therapy, which had not yet been tried. Again the BRAM was used but some changes had been made. The mapping of utility and HAQ included a quadratic term in addition to the linear which had previously been used. Compared to DMARDs the ICERs were: £34,300/QALY for adalimumab, £38,800 for etanercept, £36,200 for infliximab, £21,200 for rituximab, and £38,600 for abatacept. Rituximab dominated the TNF inhibitors. The ICER for abatacept compared to rituximab was over £100,000/QALY. Important drivers of uncertainty were found in the scenario analysis which included assumptions about HAQ progression on biologic treatments, the equation relating HAQ to QOL (utility), and for comparisons involving rituximab, the assumed time between treatments. The inclusion of adverse event costs for biologic therapy made little difference to the results.

Single Technology Appraisals

NICE have also commissioned a number of single technology appraisals (STA's). These include abatacept, rituximab, tocilizumab, and certolizumab. STAs involve critically reviewing a company's dossier; independent modelling is not performed.

TA 186

The most recent of these STAs was for certolizumab (Cimzia®) (TA 186). [114] The review group stated that 'the cost-effectiveness of certolizumab relative to other biologic DMARDs is unclear because the economic modelling undertaken may have ignored relevant effectiveness data and potential differences between trial populations, and so may have included effectiveness results that were biased in favour of certolizumab; underestimated uncertainty in the relative effectiveness of compared DMARDs; and ignored the potential influence of differences between biologic DMARDs with regard to adverse events and their related costs and health impacts'. In October 2009 the NICE committee did not recommend certolizumab as a treatment option for people with RA but requested more information from the manufacturer. Following receipt of this information and details of a PAS, NICE issued final guidance recommending certolizumab, under certain criteria, as a treatment option for people with RA. These criteria stipulated that certolizumab is used as described for other TNF inhibitor treatments in 'Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis (NICE technology appraisal guidance 130) and the manufacturer provides the first 12 weeks of certolizumab pegol (10 pre-loaded 200-mg syringes) free of charge to all patients starting treatment.'

TA 225

Golimumab is licensed for use in RA and the submission assessed the cost-effectiveness of golimumab following an inadequate response to non-biological anti-rheumatic drugs and following the lack of response to at least one anti-TNF agent. The report advised that golimumab was cost-effective if used in line with those guidelines given in TA 130 (for patients with an inadequate response to non-biological DMARDs) and TA 195 (for patients with an inadequate response to at least

one anti-TNF agent). The results were sensitive to assumptions around HAQ progression and long term benefits. An appraisal examining the use of golimumab in methotrexate naïve RA patients is currently suspended.

2.3.1.1 Economic evaluations on other biological drugs for use in RA by NICE

Roche currently have two biologic agents licensed for the treatment of RA. Rituximab (Mabthera®) is a B cell antagonist and tocilizumab (Ro-Actemra®) which is an IL-6 inhibitor. Both of these drugs have been appraised as STAs.

Rituximab

Roche submitted a micro simulation Markov model based upon the REFLEX trial for a STA. [115] The appraisal was carried out by a review group at Liverpool University. For the 'NICE-recommended' scenario and the 'sequential TNF inhibitor' scenario, the original submission reports ICERs of £14,690 and £11,601/QALY gained respectively. The review group adjusted the model assumptions to more realistic estimates and the ICERs for the NICE-recommended scenario and the sequential use of TNF inhibitor ranged from £37,002 to £80,198/QALY gained and from £28,553 to £65,558/QALY gained respectively. The guidance issued by NICE in August 2007 states that rituximab in combination with MTX is recommended as an option for the treatment of adults with severe active RA who have had an inadequate response to or intolerance of other DMARDs including treatment with at least one TNF inhibitor agent.

Tocilizumab

Roche submitted an individual sampling model based on a mixed treatment comparison which included data from four tocilizumab RCTs (OPTION, TOWARD, LITHE and RADIATE).[115] The committee recommended that tocilizumab plus MTX is not recommended for the treatment of moderate to severe active rheumatoid arthritis before, or as an alternative to, treatment with rituximab for people whose RA has responded inadequately to one or more previous TNF-α inhibitors. The

Committee also concluded that tocilizumab as monotherapy was not a cost-effective for the NHS. However, tocilizumab plus MTX as an option for people whose RA has responded inadequately to treatment with one or more previous TNF-α inhibitors and rituximab was recommended as a cost-effective option. The Committee also recommended tocilizumab plus MTX as an option for people with moderate to severe active rheumatoid arthritis whose rheumatoid arthritis has responded inadequately to one or more previous TNF-α inhibitors and in whom rituximab is contraindicated or who had rituximab withdrawn because of an adverse event.

2.3.2 Economic evaluations by CADTH

Anti-TNF agents (etanercept and infliximab)

The Canadian Agency for Drugs and Technologies in Health (CADTH) has also carried out technology appraisals for anti-TNF agents in RA (Table 4). A therapeutic review to evaluate the comparative effectiveness, harms, and cost-effectiveness of biologic response modifier agents for adults with RA was conducted in 2010.[116] The cost-effectiveness model inputs were based on the results of the clinical effectiveness review which indicated that there was no statistical difference in efficacy between the biologic agents. The review group therefore focussed mainly on differences in costs between the agents. This approach is not often used in cost-effectiveness analysis due to insufficient exploration of the uncertainties via a decision analytical model. It could be argued that due to the significant heterogeneity present among trials of anti-TNF agents, inputs from evidence synthesis should be further explored through probabilistic sensitivity analysis in an economic model. A previous CADTH appraisal, performed a systematic review on the long term clinical effectiveness, safety and cost-effectiveness of etanercept and infliximab.[117] The appraisal concluded that infliximab and etanercept based strategies were not cost-effective, based on a C$50,000 threshold value for a QALY. The agents may be cost-effective in patients with an inadequate response to DMARDs if the willingness to pay threshold was increased above C$100,000. Their results were only sensitive to estimated improvements in QOL and in order for the technologies to be cost-effective; the utility gain would have to be doubled. When long term safety was examined, the

authors found that the risk for some serious complications, such as tuberculosis and autoimmune disorders, were greater with infliximab than with etanercept. Fifty per cent of patients on infliximab and 30% of patients on etanercept discontinued therapy by the third year of treatment. The long-term impact on functionality, survival, or QOL was not demonstrated.

Golimumab

CADTH also completed common drug reviews on both golimumab and on certolizumab. For golimumab, a CMA was submitted comparing adalimumab, etanercept, infliximab, rituximab, anakinra, and abatacept. The annual cost of golimumab (C$17,365) was less than etanercept (C$18,995) and adalimumab (C$18,438). It may cost more or less than infliximab depending on patient weight, dose used and vial sharing of infliximab.

Certolizumab

The drug review of certolizumab was similar to the golimumab submission in that a cost minimisation was also submitted. The annual cost of certolizumab (C$19,271), was higher than adalimumab (C$18,943) and etanercept (C$18,388) in the first year of treatment, than subsequent years. The cost in subsequent years was lower than the other agents (C$17,277). The committee did not recommend certolizumab for reimbursement because the quality of the certolizumab trials was limited and other therapeutic options were available.

Note: C$1 = €0.72(September 2011)

2.3.3 Economic evaluations by SMC

The Scottish Medicines Consortium (SMC) evaluates all drugs under the STA process. To date they have reviewed adalimumab, etanercept, and certolizumab (Table 4). Golimumab was approved for restricted use in November 2011.

Adalimumab for RA (81/03)

Adalimumab was reviewed under the STA process in 2003. It was accepted for restricted use in accordance with guidance by the British Society of Rheumatology. This guidance stipulates that the physician must be a specialist in the treatment of RA and that the patient should be enrolled in the Biologics Register. There is no other data given by the SMC with regard to the submission.

Etanercept (305/06)

Etanercept was reviewed in 2006 under an abbreviated STA submission. This underwent an abbreviated submission as the company were applying for reimbursement of a new formulation; etanercept 50mg once weekly injection. There is no information available on the SMC website on the details of the STA submission.

Certolizumab (590/09)

Certolizumab was reviewed by the SMC in 2009/2010 and guidance was issued in April 2010. The guidance did not recommend reimbursement of certolizumab. The problems highlighted with the submission were related to the indirect comparison method and results and the assumption of long term benefit extrapolated from relatively short term data. The indirect comparison was updated with a mixed treatment comparison when requested; however this additional analysis gave a pattern of results similar to those of the indirect comparison. The data from the indirect comparison was used to drive the model and the review group considered that the uncertainty associated with these estimates in combination with the uncertainty associated with long term benefits, to be too great to recommend reimbursement. The company resubmitted certolizumab for reimbursement for RA, in combination with a PAS. The PAS follows that offered under the NICE agreement, where the first three months are supplied free of charge by the company. Additional long data was also submitted from 3 year open label studies. The review group considered the resubmission and recommended that the drug be reimbursed for RA both in combination with MTX and as monotherapy.

A search of the TUFTs cost-effectiveness database returned 31 relevant published cost-effectiveness studies completed to date in RA. Twenty eight of these were specifically on the biologic agents. The remaining studies comprised of evaluations of non-biologic DMARDs, bisphosphonate therapy, and non-pharmacological therapy. The database grades the quality of the studies on a scale from 1(low) to 7 (high). Almost all of the studies used Markov or a variation of Markov models. Two studies used a decision analytical model.[118, 119]

2.3.4 Economic Evaluations on Leflunomide

A review article of the cost-effectiveness of DMARDs in RA identified six relevant publications. [120] Schadlich et al. examined the cost effectiveness of including leflunomide in sequential DMARD therapy through a cost-utility analysis. [121] The analysis was conducted from a societal perspective using real world data from a German rheumatological database. The authors concluded that after three years a strategy including leflunomide was more effective and less costly than a strategy not including leflunomide. The addition of leflunomide to the treatment strategy extends the time patients benefit from DMARD therapy. A company representative for the manufacturer of leflunomide was one of the authors of the paper. Maetzel at al. examined the cost-effectiveness of leflunomide versus methotrexate in Canada. ICERs in comparison to MTX ranged from approximately CA$54,200 to CA$72,000 per QALY.[122] Kobelt et al. compared leflunomide to methotrexate to sulphasalazine in recently diagnosed patients. Leflunomide (LEF) dominated methotrexate and sulphasalazine. Schipper et al. estimated the cost-effectiveness of strategies aimed at inducing remission in early RA.[123] The strategies included Strategy 1: starting MTX monotherapy, followed by the addition of LEF, followed by MTX with addition of anti-TNF; Strategy 2: start with MTX and LEF combination followed by MTX with anti-TNF; and Strategy 3: immediate start with MTX and anti-TNF. The results indicated that initiation with strategy 2 or 3 was not cost effective compared to strategy 1.

Summary

To date none of the HTA agencies have carried out a MTA on all five anti-TNF agents in comparison to leflunomide. The primary areas of uncertainty in the evaluations have been around efficacy estimates and assumptions on long term benefit. While all agencies have accepted some if not all anti-TNF agents for reimbursement, it has been acknowledged that the uncertainty around the ICER estimates are significant and gaps remain as regards real life long term improvements to health related quality of life (HRQOL) and reduction in joint damage demonstrated radiologically.

Note: Quality of life data is presented in this thesis for both RA and PsA but economic modeling and associated analysis is performed for RA only.

Chapter 3

Chapter 3.0 Outcomes in Rheumatoid Arthritis

3.0 Introduction

The emphasis on the health benefit of interventions within healthcare has advanced outcome measurement to the forefront of health strategies. The focus has moved from '*what works,* to *what works most effectively in this setting'*. This is best achieved through the identification of desired outcomes associated with healthcare interventions and using these to evaluate interventions. Therapeutic strategies today promote early initiation of disease modifying anti-rheumatic drugs (DMARDs) and tight control of the disease through close monitoring of clinical outcomes. These innovative treatments have altered the path and face of RA and outcomes for patients and society.

The measurement of outcomes in a chronic disease such as RA is a multi-dimensional phenomenon. The manifestations of RA can vary from joint symptoms such as pain swelling and joint damage leading to functional impairment to more non-specific complaints such as fatigue and poor general health. This variation in presentation and course of disease has led to the development of outcome measures in an attempt to evaluate interventions used both in clinical trials and in clinical practice. [124, 125] One measure will be insufficient to capture all aspects of benefit or damage. Instead a multi-dimensional approach must incorporate aspects such as disease progression, structural damage, and QOL. This approach should be adopted at the disease diagnosis (through classification of disease) and continue through the course of the disease (measurement of treatment response). The focus for outcome measurement has largely been driven by those measured in the clinical trial setting rather than the 'real life' clinical practice setting; however due to the establishment of registries to evaluate treatments in the clinical practice setting, this is changing.

Outcome measurement in RA can be broadly classified under clinical outcomes and patient reported outcomes.

3.1 Clinical Outcome Measures

The drive towards a core set of outcomes for RA has been mainly due to the development of the international initiative known as the Outcome Measures in Rheumatology (OMERACT) group. [126] OMERACT is the acronym for an international, informally organized network initiated in 1992, aimed at improving outcome measurement in rheumatology. OMERACT was established to review and recommend a core set of outcomes which would appropriately measure the impact of treatment on chronic musculoskeletal diseases. While the remit of the group was originally clinical trials in relation to RA, it now encompasses a broad range of diseases within the musculoskeletal disease group. Data driven recommendations are prepared and updated by expert working groups and recommendations include core sets of measures for most of the major rheumatologic conditions. [126] OMERACT have succeeded in achieving consensus on core sets of measures for RA, osteoarthritis and osteoporosis, psoriasis/ psoriatic arthritis, on psychosocial measures and core set of data for cost-effectiveness evaluations. The review process must satisfy the OMERACT filter which encompasses three filters; truth (does it measure what it is supposed to), discrimination (can the measure discriminate between situations of interest) and feasibility (is it understandable and cheap).

3.1.1 American College of Rheumatology (ACR) Response

The ACR criteria specifically measure the change or improvement in disease activity of active drug against placebo.

The ACR 20 is the preliminary definition of improvement. ACR 20 can be defined as a binary categorical outcome indicating whether a patient responded or not. Using ACR 20, a patient is considered to respond when showing at least:

- 20% improvement in the tender joint count and
- 20% improvement in the swollen joint count and
- At least a 20% improvement in 3 out of the following 5:

1. Patient pain assessment
2. Patient global assessment
3. Physician global assessment
4. Patient self-addressed disability
5. Acute phase reactant (ESR or CRP)

The ACR 20 was considered for re-evaluation in 2007.[127] One of the key criticisms of the ACR 20 is the lack of sensitivity to change in comparison to a continuous measure such as the DAS. With the introduction of new therapies such as the anti-TNF agents which were more efficacious than older therapies, it became apparent that an improvement of 20% may not be enough. In light of this, the ACR 50 (50% improvement), 70 (70% improvement) and 90 (90% improvement) thresholds are now used. However, it has been suggested that the discriminating power of these outcome markers are not as well validated as the ACR 20.[128] Nonetheless, they are now commonly reported in studies.

Also new definitions of improvement were appearing in trials such as the ACR-N.[129] The ACR-N expresses change as a percentage in individual patients, making the result dependent on the initial state (e.g. a decrease of 2 of 20 tender joints is an improvement of 10%, but a decrease of 2 of 10 joints is an improvement of 20%). It is a continuous variable (like the DAS) based on the 7 accepted core set variables (as does the ACR 20) and is analysed according to a formula analogous to that used for the ACR 20, ACR50 and ACR70. A further extension of the ACR-N describes the area under the curve (AUC) which substantially increases the power to detect small differences in the treatment arm.[130] There have been some criticisms of the ACR-N and it not been through the OMERACT filter, as has the ACR 20.[131]

3.1.2 The EULAR response criteria

The EULAR response criteria include not only change in disease activity, but also current disease activity. To be classified as responders, patients should have a significant change in DAS (>1.2 for a good response and ≤1.2 and >0.6 for a

moderate response) and also low current disease activity. Three categories are defined: good, moderate, and non-responders (Figure 11)

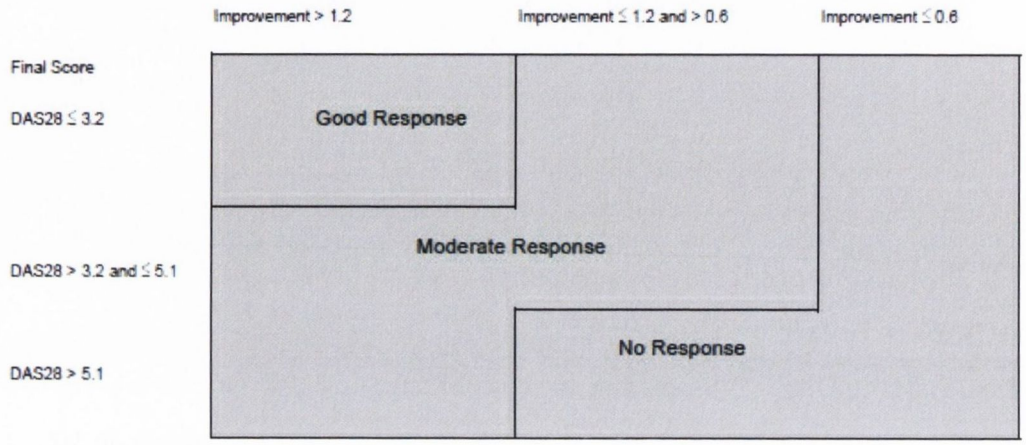


Figure 11: EULAR Response
The table indicates the response achieved which is dependent on both the quantity of change in DAS score and the final DAS score.

DAS 28[56]

The original DAS was based on a 44 joint score but validation studies recognized that the DAS 28 (modified version) can be used effectively in its place. [132] However, these measures should not be used interchangeably.

DAS 28 is a compound outcome measure comprising a number of components; 28 tender joint counts; 28 swollen joint counts; patient's general health (VAS); ESR. A DAS 28 using CRP as the laboratory measure of inflammation (acute phase reactant) has also been developed. [133]

The DAS 28 has a continuous scale ranging from 0 to 9.4. The level of disease activity can be interpreted as low (DAS 28<3.2), moderate (3.2< DAS 28 < 5.1), or high (DAS 28 > 5.1).

Remission of RA is now a realistic goal as a result of significant treatment advances, improved early diagnostic criteria, and aggressive management of inflammation. Remission is defined within the ACR criteria as ACR90 and within the EULAR criteria as DAS 28 < 2.6.

3.1.3 Radiographic Outcomes

The use of scored radiographs as an outcome measure can help estimate the progression of RA. One of the main advantages of X-rays is that they provide a permanent record of true damage, unlike measures related to disability and pain, which are subjective. Halting of radiographic damage is one of the most sought after outcomes from any intervention as evidence suggests that structural joint damage is the predominant cause of functional impairment.

The most commonly used methods are those devised by Sharp, Larsen, and van der Heijde/Sharp, and their variants. [134, 135] Methods based on the Sharp technique provide separate scores for erosion and for joint space narrowing. Larsen and modified versions, together with the Simple Erosion Narrowing Score (SENS) method, provide an overall score. [136]

The choice of scoring method depends on the time and staff available, and the required degree of reliability and sensitivity to change. In clinical practice, many patients will have x-rays at baseline and follow-up to assess joint damage. However using a formal scoring method such as described here can be time consuming and may be subject to variability.

While imaging usually involves taking X-rays of the affected joint, Magnetic Resonance Imaging (MRI), and ultrasonography are also used commonly.

3.2 Health Related Quality of Life

The growth of economic analyses and in particular CUA, which uses the QALY as a measure of outcome, has heightened the interest in the methodologies used to perform these analyses. A QALY is calculated by combining length of life with QOL. The index for translating QOL is known as the health utility. Health utilities can be measured directly or indirectly.

3.2.1 Direct HRQOL Measurement

Direct utility measurement uses techniques such as time trade off (TTO) and the standard gamble (SG).

3.2.1.1 Time Trade Off

The TTO approach offers two alternative scenarios: life expectancy of an individual with a chronic condition followed by death or healthy time followed by death. The TTO method is used to calculate the utility values which are given by the EQ-5D, an indirect measure. This was done in the valuation of the health states measured in the Measurement and Valuation for health study, in the UK(MVH).

3.2.1.2 Standard Gamble

The SG involves the patient choosing between two alternatives depending on the probability of achieving either a chronic health state preferred to death or a temporary health state preferred to death. The choice of best outcome is varied until the respondent is indifferent between the certain and uncertain prospects. This was the method used for the valuation of the SF-6D.

3.2.2 Indirect QOL Measurement

Recently, there has been an increased emphasis on the patient's outcomes through the development of a group of measures called Patient Reported Outcomes (PROs). [137] These instruments are described as indirect measures, but this is in the context of utility measurement. They directly measure a patient's QOL but in order to transform this information into a utility value, a further analysis is carried out hence referred to as indirect.

PROs primarily report the patient's HRQOL and management of disability. An increasing number of publications emphasize the importance of PRO measures of health status and HRQOL in RA. Clinicians and decision makers are recognizing the importance of measuring HRQOL to inform patient management and policy decisions.

PROs that assess HRQOL are often categorized as either generic or disease-specific. Generic measures are designed for use among diverse populations with a broad range of medical conditions, but can also be used to characterize healthy people without a particular medical condition. In contrast, disease-specific measures are designed to assess specific populations, quantify aspects of functioning, and examine the impact of particular medical conditions or treatments.

A number of studies have evaluated the reliability, validity and responsiveness in patients with RA and most instruments can discriminate between different severities of the disease. In deciding which measure to choose, it is important that the researcher or clinician considers the context; is the objective to compare outcomes with other disease outcomes or is the objective to focus on the particular attributes of the disease in question? To compare outcomes between diseases, a generic instrument may be most appropriate, whereas when focusing on attributes of a particular disease, a disease-specific instrument may be appropriate. Because generic and condition-specific measures have different attributes, and are conceptually distinct, it is sometimes useful to administer both types of instruments as part of a complete outcome assessment in clinical trials.

3.2.2.1 Disease-Specific Patient Reported Outcomes

Disease-specific PROs are very often measured in both clinical trials and in clinical practice. They are developed for a specific condition and therefore some studies have shown that they are more responsive to small changes in disease status. The most commonly used disease-specific PRO in RA is the Health Assessment Questionnaire (HAQ).

3.2.2.1.1 Health Assessment Questionnaire

The HAQ was developed as a systematic measure of outcome in patients with a wide variety of rheumatic diseases, including RA. [138] The usual form of the HAQ used is the physical disability scale (Modified HAQ) which measures function in relation to the degree of difficulty experienced in performing activities of daily living such as

dressing, rising, personal hygiene, walking, eating and ability to carry out chores (Appendix 2). The HAQ contains 20 items across 8 domains, which are scored from 0 (no difficulty) to 3 (unable to do). The scores are corrected for the use of equipment and help from carers for each domain. The scores for each domain are then summed and a score of 3 represents high dependency disability and 0 represents no disability.[138] The HAQ is a self-completed questionnaire developed as a comprehensive measure of outcome in patients with a wide variety of rheumatic diseases, including RA, osteoarthritis, ankylosing spondylitis, psoriatic arthritis, scleroderma, lupus, juvenile RA and fibromyalgia.

There are however some limitations to its use. It does not capture disability associated with sensory organ dysfunction or psychiatric dysfunction and does not directly measure patient satisfaction or social networking. The HAQ-disease index (HAQDI) is the single index score derived from scoring the HAQ. A HAQ score difference of 0.25 is said to represent the minimum clinically important difference (MCID). [139] The HAQ is commonly collected in clinical trials and also been used in many cost-effectiveness evaluations to model health states that a patient may experience over the course of their disease. [38, 39, 111]

Other PRO measures include the Rheumatoid Arthritis Quality of Life (RAQOL) and the Arthritis IMpact Scale (AIMS); however these are not used as frequently as the HAQ.

3.2.2.2 Generic Patient Reported Outcomes

Generic measures include the Short Form-6D (SF-6D), Short Form-36 (SF-36), Health Utilities Index Mark 2 (HUI2) and Mark 3 (HUI3) and the EuroQoL (EQ-5D). A number of reviews have compared the generic measures in RA. [140] These instruments are classed as multi-attribute utility theory (MAUT) instruments. MAUT instruments have a generic 'descriptive system' which is capable of describing a wide range of health states and utility weights are attached to every possible state.

3.2.2.2.1 EuroQoL (EQ-5D)

The EQ-5D (3-level) index is a preference-based index measure, where an individual provides an assessment of each component of his/her health status according to a structured health-status classification system and a single preference-based score is derived for each individual based on societal preferences.[141] It is applicable to a wide range of health conditions and treatments and provides a simple descriptive profile for health status (EuroQoL). Functioning in five dimensions is assessed: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension is assessed by one item with three response options: no problems, some problems, and severe problems (Appendix 2). Higher scores on these items indicate greater impairment. The EuroQol Group launched a new EQ-5D-5L (EQ-5D 5 level) self complete version in 2009 in order to further improve the sensitivity and reduce ceiling effects of the existing EQ-5D-3L version. [142] This is now available but the studies to elicit preferences from general populations are at the pilot stage and are not yet available.

The EQ-5D-3L is used extensively to measure QoL in inflammatory arthritis.[140, 143-145] The preferences for the scoring function was measured using the Time Trade-Off (TTO) technique on a random sample of a approximately 3000 adults of the UK population.[146] The scoring was developed using econometric modelling as opposed to multi-attribute utility theory. The EQ-5D system allows individuals to obtain negative utilities for a given health state (which are then interpreted as states 'worse than death'). There are no population values available for the Irish population. The original population scoring which assigned the weights by which to score the heath states were done in the UK but many countries have now completed population scoring. [147, 148]

3.2.2.2.2 Medical Outcome Study Short Form-36 (SF-36)

The SF-36 was designed for use in clinical practice and research, health policy evaluations, and general population surveys. It is a 36-item short-form (SF-36) which includes one multi-item scale that assesses eight health concepts: 1) limitations in physical activities because of health problems; 2) limitations in social activities

because of physical or emotional problems; 3) limitations in usual role activities because of physical health problems; 4) bodily pain; 5) general mental health (psychological distress and well-being); 6) limitations in usual role activities because of emotional problems; 7) vitality (energy and fatigue); and 8) general health perceptions. [149] The questionnaire does not give one overall score but each of the eight domains can be scored out of 100 (perfect health)(Appendix 2). Two summary component scores for physical and mental health can also be calculated. A license is required for use of the SF-36 and the associated software.

3.2.2.2.3 Short Form-6D (SF-6D)

The SF-6D is derived from and uses 11 items from the SF-36. The scoring model for SF-6D was developed using standard gamble (SG) utility measurements on a random sample (n=836) of the general population in the UK.[150] Econometric modelling was used to devise a number of different scoring models. [151] The SF-6D scoring programme was revised in 2007 using a Bayesian methodology.[152] It is known that the parametric model can over predict the value of better health states[152] and the Bayesian version overcomes some of the bias of the original regression models when assigning values to the worst health states (e.g. it yields a value of 0.203 for the worst SF-6D state compared to 0.301 using the original parametric algorithm).

3.2.2.2.4 Health Utilities Index (HUI)

The Health Utilities Index is a family of generic preference-based systems for measuring comprehensive health status and HRQoL. The dimensions assessed include vision, hearing, speech, ambulation/mobility, pain, dexterity, self-care, emotion, and cognition. Each dimension has 3-6 levels. The HUI consists of two systems; HUI MARK 2 (HUI2); HUI Mark 3 (HUI3). The HUI3 is used for primary analysis in most cases. The HUI2 may be used to provide additional information as it includes attributes such as self-care, emotion and fertility. Preferences for the HUI2 were measured on a sample of parents of schoolchildren in Canada and for this reason the HUI2 is sometimes chosen as a useful measure for assessing HRQOL in children.

While it has been used in RA, it is not used as commonly as the SF-6D or the EQ-5D.[153]

Table 5. Overview of Multi-Attribute Utility Theory (MAUT) Instrument Properties [154]

Instrument	Dimensions/domains/attributes	No. of possible Health states	Valuation Technique	Boundaries
HUI2	Sensation (vision, hearing, speech) mobility, emotion, cognition, self-care, pain	24,000	Standard Gamble	-0.03 to 1.00 -0.36 to 1.00
HUI3	Vision, hearing, speech, ambulation, dexterity, emotion, cognition, pain	972,000	Standard Gamble	
SF-6D	Physical function, role limitation, social function, pain, mental health, vitality	18,000	Standard Gamble	0.2 to 1.00
EQ-5D	Mobility, usual activities, self-care, pain, anxiety	243	Time Trade off	-0.59 to 1.00

3.3 Selecting an outcome measure for CUA of anti-TNFs

3.3.1 Choice of instrument for RA

A number of methods are used to measure disease severity and the impact of this severity on QOL. These measures include both clinical tools measuring disease activity, such as the EULAR DAS[155] and QOL instruments. A variety of QOL instruments have been used in inflammatory arthritis trials and these usually include either a generic measure, such as the EQ-5D[141] or the SF-36[149] or a disease specific instrument, such as the HAQ[156], or both.

All of the above QOL measures display some shortcomings in assessing HRQoL in inflammatory arthritis. While using generic measures should in theory allow us to compare results for a variety of different conditions, disparities have been shown to exist in the utilities derived from the EQ-5D and SF-6D, and this is attributed to the

different descriptive systems, to the valuations attached to the health states or to a combination of both.[157] Seymour *et al.* examined the association between these two measures and found that the strength of the relationship between the instruments changes across the health spectrum and is dependent on whether health improves or deteriorates. A common finding in many head- to-head studies is that the EQ-5D tends to generate higher utilities than the SF-6D in subgroups with better health (ceiling effect) , whereas the opposite occurs in less healthy groups (floor effect).[157-160] This has important implications for economic analyses on treatments such as biologic therapy, which are likely to be used for patients in more severe health states than those in mild health states.

For the majority of clinical trials, the primary outcomes chosen are those which are most relevant and meaningful for the clinical community. In RA trials, the HAQ and disease activity (via EULAR DAS 28 or ACR criteria) are almost always recorded as outcomes. One of the key problems in using the HAQ for CUA is that the HAQ does not directly produce utility values. In these cases it is necessary to derive or 'map' utilities from the outcomes measured in the trial. Previous studies have shown how QOL decreases as functional impairment increases[161, 162] thereby making the HAQ a good disease measure to correlate with utility.[38, 39, 111, 163] NICE have now included mapping in their guide to technology appraisals if the EQ-5D has not been used in the trial. [25] A number of linear transformations have been presented in economic analysis deriving utility from HAQ.[39, 163, 164] However there are some problems associated with this method. Utilities derived from other measures such as HAQ overestimate baseline values but underestimate change and this is particularly evident for the SF-6D in early and severe disease and the EQ-5D in early disease.[165, 166] In mapping onto the EQ-5D, the primary area of concern is where the responses fall within states described as worse than death (a utility score of less than zero). A recent paper discusses this fact, in particular some of the theoretical economic issues in some detail.[167]

Chapter 4

Chapter 4.0 Utility Values for an Irish RA and PsA cohort

The chapter will describe the methods used to calculate utility values for an Irish RA and PsA population. While the focus of this thesis is on RA, PsA data has been included in this chapter. This allows some comparisons as regards to be made between diseases. It will also discuss the methods used to apply a revised scoring methodology for the EQ-5D data. Finally the chapter details the methods used for mapping of utilities and how the mapping was performed for this cohort.

4.1 Background

HRQOL measures must be valid and have high reliability and responsiveness. Validity ensures that the instrument measures what it is supposed to measure. Reliable instruments are able to reproduce differences between subjects. Responsive evaluative measures are able to detect important changes in HRQOL during a period of time, even if those changes are small. They should be able to accurately measure these changes for all disease states (from very worst to very best). HRQOL measures should also be interpretable, meaning that the differences in scores that correspond to small, moderate, and large HRQOL changes are easily identifiable and match to some degree the disease state of the subjects.

The utility measure should cover the range of symptoms or aspects of health likely to be experienced by the subjects in question. The measure should be equally sensitive to the dimensions of the measure e.g. for patients with severe disease some dimensions of the SF-6D (physical functioning, role limitations and social functioning dimensions) have a significant number of respondents at the lowest level whereas this is not the case for dimensions such as pain, mental health and vitality.
The valuation method used to assign values to health should be robust and valid. The TTO and SG methods are recognized as choice based valuation techniques and are preferable to VAS.
A checklist for judging the merits of a utility measure has been published by Brazier et. al.[168] This broadly defines criteria under; practicality, reliability, validity of the

valuation technique, the descriptive validity and the empirical validity (with regard to revealed, stated and hypothesized preferences)

4.1.1 Current issues with preference based utility valuation

It is well established that the EQ-5D and the SF-6D produce different utility values in the same cohort. [169, 170] This is in part due to different definitions of perfect health. According to the 1995 Health Survey of England, the EQ-5D considers over half of the population to be in perfect health, while the SF-6D considers less than 3% to be in perfect health. [171] Therefore the SF-6D has a different criterion for perfect health than the EQ-5D. This presents decision makers with a challenge in comparing results of economic evaluations which have used different methods to calculate utility.

In a RA cohort the utility gain produced by the EQ-5D was twice that produced by the SF-6D. [20] This discrepancy between the measures has been the subject of a number of recent publications [165, 167, 172-175] which highlight the methodology of the original scoring of the EQ-5D(UK) and the manner in which *worse-than-dead* (WTD) values were adjusted. [147] In order to examine the methodology of the original EQ-5D scoring we need to describe the methods that were used in the original study.

The TTO formulation used for the MVH study was originally published in 1972 and further developed to accommodate WTD states in 1982. [176, 177] The methodology describing how individual TTO responses can be translated into valuations on an interval scale is given in these papers. Because aggregate values for groups are usually calculated as the arithmetic mean of these individual valuations, potential outliers may adversely influence the overall mean of the sample. This has been the case in a number of large studies and therefore methods have been described to deal with this anomaly. [171, 178] The most commonly used method is the transformation of WTD responses. Three such transformations have been proposed and used; the monotonic transformation put forward by Patrick and used by Dolan, a linear transformation proposed by Shaw and truncation or bounding at -1, +1 used by Dolan. [147, 178, 179]. These methods have been challenged considerably in the literature due to the manner in which they change the actual data. [167, 174, 180-182] Alternative methods have also been proposed. These involve changing the estimator and not the data as in the episodic random utility model (RUM) described for this study, using directional

statistics as proposed by Craig and Oppe and finally use of alternative summary statistics such as median or mode ratios instead of mean as suggested by Lamers and Li. [167, 182-185]

Therefore the literature presents two distinct methods for handling the problem of WTD states in TTO valuation; firstly the transformation method which has come under considerable scrutiny and challenge and secondly, the episodic estimator which is relatively new and may not yet have been subjected to full examination. Both methods present shortcomings. In this case episodic RUM is chosen to test because it presents a significantly different method from the transformation method. A pragmatic approach leads us not to use the revised scoring method exclusively but use it as a further tool to highlight the uncertainty associated with TTO derived preference based utilities. An alternative approach would be to move away from the TTO method entirely and explore the use of methods such as discrete choice analysis or ranking as some authors have explored. [186-189]

4.1.2 Scoring method for the original UK TTO

The preferences for the scoring function were measured using the TTO technique on a random sample of 2997 adults of the UK population (MVH study). [147] Dolan *et al.* devised a scoring method which assigned a single index utility value for each health state described. [147] Forty five of the health states were scored directly from the population using TTO valuation and the values of the remaining states (198) were predicted using regression estimates.

To anchor the scale, perfect health and dead were assigned scores of 1 and 0, respectively. For states described as *better-than-dead* (BTD) (>0) on the TTO, scores were calculated using the formula x/10 where x is the number of years spent in perfect health equal to 10 years in the health state. For states scored as WTD (<0) the formula given is –x/(10-x) where immediate death equates to a scenario of x years in perfect health followed by (10-x) years in the health state. For states BTD, the ratio ranges from 1 to 0 but ratios for WTD states lie between 0 and -39 (the WTD x has an upper bound at 9.75 years). The asymmetry seen between the positive and negative ratios seem to inflate the influence of the WTD responses; therefore Dolan

transformed the negative ratios to –x/10, replacing 34% of the TTO responses.[147] By bounding the negative ratio at -1, the influence of these WTD responses on the mean slope lessened and improved face validity of mean ratio estimates. An alternative method of handling these worse than death scores was proposed by Craig *et al.*.[167, 180]

4.1.3 Revised Scoring Method for the EQ-5D UK

Craig *et al.* re-examined the original data using an episodic regression model instead of a ratio regression model in the MVH study.[167] The health state valuations have been published and these are provided in Appendix 3. [180] The theoretical basis for both models was presented in a previous published paper.[187] This theory is based around the error that is present for the TTO valuation.

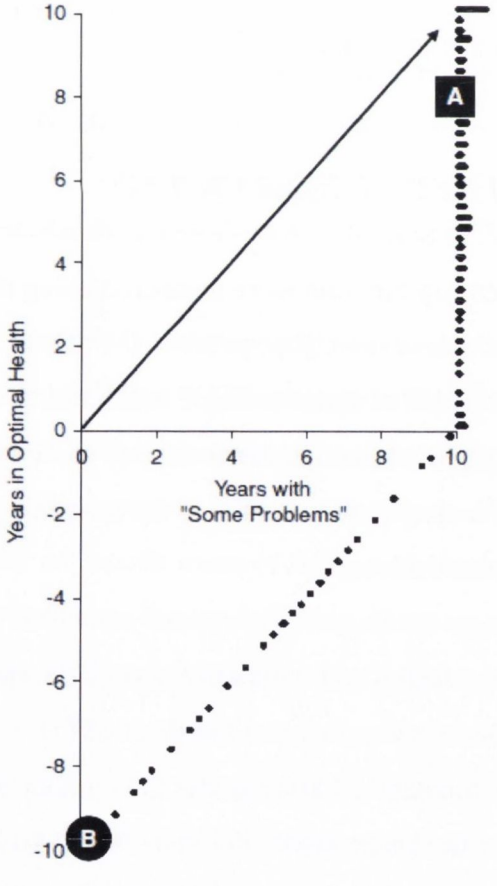


Figure 12: Error associated with Time Trade Off Responses
(based on 838 TTO responses for the 'some problems' EQ-5D state (22222) taken from the Measurement and Value of Health (MVH) study in the United Kingdom (Dolan 1997)) Figure used with permission from Craig et. al. 2010.

In the original MVH study, the responses to the questions in relation to WTD states, were positioned along the line marked 'B' in Figure 12. States better than dead are positioned along the line marked 'A'. The result of this is an over influence of the WTD states on the overall mean estimate. The episodic RUM attempts to lessen this influence of the WTD states by down weighting them.

The utility of a health state, *j*, over time, *t*, for an individual, *i, is* random and can be represented by:

$$U_{ij}(t) = \begin{cases} \mu_j t + \varepsilon_{ij} & \text{Episodic RUM} \\ \left(\mu_j + \varepsilon_{ij}\right) t = \mu_j t + \varepsilon_{ij} t & \text{Instant RUM} \end{cases}$$

Figure 13 provides a schematic on how the data is handled for the revised method in comparison to the original Dolan method.

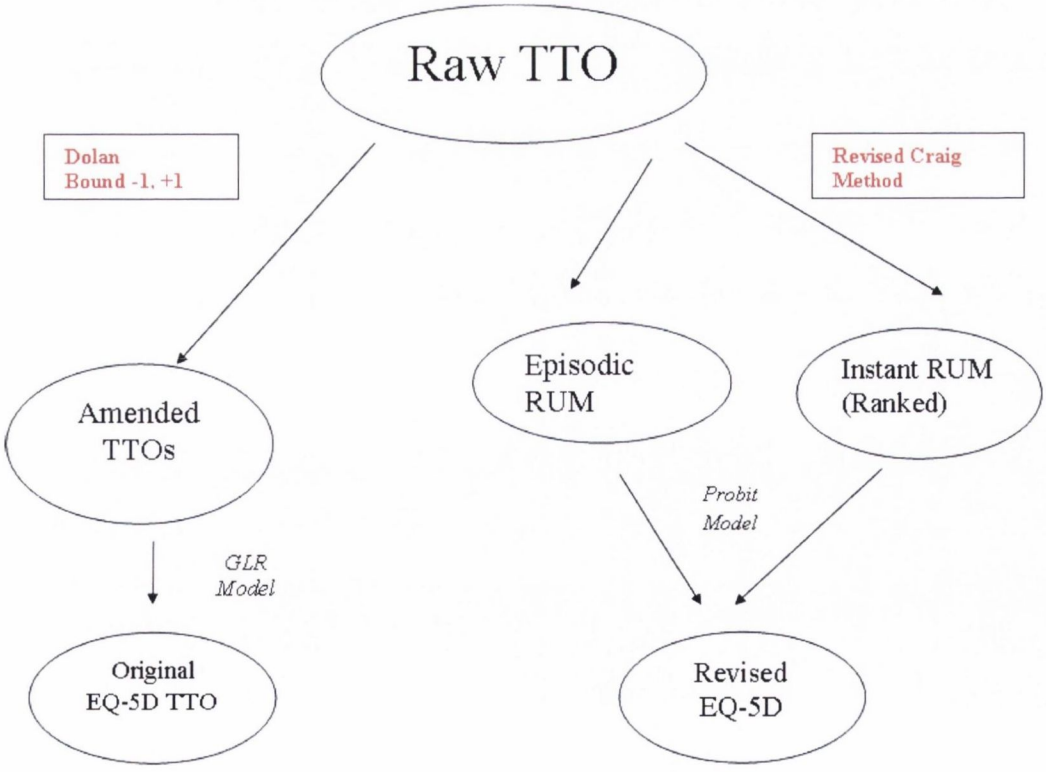


Figure 13: A schematic of the Dolan method and the revised scoring method for the EQ-5D (TTO).

4.1.4 Mapping or Cross walking QOL data

A number of models can be used to map QOL instruments to utilities. The method here uses a linear statistical model. An alternative approach may be a judgement based approach using elicitation from experts. While this approach has been criticised in the past, there have been attempts to develop new methods which allow measurement of uncertainty, using probabilities, around expert's judgements. [190] [191] Brazier *et al.* reviewed the different statistical models which can be used to map variables to other variables. [192] Details of the alternative specifications of such models are given in Table 6.

Table 6. Alternative specifications of mapping functions [192]

Model	Dependent variable	Discrete (D) or continuous (C)	Independent variables			
			Main effects	Discrete (D) or continuous (C)	Interactions	Other measures
1	Index	C	Overall score	C		For any model: squared terms, other health measures, clinical measures, demographics
2	Index	C	Dimension scores	C	Dimensions	
3	Index	C	Item levels	C	Items	
4	Index	C	Item levels	D	Items level	
5	Dimension level	C	Models 1–4	C/D	Models 1–4	
6	Dimension level	D	Models 1–4	C/D	Models 1–4	

The first model (1) is the simplest additive model, which regresses the target measure (such as the EQ-5D) onto the total score of the starting measure (e.g. SF-36, HAQDI, DAS28 etc.). This is also the most limiting specification since it assumes that the dimensions of the starting measure are equally important; all items carry equal weight; and response choices to each item lie on a similar interval scale. This is the model used for this study due to the limitations of the data available (only the final HAQDI score was recorded). Model 2 and 3 model dimension scores and item scores respectively and this allows for less rigid assumptions in relation to equality between levels. Models 4-6 use item responses and dimension scores as independent variables,

however in models 5 and 6, dimension and item scores will be treated as continuous variables and item responses are modelled as discrete dummy variables.

A recent paper created a mapping algorithm for the EQ-5D to an angina disease specific measure under a Bayesian framework. [193] The authors concluded that despite the theoretical advantages of using the more complex Bayesian models, the simple linear model performed best in the validation sample.

4.2 Aims

- To derive patient reported preference based utility scores from patients initiating on anti-TNF therapy.
- To calculate the utility scores using different patient reported HRQOL questionnaires.
- To derive a mapping function from the HAQ and the DAS 28 using Irish data in both RA and PsA.

4.3 Objectives

- To estimate the change in preference based utility score for patients prior to starting anti-TNF therapy and at follow up.
- To compare the differences in the results from using three different methods to measure QOL in both RA and PsA (EQ-5D, SF-6D and EQ-5D using a revised population scoring).
- To examine the differences between linear and non-linear models used for the mapping and to compare the findings with previous mapping studies.

In order to examine these methods for deriving utilities, we used data from an Irish cohort of patients on biologic therapy for both rheumatoid and psoriatic arthritis and compared the differences in the results from using three different instruments to measure QOL. While the focus of this project is on RA patients, we have also

analysed the PsA patients in light of the paucity of evidence with regard to utility in this cohort.

4.4 Methods

4.4.1 Data Source

Utility data was derived from a database of 504 patients from an Irish rheumatology referral centre, which records the clinical and QOL outcomes of patients on biologic therapy for RA and PsA. Patients included have a diagnosis of either RA (n=345) according to the ACR criteria and are commencing on biologic therapy (either anti-TNF α, B-cell antagonists or T-cell modulators), or a diagnosis of PsA (n=159) according to ClASsification criteria for Psoriatic ARthritis (CASPAR) criteria. While many of the patients would have failed two or more DMARDs, this is not a prerequisite for starting biologic therapy in Ireland. The demography of this patient population is described in table 7.

4.4.2 Instruments used

The QOL instruments used in this cohort are the paper versions of the EQ-5D, SF-36 (Version 1), and the modified HAQ (Appendix 2). The DAS 28[155] was collected as one of the clinical outcomes in monitoring response to treatment and disease activity.

4.4.3 Data Collection

The EQ-5D, SF-36 and HAQ were collected as part of normal clinical practice for monitoring the impact of treatment on QOL. Patients attending an outpatient biological clinic completed a standard book of questionnaires as part of their clinical assessment. When a patient is prescribed a biologic they are referred to this clinic. These patients have active disease and usually have a DAS 28 score of 5.0 or more although this is not a prerequisite. The majority of these patients have failed methotrexate therapy or are no longer responding. The clinic is attended by patients with inflammatory arthritis which cn include RA, PsA and ankylosing spndylitis.

Only the RA and PsA were used for this analysis. One of the limitations of the database is the level of data on patients who have discontinued biological therapy. Once patients have discontinued biological therapy they no longer attend the biological clinic and therefore there is no information on how their disease progressed upon discontinuation. Also because this is a clinical database, patients may not have attended at exact timepoints. For example three month patients may have attended at 14 weeks or not until 20 weeks. For this analysis data for the follow-up visit was recorded if it was either one month before or after the 12 month timepoint. This was agreed on discussion with clinicians on the basis that the disease outcomes would be unlikely to have changed in this time. Patients with missing baseline data were not included in this analysis.

The questionnaires were administered to the patients for self-completion and the data was entered onto an Excel® database. The SF-36 data was entered as both raw data and as composite scores for each domain. Only the HAQDI (final HAQ score and not the components of each domain) scores were entered onto the database. All questionnaires were measured at baseline prior to the commencement of therapy. The HAQ and EQ-5D were then repeated every three months and the SF-36 was administered on a yearly basis. The baseline and 12 month data for all instruments was analysed.

Demographic data including age, gender, disease duration, laboratory measures of inflammation, patient global assessments (VAS), joint counts, pain scores (VAS), DMARD history, and concomitant use with biological therapy.

4.4.4 Methods used to calculate utilities

Single index scores were calculated for both the EQ-5D and the SF-6D using the population scoring methods (original and the revised scoring for the EQ-5D) discussed above. The Bayesian estimates of the SF-6D utilities were calculated in Excel® (Further details of the analysis method are available at http://www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d/bayesian.html). Utility scores were calculated using both the original and revised scoring method for the EQ-5D.

The EQ-5D-3L conversion used for this study is derived from a UK population[145] and was carried out in SPSS version 16 (the SPSS syntax is available from the EuroQoL group www.euroqol.org).

Mapping was performed from the HAQ score and DAS 28 to both the EQ-5D (for both scoring methods) and the SF-6D.

4.4.5 Mapping EQ-5D and SF-6D scores from HAQDI and DAS 28

The estimated values of SF-6D and EQ-5D obtained using a dependent variable, are termed 'mapped values'. In a study by Bansback *et al.*[164] a number of different models were used to derive the best fit for a transformation from HAQ to EQ-5D and SF-6D scores. Five models were used; the best performing models were models 5 for the EQ-5D and models 2 and 4 for the SF-6D. Each of these models included either all 42 items or each of the 8 domains of the HAQ. Since only the HAQDI was available a similar mapping to model 1 of the Bansback study was performed. A mapping function derived from an Irish cohort as described above was fitted and the results were compared with the results of the model described by Bansback *et al.* (Appendix 4).[164]

There were two mapping processes performed for this study; the first mapped final utility scores of the EQ-5D and SF-6D from the HAQDI and the second mapped the final utility score of the EQ-5D and the SF-6D from the DAS 28 score. Although many studies do include HAQ as a measure of outcome in inflammatory arthritis trials, DAS 28 is often more frequently recorded as a clinical measure of disease activity in clinical practice and is often used for treatment decisions. Both mapping processes were carried out in SPSS version 16, for which the final utility score was plotted (independent variable) against the final HAQDI (dependent variable) and in the second case against the DAS 28 score (dependent variable).

4.4.6 Statistical Models Used for Mapping

General linear models were fitted for each of the measures with HAQDI and DAS28. A linear regression was performed for both, in order to derive a regression equation that could be used to calculate a mapped utility value. For the regression analysis a regression line is fitted to bivariate plots of final utility (EQ-5D and SF-6D) against final HAQDI and DAS 28 score. Quadratic and higher dimensional models were also examined. For each of the regression models standard errors, 95% confidence intervals and R^2 are shown.

Descriptive statistics were used to describe the baseline demographics; mean values, range, and standard deviation (SD) are given. A paired sample t-test was used to compare the mean utility at baseline and at follow-up and the mean change measured by the original EQ-5D UK TTO, the revised EQ-5 D UK, and the SF-6D. This approach was taken as it has been demonstrated that parametric techniques are robust to violation of the normality assumption which is common to many QOL outcomes which have discrete, bounded and skewed distributions. [194] Confidence intervals (95%) are presented around the change in utility. Statistical analysis was completed using SPSS Version 16.

4.5 Results

4.5.1 Demography of the population of interest

At baseline, the mean age at inclusion was 54 years for the RA cohort and 45 years for the PsA group (Table 7). The average disease duration was similar (RA=12 years, PsA=11 years). The mean DAS 28 score was higher in the RA group (5.39 [95% CI 5.16, 5.43]) than in the PsA group (4.91 [95% CI 4.65, 5.05]), as was the mean HAQDI (RA 1.3 [95% CI 1.26, 1.46] vs PsA 0.96 [95% CI 0.81, 1.08]). (Note: The DAS 28 is not validated to be used for PsA patients and therefore these results should be considered within this limitation).

Table 7. Baseline demographics for inflammatory arthritis cohort

Characteristic	RA (n=345)	PsA (n=159)
	Mean ± SD (range)	*Mean ± SD (range)*
Female (%)	245 (71%)	82 (52%)
Age at inclusion (yr)	54 ± 12.9 (*17,85*)	45 ± 12.8 (*15, 77*)
Duration of disease (yr)	12 ± 9.4 (*0,42*)	11 ± 10.1 (*0, 45*)
ESR	35 ± 25.8 (*2, 140*)	22 ± 21.1 (*1, 120*)
CRP	29 ± 29.5 (*2, 158*)	18 ± 22.7 (*0, 149*)
DAS 28 CRP	5.39 ± 1.18 (*1, 9*)	1.0 (*1, 7*)
Patient Global Assessment (10cm VAS)	6 ± 2.3 (*0, 10*)	5 ± 2.3 (*0, 10*)
Pain (10cm VAS)	6 ± 2.3 (*0, 10*)	5 ± 2.3 (*0, 10*)
Tender Joint Count (Range 0-28)	5 ± 2.3 (*0, 28*)	8 ± 6 (*0, 28*)
Swollen Joint Count (Range 0-28)	10 ± 6.6 (*0, 25*)	7 ±6 (*0, 28*)
Fatigue (10cm VAS)	6 ± 2.4 (*0, 10*)	6 ± 2.6 (*0, 10*)
Tender Joint Count (Range 0-66)	10 ± 6.0 (*1, 10*)	12 ± 9 (*0, 43*)
Concomitant MTX (n)	220 (64%)	56 (35%)
Previous DMARDs (n)	292*	118**
HAQDI (0-3)	1.3 ± 0.7 (*0, 3*)	0.96 ± 0.7 (*0, 2.5*)
SF-36 PCS (0-100)	30± 8.5 (*12, 57*)	34 ± 9.5 (*13, 58*)
SF-36 MCS (0-100)	45 ± 10.4 (*17, 72*)	46 ± 12.2 (*20, 66*)
SF-6D utility	0.54 ± 0.09 (*0.3, 0.7*)	0.57 ± 0.12 (*0.25, 0.80*)
EQ-5D UK TTO utility	0.43 ± 0.32 (*-0.43, 1.0*)	0.53 ± 0.32 (*-0.24, 1.0*)
Revised EQ-5D UK TTO utility	0.576 ± 0.22 (*-0.14, 0.9954*)	0.638 ± 0.19 (*0.046, 0.9954*)

*a Unless otherwise indicated **CRP** = C-reactive protein; **DAS 28** = Disease Activity Score (28 joint); **DMARDs** = disease-modifying anti-rheumatic drugs; **ESR** = erythrocyte sedimentation rate; **HAQ** = Health Assessment Questionnaire; **HAQDI** = Health Assessment Questionnaire Disease Index; **MCS** = mental component summary; **PCS** = physical component summary; **PsA** = psoriatic arthritis; **RA** = rheumatoid arthritis; **VAS** = visual analogue scale; * indicates missing data (n = 18 patients); ** indicates missing data (n = 4).*

4.5.2 Utility Scores

The mean utility scores and standard deviations (SD) are provided for each of the three methods (Table 8). At baseline the SF-6D scores are higher than those calculated using the EQ-5D indicating that the SF-6D is scoring patients who are in the worst health state at a better health utility (Table 8). The mean difference seen before and after biologic treatment is greater with the EQ-5D values than with the SF-6D partly because of this anomaly seen at baseline but it is likely that scoring of the EQ-5D may also be contributing to this. This greater change is evident for both RA and PsA.

When the revised scoring method was used, the overall change was less for the revised EQ-5D scoring than the original EQ-5D (TTO) but greater than the SF-6D in the RA group (Table 8). The change was greater in the PsA group across all three methods and a similar trend between the scoring methods as that in the RA group; greatest change was produced when using the EQ-5D, less so with the revised method and considerably less so with the SF-6D. Overall PsA appears to result in a similar decrement in utility to RA.

Table 8. Mean Utility Scores at baseline and follow up

	RA			**PsA**		
	Baseline mean ± SD (range)	**12 month mean ± SD (range)**	**change in utility (95% CI)**	**Baseline mean ± SD (range)**	**12 month mean ± SD (range)**	**change in utility (95% CI)**
SF-6D	0.54 ± 0.09 (0.29, 0.75)	0.62 ± 0.077 (0.44, 0.83)	0.08 (0.049,0.106)	0.57 ± 0.12 (0.25, 0.79)	0.66 ± 0.12 (0.32, 0.89)	0.09 (0.123,0.052)
EQ-5D	0.43 ± 0.32 (−0.43, 1.0)	0.65 ± 0.28 (−0.18, 1.0)	0.22 (0.145,0.302)	0.49 ± 0.32 (−0.24, 1.0)	0.77 ± 0.28 (−0.24, 1.0)	0.28 0.360,0.200)
Revised EQ-5D	0.56 ± 0.22 (-0.14,0.9954)	0.72 ± 0.2 (0.88,0.9954)	0.16 (0.102,0.214)	0.62 ± 0.21 (-0.14,0.9954)	0.84 ± 0.17 (0.046, 0.9954)	0.22 (0.281,0.167)

States Worse than Death

Seventeen per cent of the RA group has utilities defined as WTD at baseline and 7% at follow up. There are fewer patients falling within this utility in the PsA group, 12% at baseline, and 2% at follow up. Figure 14 compares the EQ-5D and the SF-6D at baseline and follow-up. Patients scoring pain as extreme, thereby being assigned the added weighting of what is termed the N3 constant of the EQ-5D scoring system, are strongly associated with states WTD. The other scales (mobility, anxiety and depression, self-care and usual activity) show less of an association. The change from baseline to follow up shows an improvement in pain score and therefore utilities overall.

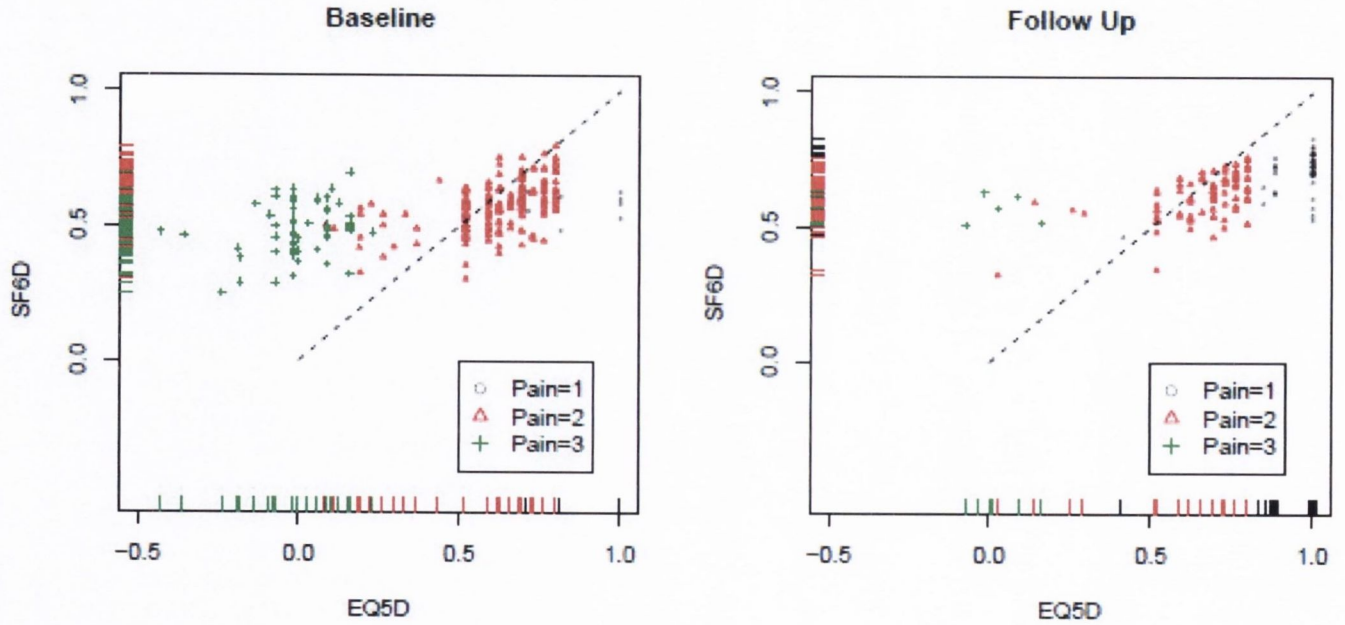


Figure 14: **Observed SF-6D and EQ-5D at baseline and follow-up.**

The graphs show the observed association between SF-6D derived utilities and EQ-5D derived utilities. The plotted lines are lines of equality, and thus the fact that the points lie mainly above the line shows that SF-6D derived scores are typically larger than EQ-5D for those with poor QOL. Overlaid on these plots are the pain scores from the EQ-5D as described in the legend. The tick marks on the axes show the marginal distribution for each measure.

4.5.3 Distribution of the utility scores for different measures

We present the distribution of the utility score for each instrument for both RA and PsA (Figure 15). The distribution of the utility score produced by the two methods of scoring for the EQ-5D differs. The marginal distribution of each of these measures is provided in Figure 16. The distribution of the scores was narrower with the revised scoring system (range -0.143, 0.995) than the original EQ-5D scoring method (range -0.429, 1.0) in this cohort. Both disease groups display a similar distribution for both the EQ-5D and the SF-6D. It is clear that the discrete jump in the middle of the EQ-5D measured range is present for both disease groups.

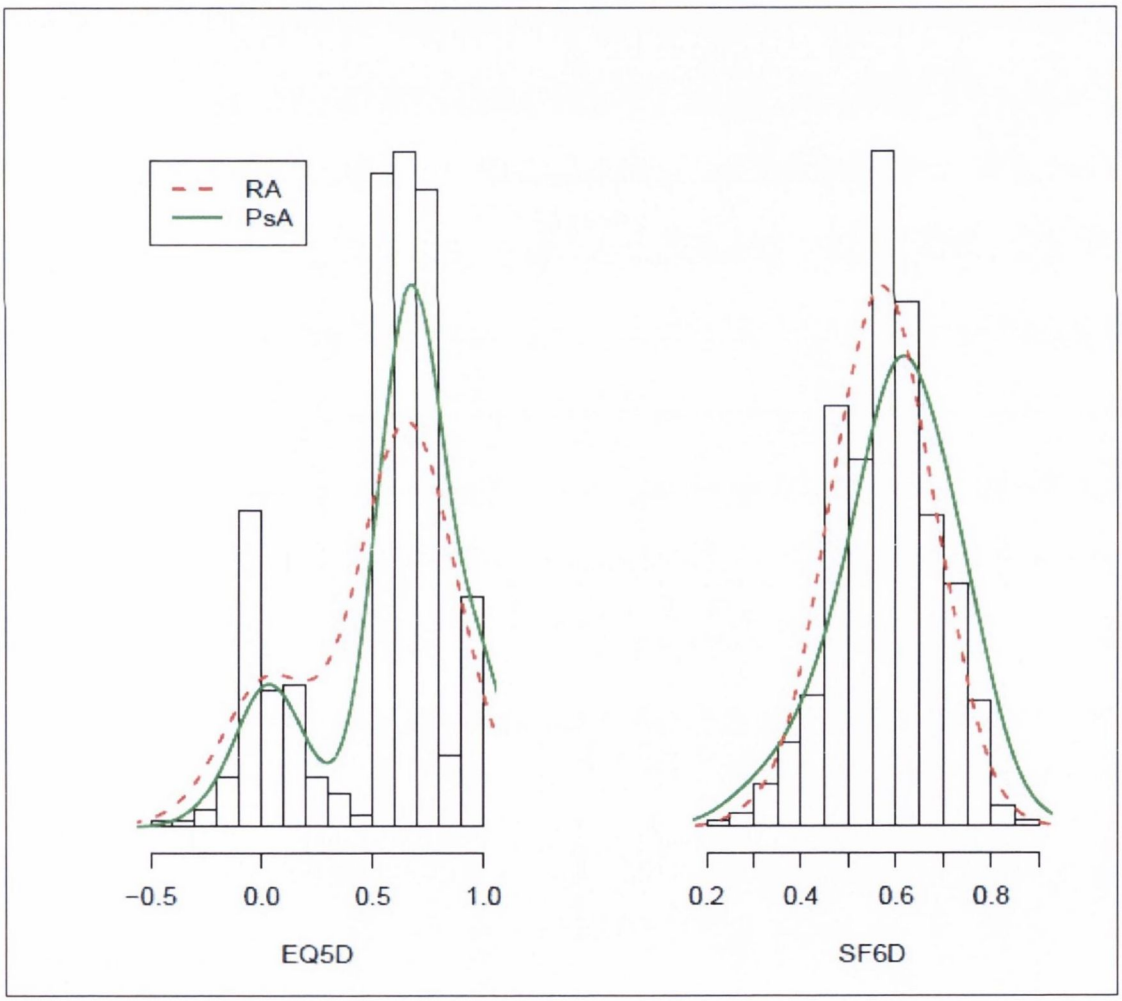


Figure 15: Histogram showing the distribution of EQ-5D and SF-6D measures overall.
Density plots for RA and PsA are overlaid. The overall distribution of utility for both disease groups is similar.

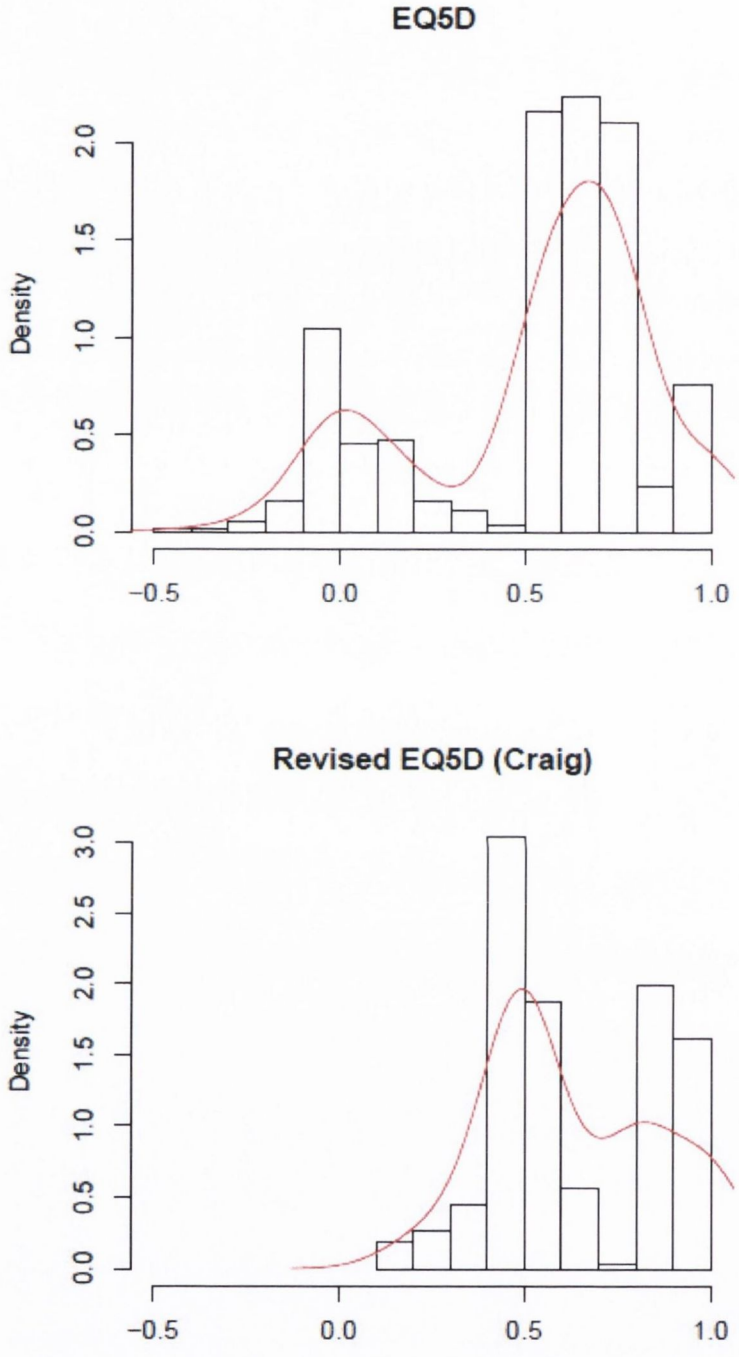


Figure 16: Histograms of the EQ-5D and revised EQ-5D
Plot shows the marginal distribution for each of these measures for this cohort. Of note in the revised version is the impact on individuals with values less than 0.

4.5.3.2 Mapping

In order to describe the relationship between the measures, regression lines are fitted. The estimates of the coefficients (B) of the regression obtained from this cohort are shown in Table 9 and the equations are presented in Appendix 5.

Individuals in states WTD have a strong influence on the best fit line with the resulting slope being more severe than that for SF-6D. Thus for the RA group, a one point reduction in HAQDI is associated with a 0.24 QALY gain per annum if using EQ-5D, but only a 0.08 QALY gain per annum using SF-6D (Table 9). In RA a one point reduction in DAS 28 is associated with a 0.084 QALY gain per annum if using EQ-5D, but only a 0.029 QALY gain per annum using SF-6D (Table 9).

In order to investigate the differences between the Irish population and a similar study[164], the best fit regression lines were plotted for comparable models on the same graph (Figure 17). The intercepts from the lines shown differ, which is compatible with heterogeneity between the studies (with the Irish cohort having lower utility overall). When considering each of the instruments separately, the slopes are similar for both studies, and it is this that is relevant when examining how a treatment impacts on change in utility as derived from HAQDI. Thus the studies are in agreement that EQ-5D derived utilities are likely to produce larger QALY gains when compared with SF-6D derived utilities for a given change in the disease specific measure (HAQDI).

Table 9. Linear Regression Equations for HAQ in RA and PsA

	RA						PsA					
	SF-6D			EQ-5D			SF-6D			EQ-5D		
	B	SE	P-value	B	SE	P-value	B	SE	P-value	B	SE	P-value
HAQ												
HAQ Index	-0.084	0.01	<0.001	-0.24	0.02	<0.01	-0.1	0.01	<0.01	-0.267	0.03	<0.01
Constant	0.669	0.01	<0.01	0.79	0.03	<0.01	0.68	0.01	<0.01	0.809	0.03	<0.01
R^2	0.395			0.27			0.284			0.333		
DAS												
DAS Index	-0.029	0.02	<0.01	-0.084	0.05	<0.01	-0.037	0.03	<0.01	-0.109	0.06	<0.01
Constant	0.704	0.003	<0.01	0.887	0.01	<0.01	0.748	0.01	<0.01	1.06	0.01	<0.01
R^2	0.237			0.196			0.263			0.333		
B= Beta coefficient, SE= Standard Error												

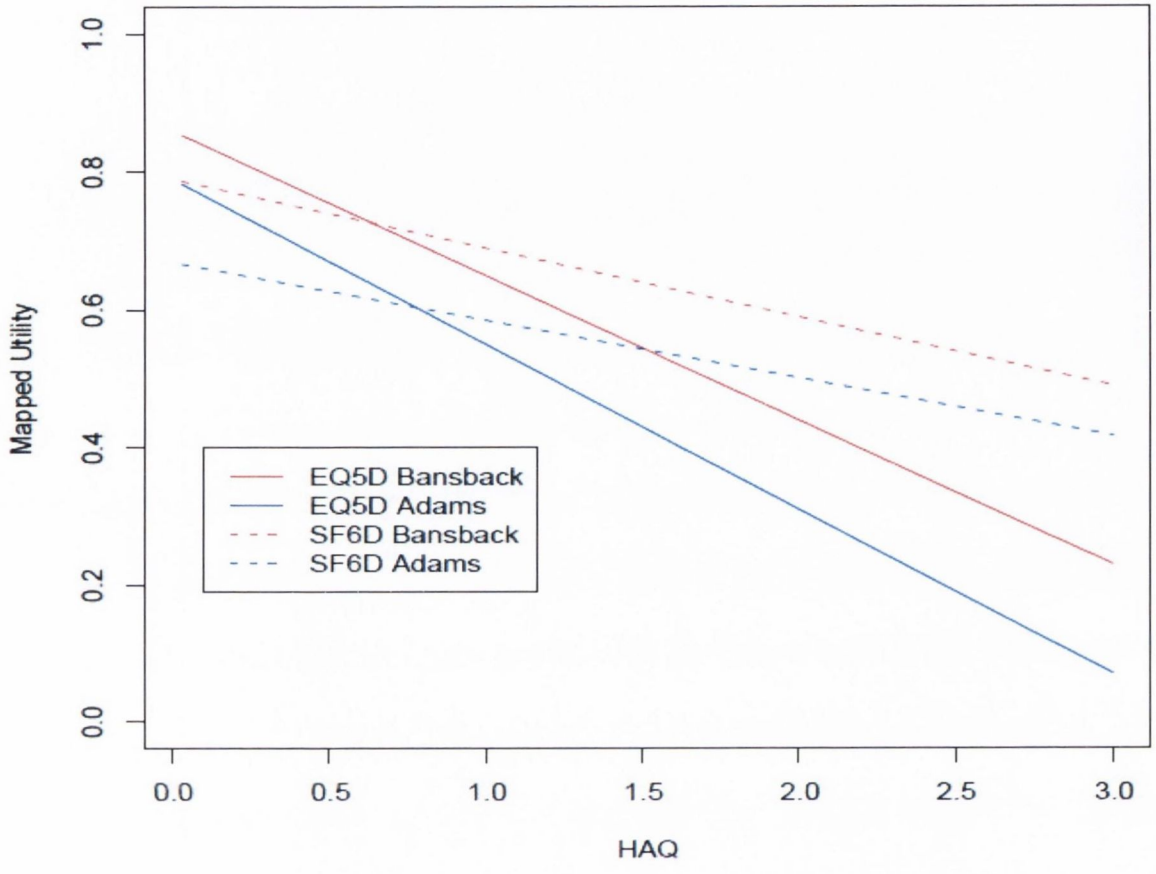


Figure 17: Comparison between the fitted lines associating the Health Assessment Questionnaire Disability Index (HAQDI) with SF- 6D and EQ-5D-derived utilities from this study and from Bansback *et al.* [164]

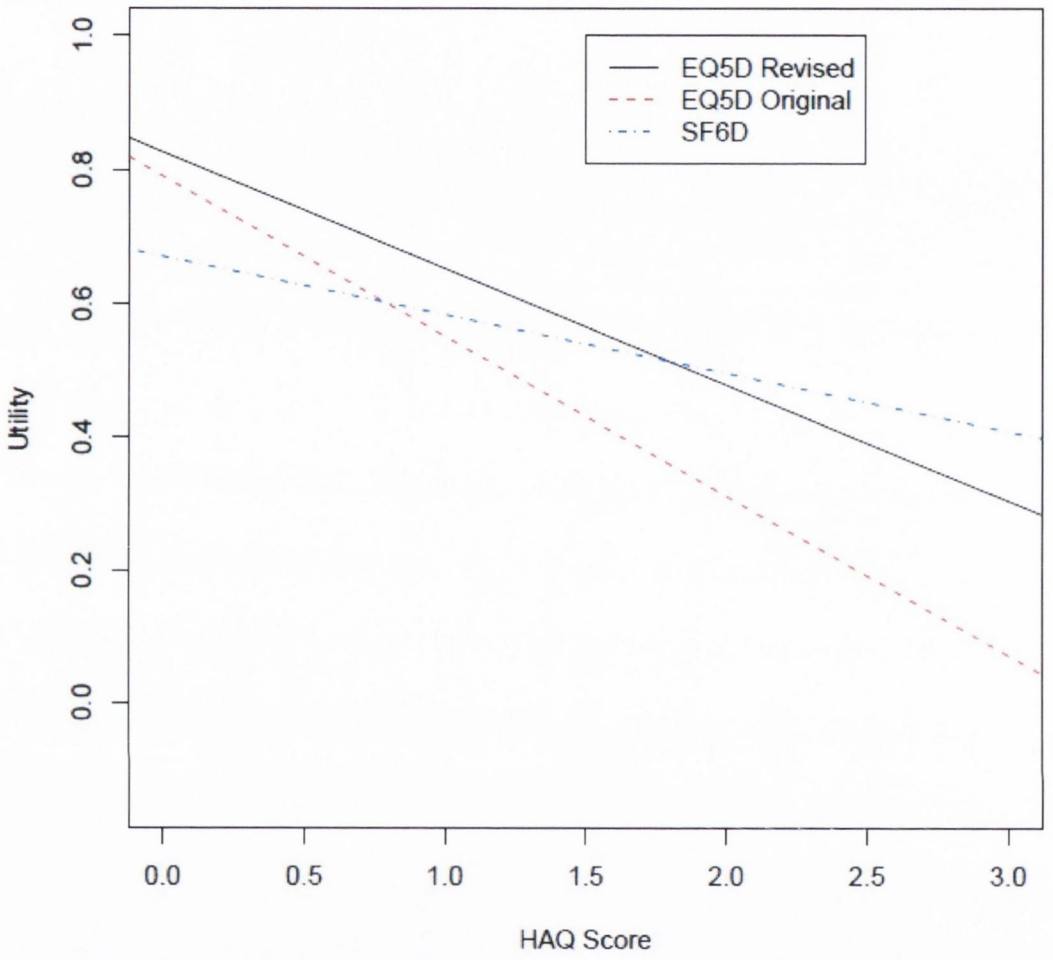


Figure 18: Comparison between the fitted lines associating mapped utility from the revised EQ-5D UK scoring, original EQ-5D UK and the SF-6D with the HAQDI score.

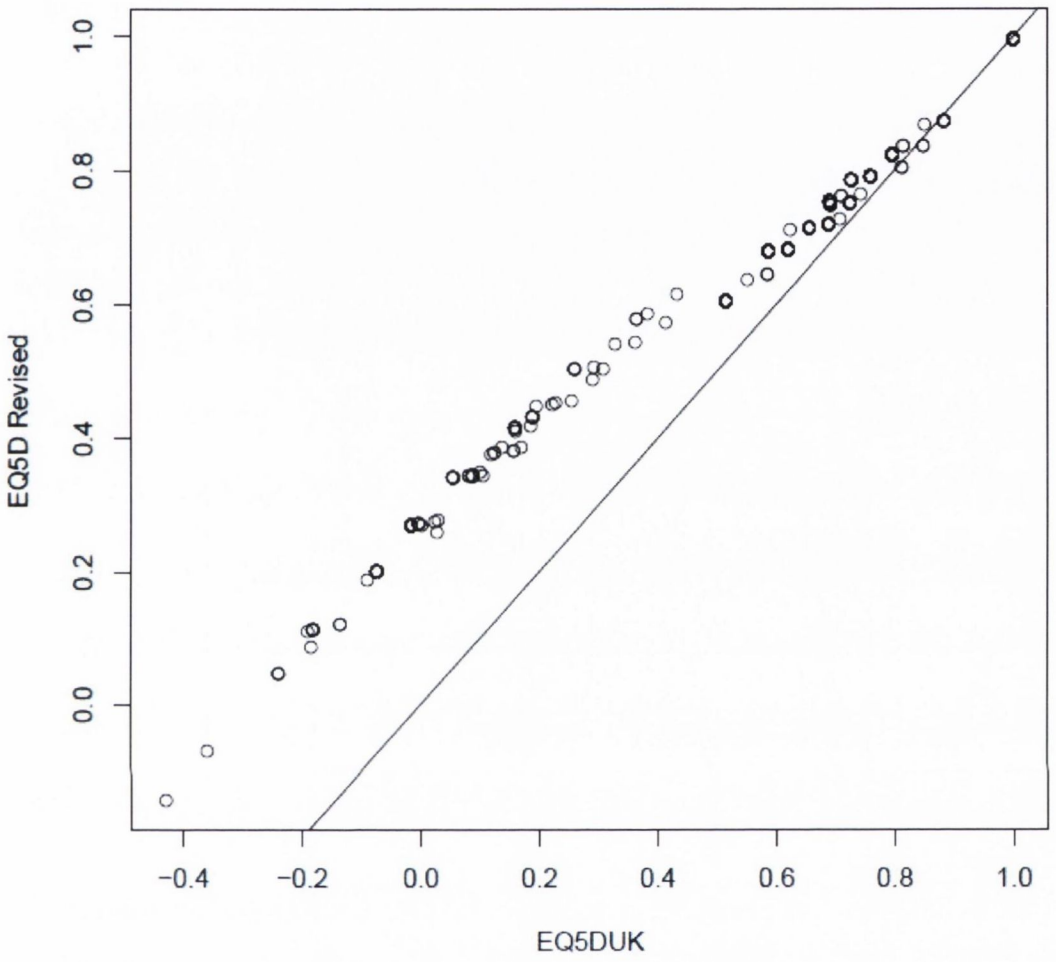


Figure 19: Scatter plot of utilities derived from the revised EQ-5D scoring and original EQ-5D scoring. *A line of equality is fitted.*

The revised scoring for the EQ-5D lessens the gap between the SF-6D and the original EQ-5D (Figure 18). The slope produced by the relationship between the HAQDI and the revised scoring is less steep than that produced by the HAQDI with the original scoring.

The results of both measures are plotted on a scatter plot. The line is the line of equality. The methods produce different results to the original method for utility scores less than 0.5. The magnitude of this difference is approximately 0.25 for scores less than 0.5 (Figure 19).

4.5.3.3 Non-Linear Mapping

In order to establish the most appropriate mapping to use as regards linear or non linear, both models were tested. The results show little difference between the two models and the quadratic term was not statistically significant for any of the utility measures (P-values EQ-5D(UK) vs. HAQDI p=0.982, SF-6D vs. HAQDI p=0.574 and for the revised EQ-5D vs. HAQDI p= 0.906) Figure 19-21)). Therefore the more parsimonious linear mapping was used in the economic model although both are again tested. These results are compatible with those of Lillegraven *et al.* from a Swedish observational cohort (n=1041).[195]

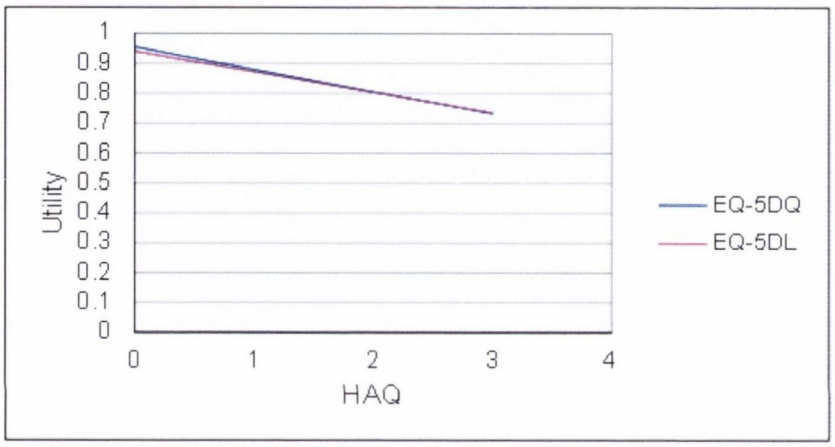


Figure 20: Fitted lines for the non-linear and linear models for the EQ-5D (original).
EQ-5D(UK) Q – quadratic model, EQ-5D (UK) L – Linear model

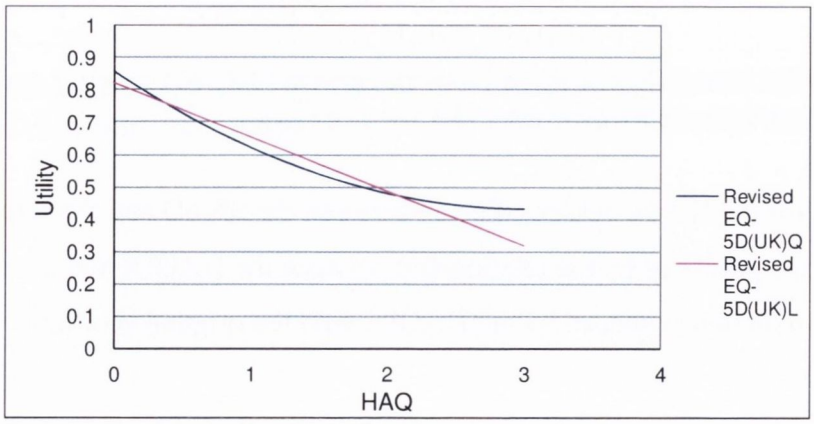


Figure 21: Fitted lines for the non-linear and linear models for EQ-5D (Revised) and HAQDI.
Revised EQ-5D(UK) Q – quadratic model, Revised EQ-5D (UK) L – Linear model

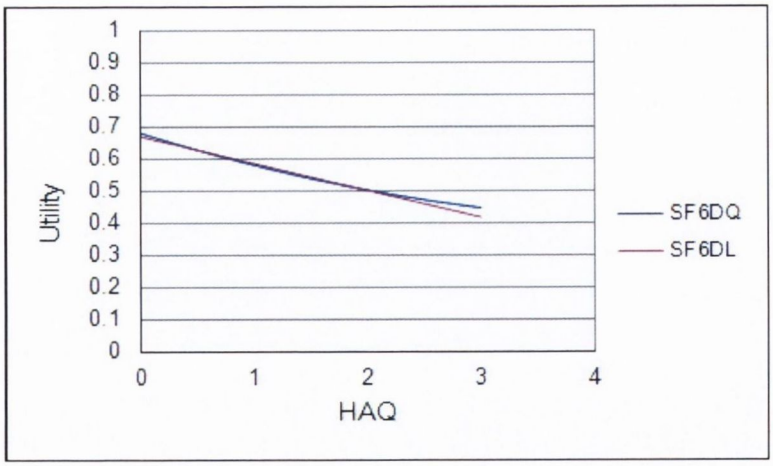


Figure 22: Fitted lines for the non-linear and linear models for SF-6D and HAQDI.
SF-6DQ – quadratic model, SF-6DL – Linear model

Table 10. Mapping Coefficients for EQ-5D (original) against HAQDI

EQ-5D Quadratic					
		95% CI	**P-value**	**Std Error**	**R^2**
Coefficient	0.806	(0.730,0.881)	<0.001	0.038	
HAQ	-0.271	(-0.404,-0.137)	<0.001	0.068	
HAQ^2	0.014	(-0.039, 0.068)	0.595	0.027	0.267
EQ-5D Linear					
Coefficient	0.792	(0.736,0.848)	<0.001	0.028	
HAQ	-0.236	(-0.277,-0.196)	<0.001	0.02	0.27

Table 11 Mapping Coefficients for SF-6D and HAQDI

SF-6D Quadratic					
		95% CI	**P-value**	**Std Error**	**R^2**
Coefficient	0.68	(0.657,0.702)	<0.001	0.012	
HAQ	-0.111	(-0.151,-0.07)	<0.001	0.021	
HAQ^2	0.011	(-0.005,0.027)	0.175	0.008	0.399
SF-6D Linear					
Coefficient	0.669	(0.652,0.686)	<0.001	0.009	
HAQ	-0.084	(-0.097,-0.072)	<0.001	0.006	0.395

Table 12. Mapping Coefficients for EQ-5D (Revised) and HAQDI

EQ-5D (Revised) Quadratic					
		95% CI	**P-value**	**Std Error**	**R^2**
Coefficient	0.858	(0.809,0.908)	<0.001	0.025	
HAQ	-0.283	(-0.408,-0.158)	<0.001	0.063	
HAQ^2	0.047	(-0.015,0.109)	0.906	0.031	0.359
EQ-5D (Revised) Linear					
Coefficient	0.822	(0.783,0.861)	<0.001	0.02	
HAQ	-0.168	(-0.196,-0.140)	<0.001	0.014	0.28

4.6 Discussion

One of the more pertinent results is the quantity of change demonstrated by each method following treatment. The greatest change in utility in this cohort is recorded by the EQ-5D using the original TTO scoring method. This is significant for researchers carrying out CUA and for decision makers. Many HTA agencies recommend the use of a preference-based measure such as the EQ-5D or the SF-6D in the calculation of utilities. [25, 26, 196, 197] However, as our results indicate, use of either of these instruments will give quite different results. A CUA using the EQ-5D in RA is likely to produce approximately twice the QALY gain for a given change in HAQ score when compared with a CUA using the SF-6D as its outcome measure. This difference is consistent across the UK and Irish populations and different disease groups. The implications of this difference between measures was also demonstrated in an analysis on the British Society of Rheumatology Biologics Register (BSRBR), where the incremental QALY gained by TNF-inhibitor therapy is almost halved if the SF-6D-derived utility is used rather than the EQ-5D instrument.[20] The pharmacoeconomic model was very sensitive to the measure used. A number of recent articles also discuss the problems associated with the EQ-5D health states.[160, 174, 198]

Previous studies have concluded that the EQ-5D is more responsive to deterioration than the SF-6D and that the SF-6D is more responsive to improvement.[154, 170, 199] One of the key issues highlighted is where utilities are valued less than 0 or WTD. The driver of this negative utility or states WTD is when pain scores are valued at their worst state. While other studies have found that there were other components influencing this WTD state,[174] scoring the other components of the EQ-5D (e.g.
anxiety/depression, mobility, usual activities, and self-care) in this cohort, at the most severe state did not result in negative utilities.

The differences in utility scores between the EQ-5D and the SF-6D are not surprising considering the different descriptive systems of the instruments. [200] According to the 1996 Health Survey of England, perfect health is largely absent from the SF-6D system and prevalent in the EQ-5D system.[201] The instruments have different periods of recall; the EQ-5D asks respondents to indicate their responses on the day of completion whereas the SF-6D standard version period of recall is in the last 4 weeks. In examining the utility associated with experiencing an adverse effect, using a recall period of 7 days compared to that over 1 day,

produced lower utilities. [202] The question format of the measures also differs considerably; the response options for the EQ-5D for a particular dimension of health are based on the current disease state severity. The SF-6D takes an alternative approach by asking respondents to consider duration of impairment for a certain time period or the extent which the impairment restricts certain activities. [200]

The lack of concordance between the two main generic QOL instruments: the EQ-5D and the SF-6D may introduce heterogeneity into economic analysis because they produce quite different results in QALY estimation. [154, 165, 170, 203]

A striking feature of this Irish cohort is that 17% of the RA patients have WTD utilities prior to starting biological therapy. This is double what has been reported in a recent UK-based cohort. [174] Because the demography of the cohorts was very similar, the most plausible reason for this was the use of UK population values for the Irish cohort which questions the face validity of the utility estimates. We have used the UK population-based scoring system to value the utility in this Irish cohort and this may not be ideal. However, there are no population values available for Ireland. At present, the EQ-5D has several national value sets or tariffs. Nevertheless, utility estimates from foreign studies are often used directly for cost-effectiveness estimates, without applying the appropriate national value set. It is unclear if this practice is advisable, due to dissimilarities between the national value sets. [204]

In the NCPE most of the economic evaluations using EQ-5D to date have been completed using the UK valuation set, as a result of the absence of an Irish value set. However a number of papers have highlighted the problems with using other population values or other societal preferences. [204, 205] A key consideration in regard to any generic HRQoL measure used for such social decision-making is the origin of its weighting system. It is generally accepted that when generic HRQoL measures are used in the computation of QALYs, then the values applied should represent the social preferences of the relevant population. [25, 206] Strictly interpreted, this requirement means that valuation systems developed in other countries should be calibrated in terms of the domestic population before being applied by decision-making agencies. The general problem faced in the majority of countries where social preferences are required for cost-effectiveness analysis is the absence of a value set based on domestic data sources.

The anomalies associated with the population scoring for the EQ-5D TTO have been discussed earlier in this chapter. The distribution of scores observed when using the revised EQ-5D scoring method is narrower than the original method. The revised scoring method uses an episodic RUM which, unlike the instant RUM, does not inflate the error of responses. This in turn lessens the influence on the mean preference score and therefore does not pull down the estimates as with the conventional approach.[167] The lowest score in this cohort with the revised method is -0.143 and with the original method is -0.43. A reasonable assumption is that the true estimate of HRQoL (via utility) may lie somewhere between both measures. We can see that the revised scoring produces an estimate that lies between the slope of the original EQ-5D and the SF-6D (Figure 18) and produces a change that is less than the original EQ-5D and more than the SF-6D (Table 8). This new scoring method is useful in corroborating some of the reasons for the discrepancies in the scoring of the EQ-5D that have been presented by other authors[167]. Dolan *et al.* replaced the negative slopes with –x/10 while Shaw *et al.* (US valuation) divided the negative slopes by a constant (i.e., 39). [147, 178] The episodic RUM reduces dependence on these arbitrary adjustments that have been made to deal with WTD valuations and provides a more robust coefficient estimator.

A further method that may deal with the issue of WTD states is the use of a lead-in time. The use of a lead-in time was introduced to overcome the discontinuity in values around zero evident in conventional methods. However, more research is required to understand the implications for states WTD and how to handle those who use up their lead-in time. [207]

In CUA, the QALY, in combining utility and efficacy represents our measure of effect. The range of ICERs estimated using three different methods of utility measurement highlights the impact that utility has on the overall result. NICE recommends the use of the EQ-5D (either directly measured using questionnaires or via mapping) as a pragmatic approach to deal with this heterogeneity.[25] However this raises the possibility that the heterogeneity surrounding utility may not be fully explored. The current method of PSA will only explore uncertainty within the limits of the instrument measured. It may be more appropriate to extend the range of uncertainty beyond the statistical limits of one instrument alone. While it is feasible to do this within a PSA, refitting a cost-effectiveness model using multiple metrics and producing a range of ICER estimates would be useful for decision makers.

Limitations

This analysis was carried out on a rheumatology cohort within a specific geographical area and therefore patient heterogeneity is not likely to be included within the cohort. This does present limitations for generalisability across a wider cohort of patients. Furthermore, the baseline scores presented here may be considerably worse than the general RA or PsA population, as this cohort represent a group of patients for whom conventional DMARDs have not been successful or who have a particularly aggressive form of the disease, as demonstrated by laboratory and disease activity scores. The utility values in this Irish cohort are lower than those seen in other studies [164, 174] and therefore results may not be generalisable across different cohorts, an issue highlighted by Harrison *et al.*[174]

The use of the episodic RUM offers an alternative approach; however there remain some limitations with the method. The combination of TTO and rank estimates merges responses from two different valuation techniques in a single estimate. It is not clear whether the variance observed with the rank responses is equal to that of TTO. It could be argued that TTO may place a larger cognitive burden on individuals than rank and therefore may result in greater errors. The combined estimator used for this rescoring includes a separate variance parameter for rank responses which describes differences between the method-specific variances.

A recent paper provides a critique on the episodic RUM as an alternative method by examining the non-monotonicity associated with the model. [208] Menzies *et al.* provides discussion on the problems that may arise as a result of using the episodic RUM. [208] The main difference between the instant RUM and the episodic RUM is the conceptualisation of the error term in the model. The limitation of the episodic RUM, according to Menzies and Salomon may be that it violates monotonicity. [208] Monotonicity is a desirable feature for an aggregate measure of social preferences. If respondents in health state valuations improve their valuation of a health state, the aggregate measure should not decline as a result. If this happens then the method displays non-monotonicity. In the case of the EQ-5D TTO, if a respondent scores an excessively low (WTD) score, the mean estimate actually gets higher. The authors estimated the percentage of observations in the Dolan MVH study[146], which failed to meet the conditions for monotonicity. In doing so it was found that the method proposed by Craig *et al.* violates monotonicity in 27% of cases. The effect that this has on the results of the population

scores may be that the mean estimator is underweighting the effect of the WTD states. This may be significant in this cohort considering the high proportion of WTD states.

The area of most concern is the WTD states. The proportion of states considered WTD by the original UK EQ-5D values is much higher in this population than in others[209] (17% WTD at baseline in the Irish cohort). Although the methodology for this rescoring is relatively new, the revised results are more concordant with the SF-6D predictions, suggesting convergent validity. While we can see how the approach changes the results in this cohort, it would be of interest to examine the effect across a number of different geographical populations and for different diseases.
The SF-6D valuation study examined standard gamble responses without WTD responses, except for one state (i.e. pits). The worse imaginable health state, "pits" was valued in a similar manner to the valuations in the Dolan paper.[147] In future research, it would be prudent to rescore the SF-6D values and compare the difference in results overall.

One of the main challenges for those carrying out CUA is deriving utilities where there are no generic QOL measures available. While mapping has provided a way around this, our results highlight the anomalies associated with this technique. Mapping the HAQDI onto the SF-6D does not produce as large a utility gain for a given improvement in disease status as the EQ-5D. The main explanation for this lies in the instrument's construction, rather than the mapping technique used. The EQ-5D displays a bimodal distribution in more severe health states in both RA and PsA.

Figure 13 highlights the pain component of the EQ-5D as the particular variable that produces the scatter below 0. The discontinuous jump observed is not exclusive to the relationship between the HAQDI and EQ-5D; this is also observed between the DAS and the EQ-5D. Therefore it is reasonable to say that this artefact is stemming from the instrument to which these measures are mapped, the EQ-5D, and not the instruments being mapped (DAS and HAQ).

Previous studies have derived utilities from HAQDI scores in order to facilitate CUA in the absence of directly derived utilities such as the SF-6D or the EQ-5D. [38, 39, 163, 210, 211] The results presented here and elsewhere indicate a number of limitations of this technique.[164, 212] The transformation completed here is a simple linear regression where the EQ-5D and the SF-

6D are inputted against the HAQDI. However, a challenge of this approach is that only the impact of the treatment on functional disability is captured, and not the psychological or pain components associated with the disease. A more sophisticated method of mapping the generic QOL measures would be to input all 42 items of the HAQ with the EQ-5D and the SF-6D as described by Bansback *et al.*[164] Deriving a mapping function for DAS 28 to the EQ-5D and SF-6D may enable us to derive utility values for studies which have not used HAQ or generic QoL measures as outcomes. While the mapping from utility onto HAQ has been done in the past, mapping onto the DAS 28 is a new method. [163, 164] The impact of these different methods for calculating utility on CUA has recently been discussed. [38, 111] Barton *et al.* found that the mean QALY gains produced by a mapping technique generated consistently lower results than actual observed QALY gains. [111] For these reasons, the practice of generating utilities from either the HAQDI or the DAS 28 should not be regarded as a substitute for using actual scores from generic preference-based utility measures such as the SF-6D or the EQ-5D. The use of different statistical models has been discussed in the introduction to this chapter. For this study two different approaches have been taken; mapping using a simple linear regression model and incorporating a quadratic term for the non-linear model. It has been proposed that the relationship between the HAQ and utility score is best described using a non-linear model and this has recently been incorporated into a model examining the cost-effectiveness of anti-TNF agents in the UK. [213] Both methods have been examined for three different utility estimates. The estimates for the non-linear model produce positive coefficients for the quadratic term, resulting in downward curved slopes which become less steep as HAQ increases. The quadratic coefficient for the revised scoring is more positive than that of both the SF-6D and the EQ-5D. In this case, there is a risk that if the coefficients fall one standard deviation above their mean values that the turn in the curve could happen within the range of the HAQ score. In this case the utility could actually increase as HAQ deteriorates further. This anomaly occurs due to the uncertainty associated with the direction of the curve and is most likely due to insufficient data to populate the mapping model.

4.7 Conclusion

The analyses in this chapter indicate that there are considerable flaws in using the EQ-5D for patients with severe RA. The UK population scoring methodology is one of the main reasons for this. In choosing an alternative method there are currently two choices; either the EQ-5D using the revised scoring methodology or using the SF-6D. While both of these methods still have shortcomings they offer an approach which at least displays face validity.

The methods described in this chapter may offer alternatives in the absence of a generic utility measure, as may be the case in older studies. In the future, a key question should be how best to handle the level of uncertainty around the estimates produced by these instruments, and further research may be required to reduce this level of uncertainty. In choosing just one QOL measure to produce a single ICER estimate we may be restricting our ability to fully explore the uncertainty within the final estimate of a cost-effectiveness analysis. The impact of decision uncertainty and the application of EVPI methods may allow consideration of whether further evidence about a parameter is needed. [32]

Decision makers using utility measures should be aware of the impact that the instrument and its scoring has on the estimate of cost-effectiveness. In the context of inflammatory arthritis the work presented here demonstrates that choice of utility measure may have a significant effect on this estimate which may therefore impact on the reimbursement decision.

4.8 Future Developments in HRQoL

Considerable debate has taken place as regards the most suitable measure to use, the most appropriate methods for eliciting preferences and feasibility of instruments; in light of this a number of developments are ongoing.

The EQ-5D-5L instrument has been developed by the EuroQOL group and the population scoring is currently in the pilot stages. A cross walk from the 3 level instrument to the 5 level instrument has also been published and allows the 5 level to be used straight away rather than await the population scoring studies which will not be available for 2 to 3 years. The 5 level was developed to reduce the ceiling effects associated with the EQ-5D.

Different techniques are being explored for preference elicitation; discrete choice modeling and conjoint analysis are the current areas of interest in this field. Discrete choice experiments (DCE) are set within a strong theoretical measurement framework and offer relatively simple judgment tasks. [181] Discrete choice valuation was compared to TTO and it was found that the DCE method produced higher values than the TTO. The pattern of response was broadly similar to the TTO and because of this comparability; the authors suggested that DCE could be substituted for TTO valuation. [181] However, there are some problems with DCE and these have also been discussed in the literature. [214, 215] DCEs produce a valuation based on a worst and best health state rather than a full health-dead scale which is used for QALY estimation.

Conjoint analysis presents hypothetical scenarios to individuals. Preferences for these scenarios are elicited from respondents using ranking, rating, or DCE. These responses can then be used to determine utilities. It can be useful in deciding what factors influence a patient or decision maker's preferences and for valuing the different attributes of a preference. [216]

A new measurement tool has been developed in the United States, under a project called the Patient Reported Outcomes Measurement Information System (PROMIS). This project was funded by the National Institutes of Health (NIH) and aims to develop a measure which would provide access to clinicians, patients and researchers to efficient, precise, valid, and responsive adult– and child–reported measures of health and well–being. The instruments are generated from item banks and measure concepts such as pain, fatigue, physical function, depression, anxiety, and social function. The studies are ongoing with the instrument and a population scoring has not yet been completed. [217]

Chapter 5

Chapter 5.0 Clinical Effectiveness of anti-TNF Drugs

In order to establish the clinical effectiveness of anti-TNF agents in RA, all available evidence, including RCT data, open label trial data and observational data, was reviewed for established RA. This chapter will describe how the evidence synthesis was performed to calculate estimates for overall efficacy and adverse effects of one anti-TNF against another. It will also describe methods and results used to combine long and short term discontinuation data for anti-TNF agents.

5.1 Efficacy of anti-TNF agents in RA from RCT

All of the published trials discussed here have achieved their primary outcomes. All used ACR outcomes to measure response to treatment. A general trend seen across the RCTs is response rates for ACR 20 of approximately 60%, ACR 50 40% and ACR 70 20%; placebo response is approximately 20% for ACR 20. The individual trial data for each of the anti-TNF agents is detailed.

5.1.1 Adalimumab

Adalimumab (Humira®) is a human-derived recombinant IgG1 monoclonal antibody engineered by gene technology. It binds to TNF-α and has a half-life of approximately 2 weeks. It was approved for use by the Federal Drugs Agency (FDA) in patients with RA in December 2002. The European Medicines Agency (EMA) licensed Humira for use in RA in September 2003. Abbott Laboratories are the license holders for Humira®.

The ARMADA trial evaluated the efficacy and safety of adalimumab in patients taking concomitant MTX (N=271). [218] This was a 24 week double blind multicentre RCT comparing adalimumab 20mg every other week, 40mg every other week, 80mg every other week and placebo in patients already receiving MTX therapy. Patients were required to be

taking a stable dose of MTX (10-25mg weekly) 4 weeks before study entry and be on MTX for at least six months prior to study entry. Patients had failed at least one DMARD besides MTX. The primary outcome was ACR 20 at 24 weeks. Secondary outcomes included ACR 50, ACR 70, significant differences in QOL score and fatigue scores. The primary endpoint was reached; the treatment groups reached a statistically significant difference over the placebo group. In the placebo group 21% reached an ACR 20 response; 52% of 20mg group, 70% of the 40mg group, and 62% of the 80mg group achieved an ACR 20 response. The absolute change in HAQ score in the placebo group was -0.27, in the 20mg group -0.54, in the 40mg group was -0.62 and the 80mg group was -0.59 ($p<0.001$).

Keystone *et al.* investigated the ability of adalimumab and MTX to inhibit the progression of joint damage, reduce the signs and symptoms of RA and improve physical function in patients who previously had an inadequate response to MTX (N=619). [219] The primary efficacy outcomes were radiographic progression at week 52, ACR 20 response at week 24 and HAQ score at week 52. Radiographic scores (Total Sharp scores) and joint erosion scores showed a statistically significant improvement in the treatment group over the placebo group. ACR 20 response rates for the treatment groups, 40mg every other week and 20mg weekly, were 63.3% and 60.8% respectively versus 29.5% in the placebo group ($p<0.001$). ACR 50 response rates were 39.1% and 41% respectively versus 9.5% in the placebo group and ACR 70 response rates were 20.8% and 17.5% respectively versus 2.5% in the placebo group. The absolute changes in HAQ scores at week 52 were -0.59 and -0.61 for the treatment groups and -0.25 for the placebo group.

Van de Putte *et al.* evaluated the efficacy and safety of adalimumab monotherapy, in patients who had failed previous DMARD therapy, in a 26 week trial (n=544). [220] The primary efficacy outcome was ACR 20 and secondary efficacy endpoints included ACR 50, ACR 70, EULAR response and HAQ. Patients treated with adalimumab 20mg every other week, 20mg weekly, 40mg every other week and 40mg weekly had better response rates than those in the control group. The ACR 20 response rates were 35.8%, 39.3%, 46% and 53.4% respectively for the treatment groups versus 19.1% in the placebo control arm ($p\leq 0.01$). ACR 50 response rates were 18.9%, 20.5%, 22.1% and 35% respectively in the treatment arms versus 8.2% in the placebo arm ($p\leq 0.05$). ACR 70 response rates were also significant better than placebo; 8.5%, 9.8%, 12.4% and 18.4% respectively versus 1.8% in the placebo arm ($p\leq 0.05$). Mean HAQ improvements were -0.29, -0.39, -0.38 and -0.49 respectively versus -0.07 in the placebo group.

The CHANGE study examined the efficacy and safety of adalimumab monotherapy in patients who had failed at least one DMARD previously (N=352). [221] Patients were discontinued from MTX 28 days prior to trial entry. The primary endpoint was ACR 20 at week 24; secondary endpoints included ACR 50 and ACR 70 and HAQ. The three treatment groups compared were adalimumab 20mg, 40mg and 80mg administered every other week versus placebo. ACR 20 response rates at 24 weeks were 28.7%, 44%, and 50.6% versus 12.6% respectively. ACR 50 response rates were 16.1%, 24.2%, and 32.2% versus 5.7% respectively and ACR 70 response rates were 10.3%, 12.1%, and 14.9% versus 1.1% respectively. The HAQ scores for the 40mg and 80mg groups did not show a statistically significant change from placebo; changes in HAQ scores for the treatment groups were -0.2, -0.2 and -0.4 respectively versus 0.1 in the placebo group.

Two smaller RCTs compared the efficacy and safety of adalimumab in patients treated with MTX. Kim *et al.* carried out a 24 week RCT Phase III study in Korean RA patients (n=128). [222] The primary efficacy endpoint was ACR 20 at week 24 and secondary endpoints included ACR 50, ACR 70 and individual ACR components. An ACR 20 response was achieved by 61.5% of the adalimumab group versus 36.5% of the placebo group. An ACR 50 and ACR 70 response was achieved by 43.1% and 21.5% of the treatment group versus 14.3% and 7.9% of the placebo group. HAQ scores were recorded using the Korean HAQ (KHAQ) and changes at 24 weeks were -0.5 in the treatment group and -0.2 in the placebo group (p=0.002).

Chen *et al.* compared adalimumab in combination with MTX with MTX alone in Taiwanese patients with RA in a 12 week RCT (N=47). [223] Primary endpoints were ACR 20, ACR 50 and ACR 70 at 12 weeks. ACR 20, 50 and 70 response rates were 54.3%, 34.3% and 14.3% in the treatment group versus 33.3%, 16.7% and 0% in the placebo group. HAQ scores for the treatment group decreased from 1.7 to 1.1 and for the placebo group from 1.8 to 1.6. The change for the treatment group was statistically significant ($p<0.05$) but this was not the case for the placebo group which failed to show statistical significance from baseline to follow-up.

Open label extension studies of the adalimumab trials in MTX non-responders

At four years, 62% of patients had remained in the ARMADA study. [224] Seventy eight per cent, 57%, and 31% had achieved ACR 20/50/70; 43% achieved clinical remission (DAS28 <2.6); and 22% had no physical function abnormalities (HAQ=0).

5.1.2 Certolizumab

Certolizumab pegol (Cimzia®) is a pegylated humanised antibody FAB fragment against TNF-alpha. The FAB fragment attached to the PEG chain increases its half life to 14 days. [225] This anti-TNF agent was licensed by the FDA in April 2008 and by the EMA in October 2009. UCB are the current license holders for Cimzia®.

The efficacy of certolizumab has been demonstrated in three main trials, RAPID 1[226], RAPID 2[227] and FAST4WARD.[228] Both RAPID 1 and RAPID 2 are combination therapy (certolizumab + MTX) trials and FAST4WARD is a monotherapy trial. All three trials had a similar patient demographic with a mean age of 52 years; patients had active disease and had failed 1 or more previous DMARDs. The combination therapy trials (RAPID 1 and 2) differed from the monotherapy trial for disease duration which was shorter (6yrs) in the combination studies than the monotherapy study (9yrs).

RAPID 1 (N=982) was a 52 week, randomised, double-blind, placebo controlled, parallel group trial. The objective of the trial was to evaluate the efficacy and safety of two dosage regimes of certolizumab (400mg at weeks 0,2,4 followed by 200mg (n=393) or 400mg every 2 weeks (n=390) as adjunctive therapy to MTX) in patients who had an inadequate response to MTX alone. These were compared to placebo + MTX (n=199). RAPID 2 (N=619) was a 24 week randomised double-blind placebo controlled, parallel group study. This trial evaluated the safety and efficacy of certolizumab + MTX with the same regimes used as RAPID 1.

The primary endpoint of the studies were ACR 20 at 24 weeks. ACR50 and ACR70 were secondary endpoints in the trials. The primary endpoints were reached in both studies. In RAPID 1, at week 24, 58.8% and 60.8% of patients achieved ACR 20 for certolizumab 200mg and certolizumab 400mg respectively, compared with 13.6% in the placebo + MTX arm

(P<0.001). The response for the treatment groups at 52 weeks dropped slightly to approximately 53.1% and 54.9%, for 200mg and 400mg respectively and and remained significant (P<0.001). The ACR 50 at 24 weeks was 7.6% for the placebo + MTX group, 37.1% for 200mg certolizumab and 39.9% for the 400mg certolizumab group. There was little change to these rates at 52 weeks. ACR 70 at 24 weeks was 3% for the placebo + MTX group, 21.4% for 200mg certolizumab and 20.6% for 400mg certolizumab. There was also little change to these values at 52 weeks except a slight increase in the rate for 400mg certolizumab to 23.2% and the placebo + MTX group to 3.5%. All were significant at 52 weeks (P<0.001).

In RAPID 2, at week 24, 8.7% of the placebo + MTX group achieved an ACR 20 response compared with 57.3% and 57.6% of the certolizumab 200mg and 400mg group respectively. For ACR 50, at 24 weeks, 3.1% of the placebo + MTX group, 32.5% of the certolizumab 200mg group and 33.1% of the certolizumab 400mg group achieved a response. For ACR70, 0.8%, 15.9% and 10.6% for the placebo + MTX, certolizumab 200mg and certolizumab 400mg achieved a response. All responses achieved statistical significance (P<0.01).

Within the RAPID trials, patients that failed to achieve an ACR 20 at week 12 and week 14 were designated treatment failure and were withdrawn from the study at week 16. Some but not all were then entered onto an open-label study at a higher dose of 400mg 2 weekly. The number of withdrawals at week 16 in RAPID 1 for the placebo group was 62.8%, the certolizumab 200mg group 21.1% and the certolizumab 400mg group 17.4%. In the RAPID 2 trial, 79.5% of the placebo group were non-responders and withdrawn, 19.9% of the certolizumab 200mg group and 18.7% of the certolizumab 400mg group were withdrawn. [227] These withdrawal rates in the placebo groups are considerable higher than those seen in other anti-TNF RCTs. One of the reasons for this may have been due to the MTX dose being relatively low in the trials (RAPID 1: mean MTX dose 13.5mg, RAPID 2: mean dose of 12.4mg) which may have led to greater drop out rates or lower response rates in the placebo group. Therefore we may be seeing an inflated response to the study drug.

The FAST4WARD trial was a 24-week, randomised, double-blind, placebo-controlled study comparing certolizumab 400 mg (n = 111) to placebo (n = 109) every 4 weeks. The primary endpoint was ACR 20 response at week 24. Secondary endpoints included ACR50 and ACR70 response.

At week 24, 45.5% of patients in the treatment group and 9.3% in the placebo group reached the primary endpoint, an ACR 20 response. Twenty two point seven per cent of the treatment group achieved ACR50 response compared to 3.7% in the placebo arm and for ACR70, 5.5% of the treatment group showed a response compared to 0% of the placebo arm. Results were significant for ACR 20 and ACR50 rates (P<0.001) and for ACR70 (P=0.013).

A further unpublished trial was considered in the EMA review process. This trial (C87014) assessed efficacy and safety of certolizumab 400mg once a month plus MTX as compared with MTX alone in the treatment of RA. [229] The ACR 20 response rate for the treatment group (certolizumab + MTX) at week 24 showed a statistically significant improvement over placebo + MTX. (45.9% vs. 22.9%) The ACR 50 response rates were 18% for the treatment group versus 5.9% in the placebo group. No patients in the treatment group achieved an ACR 70 response versus 1.7% in the placebo group. Both the ACR 20 response rate in the treatment group (45.9%) and the observed difference with the placebo group (23%) is notably lower than those achieved in other trials of anti-TNF agents. Due to the results of this trial, the monthly dosing application was not accepted for inclusion in the Summary of Product Characteristics (SPC).

5.1.3 Etanercept

Etanercept binds to, and neutralises the biological activity of TNF and lymphotoxin, competitively inhibiting the binding of both soluble and membrane bound TNF to cell surface receptors. Enbrel® was licensed by the FDA in November 1998 and by the EMA in February 2000. The current licence holders of Enbrel® are Pfizer Healthcare.

Two main studies have been completed in RA patients who had an inadequate response to MTX. Weinblatt *et al.* compared etanercept plus MTX with MTX alone over 24 weeks (N=89). [230] The primary endpoint was ACR 20 at 24 weeks; secondary endpoints included ACR 50 and ACR 70. Seventy one per cent of the treatment group achieved an ACR 20 response versus 27% in the control group (p<0.001). The ACR 50 and 70 responses rates were 39% and 15% for the treatment group versus 3% and 0% for the control group respectively. HAQ scores were also measured as one of the components of the ACR response. The absolute

median HAQ score at baseline for both groups was 1.5. At 24 weeks the median HAQ score in the treatment group was 0.8 (improvement of -0.7) and for the control group was 1.1 (improvement of -0.4). Both results were statistically significant.

Moreland *et al.* carried out a phase III randomised placebo controlled double blind trial (N=234). [231] The groups compared were etanercept monotherapy (10mg and 25mg dose) and placebo. The primary endpoints were ACR 20 and ACR 50 at 12 and 24 weeks. Secondary endpoints included ACR 70 and individual components of the ACR response index. At 24 weeks, an ACR 20 response was achieved by 51% and 59% of the 10mg and 25 mg treatment groups versus 11% in the control group. An ACR 50 response was achieved by 24% and 40% of the treatments groups respectively versus 3% in the placebo group. At 12 weeks the ACR 20 response was 45% and 62% for the treatment groups versus 25% in the placebo group. The ACR 50 response was 13% and 41% for the treatment groups and 11% in the placebo group and the ACR 70 was 9% and 15% in the treatment groups and 1% in the placebo group. The mean change in HAQ from baseline at 24 weeks was 0.58 and 0.62 for the treatment groups and 0.03 for the placebo group ($p<0.05$).

Open label trials for etanercept

Klareskog *et al.* carried out a 5 year open label study into the efficacy and safety of etanercept monotherapy in patients who had had an inadequate response to MTX (N=549). [232] The study is an extension study of two RCTs. Approximately 80% of patients achieved an ACR 20 at 6 months and maintained this up to 3 years. ACR 50 and 70 were also maintained at 3 years; approximately 50% and 25% of patients achieved ACR 50 and ACR 70 at 12 months and maintained this up to 3 years. HAQ data at 6 months had improved from a mean score of 1.8 at baseline to 1.0 at 6 months which was maintained up to 2 years; there was a slight disimprovement at 3 years to 1.1. Mean DAS score reduced from 5.2 at baseline to 2.6 at 9 months which was maintained up to 3 years.

Etanercept in MTX naïve

Two large trials have been completed in MTX naïve patients. The TEMPO trial compared etanercept plus MTX with MTX alone and etanercept alone (N=686). [233] The primary endpoint was ACR-N at 24 weeks and radiographic scores at 52 weeks. Secondary endpoints

included ACR 20, ACR 50, ACR 70 and HAQ. Primary endpoints were achieved; at week 52 the ACR 20 response rate for the combination group was 85% versus 75% in the MTX group and 76% in the etanercept group. The ACR 50 scores were 69% for the combination group, 43% in the MTX group, and 48% in the etanercept group. ACR 70 scores showed a similar relative improvement of 43% in the combination group compared with 19% in the MTX group and 24% in the etanercept group. The mean HAQ scores improved from 1.8, 1.7 and 1.7 to 0.8, 1.1 and 1.0 for the combination, MTX, and etanercept groups respectively.

The COMET trial was the first trial to include remission as a primary endpoint when comparing MTX to etanercept and MTX. It was a randomised double blind parallel treatment trial in early, moderate to severe RA patients (N=542). [234] Co-primary endpoints included remission measured with DAS 28 and radiographic non-progression (using total Sharp score) at 52 weeks. Remission was defined as DAS 28 ≤2.6. Secondary endpoints included ACR 20, 50 and 70 and HAQ. Patients were both treatment naïve and had early RA (average disease duration 9 months). At 52 weeks 50% of patients in the combination treatment group had achieved remission compared with 28% of the control group (effect difference of 22% $p<0.0001$). Radiographic non-progression was achieved in 80% of the combination group compared with 59% in the control group (effect difference 21% $p<0.0001$). The mean HAQ score improved from 1.7 to 0.7 in the combination group compared with 1.6 to 0.9 in the control group ($p<0.0001$).

5.1.4 Golimumab

Golimumab is a human immunoglobulin G1κ (IgG1κ) monoclonal antibody. It received a license from the FDA in April 2009. It received marketing authorisation in the European Union on 1st October 2009 for the treatment of rheumatoid arthritis, psoriatic arthritis and ankylosing spondylitis. Merck Sharp & Dohme are the current license holders for Simponi®.

There are three primary RCT's; GO-BEFORE conducted in a MTX naive cohort, GO-AFTER which was conducted in anti-TNF experienced cohort and GO-FORWARD. [235-237] The GO-FORWARD trial was a multi-centre, randomised, double blind, placebo controlled trial of golimumab administered to patients with active severe RA despite MTX therapy (N=444). [236] The mean duration of disease for this cohort was 8.6 years and the mean age of the cohort was

51 years. As is the case for many of the current or recently completed clinical trials in RA, disease severity and disease duration tend to be less than older trials for some of the comparators.

The primary outcomes of the study were ACR 20 response at week 14 and improvement from baseline in HAQ at week 24. Secondary endpoints included ACR50, ACR70, EULAR response, and DAS28 remission. Safety, effects of golimumab on structural damage and HRQOL (SF-36) were also assessed. There were 4 arms; Arm 1: Placebo + MTX (n=133), Arm 2: Golimumab 100mg + placebo (n=133), Arm 3: Golimumab 50mg + MTX (n=89), Arm 4: Golimumab 100mg +MTX (n=89). Both primary endpoints were reached. At week 24, 60% and 28% achieved ACR 20 response for golimumab + MTX (Arm 3) and placebo + MTX respectively. At week 24, 37% of the golimumab + MTX group (Arm 3) and 13.5% of the placebo + MTX group had an ACR50 response (statistical significance was reached in all cases $p<0.001$). The high response rate in the placebo + MTX group may be attributed to the Latin American subgroup of patients. When data from this group were removed, the ACR 20 responses were 26.1% at week 14 in the placebo group vs. ~55% in the combined golimumab + MTX groups. [238] The mean HAQ improvement in the treatment group was -0.38 and -0.13 ($p<0.001$) in the placebo group.

The GO-BEFORE study, assessed MTX-naïve patients with RA (N=637). Patients were randomised to placebo + MTX (group 1), golimumab 100mg + placebo (group 2), golimumab 50mg (group 3) or golimumab 100mg + MTX (group 4). The primary endpoint was ACR50 at week 24. A significant difference was not shown for ACR 50 between groups 1 and 4. (29.4% group 1 vs. 29.4% in group 4 ($p=0.053$). Group 3 did show a significant difference to group 1 (40.3% vs. 29.4% $p<0.05$). A significant difference was found in HAQ improvement for group 4 but not for group 3.

The GO-AFTER study included patients who had an inadequate response to anti-TNF therapy (N=.461). Patients were assigned to one of three groups; placebo; golimumab 50mg; or golimumab 100mg. DMARD therapy was permitted and approximately 70% of patients were on concomitant therapy. The primary endpoint was ACR 20 at week 14. A statistically significant proportion of patients in the golimumab 50mg and 100mg groups (35% and 38% respectively) achieved the primary outcome compared to the placebo group (17%). The result appeared to be significant for patients who had failed less than three prior anti-TNF agents. A

difference was not observed for patients who had failed three prior anti-TNF agents. A significant improvement in HAQ scores was noted in the golimumab groups at week 14 and 24.

5.1.5 Infliximab

Infliximab is a chimeric IgG1k monoclonal antibody which neutralizes the biological activity of TNF α by binding with high affinity to the soluble and transmembrane forms of TNF α and inhibits binding of TNF α with its receptors. The FDA first licensed it in November 1999 followed by the EMA in August 1999. Merck Sharp & Dohme is the current license holder for Remicade®.

The ATTRACT study was a randomized placebo controlled trial comparing different doses of infliximab in patients who had active RA despite MTX treatment (N=428). [239] The primary outcome was ACR 20 at week 30 without requiring a surgical intervention. Secondary outcomes were ACR 50, ACR 70, HAQ, and general health assessment. The primary outcome was met; 20% of patients in the placebo group achieved an ACR 20 versus 50%, 54%, 52% and 57% in the treatment groups (3mg/kg 8 wkly, 3mg/kg 4 wkly, 10mg/kg 8 wkly and 10mg/kg 4 wkly) ($p<0.001$). The ACR 50 and ACR 70 responses were 5% in the placebo group versus 27%, 29%, 31% and 26% in treatment groups and 0% versus 8%, 11%, 18% and 11% respectively. There was a statistical difference in HAQ scores from baseline to follow-up in all but one group (3mg/kg 8 wkly).

The START study compared safety of two doses of infliximab (3mg/kg and 10mg/kg) with placebo in patients with active RA despite MTX treatment (N=1084). [240] Relative risks of developing a serious adverse effect at week 22 were presented with ACR response rates. At week 22, 25.5% of patients in the placebo group achieved an ACR 20 vs 58% in the 3mg/kg group ($p<0.001$) and 61% in the 10mg/kg group ($p<0.001$). A significant difference was demonstrated between placebo and treatment groups for both ACR 50 and ACR 70. HAQ data was not available in the published paper.

Two smaller studies were carried out in Japan and China comparing infliximab + MTX with placebo + MTX (N=147 and N= 173). [241, 242] Both studies demonstrated a statistically significant improvement in the infliximab groups over the placebo groups.

Early RA patients

There are two main studies evaluating the efficacy and safety of infliximab in early RA. [97, 243] The treatment group in the ASPIRE trial achieved a higher ACR-N (primary outcome), at 54 weeks, for both MTX + infliximab doses (3mg/kg and 6mg/kg) versus MTX + placebo (38.9% and 46.7% versus 26.4%). ACR 20, ACR 50, ACR 70 and ACR 90 were achieved in both treatment groups versus the placebo group.

The BeST study evaluated four different strategies in patients with early RA (N=508); sequential DMARD therapy, step up combination DMARD therapy, initial combination therapy with tapered steroid dose and initial combination therapy with infliximab. Both ACR and DAS were measured as outcomes. Patients treated with combination therapy with tapered steroid dose and combination therapy with infliximab showed more rapid improvement than the other groups in both ACR responses and DAS 44. After 5 years, 48% of patients were in clinical remission (DAS <1.6) and 14% in drug-free remission, irrespective of initial treatment. [244]

5.2 Observational/Registry Data on anti-TNF agents and leflunomide in RA

5.2.1 Anti-TNF agents

While clinical efficacy has been established from RCTs, uncertainty remains with regard to short and long term safety and long term clinical effectiveness. In order to establish longer term efficacy and as a pharmacovigiliance tool national registries have been set up in many countries in Europe. Sweden was one of the first countries to establish a biologic register and many other countries followed suit. [245]

The initial publications on biologics included etanercept and infliximab and later studies included adalimumab. Effectiveness data from registries on golimumab and certolizumab is not yet available. Many of the original publications were from the Southern Swedish registries. One of the earliest registry reports on efficacy was for etanercept and etanercept + MTX from the STURE database. [246] For the majority of the outcomes measures (ACR response components, DAS 28, HAQ) there was no significant difference between the groups. A further

registry report from Southern Sweden on etanercept, infliximab and leflunomide presented the treatment efficacy as measured by the ACR 20, 50, 70 responses. Etanercept was significantly better than infliximab at three months ($p<0.02$) and six months ($p<0.05$) when the number of patients reaching the ACR 20 response were compared. For the ACR 50 response only the three months' registration reached significance in favour of etanercept compared with infliximab ($p<0.05$). The study found that the performance of both etanercept and infliximab complied with results in published clinical trials, albeit with a somewhat lower response rate.

The DANBIO registry in Denmark in 2010 was established to measure the rates of treatment response, remission, and the drug survival rate in patients with RA, and to identify clinical prognostic factors for response. [247] Etanercept, infliximab and adalimumab were included (N=2326). [247] The crude treatment responses were highest for adalimumab, followed by etanercept and then infliximab. A corrected response rate showed that 19% of adalimumab patients and 17% etanercept patients received an ACR 70. There were no clinically relevant differences in the distribution of age, sex, disease activity, and disease duration between the patients the treatment groups. There were fewer patients treated with etanercept who received concomitant MTX, and more patients receiving infliximab who received concomitant MTX and prednisolone.

The Dutch DREAM registry showed significant differences in the DAS 28 course over 12 months between infliximab and both adalimumab and etanercept patients ($p<0.001$) and between adalimumab and etanercept patients ($p=0.031$). Both adalimumab and etanercept showed greater improvement in DAS and HAQ than infliximab. [248]

5.2.2 Registry Evidence in relation to leflunomide

Two of the registries examined in this thesis included a control arm for the observational data. Geborek et al. included leflunomide as a control arm in comparison to etanercept and infliximab. [249] All patients included in the registry had failed at least two DMARDs, including methotrexate. The demography of the patients was similar; leflunomide patients were older (61.3 years) than the etanercept (54 years, $p<0.0001$) and the infliximab group (55.4 years, $p<0.001$). There was a statistically significant difference in the mean number of DMARDs continued in the leflunomide group (0.1) compared to the etanercept group (0.7, $p<0.001$) and the infliximab group (1.0, $p<0.001$).

The published study reports no difference as measured by percentage improvement in HAQ 20% or 50%, between etanercept, infliximab and leflunomide at three and six months. The actual values were not shown in the published report. The ACR 20 and ACR 50 did show a significant difference (etanercept was more effective between leflunomide and etanercept at both three months ($p<0.001$) and six months ($p<0.05$) and a significant benefit was also present for infliximab over leflunomide at three months ($p<0.01$) but not at six months. While the study examined data up to 12 months, the number of observations for leflunomide at 12 months made statistical comparison unreliable.

Zink et al. examined two subgroups for the control group. [250] The first was leflunomide alone and the second was leflunomide in combination with methotrexate. While the demography of the patients was broadly similar, the control group did have less erosive disease at baseline, a shorter disease duration and DAS 28 scores were also lower (control group 5.4 vs the anti-TNF groups ~6.0). The majority (~90%) of all patients had failed methotrexate. Less than 10% of patients had previous biological therapy in the etanercept and control group compared to 16.4% in the infliximab group and 30% in the anakinra group. Efficacy was not reported in this study but drug continuation rates were; these are discussed under the section 5.4.3.1.

5.3 Evidence Synthesis

5.3.1 Background

In the absence of head-to-head trials of relevant comparators, it is often necessary to combine evidence from placebo controlled trials of different treatments and thereby derive an estimate of effect of one treatment against another. A comparison of relevant data for all drugs within one model which maintains randomization allows the decision-maker to examine the relative effect of all alternatives. Pooling of direct and indirect evidence from trials can produce an estimate of the relative effect of a treatment against all other treatments in a particular network. Combining evidence in this way is increasingly used in health technology assessments around the world. [251-253]

5.3.1.1 Methods available to combine evidence

Network meta-analysis (NMA) or mixed treatment comparison (MTC) is considered a logical extension of the meta-analysis method. [254] It enables treatment effects that are not directly measured to be estimated through the use of direct evidence of linked treatments, or a network

of treatments. Bayesian meta-analysis provides more flexibility than classical methods to include more data and handle more complex modelling structures. [255]

Lu and Ades described the statistical method for performing MTC in a Bayesian framework. [256] This can deal with the uncertainty between multiple treatment groups when included in the network meta-analytic model. The method used hierarchical models, for borrowing strength from indirect comparisons when direct information from a specific comparison does not provide sufficient data for a substantial statistical analysis. An application of this in RA was described by Nixon *et al.* using different models which allowed for explicit modelling of concurrent treatments, multiple treatment arms and study level covariates. [257]

With the exception of one trial comparing abatacept and infliximab there are no head-to-head trials comparing any of the biological DMARDs. [258] Therefore it is necessary to perform evidence synthesis to derive estimates of efficacy of one biological agent against another.

In the case of no direct comparisons between treatment A and B (in this case between the anti-TNF treatments), one can perform an indirect comparison which will also include treatment C (in this case placebo). Treatment C is the link between A and B and is therefore the common comparator. This allows a comparison of AB (indirect evidence) using trials of AC and BC (direct evidence).

An unadjusted indirect comparison compares the individual arms of trials (the results in the A arm of the AC trials with the B arm of the BC trials). However this method ignores trial randomisation and is therefore not recommended. Bucher offers a widely used approach in considering indirect treatment comparisons in meta-analyses of trials for discrete data. [259] The model was developed with the odds ratio (OR) as the measure of treatment effect. While the Bucher model is a more correct method of indirect comparison than the unadjusted method where trial randomisation is not maintained, it does present with some limitations. The Bucher method is applicable when there is no correlation between the pairwise comparisons and therefore can only be applied to two arm trials. Only a simple three treatment comparison can be made (A vs. B, B vs. C) and only the OR was considered as an effect measure.

In order to handle all available data, relevant to the question (therefore not just those involving three treatments), we can use a NMA to perform an MTC. This type of meta-analysis allows comparison of one or more treatments through more than one common comparator. In NMA

more than one treatment path can be used to compare two treatments and these paths are then part of a larger network. A weighting is then assigned to these pathways and a level of agreement is obtained between the effect estimates of the paths. This agreement between the effect estimates is known as the incoherence of the network.

The NMA method offers an advantage over the Bucher method in providing a method of comparing more than one comparator to produce an indirect treatment effect estimate through the use of a network of pathways. NMA does not however account for correlations that may exist between different effect estimates when they are obtained from a single multi-armed trial. A random effects model is used when a certain degree of heterogeneity exists and this heterogeneity is not fully explained by exploration of covariates. [260] The Bayesian modelling framework, while complex, allows for this uncertainty between multiple treatment groups when included in the NMA model and makes full use of all available data (Figure 23).

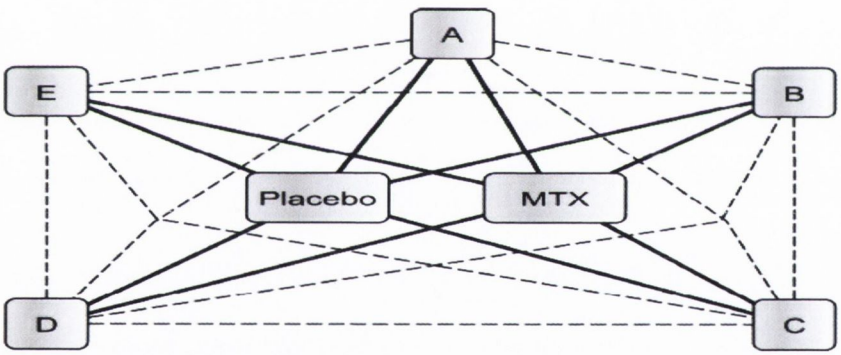


Figure 23: Network Diagram of evidence.
Solid lines show pair-wise direct pathways and the dotted lines represent indirect pair-wise comparisons.

The most sophisticated method to date for this has been presented by Nixon *et al.*[257]. The analysis followed a Bayesian framework and was carried out in WinBugs software. The NCPE has assessed the cost-effectiveness of two biological agents (golimumab and certolizumab) for the treatment of RA; both of these submissions combined evidence using indirect comparison methods. There was considerable uncertainty surrounding these estimates, with wide confidence intervals. In the critical review of these submissions, the areas of concern were the appropriateness of the studies chosen for inclusion in the analysis, the measurement, and handling of heterogeneity and the models used for the analysis. In order to include the newer drugs, golimumab and certolizumab in a complete MTC, a new analysis was performed.

The economic model, for which the relative effects calculated in this chapter will be used, examines the cost-effectiveness of anti-TNF agents versus the cost effectiveness of leflunomide. In order to take into consideration all data within a formal evidence syhthesis model, data for both the anti-TNF agents and the leflunomide should be included. However there was limited data available for leflunomide efficacy following non-response to methotrexate and therefore there was no common comparator which would allow leflunomide to be incorporated into the MTC network. While this is not a limitation to the work presented in this chapter, it is a limitation to the use of relative efficacy estimates for use in the cost-effectiveness model.

5.3.1.2 Literature review of previous meta-analyses in RA

A number of papers have performed evidence synthesis using a formal meta-analytical framework. [257, 261-267] (Table 12)

Table 13. Summary of meta-analyses of biological therapy

Author and Date	Drugs Included	Outcomes Used	Methods	Model	Measure of Effect	Study Type
Kristensen *et al.* 2007[268]	Adalimumab, etanercept, infliximab	ACR 50	Frequentist	None	NNT, OR	RCT
Venkateshan *et al.* 2009[264]	Adalimumab, etanercept, infliximab, rituximab, abatacept	ACR 20, ACR 50, ACR 70	Frequentist	DerSimonian and Laird random effects	OR	RCT
Kristensen *et al.* 2011[269]	Abatacept, certolizumab, golimumab, rituximab, tocilizumab	ACR 50 Adverse Effects	Frequentist	None	NNT, RR, RD	RCT
Salliot *et al.* 2011[266]	Adalimumab, etanercept, infliximab, abatacept, golimumab, certolizumab, rituximab, tocilizumab	ACR 50	Frequentist	DerSimonian and Laird random effects	OR	RCT
Launois *et al.* 2011[261]	Adalimumab, etanercept, infliximab, abatacept, golimumab, certolizumab, anakinra, tocilizumab	ACR 20, ACR 50, ACR 70	Frequentist and Bayesian	MTC random effects using non-informative prior	OR	RCT
Lee *et al.* 2008[262]	Adalimumab, etanercept, infliximab	ACR 20 ACR 50 ACR 70	Frequentist	DerSimonian and Laird random effects	RR	RCT
Alonso-Ruiz *et al.* 2008[263]	Adalimumab, etanercept, infliximab	ACR 20, ACR 50, ACR 70 Adverse events	Frequentist	DerSimonian and Laird random effects	RR	RCT
Lloyd *et al.* 2010[265]	Adalimumab, etanercept, infliximab	ACR 20, EULAR Response, DAS 28, HAQ	Frequentist	Random Effects and fixed effects (subgroups) Meta-regression	Percentage change from baseline	Observational
Nixon *et al.* 2007[253, 257]	Adalimumab, etanercept, infliximab	ACR 50	Bayesian	Bayesian MTC, Meta-regression	OR	RCT
Devine *et al.* 2011[270]	Infliximab, etanercept, adalimumab, golimumab, certolizumab, tocilizumab, rituximab, abatacept, anakinra	ACR 50	Bayesian	Bayesian MTC, Meta-regression	OR	RCT
Janssen *et al.* 2008[271]	MTX, Anti-TNFa, (Adalimumab, etanercept, infliximab)	HAQ	Frequentist and Bayesian	Frequentist fixed effects and random effects. Bayesian fixed effects and random effects model	HAQ change from baseline	RCT
Bergman *et al.* 2010[272]	Tocilizumab, abatacept, rituximab, anti-TNF inhibitors	ACR 20, ACR 50, ACR 70	Bayesian	Bayesian MTC fixed effects	RR	RCT

NNT= Number Needed to Treat, OR= Odds Ratio, RCT=Randomised Controlled Trials, RR= Relative Risk, RD=Risk Difference, HAQ= Health Assessment Questionnaire

Five of these studies have used a Bayesian method to combine the evidence;[257, 261, 267, 270, 272] the remaining studies used a frequentist approach to meta-analysis. Jansen *et al.* did not carry out a full MTC but used RA as a case example to demonstrate the methods described. [267] Devine *et al.* used the methodology described by Nixon *et al.* and compared all biological therapy indicated in RA. [257, 270]

Almost all of the studies concluded that there was no difference between the anti-TNF agents in the treatment of RA. Some of the studies included the anti-TNF agents as a group in the network assuming similar efficacy. [264-267, 272] Some combined a different patient population such as MTX naïve patients or early RA patients with treatment experienced or established RA patients. [257, 262-266, 270, 272]

Evidence Synthesis

5.3.2. Aim

- To perform a systematic review of the evidence to identify relevant studies.
- To estimate the relative effect of the anti-TNF agents against each other, using different outcome measures (both continuous and categorical).
- To establish if the use of different models for a MTC produces different results and whether using the results of different models for a MTC will change the result of a cost-effectiveness model.
- To combine evidence for drug continuation rates from long term observational studies.
- To combine evidence for short term discontinuation from anti-TNF agents

5.3.4. Methods

A systematic review of the literature was performed in order to identify the relevant data for extraction. Data extracted was checked by a second person (SS). The criteria used for this are given in Appendix 6. The outcomes chosen for the MTC were ACR 20, ACR 50 and percentage HAQ improvement at 6 months. These outcomes were chosen in order to establish an overall effectiveness using a well validated instrument (ACR) which was common to all the RCTs and HAQ in order to model the disease benefit for establishing an estimate of cost-effectiveness.

The literature search included published randomised controlled studies up to and including October 2010 in PubMED, EMBASE and the Cochrane Database. A number of search terms were used using papers published in the English language. (Appendix 6) Rheumatological inflammatory diseases, other than RA, such as ankylosing spondylitis, psoriatic arthritis and connective tissue diseases were excluded from the search. This analysis examined a patient cohort with established RA; therefore early RA and MTX naïve patients were excluded.

Inclusion criteria:

1. Patients with an established diagnosis of RA according to the American College of Rheumatology diagnostic criteria.
2. Patients with an inadequate response to MTX
3. Published RCTs, where patients were treated for 24 weeks (in the case where 24 week data was not available, data within 6 weeks either before or after 24 weeks was used).
4. Monotherapy and combination therapy were included.
5. All doses used in clinical trials were included.

The outcomes chosen were the ACR 20 and ACR 50 and the percentage improvement in HAQ response. The following data was extracted for the analysis; total number of patients, number of respondents achieving ACR response and the mean quantity of improvement and standard deviations (SDs) in the case of the HAQ, were extracted.

Authors were contacted in cases where the required data was not reported in the published study. In cases where no access was allowed to the required data, for studies where the mean was not reported, the median was used. In the absence of SDs, interquartile ranges (IQRs) where used to estimate SDs using a normal approximation. In all remaining cases, the maximum of clinical trial SDs was used. The doses of biological agents included are those included in the RCTs. Demographic data including age, gender, mean disease duration, baseline HAQ score, and previous number of DMARDs were recorded and a weighted mean for all studies for each drug was calculated.

5.3.4.1 Statistical Analysis

A Bayesian MTC model was fitted for each of the outcome measures. A different model was fitted for the ACR outcomes and HAQ improvement due to the nature of the outcomes. The ACR 20 and ACR 50 are binary, categorical outcomes and the HAQ improvement is a continuous outcome measure. The method used to analyse the binary data is described by Nixon *et al.* [257]

5.3.4.2 Nixon methodology

Nixon allows for explicit modelling of concurrent treatments (in this case anti-TNF and MTX), multiple treatment arms, and study level covariates. A meta-regression was also performed and included disease duration and baseline HAQ score as covariates. Five models were used, with models 1-4 using either univariate random effects or bivariate random effects. Two of the models were MTC models; model 1 assumed exchangeability between trial arms (both within and between studies), and model 3 assumed exchangeability between studies, and also allowed the MTX effect to vary between studies. Model 2 and 4 complimented the MTC models by including study level covariates as meta-regression. Model 5 assumed different random effects for two different groups of agents; anti-TNF and interleukin 1 (anakinra). Model 5 is not relevant to the analysis for this thesis because anakinra is not included as a drug for analysis. Odds ratio (OR) were published of an ACR50 event at six months if treated with a biologic in comparison to control. All models were fitted using Markov Chain Monte Carlo techniques (MCMC) using the computer package Win BUGS. [273] Vague normal priors were used.

5.3.4.2 Adaptation of Nixon model

The Nixon et al. model [257] was adapted by Susanne Schmitz in Trinity College, Dublin (Appendix 10 gives further details of work performed by Susanne Schmitz for the thesis). The code for model 4 was given in the published Nixon paper and this was adapted for the analysis performed here. A methodology described by Warn *et al.* was used to calculate risk ratio (RR) instead of OR [274]. Four models were fitted initially and one model was chosen for the final analysis (bivariate random effects). Further exploration of the impact of using different MTC models is given below. The effect of MTX is included in models so that this effect can be separated from the effect of interest (anti-TNF agent). The effect size of the MTX is only informed by studies where MTX is given in either the control arm or the treatment arm. If it is given in both arms, no information on this effect is provided.

HAQ improvement is modelled differently to the ACR 20 and ACR 50 and this was not included in the Nixon model. The HAQ improvement is dependent on the baseline HAQ score or the patient's disease severity. This is based on previous studies which have shown that the effect size of the anti-TNF depends on disease severity. [275] The methodology for this model was developed by Susanne Schmitz, Department of Statistics, Trinity College Dublin (Appendix 10).

All models were extended to a meta-regression to include the covariates duration of disease, number of previous DMARDs and HAQ score at baseline. None of these were found to have a significant impact.

Demographic of the trial populations is presented with weighted mean values. The results of the MTC are presented as relative risks with 80% credible intervals.

5.3.4.3 Exploration of structural uncertainty and model choice

Many methods of evidence synthesis exist; but usually only one is chosen. While this is a pragmatic approach, the decision-maker is left with some uncertainty as to how an alternative method may have impacted on the end result and ultimately the decision. Model uncertainty has been discussed in a number of publications on structural uncertainty. [32, 276, 277] While the uncertainty around the model chosen for the decision analysis is often explored, the model estimating the data for the parameters is often

not. In order to inform the uncertainty associated with the model, an overview of structural uncertainty is given with possible approaches to handling it appropriately. The method chosen and the rationale for the choice are detailed.

Brief Overview of Structural Uncertainty

Structural uncertainty may be described as the uncertainties associated with structural aspects of the model; these may be in relation to how aspects of the disease are handled or how aspects of the intervention are handled. For example in the case of evidence synthesis models, does the model allow for additive effects of more than one intervention or are they handled independently? Structural uncertainty has also been described as the uncertainty that is not handled under methodological or parameter uncertainty. [34]

Various approaches have been suggested for managing this uncertainty. One such approach averages the model outputs depending on the confidence in each model. [278] However this approach has not been used commonly. Model averaging has been described extensively using both frequentist and Bayesian frameworks. [279, 280] It offers an alternative to model selection where an average result is used from all models. This method offers the advantage of including different results instead of discarding a set of results from a potentially plausible model which is the case with model selection. However model averaging is a computationally challenging method and one could argue that the result is still associated with considerable uncertainty.

The model selection method chosen for this work is a pragmatic one but also allows the impact of using a different model each time to be explored. [281] Four different models were used; Model 1 was a random effects model for MTX treatment effect and anti-TNF treatment effects. Model 2 used a fixed effect model for both MTX and anti-TNF treatment effects. Model 3 was a fixed effect for MTX and random effects for anti-TNF treatment effects and model 4 used a random effects model for the MTX treatment effect and fixed effect model for the treatment effects of the anti-TNF drug. The analysis using the models described was performed using WinBUGs software as described above (analysis in WinBUGs was carried out by Susanne Schmitz). The results of the model were then used to fit a cost-effectiveness RA model. We estimated a cost per QALY for each of the treatments and presented them on a cost-

effectiveness plane (Figure 24). The willingness to pay thresholds at €20,000 /QALY and €30,000 /QALY is given to illustrate the impact of the change for the decision-maker.

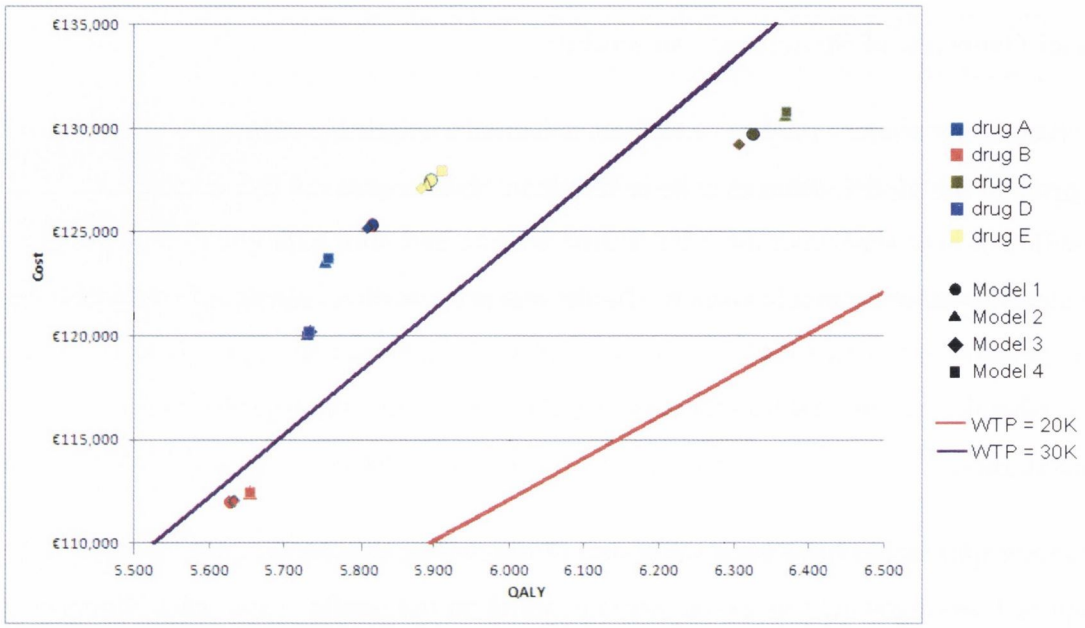


Figure 24: Cost-effectiveness plane showing estimates for each treatment from each model.
Each colour represents a drug and each shape represents a different model. A willingness to pay (WTP) threshold of €20,000/QALY and €30,000/QALY are also shown.

On the cost-effectiveness plane the estimates for each drug differs slightly for the four different models. However, the relative positioning on the cost-effectiveness plane does not differ greatly. On the CEAC (not presented), the net monetary benefit of the base case was approximately 15% lower for the RE model than for the other models used for the meta-analysis. [276]

The model chosen in this case was model 3 above. This was chosen over model 1 due to the limited trial information informing the MTX effect.

5.3.5 Results

Sixteen RCTs were selected for the analysis examining ACR 20 and ACR 50 as the outcome. Five studies met the inclusion criteria for adalimumab, four for infliximab, two for etanercept, two for golimumab, and three for certolizumab. MTX was included in all arms of 11 of the studies [218, 219, 222, 226, 227, 230, 239, 240, 242, 258, 282]; MTX was not given in four of the studies.[220, 221, 228, 231] One trial contained arms of combination and arms of monotherapy. [236] The network diagram illustrates the networks involved and the number of patients included for each pairwise comparison (Figure 24). A number of trials were excluded based on the inclusion criteria (Appendix 6).

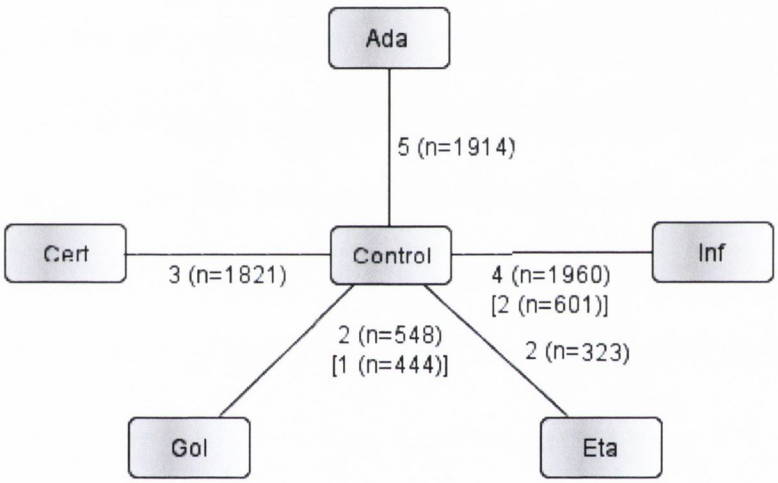


Figure 25: Network Diagram of Evidence.
Lines are labelled with the number of studies and the total number of patients included in these studies. Numbers in squared brackets refer to HAQ evidence where this differs from ACR evidence.

While the demographics of those included across the studies was broadly similar, there was some heterogeneity and in particular in relation to the severity of the disease (baseline mean HAQ score ranged from 1.3 to 1.9), dose of MTX (mean ranging from 13mg in the certolizumab trials to 18.5mg in the etanercept trials) (Table 13) and trial design. The mean age of the trial cohorts is 52 years and most are women (80.5%). The mean disease duration was 8.7 years. The disease duration appears lower for the newer anti-TNF agents (certolizumab and golimumab) but this is likely explained by differences in practice such as early referral and difficulty in recruiting anti-TNF naïve patients into trials. The rate of adverse effects was the same for the placebo and

treatment groups across the trials (5%) and there was a higher number of patients withdrawing due to lack of response in the placebo groups (19.6%) than the treatment groups (8%).

The ACR 20, ACR 50, and HAQ were examined to establish comparative efficacy of the anti-TNF agents in the treatment of RA. The data extracted from the literature following a systematic review is presented (Table 15).

Table 14. Mean Demographic of trial populations of RCTs for anti-TNF agents

Study	Total number in trial arms used (N)	Age (Yrs)	Disease Duration (Yrs)	Number of Previous DMARDs	Baseline HAQ Score	Dose of MTX (mg)
Weinblatt et al. 2003 (ARMADA)[218]	271	56	12	3	1.6	17
Keystone et al. 2004[219]	619	56	11	2	1.5	17
Van de Putte et al. 2004[220]	544	53	11	4	1.9	n/a
Miyasaka et al. 2008 (CHANGE)[221]	352	55	7	n/a	1.6	n/a
Kim et al. 2007[222]	128	49	7	n/a	1.4	16
Mean Demographic adalimumab trials		**55**	**10**	**3**	**1.6**	**17**
Maini et al. 1999 (ATTRACT)[239]	428	53	8	3	1.7	15
Westhovens et al. 2006 (START)*[240]	1,084	52	7	n/a	1.5	15
Zhang et al. 2006[242]	173	48	8	n/a	n/a	n/a
Schiff et al. 2008 (ATTEST)*[258]	275	49	8	n/a	1.7	16
Mean Demographics infliximab trials		**52**	**8**	**3**	**1.6**	**15**
Moreland et al. 1999[231]	234	52	12	3	1.7	n/a
Weinblatt et al. 1999[230]	89	50	13	3	1.5	19
Mean Demographics etanercept trials		**52**	**12**	**3**	**1.6**	**19**
Keystone et al. 2009[236]	444	51	6	n/a	1.3	15
Kay et al. 2008*[282]	104	53	7	n/a	1.5	n/a
Mean Demographics golimumab trials		**52**	**6**	**n/a**	**1.4**	**15**
Keystone et al. 2008 (RAPID 1)[226]	982	52	6	1	1.7	14
Smolen et al. 2009 (RAPID 2)[227]	619	52	6	1	1.6	13
Fleischmann et al. 2009 (FAST4WARD)[228]	220	54	10	2	1.5	n/a
Mean Demographics certolizumab trials		**52**	**7**	**1**	**1.6**	**13**

Table 15. Responders for ACR 20 and ACR 50

Study	Arm			ACR 20		ACR 50	
		Number in control arm	Number in treatment arm	Number of responders in control arm	Number of responders in treatment arm	Number of responders in control arm	Number of responders in treatment arm
Weinblatt *et al.* 2003 ARMADA	Placebo + MTX EOW	62		9		5	
	Adalimumab 20mg+ MTX EOW		69		33		22
	Adalimumab 40mg+ MTX EOW		67		45		37
	Adalimumab 80mg+ MTX EOW		73		48		31
Keystone *et al.* 2004	Placebo OW + MTX	200		59		19	
	Adalimumab 40mg EOW+placebo EOW +MTX		207		131		81
	Adalimumab 20mg OW +MTX		212		129		87
Van de Putte *et al.* 2004	Placebo	110		21		9	
	Adalimumab 20mg OW		112		44		23
	Adalimumab 20mg EOW		106		38		20
	Adalimumab 40mg OW		103		55		36
	Adalimumab 40mg EOW		113		52		25
Miyasaka *et al.* 2008 CHANGE	Placebo EOW	87		12		5	
	Adalimumab 20mg EOW		87		25		14
	Adalimumab 40mg EOW		91		40		22
	Adalimumab 80mg EOW		87		44		28

Study	Arm			ACR 20		ACR 50	
		Number in control arm	Number in treatment arm	Number of responders in control arm	Number of responders in treatment arm	Number of responders in control arm	Number of responders in treatment arm
Kim *et al.* 2007	Placebo +MTX	63		23		9	
	Adalimumab 40mg EOW + MTX		65		40		28
Maini *et al.* 1999 (ATTRACT)	Placebo + MTX	88		18		7	
	Infliximab 3mg/kg 8wkly + MTX		86		45		22
	Infliximab 3mg/kg 4wkly + MTX		86		47		25
	Infliximab 10mg/kg 8wkly +MTX		87		51		26
	Infliximab 10mg/kg 4wkly + MTX		81		49		21
Westhovens *et al.* 2006 (START)	Placebo + MTX	363		87		33	
	Infliximab 3mg/kg +MTX		360		199		110
	Infliximab 10mg/kg+MTX		361		205		119
Zhang *et al.* 2006	Placebo + MTX	86		42		22	
	Infliximab 3mg/kg + MTX		87		66		38
Schiff *et al.* 2008 (ATTEST)	Placebo + MTX	110		46		22	
	Infliximab 3mg/kg 8wkly + MTX		165		98		61
Moreland *et al.* 1999	Placebo	80		9		4	
	Etanercept 10mg BW		76		39		18
	Etanercept 25mg BW		78		46		31

Study	Arm			ACR 20		ACR 50	
		Number in control arm	Number in treatment arm	Number of responders in control arm	Number of responders in treatment arm	Number of responders in control arm	Number of responders in treatment arm
Weinblatt *et al.* 1999	Placebo + MTX	30		8		1	
	Etanercept BW+ MTX various		59		42		23
Keystone *et al.* 2009 (GO-FORWARD)	Placebo+ MTX	133		37		18	
	Golimumab 100mg 4 wkly + Placebo		133		47		26
	Golimumab 50mg 4 wkly + MTX		89		53		33
	Golimumab 100mg 4 wkly + MTX		89		53		29
Kay *et al.* 2008	Placebo +MTX*	35		13		2	
	Golimumab 50mg 4 wkly +MTX		35		21		13
	Golimumab 100mg 4 wkly +MTX		34		19		10
Keystone *et al.* 2008 (RAPID 1)	Placebo + MTX	199		27		15	
	CZP 200mg EOW* +MTX		393		231		146
	CZP 400mg EOW* +MTX		390		237		156
Smolen *et al.* 2008 (RAPID 2)	Placebo + MTX	127		11		4	
	CZP 200mg EOW* +MTX		246		141		80
	CZP 400mg EOW* +MTX		246		142		81
Fleischmann *et al.* 2009 FAST4WARD	Placebo	109		10		4	
	CZP 400mg 4 wkly		111		51		25

Table 16: Data inputs for MTC model for HAQ Improvement

Study Name	Arm	Number in control arm	Number in treatment arm	Baseline HAQ in control arm	Baseline HAQ in treatment arm	HAQ improvement in control arm	SD improvement in control arm	HAQ improvement in treatment arm	SD improvement in control arm	SD delta
Weinblatt *et al.* 2003 (ARMADA)	Placebo + MTX EOW	62		1.64		0.27	0.57			
	Adalimumab 20mg+ MTX		69		1.52			0.54	0.58	0.10
	Adalimumab 40mg+ MTX		67		1.55			0.62	0.63	0.11
	Adalimumab 80mg + MTX		73		1.55			0.59	0.53	0.10
Keystone *et al.* 2004	Placebo + MTX OW	207		1.45		0.24	0.52			
	Adalimumab 40mg EOW + Placebo + MTX		212		1.44			0.56	0.52	0.05
	Adalimumab 20mg OW + MTX		200		1.48			0.6	0.53	0.05
Van de Putte *et al.* 2004	Placebo	110		1.88		0.07	0.49			
	Adalimumab EOW 20mg		112		1.88			0.39	0.62	0.07
	Adalimumab 20mg OW		106		1.88			0.29	0.63	0.08
	Adalimumab 40mg EOW		103		1.84			0.49	0.54	0.07
	Adalimumab 40mg OW		113		1.83			0.38	0.6	0.07
Miyasaka *et al.* 2008 (CHANGE)	Placebo	87		1.39		-0.1	0.6			
	Adalimumab 20mg EOW		87		1.57			0.2	0.5	0.08
	Adalimumab 40mg EOW		91		1.64			0.2	0.6	0.09
	Adalimumab 80mg EOW		87		1.77			0.4	0.6	0.09
Kim *et al.* 2007	Placebo + MTX	63		1.3		0.2	0.5			

Study Name	Arm	Number in control arm	Number in treatment arm	Baseline HAQ in control arm	Baseline HAQ in treatment arm	HAQ improvement in control arm	SD improvement in control arm	HAQ improvement in treatment arm	SD improvement in control arm	SD delta
	Adalimumab 40mg EOW + MTX		65		1.4			0.5	0.55	0.09
Maini *et al.* 1999 (ATTRACT)	Placebo + MTX	88		1.8		0.3	0.5			
	Infliximab 3mg/kg 8wkly + MTX		86		1.8			0.3	0.5	0.08
	Infliximab 3mg/kg 4wkly + MTX		86		1.8			0.5	0.5	0.08
	Infliximab 10mg/kg 8wkly + MTX		87		1.8			0.5	0.6	0.08
	Infliximab 10mg/kg 4wk+MTX		81		1.5			0.4	0.5	0.08
Zhang *et al.* 2006	Placebo + MTX	86		1.6		0.45	0.9			
	Infliximab 3mg/kg + MTX		87		1.61			0.76	0.9	0.14
Moreland *et al.* 1999	Pbo	80		1.7		0.03	0.9			
	Etanercept 10mg		76		1.7			0.58	0.9	0.14
	Etanercept 25mg		25		1.6			0.62	0.9	0.21
Weinblatt *et al.* 1999	Placebo + MTX	30		1.5		0.4	0.9			
	Etanercept + MTX various		59		1.5			0.7	0.9	0.20
Keystone *et al.* 2004 GO-FORWARD	Placebo+ MTX	133		1.25		0.13	0.38			
	Golimumab 100mg 4 wkly + Pbo		133		1.375			0.13	0.66	0.07
	Golimumab 50mg 4 wkly + MTX		89		1.375			0.38	0.46	0.06
	Golimumab 100mg 4 wkly + MTX		89		1.375			0.5	0.46	0.06

Study Name	Arm	Number in control arm	Number in treatment arm	Baseline HAQ in control arm	Baseline HAQ in treatment arm	HAQ improvement in control arm	SD improvement in control arm	HAQ improvement in treatment arm	SD improvement in control arm	SD delta
Keystone et al. 2008 RAPID 1	Placebo + MTX	199		1.7		0.15	0.51			
	CZP 200mg EOW + MTX		393		1.7			0.55	0.85	0.06
	CZP 400mg EOW + MTX		390		1.7			0.58	0.61	0.05
Smolen 2008 RAPID 2	Placebo + MTX	127		1.6		0.14	0.45			
	CZP 200mg EOW + MTX		246		1.6			0.5	0.47	0.05
	CZP 400mg EOW + MTX		246		1.6			0.5	0.47	0.05
Fleischmann 2009 FAST4WARD	Placebo	109		1.55		-0.07	0.44			
	CZP 400mg 4 wkly		111		1.43			0.39	0.66	0.08

*CZP=certolizumab, MTX=MTX * Patients received loading doses at 0, 2, and 4 weeks.*

Table 17. Relative risks for each pair-wise comparison

(mean estimate with the 80% credible intervals). A relative risk of 1.0 can be interpreted as no difference between the drugs, if RR<1.0 drug A is less effective than drug B and if RR>1.0 Drug A is more effective than drug B.

Comparison (A vs. B)	ACR 20			ACR 50		
		80% credible interval			*80% credible interval*	
	Mean	***lower***	***upper***	**Mean**	***lower***	***upper***
Ada vs. Placebo	**2.34**	1.94	2.84	**3.47**	2.78	4.36
Inf vs. Placebo	**1.88**	1.56	2.26	**2.44**	1.98	2.96
Eta vs. Placebo	**3.60**	2.41	5.16	**5.71**	3.00	9.70
Gol vs. Placebo	**1.88**	1.40	2.52	**2.85**	1.93	4.12
Cert vs. Placebo	**4.92**	3.79	6.30	**5.90**	4.18	7.93
Inf vs. Ada	**0.80**	0.62	1.06	**0.70**	0.53	0.99
Eta vs. Ada	**1.54**	0.98	2.30	**1.65**	0.85	2.88
Eta vs. Inf	**1.91**	1.26	2.94	**2.35**	1.21	4.17
Gol vs. Ada	**0.80**	0.57	1.14	**0.82**	0.54	1.28
Gol vs. Inf	**1.00**	0.72	1.44	**1.17**	0.75	1.73
Gol vs. Eta	**0.52**	0.32	0.84	**0.50**	0.27	1.00
Cert vs. Ada	**2.11**	1.51	2.85	**1.70**	1.17	2.54
Cert vs. Inf	**2.62**	1.91	3.67	**2.42**	1.65	3.47
Cert vs. Eta	**1.37**	0.89	2.27	**1.03**	0.57	2.09
Cert vs. Gol	**2.62**	1.79	3.95	**2.07**	1.22	3.35
σ	**0.26**	0.12	0.36	**0.24**	0.04	0.35

All anti-TNF agents achieved a significant ACR response over placebo (the 80% credible intervals are higher than and do not include 1) (Table 17). The RR for certolizumab achieving ACR 20 and ACR 50 indicated improved efficacy over adalimumab, infliximab, and golimumab. The outcomes also provide evidence that etanercept is superior to infliximab and golimumab. For ACR 50, etanercept appeared approximately equal in efficacy to certolizumab (Cert vs. Eta RR 1.03) and adalimumab shows improvement over infliximab.

Table 18. Percentage HAQ improvement for each pair-wise comparison.

(Mean with 80% credible intervals). Positive mean values indicate a superiority of drug A and negative values indicate deterioration in comparison to drug B; σ is the between trial variance parameter.

Comparison (A vs. B)	*80% credible interval*		
	Mean	***lower***	***upper***
Ada vs. Placebo	**0.20**	0.17	0.23
Inf vs. Placebo	**0.10**	0.06	0.15
Eta vs. Placebo	**0.31**	0.25	0.37
Gol vs. Placebo	**0.23**	0.17	0.29
Cert vs. Placebo	**0.25**	0.22	0.29
Inf vs. Ada	**-0.10**	-0.15	-0.05
Eta vs. Ada	**0.11**	0.04	0.17
Eta vs. Inf	**0.21**	0.13	0.28
Gol vs. Ada	**0.02**	-0.05	0.08
Gol vs. Inf	**0.12**	0.05	0.20
Gol vs. Eta	**-0.09**	-0.17	0.00
Cert vs. Ada	**0.05**	0.00	0.09
Cert vs. Inf	**0.15**	0.10	0.21
Cert vs. Eta	**-0.06**	-0.13	0.01
Cert vs. Gol	**0.03**	-0.04	0.09
σ	**0.03**	0.01	0.05

The model detected some heterogeneity across the trials, which was captured in the between trial variance parameter σ^2 (σ = 0.26 for ACR 20 and 0.24 for ACR 50); the heterogeneity is likely to be attributed to demographic characteristics such as background MTX dose and trial design.

For HAQ response, again, all anti-TNF agents show significant improvement over placebo, etanercept reaching the highest improvement (m = 0.31) (Table 18). All anti-TNF agents provide evidence of improvement over infliximab. Certolizumab and etanercept appear superior to adalimumab. Golimumab shows slight improvement in efficacy over adalimumab but the result is not significant. The between trial standard deviation in the HAQ model is estimated to be σ = 0.02. Both ACR outcomes and HAQ are illustrated in a forest plot (Figure 26).

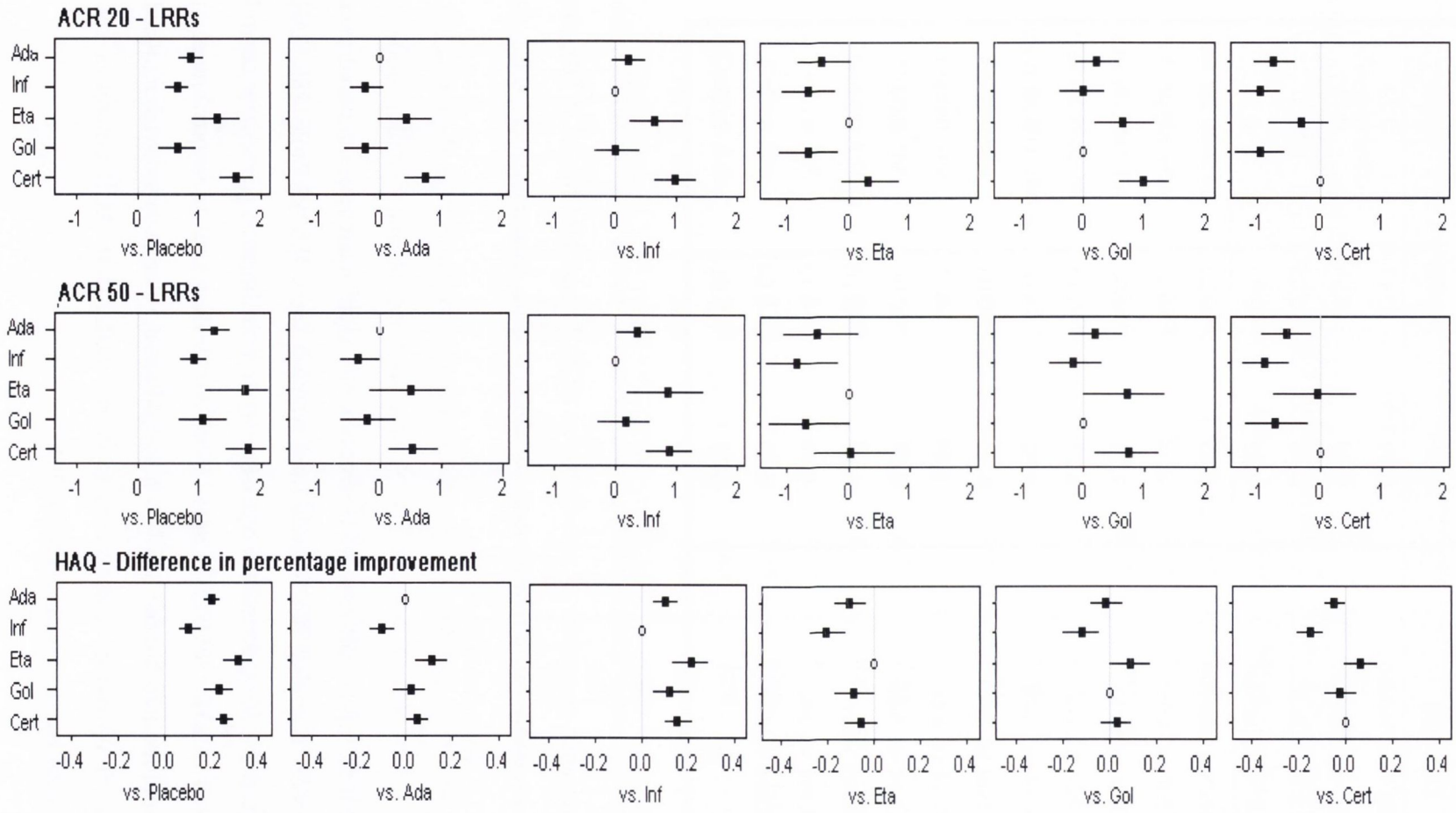


Figure 26: **Pairwise log risk ratios (LRRs) for ACR 20 and ACR 50 outcome and estimated HAQ improvement multiplier of anti-TNF against placebo and one another.**

5.4.1.6 Discussion

There have been a number of indirect comparisons performed to date [253, 261, 262, 266, 267, 283]; this analysis adds to this evidence base by examining the HAQ multiplier in a Bayesian framework. The value of MTC in evidence based healthcare evaluations was recently highlighted.[284] It is acknowledged that the MTC framework allows inclusion of evidence that may not be possible using classical methods; the inclusion of such evidence could in turn reduce the uncertainty around the estimates of effectiveness.

In accordance in what was found in the individual trials, all anti-TNF agents show a significant improvement over placebo across all outcome measures. Etanercept and certolizumab show high efficacy throughout; other indirect comparisons have found a similar result in relation to etanercept.[285] Improved outcomes with etanercept and certolizumab may relate to a reduced immunogenicity as compared to the antibody therapies. It may also be affected by RCT attributes such as the early withdrawal rates in the placebo arms of the RAPID trials. With certolizumab, it may be the pegylated formulation which allows less exposure of the molecule and less opportunity for an immune response. The HAQ outcome data provide evidence that all anti-TNF agents show improvement over infliximab. This effect is not found with the ACR outcomes for adalimumab and golimumab. Furthermore, the HAQ model indicates a superiority of etanercept over adalimumab. The evidence of certolizumab providing improvement over golimumab, which can be found in the ACR outcomes, is not apparent in the HAQ outcomes. This may be attributed to the fact that only one of the golimumab trials was included for this model. Therefore there is not enough power to detect these differences.

This enhanced significance for the continuous outcome measure (HAQ) underlines the lack of sensitivity to change in binary outcome measures (ACR). One of the key problems highlighted previously with the ACR response, is its binomial nature. It estimates the proportion of patients achieving a certain percentage improvement, and hence provides an adequate measure for a clinical trial. However, no difference is made between different response levels. For instance, using the ACR 20 response, no difference is made between patients with a 20% improvement and patients with 80%

improvement. It may be the case, in MTCs, where the goal is to detect differences, an outcome measure which is sensitive to change is more appropriate. The HAQ multiplier provides one such measure, but other continuous measures may also be suitable (e.g. ACR hybrid measure[127] or DAS 28).

In this application the differences due to model choice are small, the relative positioning of the estimates remains and hence the decision on cost-effectiveness is robust to model choice. This may not always be the case; the impact of model choices changes in different applications and different models can also change the decision on cost-effectiveness. A similar conclusion was reached in a recent study by Oppe *et al.* [276] In chronic obstructive airways disease, Oppe *et al.* found that the Bayesian random effects model led to most change in the cost-effectiveness estimates both in the point estimates and the uncertainty around them.

Diversity or heterogeneity can exist across studies clinically and methodologically. Statistically, the heterogeneity in this analysis is partially represented by σ^2, the between trial variance. The between trial variance only captures heterogeneity between study populations of the same drug; variation between drugs is not captured. It is likely, that there is some heterogeneity between trials of different drugs; these trials were conducted over a 10 year time scale (publications dates range 1999-2009). Differences between trial design and statistical analysis of the data exist. Statistical methods for handling missing data differed over this time; recent certolizumab trials used non-responder imputation for some of the analyses whereas earlier trials used last observation carried forward. There were also some differences in the manner in which non-responders were handled between the earlier and latest trials.

The demographic of the patients also differs (Table 13). While this study does not include other clinical outcomes, such as radiographic scores, it should be acknowledged that higher radiographic scores at baseline may impede the degree of improvement that can be achieved in HAQ scores. This is likely to apply to the earlier studies where effective treatment strategies were newly developed. In a meta-regression the impact of different trial demographics including year of publication was examined and found not to be significant. There has also been some discussion in the literature about the problems of comparing trials of anti-TNF agents. [261, 286, 287]

Doses across treatments and treatment arms varied, which has not been accounted for in this analysis. The aim of this analysis was to compare the overall efficacy of one anti-TNF against another and therefore all doses were included in the analysis. Generalizing the model to include a meta-regression for dose would raise difficulties of comparability; doses across treatments are hard to compare as well as within one treatment when the same dose is given, but in different instances.

A MTC allows for the added effect of an additional medication, which in this case is MTX. The estimation of MTX efficacy could have been improved by including trials comparing MTX against placebo, however the inclusion criteria for the analysis meant that many of these study types were excluded. Comparing trials where patients were MTX naive would have led to confounding issues. Since the interest was in estimating relative efficacy of anti-TNF agents only, the MTX parameter was treated as an explanatory variable, and not as an actual effect estimate.

5.4 Withdrawal Rates

Since the previous NICE assessment using the BRAM 2006, a considerable amount of registry data has been published. [247, 249, 250, 288, 289] Also when assessing the newer anti-TNF drugs in the NCPE, altering withdrawal rates had a considerable impact to the estimates of cost-effectiveness. [41] A Bayesian survival analysis for the long term data was fitted in WinBUGs and a frequentist approach was used for the short term data. Observational data was not available for golimumab or certolizumab; therefore a pooled estimate is calculated from all drugs and this is used.

5.4.1 Meta-analysis of long term withdrawal and adverse effects

The length of time a patient remains on a biological agent can be one of the key drivers of the cost-effectiveness of the agents. [290] Observational studies were chosen as they would be considered a truer estimate of clinical practice and these studies have a longer duration.

5.4.1.1 Aims

- To perform a literature review of long term observational studies on the use of anti-TNF agents.
- To combine the data extracted, using evidence synthesis methods.
- To estimate a long term withdrawal rate for each anti-TNF agent, leflunomide and an overall withdrawal rate for anti-TNF agents as a group.
- To estimate short term withdrawal rates for each drug

5.4.1.2 Methods

A literature review was carried out, to identify observational studies on the use of anti-TNF agents. Studies included had to report recorded numbers of patients remaining on drugs at specific time points, data for individual drugs and data beyond 6 months. These were in the form of published registry data. Data was extracted in numeric form if given and otherwise was extracted from survival plots given in the published paper. The data was extracted at six weeks, three months and at three monthly intervals from Kaplan Meier plots up to five years. The number of patients remaining in the study at each time point was recorded. In the cases where treatments were broken down into groups such as etanercept, etanercept and MTX and etanercept and DMARDs, a weighted average was used.

The data was combined within a Bayesian framework and was done in WinBUGs (technical analysis carried out in collaboration with Susanne Schmitz, Trinity College Dublin (Appendix 10)). [291] A Weibull curve was then fitted to the extracted data. The Weibull distribution has 2 parameters; shape or lambda (λ) and either scale (gamma) or rate (1/gamma). The WinBUGs outputs specify beta, mu and r. The BRAM uses shape and scale parameters. For the model these are the 'a' (shape) and 'b' (scale) parameters. In order to parameterize these for the model; a=r and b=1/mu.

Data for short term withdrawal rates was also extracted (read from Kaplan Meier plots). A meta-analysis using a frequentist framework was used to combine this data. The analysis was done in R® software. Both a random effects and fixed effect model were fitted and heterogeneity was assessed using the I^2 statistic.

There are a number of formal methods used for measuring statistical heterogeneity; Cochrane's Q and the I^2 statistic. The Q statistic leads to a P value and a P value of 0.10 is used as a threshold for significance. The I^2 statistic leads to a percentage and describes the variability in effect estimates due to heterogeneity rather than to chance (sampling error). The Cochrane Handbook gives the following values to classify the inconsistency of the effect measures across studies: 0-40%: might be important, 30-60%: may represent moderate heterogeneity, 50-90%: may represent substantial heterogeneity, 75-100%: considerable heterogeneity. [292]

The significance of the I^2 statistic depends on the magnitude and direction of the effects; if the I^2 is > 50%, all the studies are in the direction of benefit, and the random effect meta-analysis gives highly significant benefit, then we are uncertain about the amount of benefit but not about its existence. In this case it is safe to assume that the treatment is beneficial. If the I^2 statistic is indicating substantial heterogeneity and wide confidence intervals there are three possible solutions: (a) avoid meta-analysis (b) explore heterogeneity further (via subgroup analysis) (c) perform a random effect meta-analysis.[292]

For this analysis, where heterogeneity was present, the estimates from the random effect model were chosen. Confidence intervals (95%) are also given.

5.4.1.3 Results

The data extracted for the analysis is shown in table 19. Five studies were included from four different countries. The length of the studies ranged from 1 year to 8 years. Data was only available on the older anti-TNF agents (etanercept, infliximab and adalimumab) as the newer agents' golimumab and certolizumab did not yet have long term data.

Summary of the demographic of the registries

The demographics of the registry data is given in table 18.

Zink 2002 et al. [250]

The analysis for this study included 924 patients who were on etanercept, infliximab or anakinra. The patients were included in the German registry between May 2001 and September 2003. The registry did include a control group (n=599). The control groups included patients who were on leflunomide alone (n=120) or leflunomide with MTX (n=141). Patients were aged 18-75 years and all met the ACR criteria for diagnosis of RA. Entry into the database required patients to be initiated on etanercept, infliximab or anakinra for the treatment group and a new DMARD treatment was begun after failure of at least one previous therapy for the control group. A number of outcomes were measured at baseline, and at each visit; the minimum dataset required was baseline characteristics, the start and end of DMARD or biological treatment, reasons for treatment termination and details on adverse events. Kaplan Meier plots are presented in the published paper as well as numbers of patients remaining on treatment at 6 months. Data from this registry indicates that patients remained on anti-TNF therapy longer than leflunomide. Continuation rates after 12 months were 81.3% for infliximab, 87.4% for etanercept, leflunomide 67.8% and leflunomide plus methotrexate 62.4%.

Kristensen et al. 2006[288]

This study, from a six year observational Swedish cohort (n=1,161) examined the impact of concomitant DMARD therapy on adherence to anti-TNF agents (infliximab and etanercept). Entry onto the registry database was determined by the clinical judgment of the treating physician. Patients were treated at eight hospital centres, between March 1999 and December 2004. While no formal disease activity measurement was required, patients should have received at least two DMARDs one which was MTX. Only biologic naïve patients were enrolled. Patients were followed up every 6 months and 2% of patients remained with unknown status on retrieval of data. Kaplan Meier plots were presented in the study. Numbers of patients remaining on treatment were not presented separately and therefore these were read exclusively from the graphs.

Du Pan et al. 2009[289]

This study reported on a Swiss longitudinal, population-based cohort (n=2,364) to compare drug retention rates and causes of discontinuation between anti-TNF agents. The registry follows patients with a diagnosis of RA; it was estimated that the registry included 70-80% of all RA patients receiving biological agents in Switzerland. Data

included in the publication was collected between January 1997 and December 2006. The study did not consider drug interruptions of less than 6 months a drug discontinuation. Data was presented in the form of Kaplan Meier plots for etanercept, infliximab and adalimumab up to 3 years.

Hetland et al. 2010[247]

The results of the DANBIO Danish registry were published in 2010. This was a nationwide study which collected data on rheumatology patients receiving routine care. The data used for the published study included patients who were biologic naïve at baseline (n=2,326). Adalimumab, etanercept and infliximab were included and data was collected between October 2000 and December 2007. The study reported treatment responses, remission rates and drug adherence. Drug adherence was reported in Kaplan Meier plots and number of patients remaining on each drug was also recorded.

Geborek et al. 2002[249]

This study presented the clinical experience of treatment with etanercept, infliximab and leflunomide over a 2 year period in a Swedish registry (n=404). Patients were treated at seven centers and were required to have an inadequate response to two or more DMARDs including MTX and a diagnosis of RA according to the treating doctor. Similar to the Kristensen study, no formal disease activity level was required for inclusion. Patients were included between March 1999 and November 2000. Kaplan Meier plots were presented and the data extracted was read exclusively from these.

Kievit et al. 2008[248]

The Dutch Rheumatoid arthritis Register (DREAM) was established to evaluate the effects of adalimumab, etanercept and infliximab on disease activity, functional ability, QOL and medication costs in a real life setting (n=916). Patients were reviewed every three months and the outcomes were DAS 28, HAQ, EQ-5D and SF-36. The data was collected for one year; survival curves are extended to three years and data was extracted up to three years from the Kaplan Meier plot.

Table 19. Baseline demographic of registry populations

Study	Drugs	Control Arm	Mean age in biologic group (years)	Mean Disease duration (years)	Number of Previous DMARDs	DAS at baseline	Concomitant MTX*
Zink 2005[293]	Etanercept Infliximab	Yes (LEF or LEF + MTX)	54 54	9 8.5	3.9 3.7	6.1 6.0	33%* 64%*
Kristensen 2006[288]	Etanercept Infliximab	Not reported	55 58	13.8 13.6	3.9 3.8	5.7 5.6	41%* 69%*
Du Pan 2009[289]	Etanercept Infliximab Adalimumab	No	54 53 55	10 10 10	n/a	4.23 4.27 4.14	55% 74% 61%
Hetland 2010[247]	Etanercept Infliximab Adalimumab	Not reported	58 57 56	8 9 9	3 3 3	5.4 5.4 5.3	61% 87% 70%
Geborek 2002[249]	Etanercept Infliximab	Yes (LEF)	54 55.4	14.9 14.1	4.5 4.0	5.8 5.6	Not reported
Kievit 2008[248]	Etanercept Infliximab Adalimumab	No	54.6 57.8 55.1	6 7.7 7.7	3 3 3	5.5 5.2 5.3	78%** 85%** 87%**

LEF *Leflunomide* ***MTX*** *Methotrexate* ***DAS*** *Disease Activity Score*

*Indicates percentage of the individual anti-TNF group (e.g. etanercept or infliximab) rather than a percentage of the group combined (etanercept + infliximab) ** Only the adalimumab includes specifically concomitant MTX, figures for etanercept and infliximab are concomitant DMARDs in general and not specifically MTX.

Weibull curves, illustrating the long term discontinuation rates of each anti-TNF agent are presented (Figure 27). Infliximab demonstrates the shortest discontinuation rate and adalimumab the longest, indicating that patients are more likely to remain on adalimumab longer than if they are on infliximab. The black line on the graph represents the average continuation rates of the three anti-TNF agents. Each point on the line indicates the number of patients remaining on the drug at that time point. For all anti-TNF agents there are less patients remaining on the drug at five years.

Table 20. Data for meta-analysis of observational data for long term drug survival

Drug	Study	Months *Number of Patients remaining on drug*									Length of Study	Study Type
		0	**3**	**6**	**9**	**12**	**24**	**36**	**48**	**60**		
Etanercept	Zink 2005	254	n/a	181	n/a	164	n/a	n/a	n/a	n/a	1 year	German Observational Registry
Etanercept + MTX		167	n/a	137	n/a	120	n/a	n/a	n/a	n/a		
Etanercept + other DMARD		90	n/a	76	n/a	67	n/a	n/a	n/a	n/a		
Total Etanercept		**511**	**460**	**394**	**358**	**351**	**n/a**	**n/a**	**n/a**	**n/a**		
Infliximab		36	n/a	24	23	16	n/a	n/a	n/a	n/a		
Infliximab + MTX		219	n/a	169	n/a	145	n/a	n/a	n/a	n/a		
Infliximab + other DMARD		88	n/a	70	n/a	63	n/a	n/a	n/a	n/a		
Total Infliximab		**343**	**292**	**263**	**233**	**224**	**n/a**	**n/a**	**n/a**	**n/a**		
Etanercept + MTX	Kristensen 2006	179	175	172	161	159	152	141	134	122	6 year	Swedish Observational Registry
Etanercept + Other DMARDs		68	65	64	59	58	55	54	48	46		
Etanercept		193	185	170	154	145	131	122	106	97		
Infliximab + MTX		501	471	431	401	351	281	251	210	190		
Infliximab + Other DMARDs		116	102	93	80	67	51	37	35	n/a		
Infliximab		104	92	79	62	52	31	26	18	n/a		
Adalimumab	Du Pan 2009	882	838	820	785	759	591	485	n/a	n/a	3 year	Swiss Observational Registry
Infliximab		595	565	536	518	476	357	274	n/a	n/a		
Etanercept		887	843	825	789	745	585	444	n/a	n/a		
Adalimumab	Hetland 2010	544	522	457	408	403	348	316	283	272	8 Years	Danish Observational Registry
Etanercept		425	412	366	357	349	306	272	242	221		
Infliximab	Geborek 2002	908	854	699	636	590	472	409	381	345	20 months	Swedish Observational Registry
Etanercept		166	161	156	146	141	n/a	n/a	n/a	n/a		
Infliximab		135	123	117	108	101	n/a	n/a	n/a	n/a		
Adalimumab	Kievit 2008	267	246	224	214	211	195	179	n/a	n/a	3 years	Dutch Observational Study
Etanercept		289	260	237	225	225	194	188	n/a	n/a		
Infliximab		151	121	112	107	104	94	94	n/a	n/a		

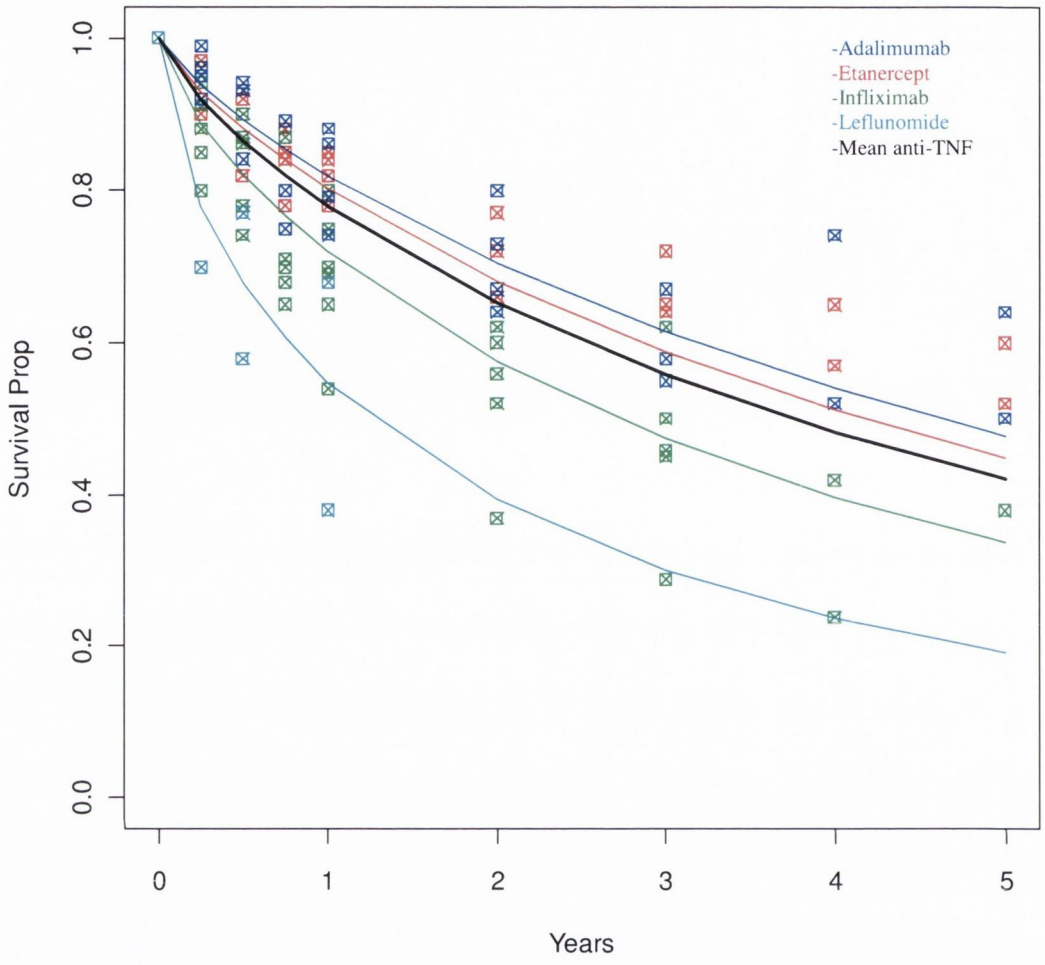


Figure 27: **Weibull curve for discontinuation rates of anti-TNF agents and leflunomide up to 5 years.**
Each colour represents a different drug and each square represents a different study at individual time points (adalimumab (blue), infliximab (green), etanercept (red), leflunomide (turquoise) and mean anti-TNF (black)).

5.4.2 Meta-Analysis of short term withdrawal rates

5.4.2.1 Aims

- To extract short term discontinuation data from observational studies.
- To identify the reasons for withdrawal
- To combine data using classic meta-analytical techniques.

5.4.2.2 Methods

Data was extracted from Kaplan Meier plots or from tables if available in the published papers. The number of patients withdrawing from treatment due to toxicity was recorded up to six weeks and between 6 and 24 weeks. The number of patients withdrawing due to lack of response between 6 and 24 weeks only was recorded; up to six week data was not recorded for inefficacy as it was deemed unlikely that patients would withdrawn at such an early stage due to lack of response.
Data was combined using the R software package. The R function used for the analysis was 'metaprop' which can calculate an overall proportion from studies reporting a single proportion. The syntax for the analysis is provided (Appendix 7). Forest plots were also produced in R. Both fixed and random effects models are fitted; both estimates are provided.

There was no registry data available for golimumab and certolizumab at the time of the analysis; therefore data is combined for an overall estimate of the anti-TNF agents and this estimate is used for certolizumab and golimumab in the model. Data is also combined for an overall estimate of all short term discontinuation (due to toxicity and lack of response) for each drug.

5.4.2.3 Results

The extracted data is presented in Table 21. The results indicate that there are some differences between the drugs as regards short term withdrawal rates. Specific data on withdrawal was not presented in the DREAM study and therefore this was excluded. This is particularly the case for withdrawal due to toxicity from 6 weeks and 24 weeks. One of the studies (Zink 2005) consistently has higher probabilities of withdrawal from treatment in all three categories for drugs included in the study (infliximab and etanercept). It is not clear why this is the case. This is presented graphically on the forest plots (Figure 26-29).

Withdrawal from anti-TNF therapy up to six weeks

The number of patients withdrawing up to six weeks was assumed to be due to toxicity and was broadly similar for all drugs (Figure 28). The overall probability of withdrawing from anti-TNF agents (adalimumab, etanercept and infliximab) is 0.02. When the individual drugs are examined, the result is 0.02 for adalimumab, 0.03 for etanercept and 0.03 for infliximab (Figure 28-30). Significant heterogeneity was present for all analysis and therefore the random effects estimates are used.

Table 21. Data for meta-analysis of observational data for short term drug discontinuation

Drug	Study	Number withdrawing due to toxicity 6 week	Number withdrawing due to toxicity 6-24 weeks	Number withdrawing due to inefficacy 6-24 weeks	Length of Study	Study Type
Etanercept	Zink 2005	n/a	23	47	1 year	German Observational Registry
Etanercept + MTX		n/a	18	18		
Etanercept + other DMARD		n/a	7	7		
Total Etanercept		n/a	44	72		
Infliximab		n/a	6	8		
Infliximab + MTX		n/a	32	21		
Infliximab + other DMARD		n/a	9	11		
Total Infliximab		n/a	47	40		
Etanercept+ MTX	Kristensen 2006	2	4	5	6 year	Swedish Observational Registry
Etanercept + Other DMARDs		1	5	3		
Etanercept		4	19	15		
Infliximab+MTX		0	50	15		
Infliximab+ Other DMARDs		5	19	3		
Infliximab		4	23	8		
Adalimumab	Du Pan 2009	4	44	9	3 year	Swiss Observational Registry
Infliximab		3	30	6		
Etanercept		4	27	9		
Adalimumab	Hetland 2010	22	n/a	15	8 years	Danish Observational (Danbio)
Etanercept		21	n/a	19		
Infliximab		54	n/a	29		
Etanercept	Geborek 2002	8	8	3	20 months	Swedish Observational Registry
Infliximab		7	12	7		

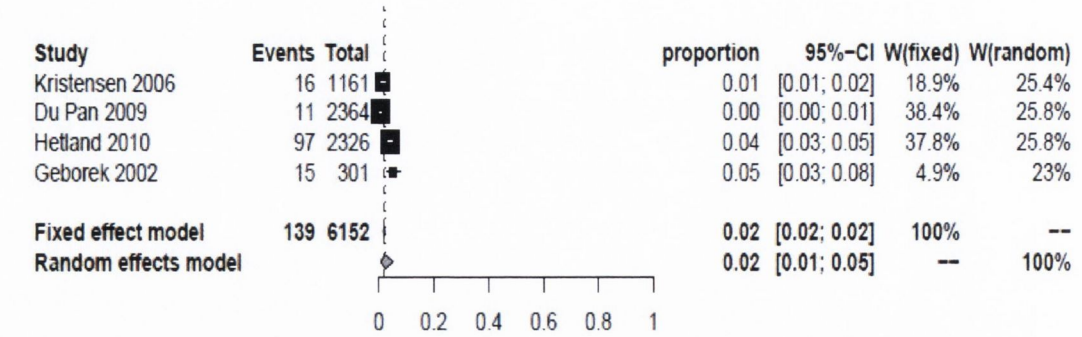


Figure 28: Forest Plot of proportion of patients withdrawing from drug treatment at six weeks, due to toxicity, for all drugs.
(pooled data for etanercept, infliximab and adalimumab). $I^2 = 97\%$

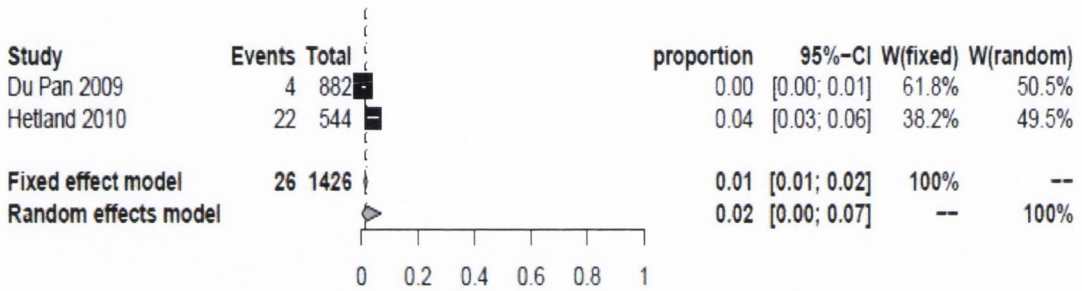


Figure 29: Forest Plot of proportion of patients withdrawing at six weeks due to toxicity for adalimumab.
$I^2 = 95.8\%$

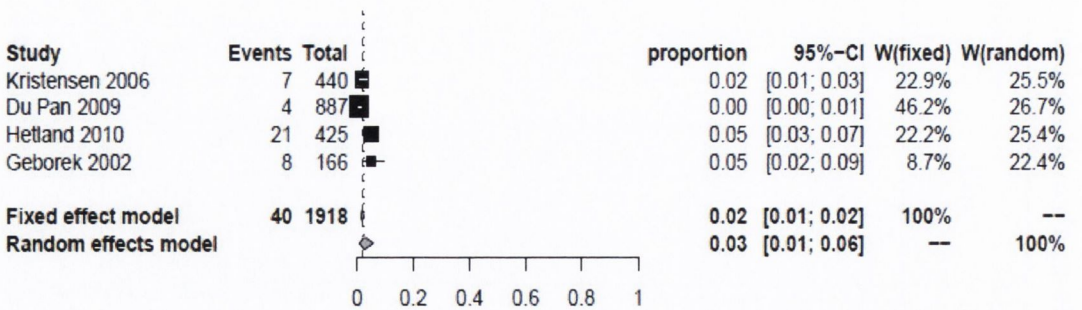


Figure 30: Forest Plot of proportion of patients withdrawing at six weeks due to toxicity for etanercept.
$I^2 = 91.3\%$

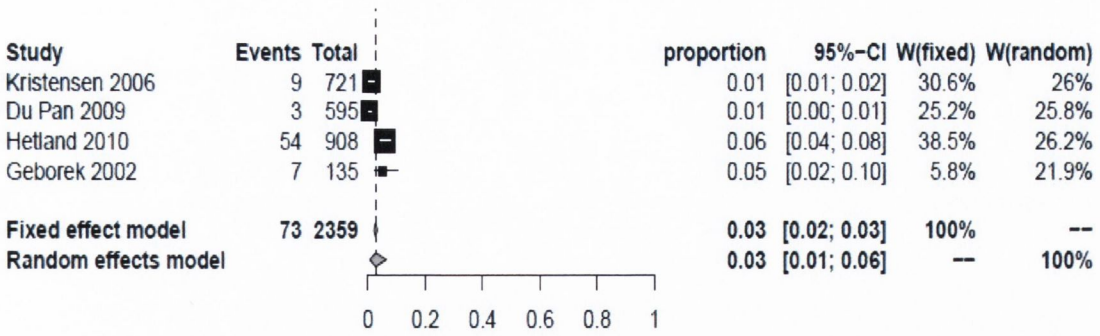


Figure 31: Forest Plot of proportion of patients withdrawing at 6 weeks due to toxicity for infliximab.
I^2 = 94.4%

Withdrawal due to toxicity between 6 and 24 weeks

There are greater differences between the anti-TNF drugs in terms of withdrawals due to toxicity between 6 and 24 weeks. When data for each of the three anti-TNF agents (adalimumab, etanercept and infliximab) is pooled, the probability of withdrawal due to toxicity is 0.08 (random effects) (Figure 32). Etanercept shows the lowest probability (0.03) for withdrawal due to toxicity followed by adalimumab (0.05) and infliximab shows the highest probability of withdrawal (0.10) (Figure 33-35). Only one study informs the estimate for adalimumab (Du Pan).

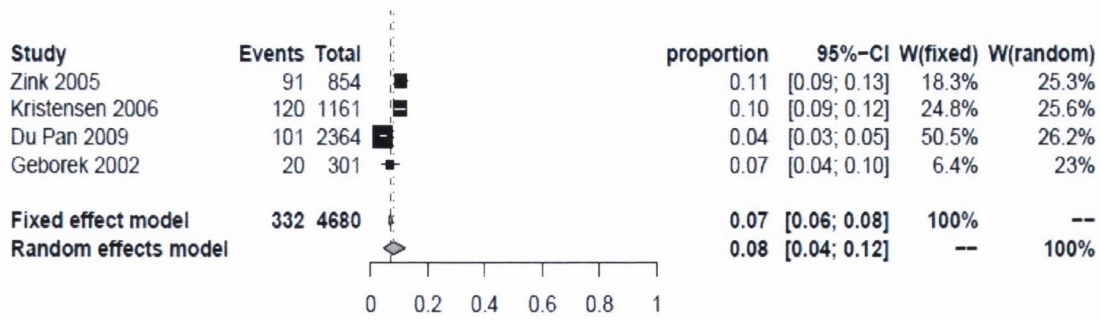


Figure 32: Forest Plot of proportion of patients withdrawing between 6 and 24 weeks due to toxicity for all drugs.
(pooled data for etanercept, adalimumab and infliximab). I^2 = 95.3%

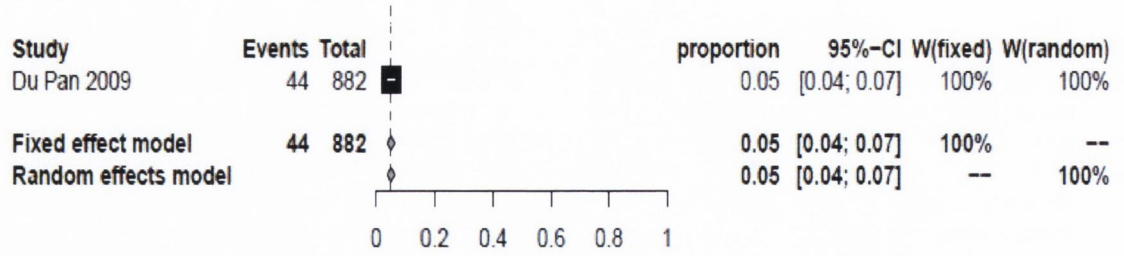


Figure 33: Forest Plot of proportion of patients withdrawing between 6 and 24 weeks due to toxicity for adalimumab.
$I^2 = N/A$

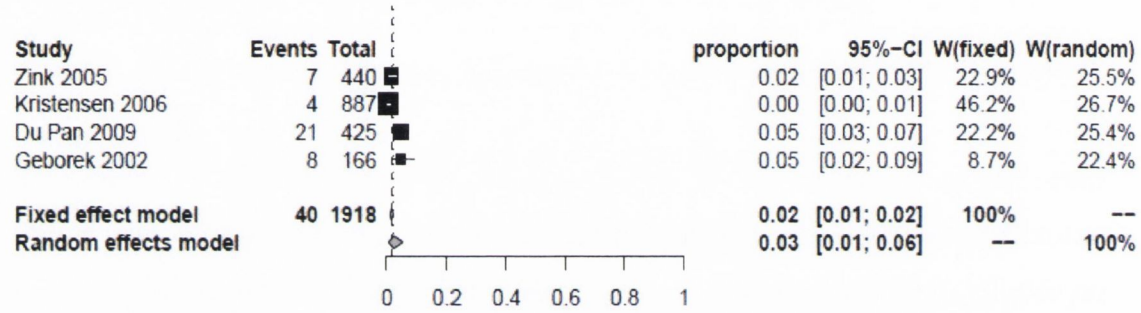


Figure 34: Forest Plot of proportion of patients withdrawing between 6 and 24 weeks due to toxicity for etanercept.
$I^2 = 85.8\%$

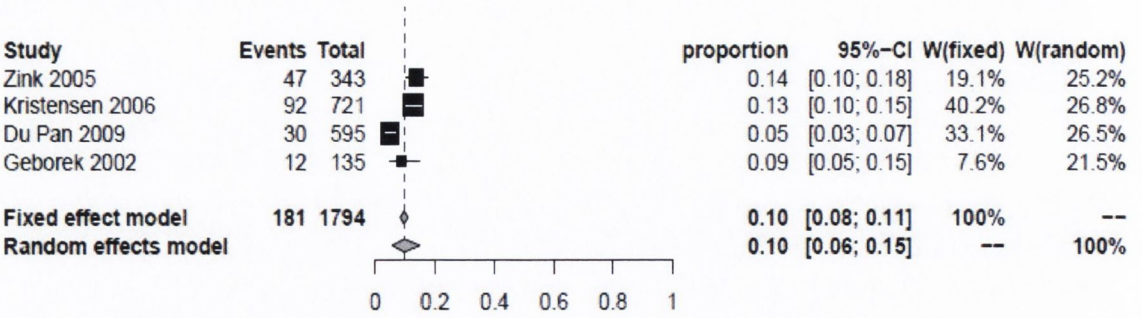


Figure 35: Forest Plot of proportion of patients withdrawing between 6 and 24 weeks due to toxicity for infliximab.
$I^2 = 90.4\%$

Withdrawal from anti-TNF agents due to inefficacy between 6 and 24 weeks

Withdrawal due to inefficacy is similar between the drugs. The pooled estimate of probability for anti-TNF's (etanercept, infliximab and adalimumab) is 0.04 (random effects model) (Figure 36). The probability of withdrawal for inefficacy is lowest for adalimumab (0.02) and but higher for both etanercept and infliximab (0.05). As with the previous analyses for short term withdrawal, adalimumab is informed by the least number of trials.

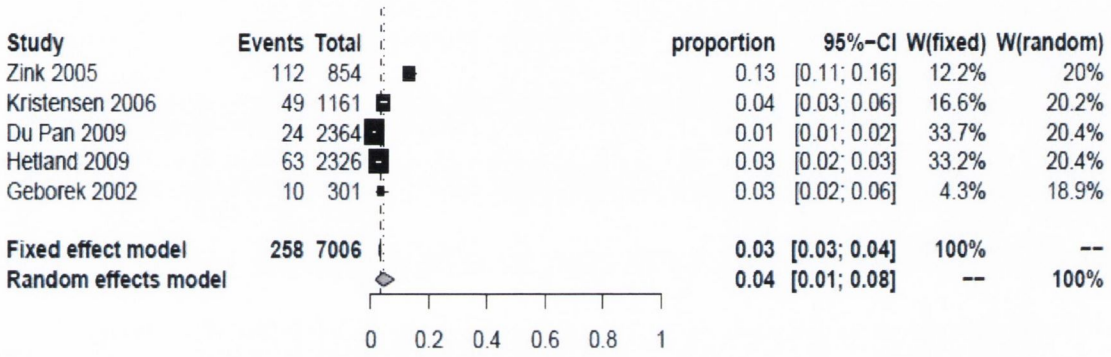


Figure 36: Forest Plot of proportion of patients withdrawing between 6 and 24 weeks due to inefficacy for all drugs.
(pooled data for etanercept, adalimumab and infliximab). $I^2 = 97.9\%$

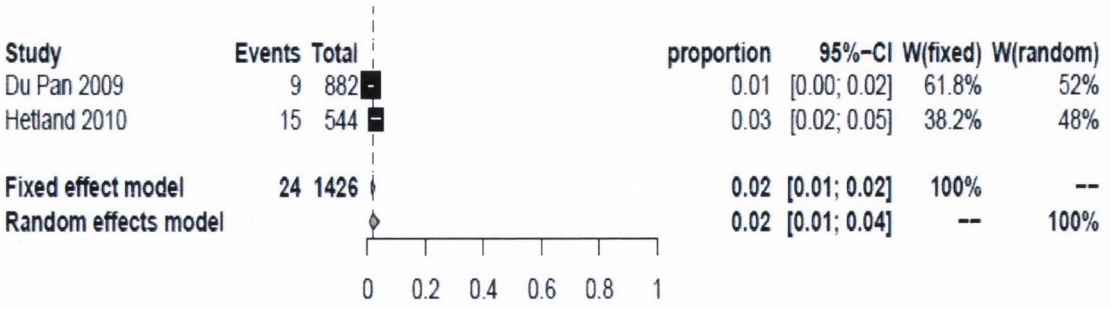


Figure 37: Forest Plot of proportion of patients withdrawing between 6 and 24 weeks due to inefficacy for adalimumab.
$I^2 = 82.7\%$

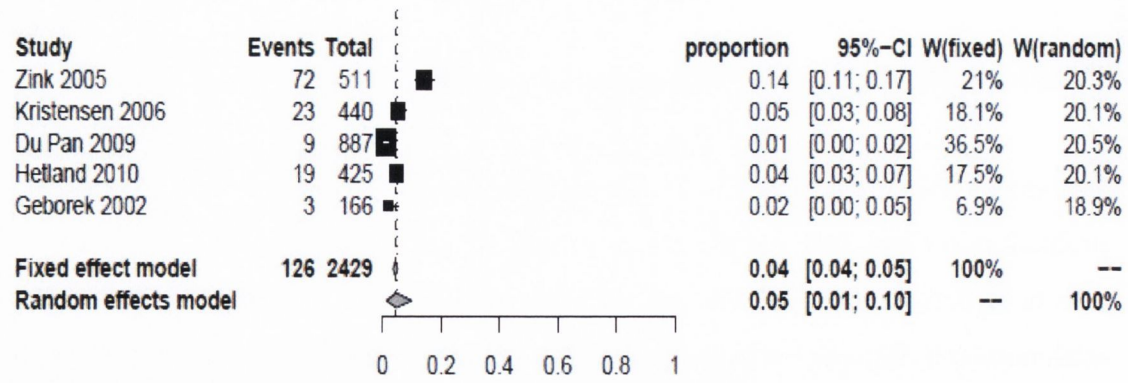


Figure 38: Forest Plot of proportion of patients withdrawing between 6 and 24 weeks due to inefficacy for etanercept.
I^2 = 96.3%

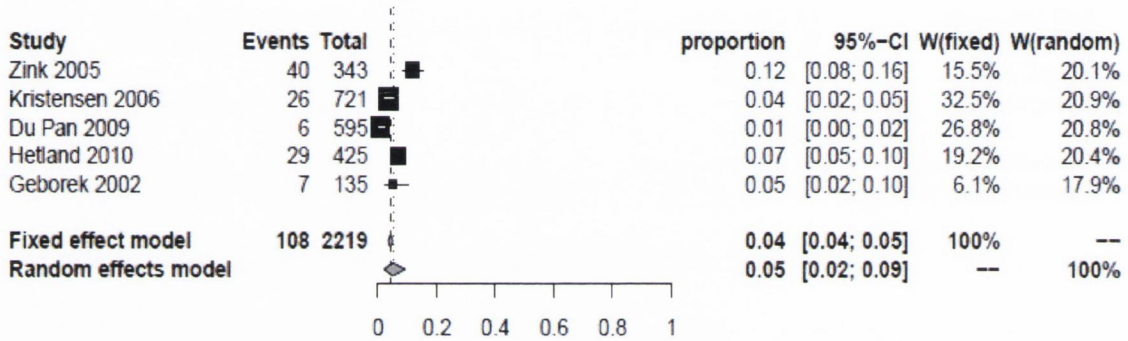


Figure 39: Forest Plot of proportion of patients withdrawing between 6 and 24 weeks due to inefficacy for infliximab.
I^2 = 93.4%

Total withdrawals between 6 and 24 weeks (any reason)

The BRAM handles the data as total withdrawals between 6 and 24 weeks for any reason and the proportion of these due to toxicity is also a data input. Therefore the withdrawals due to toxicity and inefficacy were combined (Figure 40). The results for the individual drugs are presented (Figure 41-44).

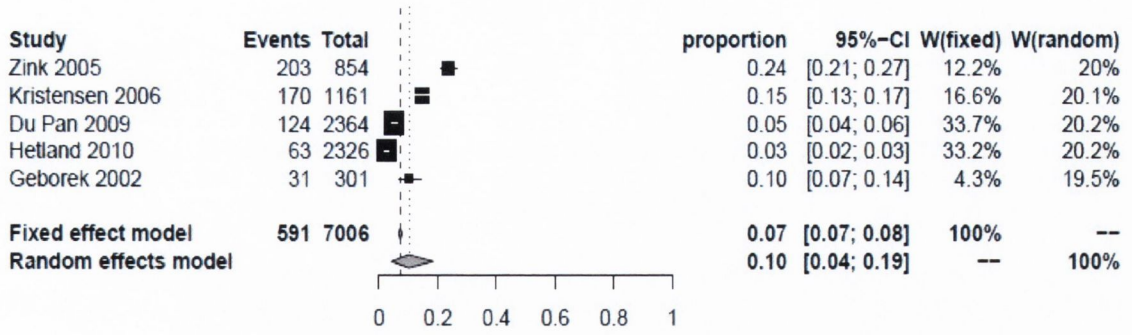


Figure 40: **Forest Plot of proportion of patients withdrawing between 6 and 24 weeks for any reason for all drugs.**
(pooled data for etanercept, adalimumab and infliximab). $I^2 = 99\%$

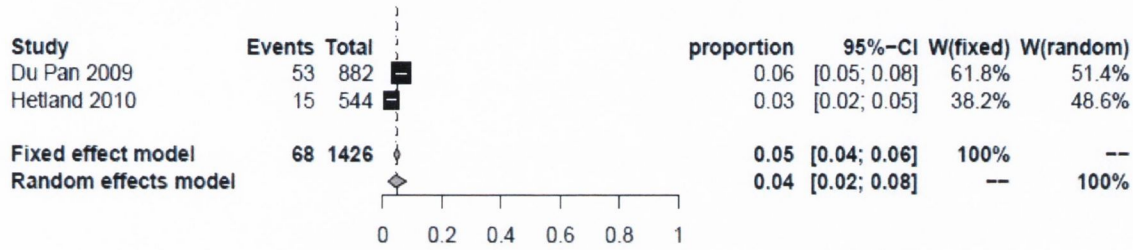


Figure 41: **Forest Plot of proportion of patients withdrawing between 6 and 24 weeks for any reason for adalimumab.**
$I^2 = 88.2\%$

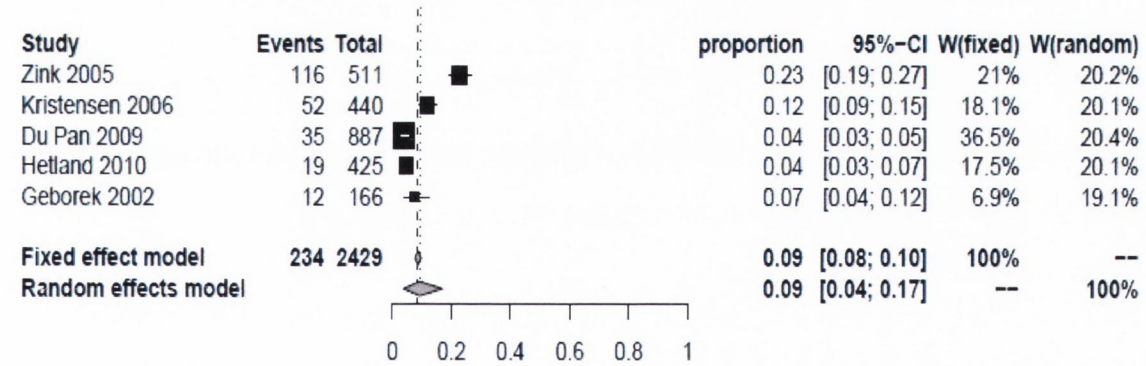


Figure 42: **Forest Plot of proportion of patients withdrawing between 6 and 24 weeks for any reason for etanercept.**
$I^2 = 97\%$

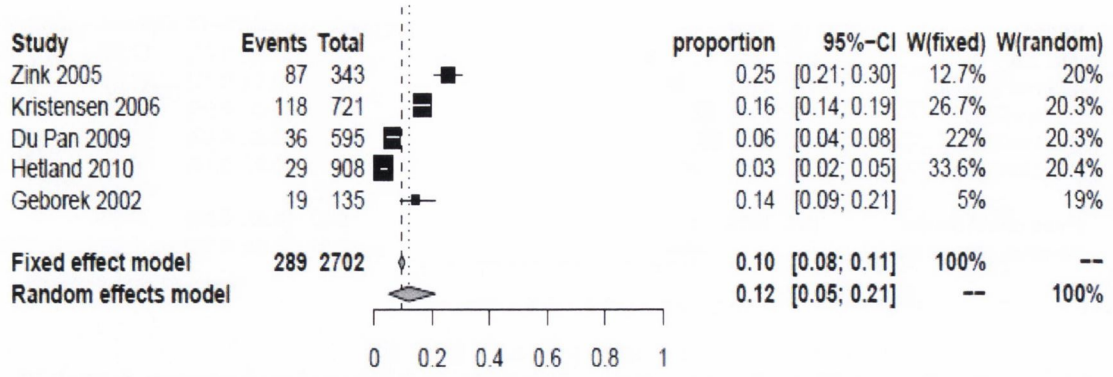


Figure 43: Forest Plot of proportion of patients withdrawing between 6 and 24 weeks for any reason for infliximab.
$I^2 = 97.7\%$

5.4.3 Discussion

In the real world use of anti-TNF agents, non-response (both primary non-response and secondary drug resistance) and drug related adverse events are common issues. [294, 295] Kristensen *et al.* examined the possible predictors of premature treatment termination of these agents and found that patients on monotherapy and patients not receiving MTX were at increased risk of terminating therapy earlier.[288]

There appears to be a difference between the drug retention rates of the anti-TNF agents. Patients on infliximab therapy appear to be at a higher risk of earlier treatment termination as compared with adalimumab or etanercept. A number of reasons have been proposed for this difference in retention rates. These include chimeric autoantibody formation and infusion related reactions. Adverse effects and non-response are the most common reasons for drug discontinuation. Du Pan specifically examined the differences between anti-TNF agents and found little difference between the treatments when non-response was the reason for discontinuation. [289] However, when adverse effects were the reason given, differences were found between the agents. More patients on infliximab in the Du Pan study discontinued due to adverse events than etanercept and adalimumab (HR 1.4, 99% CI 1.003–1.96). Infusion related reactions associated with infliximab may

be explained by the chimeric nature of the molecule or the presence of human chimeric auto antibodies. [289]

The reasons for early withdrawal were not the focus of this analysis but would be relevant to a cost-effectiveness study examining sequential therapy from anti-TNF agents. Primary non-responders may be more susceptible to poor response to follow on therapy than those withdrawing due to adverse effects. [296] There is limited data on adalimumab withdrawal because of its later entry into the market and no data available on golimumab and certolizumab.

5.5 Conclusion

Evidence synthesis indicates that etanercept is the most effective drug for improving HAQ scores and certolizumab is the most effective drug at producing an ACR response. Long term discontinuation data indicates that patients remain on infliximab for the least amount of time and adalimumab the longest. The number of short term withdrawals is small overall but is greatest for infliximab.

Of particular interest in examining the effectiveness of these agents is whether the highly effective results seen in clinical trials are reproducible in clinical practice. A number of studies have examined this issue and have found that response rates are not as high in clinical practice. [81, 293, 297] There are many reasons suggested for these including eligibility criteria (in general RCT eligible patients have higher response rates than ineligible patients due to selection towards high disease activity) and the use of co-medication in RCTs. When using RCT trial data for modelling chronic disease, clinical efficacy data from RCT's will be an ongoing source of uncertainty. The practice of combining observational data with RCT data in a Bayesian framework, where observational data is used as a prior is one solution to this. While such analyses have been done the methods are still being developed. [298]

Chapter 6

Chapter 6.0 Modelling the cost-effectiveness of anti-TNF drugs in RA

The cost-effectiveness of the anti-TNF agents has been the subject of many HTAs, which inform decision making, usually from payers perspective. A review of these assessments indicates that these agents are cost-effective under certain thresholds, but not all. This chapter draws from the analysis presented in previous chapters (primarily chapters four and five) in order to perform an analysis of the cost-effectiveness of these agents for the treatment of RA patients in Ireland.

The assessment differs from previous work in so far as it includes all the anti-TNF agents currently reimbursed in Ireland. Previous assessments have been in the form of single technology appraisals, where company dossiers for golimumab and certolizumab have been critically reviewed. An independent economic evaluation multiple treatment assessment has not been performed to date in Ireland. This assessment examines the use of each anti-TNF agent followed by a sequence of DMARD therapy, once there is an inadequate response. The comparator arm is a stand alone DMARD (leflunomide) followed by a sequence of DMARDs, once there is an inadequate response.

6.1 Research question and objectives

Research Question:

Are each of these anti-TNF agents cost-effective when used in RA, for patients who have not adequately responded to MTX compared to an alternative DMARD treatment in the Irish setting?

Objectives of the assessment:

- To calculate the ICER of each anti-TNF agent in comparison to leflunomide.
- To estimate the comparative effectiveness of these agents in the treatment of established RA in MTX non-responders.

- To calculate the gross health gain and resource use associated with the use of anti-TNF agents.
- To present the relative cost-effectiveness of the anti-TNF agents on a cost-effectiveness plane.
- To explore the uncertainty associated with the assumptions made for the model and how analysis from previous chapters has impacted on this.

Note: Sequential biological DMARD therapy is not assessed in this work.

6.1 RA Model Overview

The Birmingham Rheumatoid Arthritis Model (BRAM) was chosen for this analysis as the model has now been used for three different assessments for NICE and has undergone considerable review since its original development. [77, 111, 213, 299] The BRAM was initially developed for the cost-effectiveness evaluation of etanercept and anakinra in rheumatoid arthritis. [299] The initial model (Birmingham Preliminary Model (BPM)) was built in TreeAge Data 3.5. The model was run over a patient's lifetime in line with the chronic nature of RA. Patients first entered the model from the time of initiating DMARD therapy; DMARD therapy was not confined to one drug, rather a sequence of treatments, as would be the case in clinical practice. The model differed from Markov models in that it used tracker variables to alter the time spent on a particular drug, influenced by clinical course of their disease and medical history. The tracker variables also allowed events to occur at any time and not at a fixed cycle time as is the case in Markov models.

Anti-TNF therapy was included in the sequence of DMARD therapies and the patient can either have a sequence of including anti-TNF or not. This represents the two pathways compared; treatment with DMARD therapy, failure of DMARD therapy and addition of anti-TNF or treatment with DMARD therapy, failure of DMARD therapy and switch to alternative DMARD therapy. The patient pathway is presented in Figure 44.

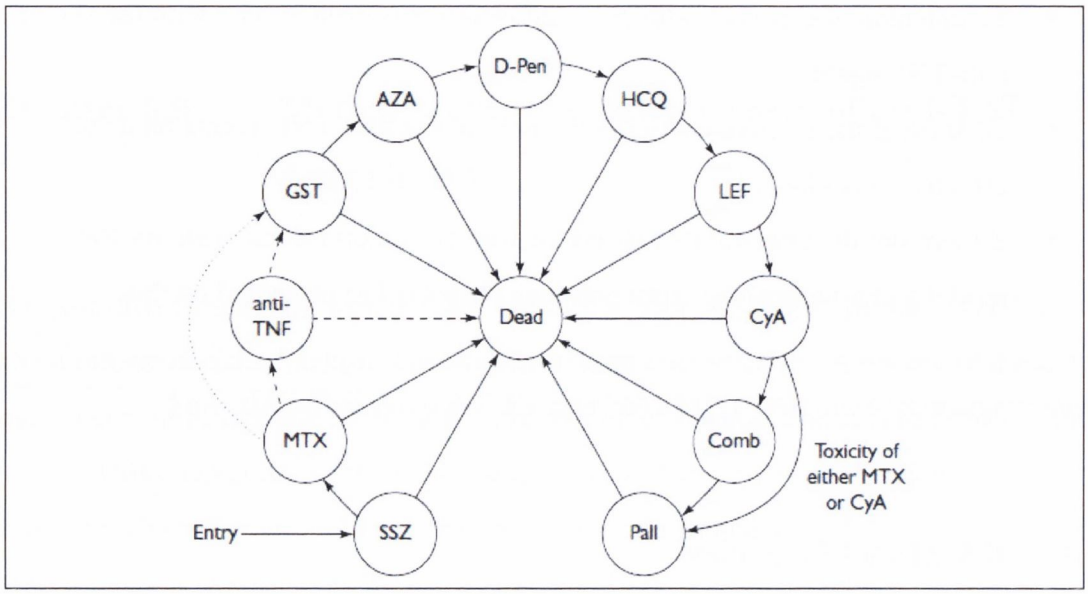


Figure 44: Patient Pathway in the BPM [299]

SSZ *Sulphasalazine,* ***MTX*** *MTX,* ***GST*** *Gold,* ***AZA*** *Azathioprine,* ***D Pen*** *Penicillamine,* ***HCQ*** *Hydroxychloroquine,* ***LEF*** *Leflunomide,* ***CyA*** *Cyclosporin,* ***Comb*** *Combination DMARD,* ***PALL*** *Palliative care*

There were a number of limitations to the BPM broadly divided into structural and data limitations. The model assumed that the effect pattern on QOL by DMARDs was fixed and therefore possible effects of DMARDs on mortality were not accounted for. It did not account for influence of DMARDs on joint replacement, hospitalisation, or disease progression.

The Birmingham Rheumatoid Arthritis Model (BRAM)

The BRAM is an individual sampling model, which assesses the cost-effectiveness of adding a TNF-alpha inhibitor to a sequence of DMARDs when compared with the same sequence of DMARDs without a TNF-alpha inhibitor. In the model, the initial age and sex distributions, as well as the starting distribution of HAQ scores, were based on observational data from the Norfolk Arthritis Register, a primary-care-based cohort of patients with inflammatory polyarthritis. [300] Change in HAQ score was modeled as a multiplier of the starting HAQ score and both were sampled from distributions rather than being constant. Utilities were estimated based on a mapping

process whereby HAQ scores in the trial were mapped via an algorithm to EQ-5D scores in order to derive estimates of utility. The model included a proportion of people stopping treatment at 24 weeks due to toxicity and lack of efficacy. Joint replacement and associated costs were included in sensitivity analyses.

RA is a chronic disease characterized by periods of response to treatment followed by unpredictable loss of therapeutic effect and switches to alternative treatments. The model, therefore, compared sequences of treatments rather than single agents. The individual patient simulation model follows a structure described by Bansback, with individual patients' outcomes sampled at 6-month intervals. [39] For the analysis, 1000 patients were randomly simulated to experience several alternative sequences of therapies. Patients were entered into the model at a baseline degree of disability, represented by a score of 1.5 on the HAQDI, which was the mean baseline score of patients in the adalimumab PREMIER trial. [301] The PREMIER trial was the first head-to-head trial of MTX-naïve patients with early RA (< 3 yrs). PREMIER evaluated a TNF antagonist plus MTX versus the TNF antagonist alone and versus MTX alone. [301] As patients progress through sequenced therapy, the HAQ score was modelled to deteriorate over time, with periods of response to treatment bringing the benefit of a slower rate of disease progression. At each initiation of a new therapy, patients who responded were modelled to also receive a one-time reduction in HAQ score, a benefit that was retained until loss of efficacy occurred.

6.1.1 The BRAM 2009

The BRAM 2009 has been updated to incorporate a non-linear function for the relationship between HAQ and utility and also further coding has been added to allow for probabilistic sensitivity analysis. The model simulates virtual patient histories and these are associated with differing costs and QALYs.

The BRAM has been used in a number of assessments for RA by NICE. The BRAM offers a number of advantages. It does not have defined cycles in which a patient must remain once a responder or non-responder. This alleviates the problem of a patient remaining in a cycle of non-response with associated accrual of cost and utility decrement. The use of a continuous measure such as the HAQ to assign disease

change is more sensitive to changes than a categorical measure such as the ACR response. The incorporation of the HAQ as a multiplier, thereby 'weighting' the disease response against their baseline disease activity incorporates the influence of this variable into the disease response estimate. As regards adapting the model, it is a relatively easy model to adapt. As mentioned the BRAM has been subjected to much review through the assessments commissioned by NICE; as a result improvements have been made through practical testing of the model.

The BRAM has also been subjected to criticism. This has mainly been in relation to the design and transparency of the model for those who are not familiar with it. The model is not intuitive and requires considerable explanation as to the workings. In this case I was fortunate to have been able to visit the models developer, Dr. Pelham Barton, for training on the BRAM.

6.2 Methods

6.2.1 The Model

The model was used with the permission of Dr. Pelham Barton, University of Birmingham, who developed it. While the inputs for the model were derived from local Irish data, the basic structure was retained (Figure 45). The model assumes that patients entering will have a certain sequence of treatments. The patient initiates therapy, remains on it for a certain amount of time and then either dies or is a non-responder; in which case, they may then start on an alternative treatment. Patients enter the model following non-response to previous DMARD therapy. The final treatment sequence of the model is palliative care.

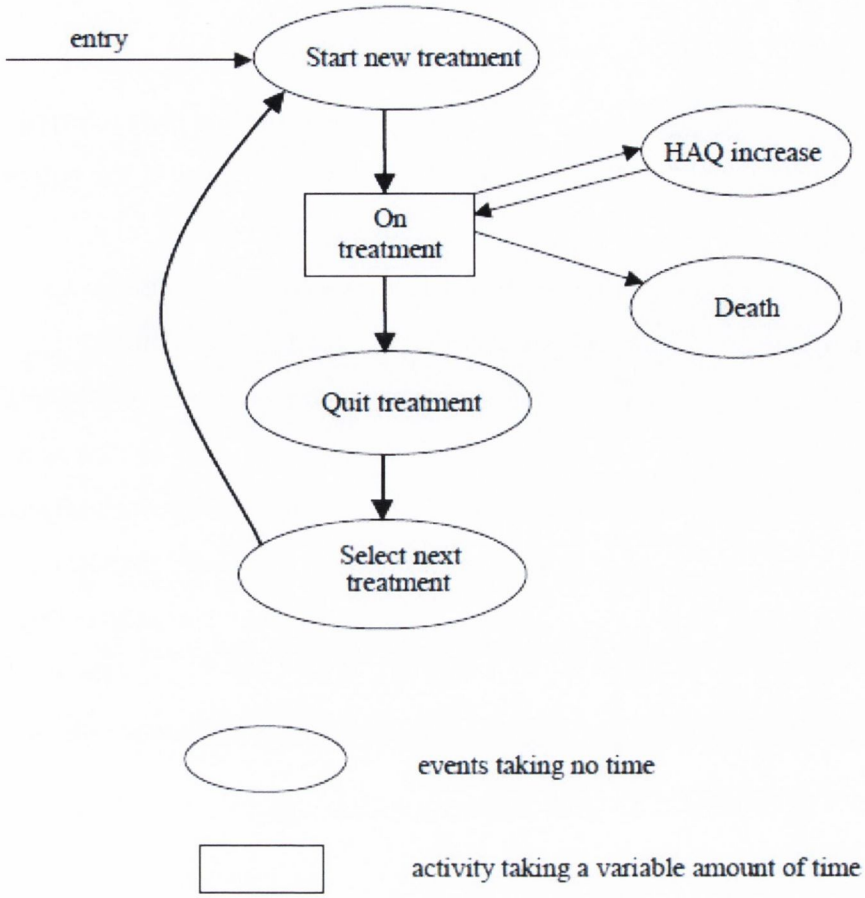


Figure 45: Structure of the BRAM 2009
(taken from Fig. 95, pg.207 of the HTA report) [213]

Model Adaptation:

The BRAM model code is written in Delphi programme language and is run by selecting parameter inputs from a series of comma separated value (CSV) files through a model interface. The interface of the model is used to select the number of iterations, size of patient cohort, number of options to be compared and alternative settings such as inclusion of offset costs or setting the model to disallow negative utility values. This version of the BRAM has been adapted to the Irish setting (Table 23).

Strategies used:

The current practice in Ireland is to use DMARD therapy first line and the preferred choice is MTX. If a patient stops responding to MTX or cannot tolerate it, the patient may be tried on an alternative DMARD treatment, or they may be started on a biologic agent such as an anti-TNF. The question of interest for this economic evaluation is whether anti-TNF agents (adalimumab, etanercept, certolizumab, golimumab, and infliximab) are cost-effective as compared to an alternative DMARD (leflunomide) following treatment failure with MTX. Leflunomide was chosen as a reasonable alternative following the failure of MTX. The population for the analysis includes patients who have had an inadequate response to conventional DMARD therapy (conventional DMARD therapy is considered to be MTX). The scope of this evaluation does not include sequential anti-TNF therapy. The patient enters the model following inadequate response to MTX. The patient then follows a sequence of treatments as described in Figure 46.

Table 22. Treatment sequences used in the model.

Strategy Name	**ADA**	**ETN**	**IFX**	**GOL**	**CZP**	**LEF**
1st ↓	ADA + MTX	ETN + MTX	IFX + MTX	GOL + MTX	CZP + MTX	LEF
2nd ↓	LEF	LEF	LEF	LEF	LEF	GST
3rd ↓	GST	GST	GST	GST	GST	CyA
4th ↓	CyA	CyA	CyA	CyA	CyA	AZA
5th ↓	AZA	AZA	AZA	AZA	AZA	Pall
6th	Pall	Pall	Pall	Pall	Pall	

*(**ADA** adalimumab, **ETN** etanercept, **IFX** infliximab, **GOL** golimumab, **CZP** certolizumab, **LEF** leflunomide, **GST** Gold (injectable), **CyA** ciclosporin, **AZA** azathioprine, **Pall** palliative care)*

The sequence follows that used for the NICE assessment using BRAM 2009.[302] The sequences were maintained the same for all anti-TNF treatments and therefore the efficacy of the agents beyond the anti-TNF treatment should be the same. It was difficult to establish a defined sequence of treatments which would be used in the

Irish setting as many rheumatologists will switch between anti-TNF treatments. Therefore a pragmatic decision was made to use the same non-biological DMARDs for all sequences. Sulphasalazine is excluded as it was established as the most commonly used 1st or 2nd line DMARD in RA and therefore patients would already have failed on sulphasalazine prior to starting on anti-TNF therapy. MTX is used in combination with anti-TNF only and is not used in DMARD cycles following failure of anti-TNF therapy or in the comparator cycle which starts with leflunomide. Leflunomide was not considered before anti-TNF therapy for this thesis as the current practice would be to choose an anti-TNF agent following the failure of MTX. In order to assign a comparator for these sequences leflunomide was chosen as the next most effective non-biological treatment after methotrexate. It has not been possible to adapt the model to include leflunomide before anti-TNF treatment. This would have required additional coding which would need to be carried out by the models developer, Dr. Barton.

Table 23. Comparison of BRAM 2006, 2009 and Irish adaptation

Parameters	BRAM 2006	BRAM 2009	BRAM 2009 for Irish Setting
Comparator	DMARD sequence	DMARDs sequence (initiating with leflunomide) and each other	DMARD sequence (initiating with leflunomide) and each other
Utility	Mapped HAQ to EQ-5D utility (UK cohort)	Mapped HAQ to EQ-5D utility (UK cohort)	Mapped HAQ to EQ-5D, SF-6D and a revised EQ-5D (Irish cohort)
Model for Mapping	Linear	Linear incorporating quadratic term	Linear (sensitivity analysis with quadratic)
Utility Measure	EQ-5D	EQ-5D	EQ-5D (original and revised scoring) and SF-6D
HAQ change on initiation of biologic	Data from individual RCTs – HAQ improvement modelled as a multiplier of the baseline HAQ. DMARD data from observational studies.	Data from observational studies – HAQ improvement modelled as for 2006. DMARD efficacy data from 2006 model used but halved due to differing patient cohort	MTC of HAQ improvement data from RCT's with inclusion of baseline HAQ in the MTC model. Estimate for leflunomide from Kremer 2002; other DMARD used as for BRAM 2009
HAQ change on treatment	0.03 on biologic treatments (as for general population) 0.045/year on conventional DMARDs 0.06/yr on PC	0 on biologic treatments 0.045/year on conventional DMARD 0.06/year on PC	0 on biologic treatments 0.045/year on conventional DMARD 0.06/year on PC
Long term discontinuation	Data from observational cohort (either Geborek 2002 or the GPRD database)	Data for anti-TNF from BSRBR NICE submission (observational UK cohort)	MTA of Weibull estimates from observational data from 5 observational cohorts (Sweden, Germany, Denmark, Switzerland)
Short term discontinuation	Observational data (Geborek 2002) used for etanercept and infliximab; adalimumab assumed same as infliximab. Observational data used for DMARD estimates	Observational data (combined 2 studies for etanercept only Observational data used for DMARD estimates	MTA of registry data for each anti-TNF agent

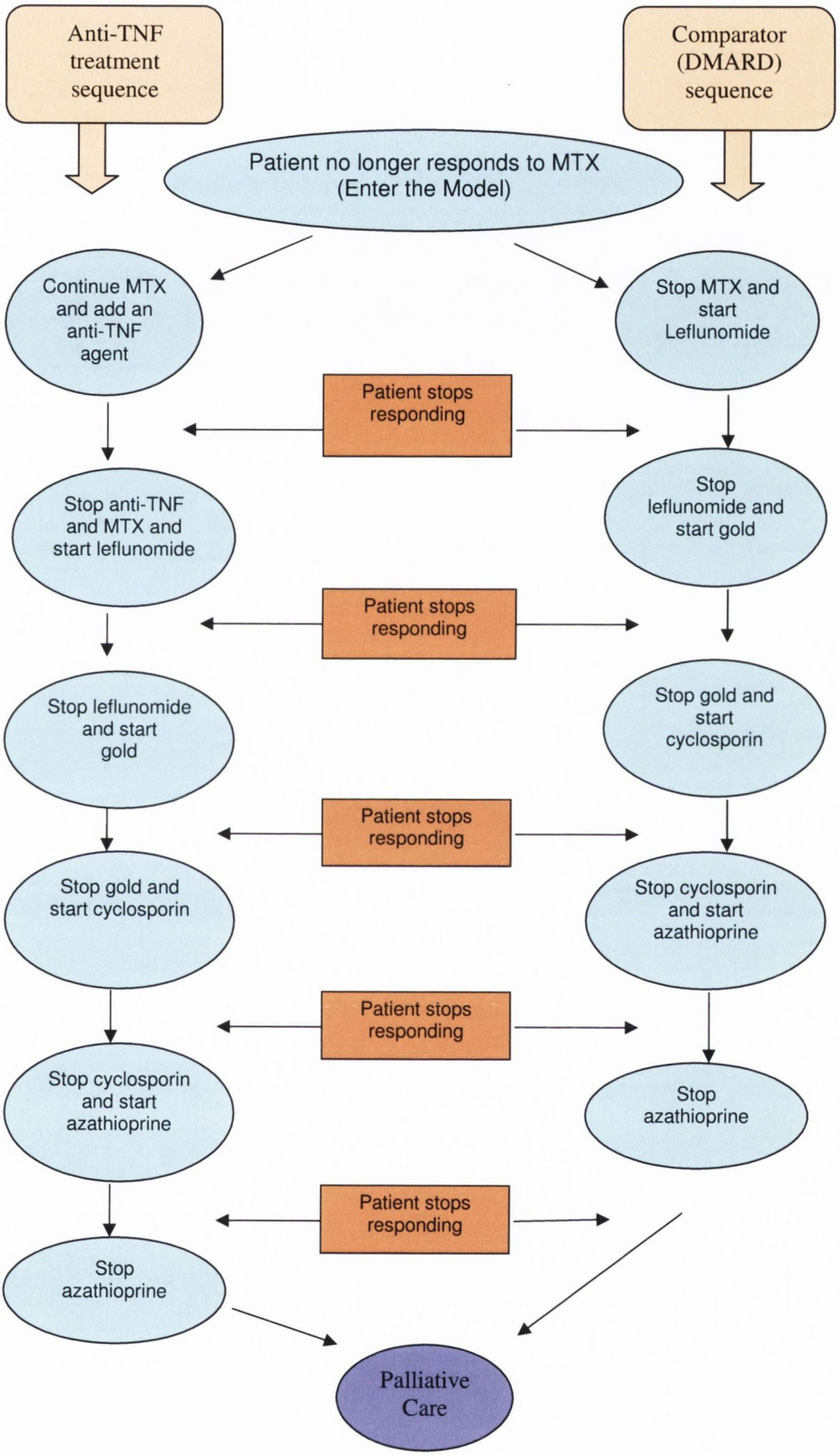


Figure 46: Treatment Sequence for patients

For the anti-TNF strategies, a patient starts on anti-TNF + MTX (1st cycle); and following failure of 1st cycle patient moves onto leflunomide (2nd cycle) and so on.

6.2.2 Outcomes

The model uses HAQ scores as an indicator of disease activity. Patients with more severe RA will have higher HAQ scores, and response to treatment can be assessed by a decrease in this score. The scale for HAQ is 0 to 3, 0 meaning no disability and 3, most disability. The score moves up and down in blocks of 0.125. When a patient in the model starts treatment and responds, the HAQ score can improve and then remain stable or it can decline slowly over time; for anti-TNF therapy the former is assumed, and for DMARDs the latter. When a patient stops responding to therapy the HAQ score will drop back to the original starting HAQ score (prior to starting therapy). Figure 47 illustrates the trajectory of HAQ for a patient.[213]

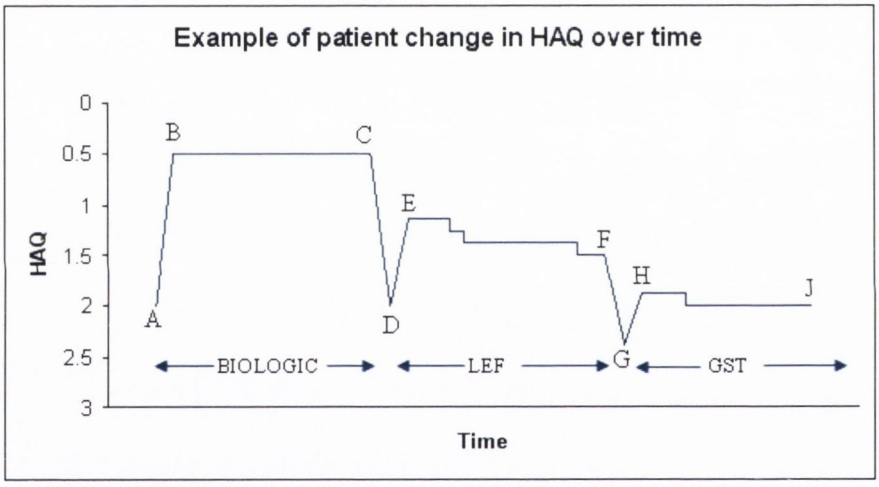


Figure 47: Trajectory of HAQ change while on treatment.
Initial improvement on a biological agent (AB) is lost on quitting the treatment (CD). A smaller improvement (DE) on starting LEF is similarly lost on quitting (FG) and followed by a gain (GH) on starting GST. In this case the patient dies of other causes (J) while still responding to GST. There is a gradual deterioration in HAQ from E to F and from H to J, but not from B to C in the reference case analysis. In some cases, the time spent on a conventional DMARD is not long enough for any deterioration in HAQ to occur.

(Source: National Institute for Clinical Excellence. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor. NICE technology appraisal guidance 195. 2010)

6.2.3 Data Inputs

6.2.3.1 Utility Estimates

Disease changes are driven by changes to patients' HAQ scores, and therefore it is the HAQDI and utility mapping coefficients that were used as inputs (See chapter 4). The utility estimates were calculated using a mapping coefficient from HAQ to utility. The model incorporated a linear equation to model the utility change as mapped from the HAQ. The 2009 BRAM used a quadratic equation to describe the relationship between utility and HAQ. In the Irish case, there was no significant statistical difference between the quadratic model and the linear model in this analysis (chapter 4). The quadratic model was used in the sensitivity analysis and the coefficient estimates are presented (Table 24).

The mapping coefficients were derived from a cohort (n=345) in which EQ-5D, SF-6D and HAQ were recorded as part of clinical monitoring of disease response to treatments.[175] The cohort has been described in detail in chapter 4. Three different utility measures were used against HAQ (EQ-5D preference based estimates, a revised EQ-5D scoring method and SF-6D estimates).[303] The mapping coefficients for the Irish biologic cohort are shown in Table 24.

Table 24. Linear Regression models between HAQ and Utility in RA

	Mean	95% CI	P-value	Std. Error	R^2
EQ-5D Linear					
Coefficient	0.792	(0.736, 0.848)	<0.001	0.028	
HAQ Constant	-0.236	(-0.277,-0.196)	<0.001	0.020	0.270
SF-6D Linear					
Coefficient	0.669	(0.652,0.686)	<0.001	0.009	
HAQ Constant	-0.084	(-0.097, -0.072)	<0.001	0.006	0.395
Revised EQ-5D Linear*					
Coefficient	0.822	(0.783, 0.861)	<0.001	0.020	
HAQ Constant	-0.168	(-0.196, -0.140)	<0.001	0.014	0.280

* Basecase

This mapping study is based on a cohort of 345 patients with RA. Data was collected in a biologic clinic for all patients prior to starting a biologic and at follow-up. Patients had high disease activity- mean DAS28 score of 5.39 at baseline, mean HAQ 1.3.

6.2.3.2 Initiating patient data

The demographics of the Irish rheumatology cohort (used for describing the initial patient cohort) were similar to those used for the UK model, and therefore they were deemed appropriate to use (Table 25). [213] The initial patient cohort describes patients as they enter the model i.e. immediately prior to starting anti-TNF therapy. A distribution of initiating HAQ scores was obtained from the Irish cohort and this was used in the model (Table 26).

Table 25. Initial age and gender distribution

Age (yrs)	15-24	25-34	35-44	45-54	55-64	65-74	75-84	Total %
Male	1%	2%	3%	5%	9%	7%	0%	27
Female	3%	3%	9%	19%	24%	12%	3%	73
Total	4%	5%	12%	24%	33%	19%	3%	100%

Table 26. HAQ distribution of starting HAQ scores

HAQ	0	0.125	0.250	0.375	0.5	0.625	0.75	0.875	1.0
%	-	2.4	3.5	3.8	4.7	6.8	6.8	2.6	5.6
HAQ	**1.125**	**1.250**	**1.375**	**1.5**	**1.625**	**1.75**	**1.875**	**2.0**	**2.125**
%	5.6	5.6	8.2	6.1	4.9	6.6	4.7	3.5	4.0
HAQ	**2.25**	**2.375**	**2.5**	**2.625**	**2.75**	**2.875**	**3.0**		
%	2.4	0.5	1.9	0.7	n/a	0.2	0.2		

6.2.3.3 HAQ Multiplier

The response to each treatment is incorporated via the use of a HAQ multiplier.

To account for this the HAQ improvement outcome is modelled as a multiplier (m) to the baseline HAQ score on study arm level. If, for example, this multiplier is estimated to be m=0.23, the estimated improvement in HAQ score for a patient with an initial HAQ score of 2 is therefore calculated as follows:

HAQ improvement = m x Baseline HAQ = 0.23*2 = 0.46

This method of estimating the HAQ multiplier represents the percentage improvement from baseline disease severity. A full description on the calculation of the HAQ multiplier using mixed treatment comparison methods, and subsequent results is given in chapter 5. For the purposes of PSA the beta distribution is fitted to the intervals obtained from this analysis (Table 27). The estimate used to calculate the HAQ multiplier for leflunomide was from one published paper. Due to limited data being available for leflunomide in patients who have failed methotrexate or in anti-TNF RCTs, it was only possible to use this estimate. Using a single estimate which has not been incorporated into the MTC is treating it as an estimate from unadjusted meta-analysis (randomisation has not been maintained). This in turn could either under or overestimate the effect of leflunomide. For this reason the HAQ multiplier has been varied to a mean estimate of 0.25 for leflunomide to reflect the uncertainty around this parameter. While the efficacy evidence from the registry did not indicate that anti-TNF agents were significantly more effective than leflunomide, it was pragmatic to explore some reduction in benefit.

Table 27. Beta distributions for HAQ multipliers (point estimates)

Treatment	α	β	Mean	Source
Adalimumab	57.81	225.46	0.20	MTC Random Effects
Infliximab	5.69	49.66	0.10	MTC Random Effects
Etanercept	24.91	55.15	0.31	MTC Random Effects
Golimumab	14.75	50.62	0.23	MTC Random Effects
Certolizumab	50.66	148.98	0.25	MTC Random Effects
Leflunomide	25.05	58.46	0.3	Kremer 2002 *calculated from 95% CI from published paper
Gold (Injectable)	0.225	0.925	0.20	Taken from NICE report 2009 – effectiveness halved from values used in 2006 report
Cyclosporin	0.065	0.325	0.17	
Azathioprine	0.10	0.9	0.10	

For probabilistic sensitivity analysis, the values a and b are drawn from normal distributions with standard deviation 0.1 times the point estimate.

6.2.3.4 Continuation times on treatments

Short term (early quitting)

Short term quitting is defined as stopping treatment before 24 weeks. This may be due to either toxicity or inefficacy. Patients may fall into one of four categories defined as; stopping before 6 weeks due to toxicity; stopping 6-24 weeks due to inefficacy; stopping between 6-24 weeks for toxicity; or remaining on therapy. The proportion of patients in each category is derived from a meta-analysis of the registries up to 24 weeks (see chapter 5) and is presented (Table 27).

Table 28. Probability of short term withdrawal for anti-TNF agents

Treatment	Withdrawal ≤ 6 weeks (Toxicity) *(95% CI)*	Total withdrawals between 6 and 24 weeks (95% CI)	Withdrawal 6-24 weeks (Toxicity) *(95% CI)*	Withdrawal 6-24 weeks (Inefficacy) *(95% CI)*	Source
Adalimumab	0.0187 *(0.0000; 0.0701)*	0.0435 (0.0175; 0.0806)	0.0499 *(0.0365; 0.0664)*	0.0182 *(0.0051; 0.0393)*	Meta-analysis (random effects) of observational studies*
Infliximab	0.0275 *(0.0059; 0.0642)*	0.1188 (0.0493; 0.2131)	0.0994 *(0.0572; 0.1516)*	0.0515 *(0.0210; 0.0945)*	
Etanercept	0.0257 *(0.0064; 0.0571)*	0.0924 (0.0361; 0.1712)	0.0564 *(0.0314; 0.0882)*	0.0464 (0.0122; 0.1011)	
Golimumab	0.0236 *(0.0060, 0.0522)*	0.1022 (0.0414; 0.1859)	0.0781 (0.0439; 0.1211)	0.0420 (0.0148; 0.0823)	Pooled estimate for all drugs
Certolizumab	0.0236 *(0.0060, 0.0522)*	0.1022 (0.0414; 0.1859)	0.0781 (0.0439; 0.1211)	0.0420 (0.0148; 0.0823)	Pooled estimate for all drugs

*See chapter 5 for methodology and results

Table 29. Beta distributions for short term withdrawal

Withdrawal at 6 weeks	Adalimumab	Infliximab	Etanercept	Golimumab	Certolizumab
Mean	0.0187	0.0275	0.0257	0.0236	0.0236
SD	0.018	0.015	0.013	0.012	0.012
Beta	361.07	107.57	146.86	160.27	160.27
Alpha	83.05	3.04	3.87	3.87	3.87
Withdrawal 6-24 weeks for all reasons					
Mean	0.0435	0.1188	0.0924	0.1022	0.1022
SD	0.016	0.040	0.034	0.036	0.036
Beta	143.68	53.72	65.54	59.20	59.20
Alpha	6.53	7.24	6.67	6.74	6.74
Withdrawal 6-24 weeks due to toxicity					
Mean	0.0499	0.0994	0.0564	0.0781	0.0781
SD	0.01	0.024	0.014	0.019	0.019
Beta	653.99	147.68	272.71	192.41	192.41
Alpha	34.34	16.30	16.30	16.30	16.30

Long term Withdrawal

The length of time on treatments was gathered from observational studies. A full description of the methods and results of this analysis are also given in Chapter 5.

Table 30. Long term continuation on treatment (yrs)

Treatment	a	b	Overall Mean (yrs)	Source
Etanercept	0.83	4.71	5.20	MTA of observational studies
Infliximab	0.76	3.05	3.59	MTA of observational studies
Adalimumab	0.84	5.23	5.73	MTA of observational studies
Golimumab	0.79	3.97	4.54	Overall estimate used
Certolizumab	0.79	3.97	4.54	Overall estimate used
Overall anti-TNF	0.79	3.97	4.54	MTA of observational studies
MTX	0.51	15.73	30.35	GPRD database (NICE 2006 appraisal)[77]
Leflunomide	1	5.98	5.98	GPRD database (NICE 2009 appraisal)[213]
Gold (injectable)	0.48	1.81	3.91	GPRD database (NICE 2009 appraisal)[213]
Azathioprine	0.39	4.35	15.53	GPRD database (NICE 2009 appraisal)[213]

Normal distributions used for shape parameter a; lognormal for scale parameter b. If a< 1 the hazard decreases with time and if a > 1 the hazard increases with time. The MTA is described in chapter 5.

6.2.3.5 Costs

Costs are attached to each drug treatment and include drug acquisition costs and associated costs with the drug. The associated costs include screening and monitoring of treatment. The costs are initially higher for each of the drugs because of the additional screening that is undertaken prior to starting anti-TNF therapy and also for some drugs there are loading doses (Appendix 8). Therefore the costs are divided into two sections; start-up costs and on-going costs. This is calculated by estimating the costs in the first year and the second year separately and using the difference as the start-up cost.

The drugs included in the analysis are reimbursed under different reimbursement schemes and therefore the methods used to cost them are provided separately.

6.2.3.6 Treatment costs

6.2.3.7 Drug Costs

Doses

Where a range of doses is available the average dose is used for DMARDs; or in the case of MTX, the dose most commonly used in practice (based on a recent audit of MTX prescriptions from community pharmacies). If the dose of the drug is weight based, the dose for a 70 kilogram patient is used. Loading doses only apply to the anti-TNF agents and these are costed within the start up costs.

HTDS

Drugs reimbursed under this scheme included adalimumab, certolizumab, etanercept, golimumab and ciclosporin. Drug costs are calculated using the reimbursement price from the payer to the wholesaler. This price includes the ex-wholesale price (i.e. price to the wholesaler plus wholesale margin (10%) and a set patient care fee of €62.03 per patient per month (12 fees are included per year). Value added tax (VAT) is not included in the cost utility analysis. Dose escalations are not included in the model.

Reimbursement for certolizumab was granted in 2010 and the conditions of this reimbursement included a patient access scheme (PAS). The details and conditions of this scheme are not publicly available, and therefore the PAS has not been included in the basecase analysis. However a scenario is included in the results, which describes a PAS where the first three months (10 injections) of the drug are free to the payer.

GMS and DPS

Drugs reimbursed under these schemes include MTX (tablets and injection), azathioprine, leflunomide, prednisolone, and diclofenac. If generic brands are also available for a drug, then the average price of branded and generic is taken as the price for the model. These drugs include a dispensing fee which is paid to the pharmacist. The dispensing fee changed in 2009 to a sliding scale of fees which is not possible to apply as it based on number of items dispensed in individual pharmacies. For this reason a fee of €4.50 per item is added to the cost of the drug (average dispensing fee paid between September and December 2009).

There are differences between the GMS and DPS, in fees paid by the payer; the pharmacy reimbursement price for the GMS is determined from the pharmacy purchase price + dispensing fee and for the DPS the pharmacy reimbursement price comprises the pharmacy purchase price + 20% mark-up + dispensing fee.

6.2.3.8 Other Treatment Costs

Palliative care is calculated based on treatment with an anti-inflammatory (diclofenac 75mg twice daily), a corticosteroid (prednisolone 5mg once daily), and 7 days inpatient stay per year. There is only one inpatient rheumatology rehabilitation unit in the country, and this is the Bone and Joint Unit at Our Lady's Hospice, Dublin. A weekly cost for Our Lady's Hospice, is available under the Public Nursing Home costs available from the HSE. [304] A one week rehabilitation stay was agreed in consultation with consultant rheumatologists (St. Vincent's University Hospital).

6.2.3.9 Screening and monitoring costs

All patients must be screened before either a conventional DMARD or a biological DMARD can be initiated. A number of countries have guidelines for screening and monitoring of DMARD therapy (both non-biological and biological). The Irish Society for Rheumatology produced guidelines in 2006. [305] Screening measures before initiating biological therapies are outlined in Table 30.

Table 31. Screening and monitoring for anti-TNF therapy

Test	Unit Cost per test	Year 1			Year 2		Source
		Pre-initiation (Screening) (No of tests)	Frequency first year	Total Cost for first year	Frequency for subsequent years	Cost for subsequent years	
FBC	9.26	1	3	37.04	2	18.52	Dublin teaching hospital (SVUH) 2011
ESR	8.32	1	3	33.28	2	16.64	
U&E	7.85	1	3	31.40	2	15.70	
LFT	12.20	1	3	48.80	2	24.40	
CRP	12.74	1	3	50.96	2	25.48	
Urinalysis	0.51	1	3	2.04	2	1.02	
Hepatitis B screen	26.00	1	0	26.00	0	0	
Tuberculosis screen	22.21	1	0	22.21	0	0	
Chest x-ray*	112.40	1	0	112.40	0	0	
X-ray * (limbs, hip &shoulder)	146.12	1	0	146.12	0	0	
GP visit	25.10	0	4	100.40	4	100.40	Madden Report[306]
Outpatient visit	168.60	1	3	674.40	2	337.20	Ready Reckoner 2009**
Education (CNS)	36.33	1	0	36.33	0	0	1 hour CNS ***

Cost includes cost of reporting. average cost per case of outpatient attendance *** Mid-point of **CNS** (clinical nurse specialist) (DoH&C 2008) including 70/30 pay non pay rule and budget 2010 deductions for public sector pay (37.5 hours per week) **FBC** Full Blood Count, **ESR** Erythrocyte Sedimentation Rate **CRP** C reactive protein, **U&E** Urea and Electrolytes **LFTs** Liver Function Tests*

Table 32. Drug Costs (initial 12 months base case)

Drugs Costs	Mode of Administration	Unit cost (€)	Number of doses per year	Annual Costs (€)	Pharmacy Fee (€)	Total cost Monotherapy (€)	Anti-TNF+ MTX (€)
Adalimumab	Subcutaneous injection	562.59	26	14627.34	62.03	15371.7	15482.39
Etanercept	Subcutaneous injection	257.93	52	13412.23	62.03	14156.59	14267.28
Golimumab	Subcutaneous injection	1105.5	12	13266	62.03	14010.36	14121.05
Certolizumab	Subcutaneous injection	515.9	29	14961.1	62.03	15705.46	15816.15
Infliximab 3mg/kg (3 vials)	Infusion	2071.29	9	18641.61	0	22568.86	22679.55
Leflunomide 10-20mg tablets	Oral	2.26	365	823.56	4.3	875.16	-
Azathioprine 50mg (150mg dose)	Oral	0.876	365	319.74	4.3	371.34	-
MTX tablets 2.5mg (Lederle) 15mg	Oral	0.82	52	59.09	4.3	110.69	-
Gold IM injection (Myocrisin)	Oral	4.61	52	239.51	4.3	291.11	-
Cyclosporin 200mg (Neoral®) (based on 3mg/kg dose for 60-70kg)	Oral	5.99	365	2184.89	4.3	2236.49	-

Table 33. Drug Costs (Subsequent 12 months base case)

Drug and Dose	Form	Unit cost (€)	No. doses per year	Annual Costs (€)	Pharmacy Fee (€)	Total Costs (€)	Anti-TNF+ MTX (€)	Source/ Brand
Adalimumab 40mg	PFS	562.59	26	14627.34	62.03	15371.70	15482.39	PCRS April 2011
Etanercept 50mg	PFS	275.89	52	14346.28	62.03	15090.64	14267.28	PCRS April 2011
Golimumab 50mg	PFS	1,182.48	12	14189.76	62.03	14934.12	14121.05	PCRS April 2011
Certolizumab 200mg	PFS	515.90	26	13413.40	62.03	14157.76	14268.45	PCRS April 2011
Infliximab 3mg/kg (3 vials)	Vial	690.43	6.5	13463.39	0	13463.39	13463.40	MIMS 2011
Leflunomide 10-20mg tablets	Tab	2.26	365	823.56	4.30	875.16	-	PCRS May 2011
Azathioprine 50mg (150mg dose)	Tab	0.88	365	319.74	4.30	371.34	-	PCRS May 2011
MTX tablets 15mg	Tab	0.82	52	42.68	4.30	94.28	-	PCRS May 2011 Lederle
Gold IM injection (Myocrisin)	Vial	4.61	52	239.51	4.30	291.11	-	PCRS May 2011
Cyclosporin 200mg (Neoral®) (based on 3mg/kg dose for 60-70kg)	Oral	5.99	365	2184.89	4.3	2236.49	-	PCRS May 2011

Table 34. Start-up and annual costs associated with treatment

Treatment	Start-up (€)	Annual Use (€)	Source
Adalimumab	573.79	15956.15	PCRS 2011
Infliximab	11715.61	15938.73	MIMS 2010
Etanercept	573.79	15675.09	PCRS 2011
Golimumab	573.79	15518.57	PCRS 2011
Certolizumab	2121.49	15938.73	PCRS 2011
MTX	573.79	94.28	PCRS 2011
Leflunomide	573.79	875.16	PCRS 2011
Gold	573.79	291.11	MIMS 2009
Cyclopsorin	573.79	2820.94	PCRS 2011
Azathioprine	573.79	371.34	PCRS 2011
Palliative Care	0	2887.3	Nursing home costs + drug Tx costs

6.2.4 Sensitivity Analysis

Two forms of sensitivity analysis were carried out; one way sensitivity analysis and probabilistic sensitivity analysis. Scenarios varied were; length of time on anti-TNF therapy (use the same time for all agents), HAQ change while on anti-TNF therapy and when stopping therapy, mapping function from HAQ to utility using a quadratic term, mapping to the SF-6D and the EQ-5D original scoring method. The cost parameters varied were; inclusion of offset costs and inclusion of hospital discounts for hospital administered drugs.

6.2.4.1 Probabilistic Sensitivity Analysis

The PSA involved running the model for 1000 parameter sets for 5000 patients. For each simulation, the model selected a value for each parameter from its probability distribution. Standard deviations or standard errors are used as inputs to the model for the uncertainty associated with the point estimate. The model uses beta, normal and log-normal distributions. The choice of the appropriate distribution depended on the properties of the parameters. Beta distributions were calculated for Weibull distribution and short term discontinuations due to the binomial nature of the data. Lognormal was chosen for skewed distributions; the parameters of the beta distribution. The distribution on costs was fixed. Parameters were not correlated in the model.

6.2.4.2 One Way Sensitivity Analysis (OWSA)

The OWSA involves fixing all results other than the parameter being tested in the sensitivity analysis. The analysis here changed the parameters individually (Table 40) and the deterministic model results were recorded. The deterministic model uses only the point estimates of the inputs.

6.2.5 Discounting

Both costs and utilities are discounted at an annual rate of 4% for all patients in the model. This discount rate is based on an estimate of the Social Rate of Time Preference in Ireland and is recommended in the Economic Evaluation of Health Technologies in Ireland Guidelines.[15] Discounting varies from country to country and from time to time depending on the economic status of the country; therefore this rate is varied in the sensitivity analysis to a rate of 0% and 6% for both costs and outcomes.

6.3 Results

Results are presented for the basecase model (Table 35 and Table 36). Both the deterministic and mean of the PSA are presented. The BRAM uses a cohort of 5000 patients for the deterministic and probabilistic models. However there is considerable variation in the results produced by the deterministic model at this sample size. In order to reduce this variation, the cohort size was increased to one million patients to give a confidence interval width for the ICER of approximately €3,000. For the probabilistic model the cohort was maintained at 5000 as the uncertainty associated with the patient cohort is less than that associated with the parameters. In order to establish congruency for the PSA results, five PSAs were compared by calculating a standard deviation between the mean results of the each of the PSAs. As the BRAM uses random numbers to select a parameter set, a fixed seed setting is included in the interface to ensure consistency in the results from the PSA. For the comparison between PSAs this fixed seed was varied for each of the five PSAs. This comparison is shown in Appendix 9.

Results were obtained for each strategy set described. An ICER is calculated for each drug against leflunomide. The lifetime discounted costs and QALYs for each drug from the deterministic model are presented (Table 374); Table 36 shows the mean costs and QALYs for each drug calculated from the mean of the PSA model; the mean difference in costs and QALYs between the anti-TNF agents and DMARDs are shown in table 36.

All anti-TNF agents were associated with QALY gains when compared to the alternative DMARD, leflunomide. The DMARD option incurs the least cost and the least benefit. Adalimumab incurs the most incremental costs of the anti-TNF agents, followed by etanercept. Infliximab accrues the least number of QALYs of the five anti-TNF agents. The credible intervals for the mean QALY are wide, representing uncertainty in the quantity of benefit of these agents.

Table 35. Basecase Deterministic Results (1 million patients)

Comparison	Diff Cost (€)	Diff Utility	ICER (95% Credible Interval)
ETN - DMARD	56323	0.54	104000 (102000,105000)
CZP - DMARD	48593	0.41	118000 (117000,120000)
GOL- DMARD	49784	0.38	130000 (130000,132000)
ADA – DMARD	64195	0.44	145000 (143000,147000)
IFX - DMARD	53855	0.19	277000 (269000,285000)

Table 36. Basecase results of mean costs and QALYs (taken as a mean of results of PSA)

Treatment	Mean Cost (€)	95% Credible Interval		Mean QALY	95% Credible Interval	
Adalimumab	94570	74388	119118	8.246	7.117	9.368
Etanercept	86259	72137	101410	8.345	7.269	9.489
Infliximab	83625	78945	88226	7.996	6.816	9.128
Certolizumab	78459	73524	82867	8.212	7.061	9.336
Golimumab	79576	74846	83973	8.182	7.034	9.342
DMARD (leflunomide)	29775	26862	32621	7.795	6.599	8.976

Table 37. Basecase results of mean difference in costs and QALYs (mean results of PSA)

Anti-TNF treatment Vs. DMARDs	Mean Diff Cost (€)	95% Credible Interval		Mean Diff QALY	95% Credible Interval	
Etanercept	56500	43100	71800	0.55	0.365	0.767
Adalimumab	64800	45200	89600	0.45	0.285	0.643
Certolizumab	48700	44500	51700	0.42	0.31	0.52
Golimumab	49800	50000	53100	0.39	0.28	0.49
Infliximab	53900	50000	57000	0.20	0.111	0.289

The mean PSA ICERs for the basecase are presented in Table 38. All the ICERs are above the willingness to pay threshold of €20,000 and €40,000 per QALY.

Table 38. ICERs of anti-TNF agents against DMARD base case (Mean PSA results)

Comparison	ICER	95% Credible Interval		Probability of cost-effectiveness at threshold €40,000	Probability of cost-effectiveness at threshold €20,000
		Lower	Upper		
ETN – DMARD	102700	85300	131200	86%	22%
CZP – DMARD	117000	94500	155500	12%	61%
GOL- DMARD	129000	101200	175800	2%	14%
ADA – DMARD	143700	115400	185100	0%	2%
IFX – DMARD	268300	187200	480700	0%	0%

Each incremental value compares value of each anti-TNF to an alternative DMARD (Leflunomide). Discount applied of 4%.

The probability of cost-effectiveness for the basecase is presented in Figure 48. At a WTP up to €20,000, certolizumab has the highest probability of cost-effectiveness, followed by golimumab, adalimumab, and infliximab. At a WTP above approximately €40,000, etanercept has the highest probability of cost-effectiveness (Table 38).

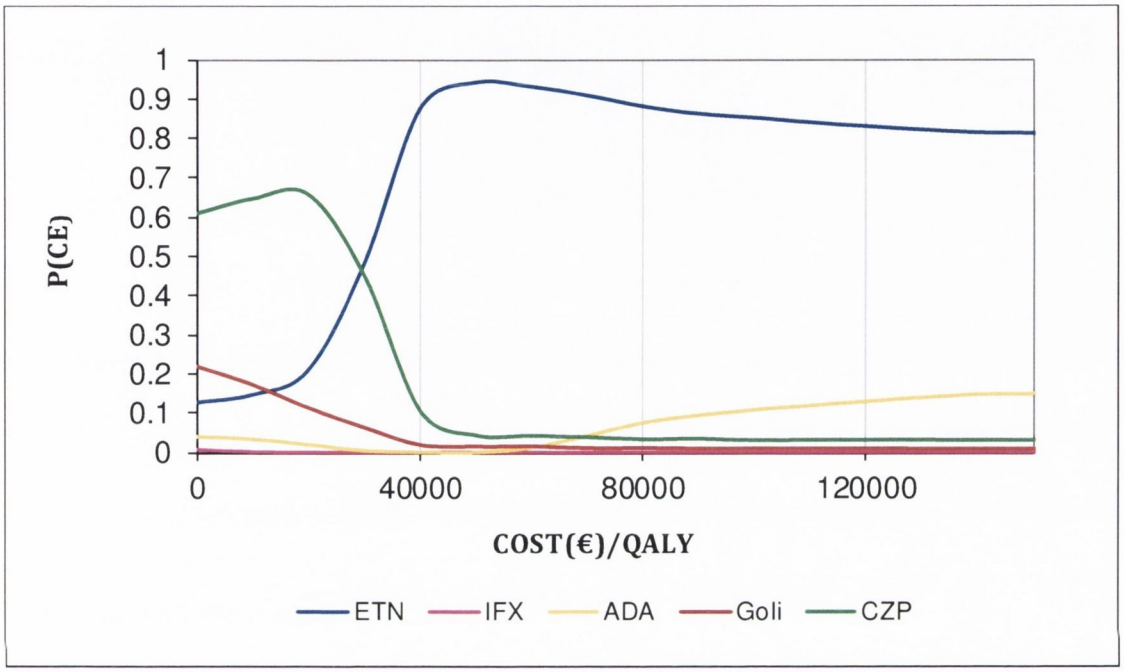


Figure 48: CEAC of each anti-TNF vs. all other anti-TNFs. The origin is represented by the leflunomide.

The cost-effectiveness plane for the basecase is presented both as a scatterplot and as mean estimates, represented by an ellipse on the plane (Figure 49 and Figure 50). Each of the agents overlaps; showing that the costs and benefits are similar for each of the anti-TNF drugs (Figure 49). Comparison of the mean does indicate there are some differences (Figure 50). The scatter around adalimumab and etanercept are greater than other agents which may be due to the longer amount of time patients remain on these drugs (Weibull estimates).

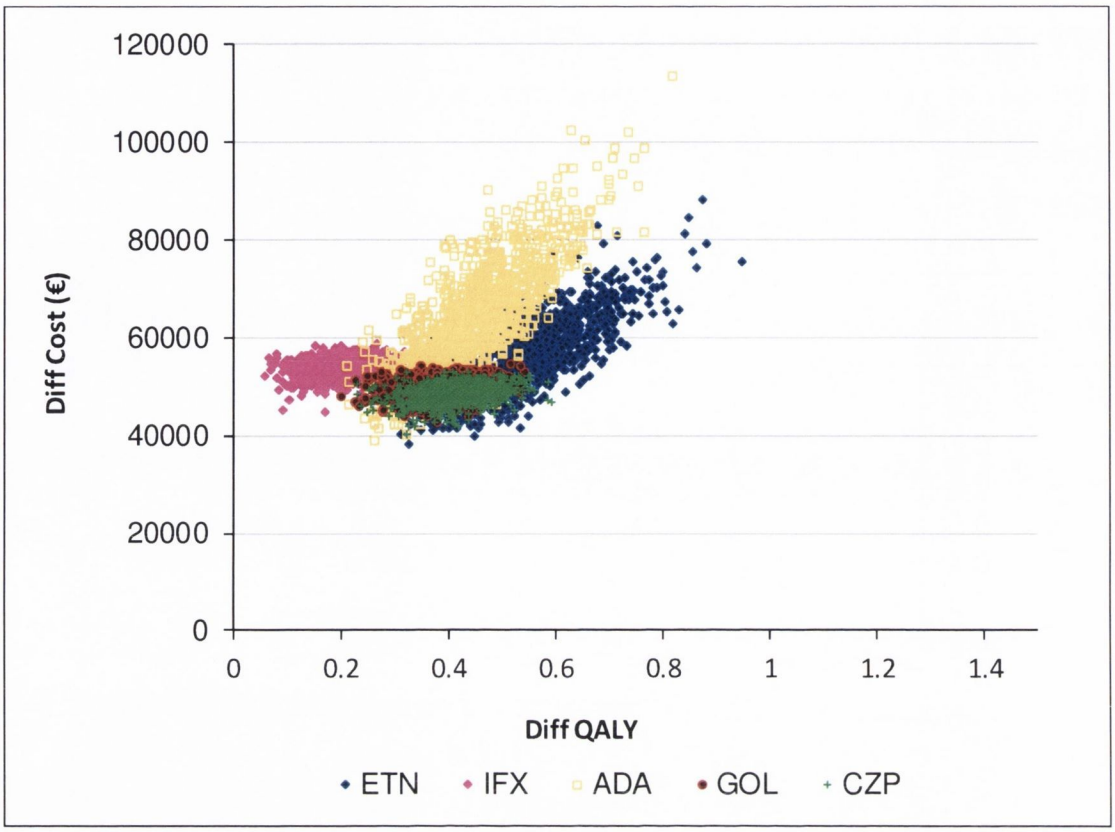


Figure 49: Results of PSA on CE Plane showing overlap of the anti-TNF agents.

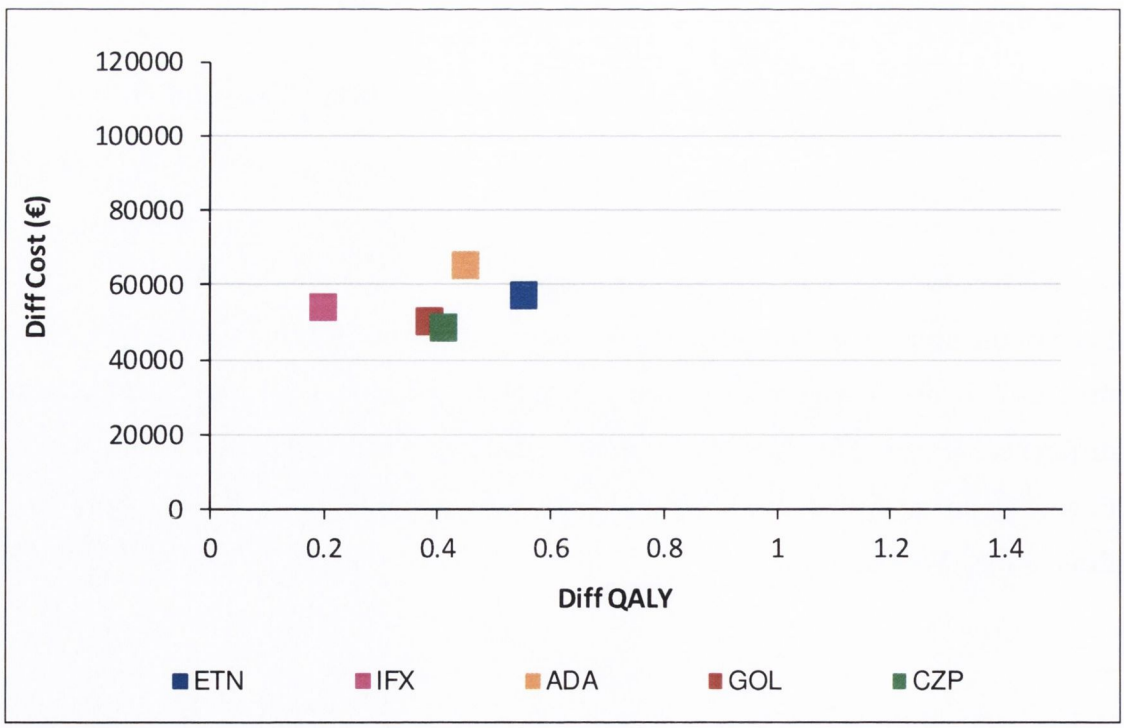


Figure 50: Cost-effectiveness Plane (Mean estimates of PSA).
Infliximab and adalimumab are dominated.

The cost-effectiveness frontier plots the probability that the optimal option is cost-effective at different WTP thresholds (Figure 51). The CEAF shows that at a willingness to pay up to €30,000, certolizumab presents the highest probability of being the optimal cost-effective option. At a WTP threshold above €30,000, etanercept is the optimal cost-effective option. If leflunomide is included in the CEAC, leflunomide presents the highest probability of cost-effectiveness.

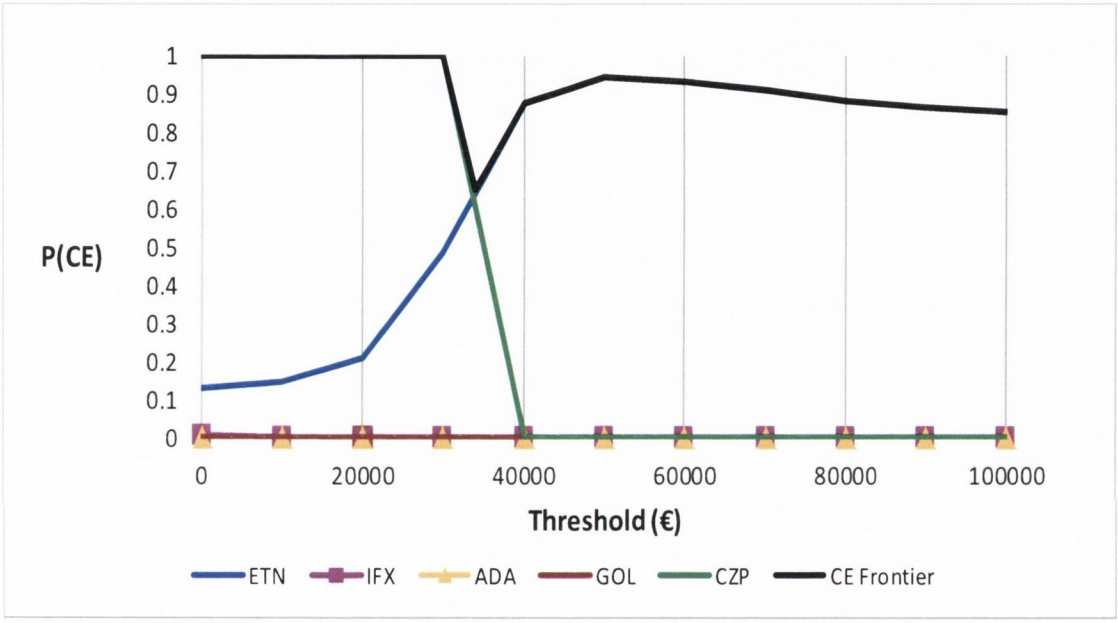


Figure 51: CEAF of anti-TNF agents (basecase).
The frontier indicates that certolizumab presents the optimal effectiveness at a threshold up to €30,000 and at a threshold above this etanercept is the optimal cost-effective option.

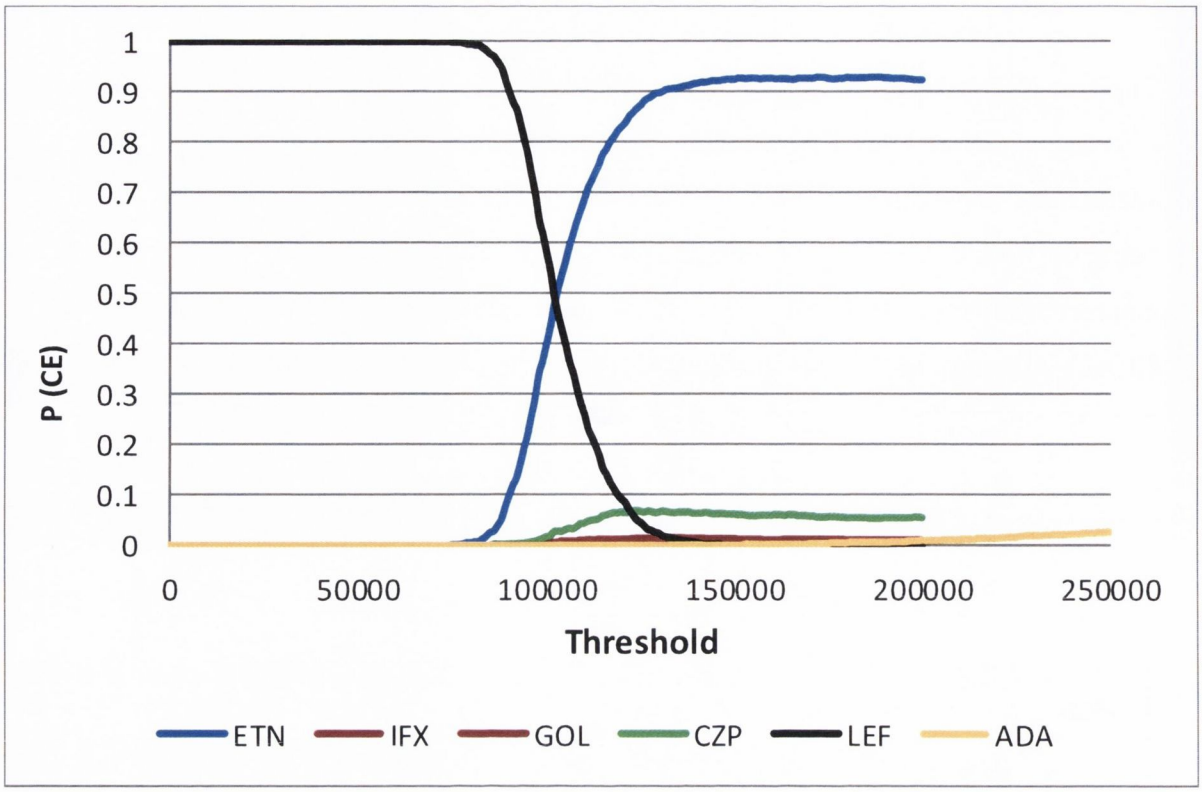


Figure 52: CEAC of anti-TNF agents and leflunomide versus no treatment (origin).

Leflunomide is the treatment of choice at a willingness to pay up to approximate €100,000/QALY

The CEAC presented in figure 52 is relevant if the choice of treatment also includes leflunomide and all are compared to no treatment which would be represented by the origin. The CEAC demonstrates that at a willingness to pay of up to approximately €100,000/QALY leflunomide would be the most cost-effective strategy to use.

6.4 Results of sensitivity analysis

A number of sensitivity analyses were carried out by running the BRAM with different parameter sets (Table 39).

Table 39. Parameters for Basecase

Model Parameters
Utility: Linear Relationship between HAQ and EQ-5D (using a revised scoring method)
Efficacy: HAQ multiplier estimates calculated from MTC
Long term Discontinuation: MTC of Observation studies for each anti-TNF – mean estimates from other anti-TNF agents used as estimate for certolizumab and golimumab
Mortality: SMR of 1.33
Time to effect and loss of effect (0.2 year)
Irish life tables
Drug costs: Reimbursement price (discounts, rebates, PAS not included), loading doses included
Screening, monitoring costs included

Table 40. Parameter sets for sensitivity analysis

Parameter Set	Parameters Varied
1	***Basecase: Revised scoring utility coefficients using linear term***
2	Linear Utility Mapping coefficients from original EQ-5D scoring methods (Irish Cohort)
3	Linear mapping coefficients from SF-6D (Irish cohort)
4	Quadratic mapping coefficients from original EQ-5D (Irish cohort)
5	Quadratic mapping coefficients from SF-6D (Irish cohort)
6	Linear mapping coefficients from EQ-5D mapping study (UK cohort)[145]
7	Linear mapping coefficients from EQ-5D mapping study (UK alternative cohort)[20]
8	Same time is spent on the treatments (Weibull distributions same for all treatments)
9	HAQ change on anti-TNF treatment
10	Drug price reduction (50%)
11	Include offset costs
12	Setting utility estimates to zero
13	Discount rate varied from 0% to 6% for both costs and outcomes
14	Price Reductions on individual drugs
15	Lower efficacy of leflunomide (HAQ multiplier 0.25)
16	Long term discontinuation rate increased for leflunomide (patients remain on leflunomide for a shorter time period)

Table 41. ICERs for One Way Sensitivity Analysis

Parameters		ETN	IFX	ADA	GOL	CZP
Basecase		*104000*	*277000*	*145000*	*130000*	*118000*
Linear mapping EQ-5D (Orig. TTO) (Irish Cohort)		78700↓	213000↓	111000↓	99800↓	90100↓
Linear mapping SF-6D (Irish cohort)		184000↑	480000↑	255000↑	229000↑	210000↑
Quadratic mapping EQ-5D (Irish cohort)		58100↓	153000↓	81000↓	73000↓	66000↓
Quadratic Mapping SF-6D (Irish cohort)		112000↑	282000↑	153000↑	138000↑	126000↑
Linear Mapping EQ-5D (UK cohort)[145]		58500↓	160000↓	83000↓	74500↓	67100↓
Quadratic mapping equation (UK cohort)[213]		**49200↓**	**121000↓**	**66600↓**	**60600↓**	**55000↓**
Linear mapping EQ-5D (UK biologic cohort)[20]		85600↓	231000↓	121000↓	108000↓	98000↓
Same long term discontinuation rate for all anti-TNF		118000↑	351000↑	177000↑	168000↑	146000↑
HAQ change on treatment (0.03 deterioration)		390000↑	5040000↑	812000↑	638000↑	532000↑
Drug Price reduction 50%		**47400↓**	**170000↓**	**71200↓**	**59700↓**	**60000↓**
Offset costs included		97100↓	271000↓	139000↓	124000↓	112000↓
Utility estimates to zero		104000=	277000=	145000=	130000=	118000=
Lower efficacy of leflunomide (HAQ multiplier 0.25)		101000↓	259000↓	139000↓	125000↓	114000↓
Long term discontinuation rate for leflunomide increased		97000↓	241000↓	133000↓	120000↓	109000↓
Discount rate for cost and outcomes	0%	63500↓	119000↓	79600↓	72900↓	67700↓
	6%	126000↑	433000↑	190000↑	168000↑	150000↑

* ↑ indicates an increase from the basecase ICER, ↓ indicates a decrease from the basecase ICER, = indicates no change to the basecase ICER. Parameters marked in **bold** are those that bring the ICER within range of cost-effectiveness.

The results of the deterministic OWSA indicate that the model is sensitive to a number of parameters (Table 41). The individual results of the sensitivity analaysis are presented in Appendix 11. The greatest change occurred when assumptions

regarding HAQ progression while on anti-TNF therapy are altered. The basecase assumes that there is no progression of HAQ while on treatment. However, if this assumption is challenged to a HAQ progression of 0.03 while on treatment, the ICER increases three fold. The relationship between utility and mapping also changes the ICER considerably. It appears that if a quadratic model rather than a linear model, is used to describe the relationship, the ICERs appear lower and the credible intervals wider. This is consistent for the Irish and UK data. More notable is the mapping relationship used from different cohorts. If mapping coefficients derived from a biologic cohort are used as inputs for the model, the ICERs increase from estimates using a general RA population for the model. If the model does not include utility estimates below zero or WTD states, there is no change from the basecase. This is an expected result in line with the revised scoring method for the EQ-5D. A reduction in costs by 50% does decrease the ICERs by almost 50%. Inclusion of offset costs in the model did not change the ICERs greatly; however the credible intervals are very wide and therefore there may be considerable uncertainty associated with this. The efficacy and the long term discontinuation rate were varied for leflunomide due to the uncertainty around the inputs for this parameter. While the change did reduce the ICER, it did not bring it within range of cost-effectiveness (Table 40). The results of the PSA using different parameter sets are also presented (Appendix 9 Table 41-49).

6.4.1 Cost-effectiveness with Price Reduction

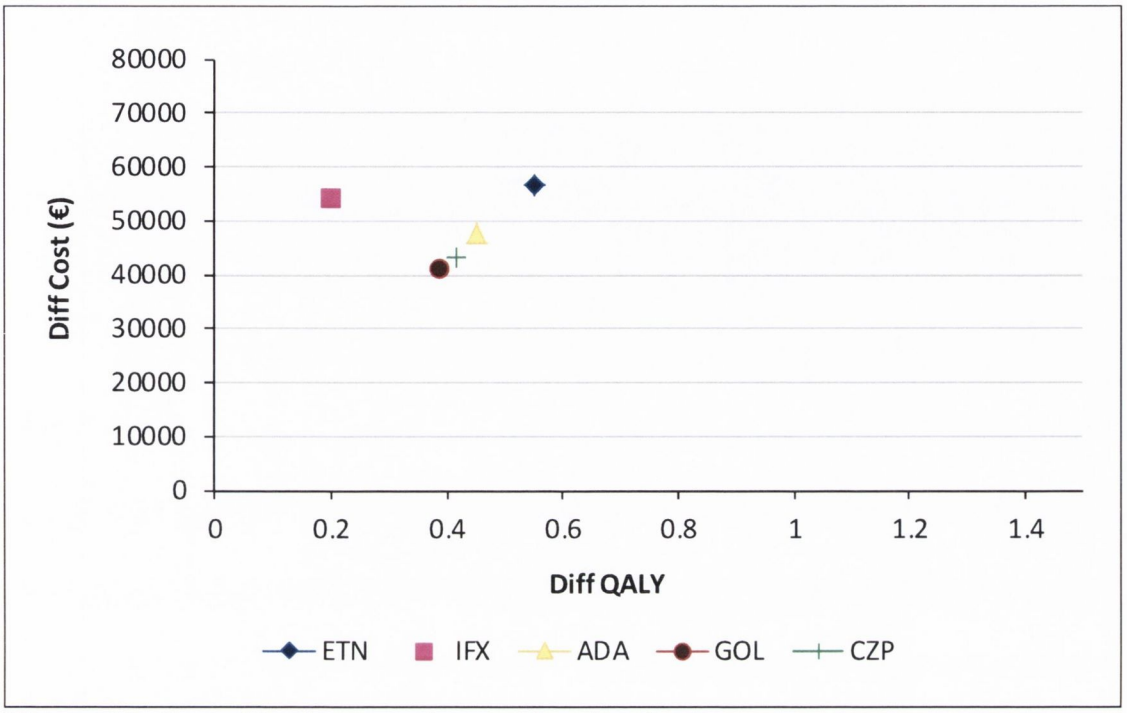


Figure 53: Cost-effectiveness plane with price decrease

(Golimumab (↓12%), adalimumab (↓26%) and certolizumab (↓8% + initial 10 doses free of charge)).

In order to bring other anti-TNF agents on the HTDS in line with etanercept, a price reduction is necessary in the range of 26% for adalimumab, 12% for golimumab, and a further 8% from certolizumab in addition to a scheme where the first three months of the drug would be free of charge to the payer (Figure 53). The CEAC and the CEAF are shown in Figure 54; at a willingness to pay below €20,000 golimumab would be the optimal choice; at a threshold of €20,000 to €50,000 certolizumab would be optimal and above this etanercept would be the optimal option.

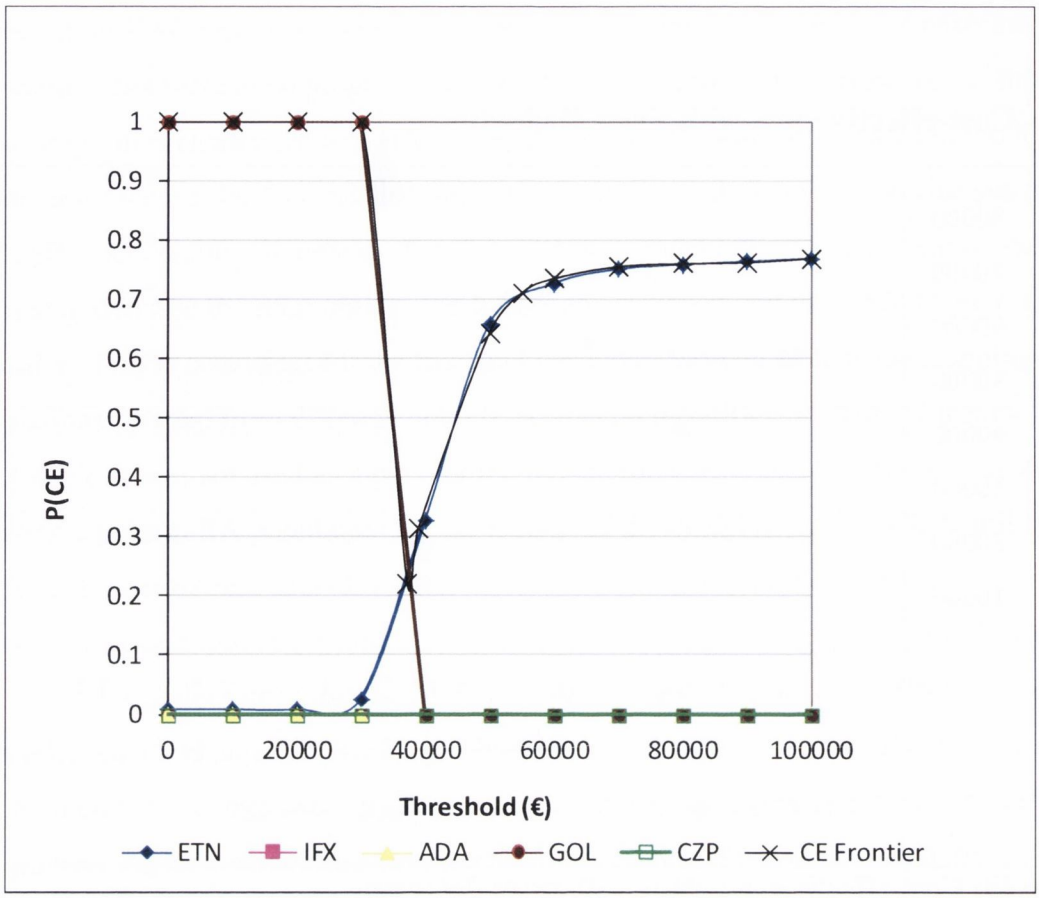


Figure 54: **Cost-effectiveness Acceptability Curves and Cost-effectiveness Frontier for anti-TNF agents with price reduction**
(Golimumab (↓12%), adalimumab (↓26%), and certolizumab (↓8% + first 10 syringes free of charge)).

6.5 Discussion

The aim of this economic evaluation is to assess whether the anti-TNF agents are a cost-effective option following failure of conventional DMARD therapy. This is a fundamental question to establish a baseline for cost-effectiveness of these agents in RA in Ireland. To date a multiple treatment assessment has not been performed in the Irish setting.

The basecase results of this analysis show that the anti-TNF agents are not cost-effective at a willingness to pay threshold of less than €100,000. Previous

assessments using this model which asked a similar question, have found these drugs to be cost-effective under £30,000 threshold, approximately €45,000.

The research work before this analysis was performed identified many of the areas of uncertainty in previous models and attempted to improve on this; however limitations remain. Some of the limitations stem from the model structure itself, some from the parameters chosen and some due to lack of robust long term data. The results of the economic analysis, improvements to the modelling process and the remaining limitations are discussed.

Incremental Cost-effectiveness

The incremental cost-effectiveness of these agents in comparison to an alternative DMARD (leflunomide) is well above what has been an acceptable willingness to pay threshold. The basecase results from the deterministic model indicate that the lowest ICER is for etanercept at €102,700/QALY; the highest ICER is for infliximab at €268,300/QALY. The ICERs for adalimumab, golimumab, and certolizumab are €143,700/QALY, €129,000/QALY, and €117,000/QALY respectively. These ICERs are considerably higher than those estimated in other HTA assessments using the BRAM. [40, 299] In 2006, NICE estimated the ICER for etanercept +MTX to be £23,800/QALY, for adalimumab + MTX to be £29,700/QALY and infliximab to be £37,900/QALY. [40] These ICERs were calculated on the basis that the anti-TNF agent was used last in the sequence. The Irish assessment differs considerably from the UK 2006 assessment and these differences, including the utility mapping, inputs for long term discontinuation and costs have had a strong influence on the results of the model.

Certolizumab and golimumab have not as yet been included in a MTA for anti-TNF drugs in RA; however the ICERs for STA's are also very different from what is presented here. In 2010, the NCPE assessed a company model for golimumab and recommended that golimumab was not cost-effective for the treatment of RA. The ICER for golimumab plus MTX versus MTX alone was estimated at €31,212/QALY. [41] Following price review the drug is now reimbursed by the payer. This model was based on the HAQ response mapped to utility and built in Excel®. The mapping equation used was that of Hurst *et al.* and other inputs such as long term

discontinuation and costs also differed greatly from the Irish MTA assessment. [145] Also in 2010, certolizumab was assessed by the NCPE through the STA process. [307] The company model estimated the ICER to be approximately €27,900/QALY in comparison to placebo + MTX. The main ICERs presented were compared to the other anti-TNF agents; certolizumab vs. etanercept €116,000/QALY, certolizumab vs. adalimumab €17,606/QALY and certolizumab vs. infliximab €25,545/QALY. This Excel® model used ACR response to model response to treatment. The model assumes that the initial response to treatment is maintained in the long term. The NCPE did not recommend certolizumab as cost-effective. Reimbursement was granted following a price review and incorporation of a PAS.

In other jurisdictions, CADTH examined etanercept and infliximab and deemed them not to be cost-effective at a threshold of CA$50,000 and estimated ICERs to be in excess of CA$125,000/QALY. [308] The result was only sensitive to QOL. A follow-up review of these drugs' long term effectiveness and toxicity suggested that they were only cost-effective if used with MTX, and if society was willing to pay more than CA$100,000/QALY.

Uncertainty associated with cost-effectiveness estimates

Uncertainty was explored using both OWSA and PSA. Assumptions in relation to QOL had the most impact on the ICERs. The long term progression of HAQ had most influence on the ICER with most influence on the ICER for golimumab, increasing from the basecase of €130,000 to €638,000 when ongoing progression of 0.03 while on treatment is assumed.

The relationship between utility and utility preferences also has a significant influence on the ICER estimates. In this case, using a quadratic model reduces the ICERs when compared with ICERs presented using a linear mapping model in this model. This was demonstrated in both the Irish cohort and the UK cohort. [40, 213] Chapter 4 provides a rationale for using the linear model over a quadratic model. The choice of population clearly influences the ICER estimates and this is shown when both the Irish and UK cohort are compared and when the UK general RA population is compared with a UK severe disease population. [20, 145]

The reduction in drug costs by 50% reduced the ICERs considerably, but did not reduce the ICERs below a threshold of €45,000/QALY. Discounting both costs and outcomes at 0% and 6% influenced the ICER greatly.

One of the main arguments for promoting the use of anti-TNF agents, is in relation to the savings that can be made or the costs that be offset; however when costs are offset in this model there are only small reductions in the ICERs.

The influence of negative or WTD utilities was also explored and this had no effect on the ICER. The basecase does use the revised scoring for the EQ-5D which reduces the influence of WTD.

The PSA was run for a total of 1,000 iterations. The results of the PSA were compared with the deterministic model which was run for a total of one million patients. There was very little difference in the results from each of the models. The probability of cost-effectiveness is given relative to each of the other anti-TNF agents, as there would be zero probability of cost-effectiveness in comparison to a DMARD at a threshold below €50,000. For the basecase the highest probability of cost-effectiveness is for certolizumab followed by etanercept if the willingness to pay is above €30,000/QALY.

Relative efficacy of treatments

The inputs for the relative efficacy of the anti-TNF agents are in the form of HAQ multipliers. These were calculated from a MTC of the five anti-TNF agents in a Bayesian framework. [309] The advantage of modelling efficacy in this way allows the improvement in HAQ score to be influenced by the baseline disease state (as measured by HAQ). A study of the BSRBR found that patients with worse baseline HAQ scores have a greater HAQ response to treatment with anti-TNF agents. [310] Previous assessments have used the HAQ multiplier but have not combined data from various studies in the framework used here. The estimates for HAQ multiplier are different to those used in the BRAM 2006 for adalimumab, etanercept, and infliximab. In this analysis, the HAQ multiplier for etanercept was higher (more effective), for infliximab was lower (less effective) and was not greatly different for adalimumab. The possible reasons for infliximab being so low are described in

chapter 6, and relate to the differing disease states of the trial cohorts. Most data was included for infliximab and adalimumab and the least amount of data was available for etanercept. A number of the larger etanercept studies (TEMPO[311] and COMET[234]) included either DMARD naïve patients or early RA patients. These were excluded due to possible confounding issues.

The efficacy of leflunomide is taken directly from a clinical trial as it was the only trial which used leflunomide following the failure of MTX. [101] Ideally leflunomide should have been included in the MTC of the anti-TNF drugs in order to derive an estimate of relative efficacy without breaking randomisation. However, due to a lack of efficacy data and in particular a link to the network diagram for the MTC (chapter 6) it was not possible to do this. In order to allow for this uncertainty observational data were used to inform an estimate of effectiveness for sensitivity analysis and long-term discontinuation rates. The result of this sensitivity analysis indicates that this was not a key driver for the results of the cost-effectiveness analysis and therefore had little impact on the results.

The consideration of long term effectiveness takes into account the question of discontinuation. In the model this is estimated from registry data from five different countries. Unfortunately, Ireland was not included due to the lack of registry data available here. A survival analysis was used to combine this data and an estimate was calculated for adalimumab, etanercept, and infliximab. Previous assessments did not have this quantity of registry data available and therefore the analysis presented here is more comprehensive due to the additional data. A limitation for the model is the lack of data for certolizumab and golimumab and the use of a mean estimate from the other anti-TNF agents. The assessments on these newer agents undertaken by the NCPE highlighted this as a key driver of the models. For this reason the mean Weibull assessment has been used in the scenario analysis for all anti-TNF agents. Under the assumption that there is no difference between the long term discontinuation rates, the ICERs all increased from baseline (etanercept increased from €104,000 to €118,000).

Toxicity is included in the model in the form of short term discontinuations. Data was extracted from observational studies and pooled using meta-analysis. A general

assumption has previously been made in company submissions that there is no significant difference in adverse effects between the anti-TNF agents. [41] However this may not be the case. The analysis presented in chapter 5 demonstrates this. The rate of discontinuation due to toxicity from 6-24 weeks is largely similar for adalimumab and etanercept but is greater for infliximab. An assumption is made for very short term withdrawal (< 6 weeks) that it is due to toxicity only; as it is too soon to identify that it is due to inefficacy from a clinical perspective. Evidence indicates that anti-TNF agents are well tolerated in the short term; clinical trial data indicates increased adverse effects in the treatment groups but with relatively few withdrawals due to toxicity (chapter 5). There is concern about the long-term safety with respect to infections, lymphomas, autoimmunity, and demyelination. While some studies have indicated that the rates for these are low, a number of meta-analyses show an increased hazard ratio for serious infections and cancer over the untreated RA population. [312, 313] A distinction is not made in the model between different adverse effects and while this may be a limitation, the RCT, and observational data indicates that the type of adverse effects experienced, are largely similar for each of the agents.

HAQ and Utility Assumptions

The fundamental driver of this model is the relationship (mapping) between utility and HAQ. Data used for the mapping model from HAQ to utility differs considerably from that used for both the BRAM 2006 and the BRAM 2009. When compared, the economic model using the Irish mapping estimates produces higher ICERs than those produced when the data by Hurst *et al.* was used (BRAM 2006 and BRAM 2009). [145] The cohort for the Irish mapping study has been described previously (chapter 4) and represents the cohort in which anti-TNF agents are initiated in Ireland. Other than a description of the functional class of the RA patients used for the Hurst paper (n=224), there is very little detail on the demographic of the cohort. The paper was published in 1997 and it could be argued that the management of the RA patient has changed considerably since then with an emphasis on early diagnosis and aggressive DMARD treatment. More recently an independent report was carried out by the University of Sheffield, on the British Society of Rheumatology Biologics Register (BSRBR). [20] For this report a mapping exercise was carried out on patients with RA. The description of the cohort is very similar to the cohort described in the Irish study. When the estimates for model 1 of this study were used in the economic model, the

ICER results were considerably higher than those produced when using the Hurst equation. This supports the hypothesis that there are considerable differences in the patient population used and it may be more appropriate to use estimates from a mapping study in patients suitable for biologic therapy as is pertinent here.

The differences produced in decision model estimates due to the use of different instruments for QOL measurement have been the subject of much discussion in the literature (chapter 4). These differences are also evident in this analysis. The ICER estimates between the EQ-5D and the SF-6D are not surprising due to the differing descriptive of each measure and the differing methods used to obtain preferences (TTO or SG). The large quantity of WTD states is a concern due primarily to the manner in which the data was handled in the original valuation study. [147] The revised scoring method of the EQ-5D presented in chapter 4, offers an alternative method of handling these states. The ICER estimates produced using this mapping equation are greater than that produced by the original EQ-5D scoring and less than that produced when the SF-6D preferences are applied. This is in agreement with the coefficients of the mapping exercise which indicates that less QALYs are produced per unit change in HAQ with the revised scoring method than the original. In light of the possible overestimation of benefit produced by the EQ-5D and the underestimation produced by the SF-6D, a pragmatic choice was made to use the revised EQ-5D, which lies somewhere between the over and underestimation, for the basecase.

HAQ progression on treatment

Following the initial increase in HAQ or response, an assumption is made for the basecase that no further improvement or worsening occurs in the HAQ score. In order to explore this further, a scenario is presented where there is an ongoing disimprovement in HAQ equal to that experienced by the general population. This scenario increases the ICERs significantly and is one of the main areas of uncertainty in the model. While it remains unclear from the literature if patients maintain zero HAQ progression, it would seem a more likely assumption.

One of the main limitations of this cost utility analysis is the lack of data on the long term effectiveness of these agents for the treatment of RA. This has also been an

issue with other HTA undertaken in this speciality. [113, 251] A report examining the cost-effectiveness of the anti-TNF agents infliximab and etanercept concluded that while these drugs were moderately effective at 1 year, they were not cost-effective under a threshold of CAN$100,000. [117] The majority of the clinical trials report six monthly data showing greater efficacy over placebo. The assumption made by the company assessments submitted to the NCPE has been that much of this effect is maintained. [41, 307] However both the CADTH review and registry data provide evidence that this may not be the case.

The two primary reasons for discontinuation cited from both long term and short term studies has been adverse effects and loss of efficacy. In the model short term withdrawal is based on toxicity and inefficacy. The long term discontinuation analysis performed for this model indicates that there are differences between the agents.

Model Limitations

The BRAM models disease response through improvement or deterioration in HAQ. Other models have used ACR response for modelling improvement.[307] The ACR does not include baseline severity as an indicator as does the HAQ multiplier and therefore may overestimate the benefit of the drugs. The goal of treatment for RA is to stop progression of the disease and thereby halt ongoing damage and long term complications. Long term progression and damage can lead to joint replacement and chronic pain. If radiological evidence supports halting progression, there may be an argument for incorporating radiological progression in a decision model for RA.

The model coding does not incorporate correlation between parameters. In the case of HAQ, there is likely to be some interdependence between HAQ and age; this is a limitation of the model.

Chapter 7

Chapter 7.0 Implications of Research for Practice and future recommendations

Previous analyses in other jurisdictions on the treatment of RA with anti-TNF agents have indicated that they are cost-effective. However in Ireland cost-effectiveness has not been proven. The work presented here has added to the evidence base and highlights a number of methodological issues associated with carrying out such an evaluation. The results of the analysis imply that the payer is not getting value for money in this area; however there are challenges as to how this could be improved upon.

7.1 Methodological Recommendations

7.1.1 Utilities

The results of this thesis provide some interesting evidence to a number of key areas. The utility analyses presented in chapter 4 and 5 indicate that there are significant concerns associated with using the EQ-5D and the TTO population scoring in its current format. The evidence presented in this thesis and published scientifically indicates that the use of the EQ-5D and the original scoring deflates the ICERs in economic analysis for RA. The revised scoring presented here attempts to correct some of this deflation by changing the manner in which worse than death states are handled. The EuroQOL group have produced a five level EQ-5D which is currently undergoing population scoring studies. While it is hoped that this work will correct some of the methodological flaws in the original EQ-5D population scoring, recommendations should be put in place for HTA agencies using the EQ-5D to allow for potential underestimation of ICERs and therefore potentially incorrect decisions being made.

7.1.2 Evidence Synthesis

The analysis presented in chapter 5 and the subsequent publication in the Annals of the Rheumatic Diseases used an MTC to allow ranking of the anti-TNF treatments. Previous indirect comparisons of these agents concluded that there was no statistical difference between the anti-TNF agents at 5% significance level which had led some jurisdictions to conclude that there was no difference between the anti-TNF agents and therefore cost-minimisation analyses are sufficient for economic analysis. Incorporating the full distribution of the evidence from the MTC, despite lack of a significant difference at the 5% level, demonstrated differences in the effect through cost-effectiveness analysis; this would not be the case for cost-minimisation analysis. The work in this thesis refutes this and caution is therefore advised with such an approach.

Finally the work produced in chapter 6 where the cost-effectiveness of anti-TNF agents is modelled provides a number of recommendations for practice. This are outlined below.

7.1.3 Strategy Choice

7.1.3.1 DMARD Choice

The strategy choice should reflect clinical practice in the population to be modelled. The current treatment pathway in Ireland recommends MTX as first line treatment for newly diagnosed RA patients unless contraindicated. For established RA patients MTX is also recommended if not previously tried. Once patients are deemed no longer responsive to MTX an anti-TNF agents can be introduced. The population modelled in this case (established RA patients), therefore, cannot be compared to MTX as it would previously have been used unsuccessful. Leflunomide is therefore used as a comparator strategy to anti-TNF therapy. It could be argued that leflunomide is not a realistic comparator in the Irish healthcare setting. However it is realistic to consider leflunomide as a comparator in this patient population. The question for the Irish healthcare payer is whether anti-TNF agents are cost effective in comparison to non biologic therapy in an established population. The cost-effectiveness of these agents (as a multiple technology assessment) had not previously

been assessed in Ireland and therefore establishing a baseline for the use of these agents was crucial to the overall assessment.

7.1.3.2 Anti-TNF therapy choice

Currently five anti-TNF agents are reimbursed in Ireland; adalimumab, certolizumab, etanercept, golimumab and infliximab. For this analysis each anti-TNF agent has been compared as a starting point of a treatment sequence. The sequence following the failure of the anti-TNF agent includes only non-biological DMARDs. Therefore, the anti-TNF agents are not included as sequential i.e. this analysis does not consider etanercept followed by adalimumab or vice versa. Clinical practice has been to use an alternative anti-TNF agent following the failure of the initial anti-TNF. Recent guidance indicates that this practice may not be cost-effective and it is more cost-effective to move to an alternate treatment group such as B-cell therapy (rituximab) following the failure of the first anti-TNF agent. [302] In order to inform best choice for the initial anti-TNF therapy, the relative cost-effectiveness of the anti-TNF agents when used first line was estimated. By comparing identical sequences starting with different anti-TNF agents the estimate of cost-effectiveness for each of the anti-TNF agents can be compared. By placing leflunomide at the origin rather than a do nothing strategy the probability of cost-effectiveness given the choice of only anti-TNF agents can be calculated.

7.1.3.3 Sequences

This model assumes divergence after the failure of MTX which is in line with the patient population. Other scenarios are possible; the most likely one to consider would be assuming that patients in the anti-TNF strategies also fail leflunomide before starting anti-TNF therapy. However, in order to consider a comparator for the patient population, leflunomide needs to be a viable treatment option for all arms. A further consideration might be to remove leflunomide from the treatment sequence after anti-TNF therapy. The result of including the leflunomide in the anti-TNF sequences are that consequences associated with leflunomide are included in these arms and therefore the difference between the anti-TNF arms and the comparator arm

will not be large. A counter argument to this would be that the difference between the arms is due solely to the anti-TNF agent and not to additional DMARD therapy following failure of treatment. The approach taken here answers the questions of interest for the decision maker; what is the cost-effectiveness of anti-TNF agents in comparison to a non-biologic and what the relative effectiveness of these agents is, against each other. The question in relation to sequencing of anti-TNF agents may be appropriate to ask following these initial questions. However considering anti-TNF agents have not been proven to be cost-effective against non-biological therapy, the question of sequencing of treatments becomes less important.

7.1.3.4 Choice amongst anti-TNF agents

The question which arises in addition to the baseline cost-effectiveness of these anti-TNF agents is whether there are differences in cost-effectiveness between them. Chapter 5 details the differences in both efficacy and discontinuation rates. The analysis in chapter 6 indicates that there are also differences between the cost-effectiveness. Under the assumption that anti-TNF drugs present the only treatment option, etanercept is likely to be most cost-effective. The ICERs are all above current willingness to pay thresholds so this choice can only be considered relative to the other anti-TNF agents.

7.1.3.5 Use of real world data

The efficacy parameters for this thesis were drawn from six month RCTs. One of the main drivers for cost-effectiveness of anti-TNF agents in RA patients is in relation to disease progression in the long term. Short term clinical trial data does not inform on this and in a recent publication indicates that the short term benefit does not appear to be maintained in the long term. [314] Many countries have developed biological registries to gather real world outcomes and monitor adverse effects (UK, Germany, Sweden, Netherlands, and Switzerland). In the Netherlands a registry was specifically established to monitor clinical outcomes and costs associated with anti-TNF therapy and reimbursement was dependent on this. [248] Data from this registry demonstrated that real life outcomes were less than those outcomes achieved in RCTs. When considering the cost-effectiveness of a technology, the effectiveness of the technology is as important as the efficacy data. For this reason it is vital that cost-effectiveness

studies consider real world data. This study included real world data for discontinuation rates and also to inform the effectiveness of leflunomide in this cohort. Incorporating real world data within the MTC analysis would give a more robust estimate of effectiveness for the anti-TNF agents. However as stated previously there are methodological challenges in incorporating such data.
Data available from the Irish setting was not registry data and was only from one source. The establishment of a formal registry for RA patients, providing outcome data on a national level, would strengthen the cost-effectiveness estimates calculated for Ireland.

7.2 Implications for Reimbursement of anti-TNF agents in Ireland

All of these agents are now reimbursed by the Irish healthcare payer. The results indicate that these are not cost-effective under a willingness to pay threshold of less than €100,000/QALY. These drugs are currently under review by the Irish payer mainly due to the current spending on these drugs reaching approximately €100 million per year. However almost 10,000 patients are currently on biological drugs for RA in Ireland and therefore the options available for cost containment may be limited to pricing. Strategies for this might include schemes such as risk sharing in the form of PAS, service provision, fair price strategy, price reductions or tendering for a preferred drug for use. An assessment of the options available is presented.

7.2.1 Risk Sharing via PAS

Risk sharing has been the subject of much discussion among healthcare payers and pharmaceutical companies over the past number of years. [315] A risk sharing practice allows payers and pharmaceutical companies to build clinical experience with new medicines which might not otherwise have been eligible for reimbursement. The practice is already in place in a number of European countries including Netherlands, Belgium, UK, and other jurisdictions. In the Netherlands a time scale of three years is granted to gather evidence to answer questions from a cost-effectiveness review. If the evidence gathered proves benefit with an acceptable ICER, the drug is funded indefinitely. If the corollary occurs, funding is withdrawn. This system has been in

place since 2006 and 23 drugs have been subject to the practice. A biologics review of the DREAM database found that the effectiveness of the anti-TNF agents for RA was less in clinical practice than that observed in clinical trials. [297]

Belgium established a conditional reimbursement system which leads to a reassessment after 18-36 months. The additional data collection is often undertaken by a sponsoring pharmaceutical company. When identifying a drug eligible for this system, factors such as effectiveness in clinical practice, pharmacoeconomics in clinical practice, number of eligible patients, sales volume and reimbursement elsewhere are taken into account. If data does not demonstrate added value, changes may be recommended such as restricting to certain subgroups or restricting to certain prescribers. By 2007, 18 drugs were re-appraised and only one was withdrawn.

One of the first risk-sharing schemes in the UK was for multiple sclerosis (MS). The study aimed to gather data on the long term benefits of MS drugs in clinical practice. The study was launched in 2002 and was reviewed in 2008. The interim analysis did not meet the pre-defined cost-effectiveness level. However, the methodological difficulties of undertaking an observational cohort study were highlighted and in particular the question of historical comparator. [316] A slightly different risk sharing scheme was introduced for bortezomib for multiple myeloma. In this case a 'response rebate' scheme was established where patients who showed no or minimal response were taken off treatment, and drug costs were refunded by the manufacturer.
The details of risk sharing schemes in Ireland are not in the public domain, and discussion is beyond the scope of this thesis. However, recently a PAS was agreed upon for certolizumab.

A risk sharing scheme, can provide an innovative solution to the problem of assessing long term benefits or risks of new drugs. The issues around governance and methodologies of such schemes remain a challenge. Risk sharing in the form of a PAS does have implications for European pricing. The agreement of PAS can often maintain a high reimbursement price which is then considered for the European basket pricing system. The undisclosed details of PAS' in Ireland, has further implications for future HTA's in choosing comparator prices. In order to maintain

transparency in the process of cost-effectiveness evaluation and pricing, it would seem reasonable that details of PAS should be made publicly available.

7.2.2 Pricing

Value based pricing sets prices based on the value achieved to the customer or in this case the payer rather than the price set by competitor drugs or market prices. Value based pricing has been introduced formally in the UK. This method of price setting has been in place in an informal way in Ireland for some time. Value based pricing is to be introduced in the UK in a formal mannerin 2014. [317]

In pharmaceutical reimbursement, price premiums are usually only given if there is evidence to indicate that the drug offers more benefit than those it is competing with. A system like this is formally in place in Germany where a price premium is awarded based on the additional benefit the drug delivers. A drug which delivers less benefit will not be reimbursed and a drug which delivers equal benefit will be priced equally. In the case of the anti-TNF agents, a price premium exists for all anti-TNF agents over conventional DMARDs which are fair, due to the additional benefit they provide. There are currently five anti-TNF agents in use and this analysis indicates that they do not deliver equal benefit. In the case of adalimumab a price premium is paid for a drug which is less effective than its comparator; etanercept. Equally a price premium is paid for golimumab despite its inferior efficacy to certolizumab and finally infliximab has a price premium over both certolizumab and golimumab despite displaying the least benefit. A reasonable approach would be to offer a price relative to the benefit shown. In this case a decrease in price of the less effective agents to a price below the next most effective one. This could be illustrated by bringing all of the points on the cost-effectiveness plane onto a line drawn from etanercept to the origin. This is presented in Figure 53.

7.2.3 Price Parity

A further strategy, employed in for cost containment, would be price parity. In this case the payer assumes equal efficacy across agents and prices all agents the same.

The challenge with this system in this case is the differing efficacy of the agents which has been demonstrated in chapter 5 and in Figure 53 of this chapter.

7.2.4 Programme Preferred Drugs Schemes

Programme preferred drug strategies have been used most commonly in the US under the Medicaid and Medicare health insurance schemes. A review of the literature on programme preferred drugs appears to indicate many different choices involved in such programmes ranging from choosing the cheapest drug from a group of similar efficacy to choosing the most cost-effective drug for use for a condition. A 2006 American review of drug formulary policies asked whether these policies reflected value for money. [318] In theory the choice of a programme preferred drug should be based on a drugs overall value or cost-effectiveness, however this is not always and the case. In practice it is often the cost of the drug which drives the choice. [319] A programme preferred drug in the case of anti-TNF agents may be of benefit if considered in terms of most cost-effective drug rather than cheapest drug. The results presented in this thesis demonstrate differences in efficacy and therefore this must be taken into consideration.

7.3 Conclusion

All five of these drugs are currently reimbursed by the HSE, four of these through the HTDS and one through the hospital system. The results of this project demonstrate that disparity exists in the pricing strategy chosen for these drugs, where less effective drugs are paid a price premium. A further challenge exists in that these drugs have not demonstrated, in this analysis, to be cost-effective in a willingness to pay threshold of less than €100,000.

7.4 Future Research

The results of this economic analysis are influenced greatly by the effect parameters; both utility estimates and relative effectiveness. The uncertainty surrounding the longer term use of anti-TNF agents stems from two main aspects. The first is

concerned with the impact on the HAQ scores following the initial improvement. It is not clear whether disimprovement in HAQ is halted or slowed. There are significant implications of assuming that there is ongoing damage while on anti-TNF treatment. Ongoing collection and analysis of this data is required for the Irish population.

The results of the MTC, while useful, are limited to the initial response to anti-TNF agents. The incorporation of observational data through a Bayesian MTC, where the observational results are used as the prior estimate would aid the estimation of effectiveness in the long term. The lack of an appropriate control group is often the challenge when analysing observational data. The methodologies required to perform such an analysis without introducing confounding are still in development.

This research examined closely the methods used to calculate utility values for the QALY. The results are in agreement with other research in this area. The EuroQOL group has also recognised that there may be problems with the original methodology used to elicit preference based utilities and are developing protocols using alternative methodologies such as DCE and lead time. Ireland as yet has not valued population health using direct techniques. Utility weighting for Irish economic evaluations has predominately used the preferences of the United Kingdom. The results presented here demonstrate that this is not ideal. While the area of HTA has undergone substantial development in the past decade in Ireland, relatively little work has been done in health state valuation. Such a study would allow Irish citizens, through health valuations, to influence decision-making.

The model presented in this evaluation uses the HAQ score to model response to treatment over time. There is an argument that HAQ scores may not fully demonstrate the impact of these treatments both positively and negatively. Radiographic scores have been suggested as a more predictive outcome measure, however there are challenges in the use of radiographic data; considerable resources to assign validated scores are required.

This research question posed in this research was tailored to the current situation in Ireland. All anti-TNF agents are reimbursed but only two have been assessed through a formal process. In order to assess all of these agents it was pertinent to examine

them as they are used primarily: initial anti-TNF in patients who no longer respond to MTX. The manner in which these agents are reimbursed currently does not distinguish between second line or third line use or sequential use. However this is an important question and it would be useful to examine the sequential use which would include all biological drugs including rituximab, tocilizumab, and abatacept.

The results of this economic evaluation indicate that these agents are not cost effective under a willingness to pay threshold of €100,000. The analysis includes an aggregate of all patients treated in Ireland. Some of these patients may have more aggressive disease than others or more permanent joint damage. Restricting access to these agents for all RA patients would not be appropriate for this reason. Irish guidelines on the use of anti-TNF agents for patients with RA were produced in 2005. These guidelines should be updated and could be guided by subgroup analysis to identify which group these agents demonstrate most cost-effectiveness.

References

1. Department of Health and Children. Commission on Financial Management and Control Systems in the Health Service. [http://www.dohc.ie/publications/pdf/brennan.pdf?direct=1][Accessed Sept 2010]. 2003.
2. Department of Health and Children. Economies in Drug Usage in the Irish Healthcare Setting. http://www.indi.ie/docs/878_economies_drug_usage_(2).pdf [Accessed Sept 2010]. 2009.
3. Irish Pharmaceutical Healthcare Association. IPHA/HSE Agreement on the Supply of Medicines to the Health Services www.ipha.ie/alist/ipha-hse-agreement.aspx [Accessed 13 May 2011]. 2010.
4. Office of Technology Assessment. Development of medical technology: opportunities for assessment. Washington DC: U.S. Government Printing Office. 1976.
5. Hutton J, McGrath C, Frybourg J-M, Tremblay M, Bramley-Harker E, Henshall C. Framework for describing and classifying decision-making systems using technology assessment to determine the reimbursement of health technologies (fourth hurdle systems). Int. J. Technol. Assess. Health Care 2006;22(1):10-18.
6. Bootman JL, Townsend RJ, WF. M. Principles of Pharmacoeconomics. 2nd ed. Cincinnati: Harvey Whitney Books Company; 1986.
7. Usher C, Tilson L, Olsen J, Jepsen M, Walsh C, Barry M. Cost effectiveness of HPV vaccine in reducing the risk of cervical cancer in Ireland due to HPV types 16 & 18 using a transmission dynamic model. Vaccine 2008;26(44):5654-5661.
8. Bennett K, Kabir Z, Barry M, et al. Cost-Effectiveness of Treatments Reducing Coronary Heart Disease Mortality in Ireland, 2000 to 2010. Value Health 2009;12(1):10-15.
9. Tilson L, Usher C, Butler K, et al. Economic Evaluation of a Universal Childhood Pneumococcal Conjugate Vaccination Strategy in Ireland. Value Health 2008;11(5):898-903.
10. Tilson L, Barry M. Recent developments in pharmacoeconomic evaluation in Ireland. Expert Review of Pharmacoeconomics & Outcomes Research 2010;10(3):221-224.
11. National Centre for Pharmacoeconomics. Template for rapid review assessment submissions from manufacturers http://www.ncpe.ie/document.php?cid=26&sid=145&docid=200 [Accessed Nov 2010]. 2010.
12. National Centre for Pharmacoeconomics. Recommendations on the Reporting Format and Layout of Pharmacoeconomic Submissions to the National Centre for Pharmacoeconomics http://www.ncpe.ie/u_docs/doc_204.pdf [Accessed Nov 2010]. 2010.
13. National Institute for Clinical Excellence. Rheumatoid arthritis - certolizumab pegol: final appraisal determination. Available from http://www.nice.org.uk/guidance/index.jsp?action=download&o=47135 [Accessed 31 May 2010]. 2009.

14. Primary Care Reimbursement Services. Statistical Analysis of claims and payments. www.hse.ie [accessed 13 May 2011]. 2009.
15. Health Information and Qaulity Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. 2010.
16. Briggs AH, O'Brien BJ. The death of cost-minimization analysis? Health Econ. 2001;10 (2):179-184.
17. Canadian Agency for Drugs and Technologies in Health. Common Drug Review: Certolizumab pegol. 2010 Available from www.cadth.ca/media/cdr/complete/cdr_complete_Cimzia_May-28-2010.pdf [Accessed 31 May 2011].
18. Canadian Agency for Drugs and Technologies in Health. Common Drug Review: Golimumab. 2010 Available from www.cadth.ca/media/cdr/complete/cdr_complete_Simponi-RA_March-17-2010_e.pdf [Accessed 31 May 2011].
19. Doan Q, Chiou, CF, Dubois, RW. Review of eight pharmacoeconomic studies of the value of biologic DMARDs (adalimumab, etanercept, and infliximab) in the management of rheumatoid arthritis. Journal of Managed Care Pharmacy 2006;12(7):555-569.
20. Brennan A, Bansback N, Nixon R, et al. Modelling the cost-effectiveness of TNF-a inhibitors in the management of rheumatoid arthritis: Results form the British Society for Rheumatology Biologics Registry - An Independent Report. Rheumatology (Oxford). 2007;46:1345–1354.
21. McCullagh L, Tilson L, Walsh C, Barry M. A Cost-Effectiveness Model Comparing Rivaroxaban and Dabigatran Etexilate with Enoxaparin Sodium as Thromboprophylaxis after Total Hip and Total Knee Replacement in the Irish Healthcare Setting. Pharmacoeconomics 2009;27:829-846.
22. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). The Lancet 2002;360(9334):711-715.
23. Akehurst R, Anderson P, Brazier J, et al. Decision analytic modelling in the economic evaluation of health technologies. Pharmacoeconomics 2000;5:443 - 444.
24. Health Technology Board for Scotland S. Scottish Medicines Consortium.
25. NICE. Guide to the Methods of Technology Appraisal:London. 2008.
26. Pharmaceutical Benefits Board Sweden. General guidelines for economic evaluation from the Pharmaceutical Benefits Board,Sweden. 2003.
27. Eichler H-G, Kong SX, Gerth WC, Mavros P, Jönsson B. Use of Cost-Effectiveness Analysis in Health-Care Resource Allocation Decision-Making: How Are Cost-Effectiveness Thresholds Expected to Emerge? Value Health 2004;7(5):518-528.
28. Weinstein M, Zeckhauser R. Critical ratios and efficient allocation. J Public Econ 1973;2:147–157.
29. Barry M, Usher C, Tilson L. Public drug expenditure in the Republic of Ireland. Expert Review of Pharmacoeconomics & Outcomes Research 2010;10(3):239-245.
30. Barry M, Tilson L. Recent developments in pricing and reimbursement of medicines in Ireland. Expert Review of Pharmacoeconomics and Outcomes Research 2007;7(6):605-611.

31. Barry M, Usher C, Tilson L. Public drug expenditure in the Republic of Ireland. Expert Review of Pharmacoeconomics and Outcomes Research 2010;10(3):239-245.
32. Claxton K. Exploring uncertainty in cost-effectiveness analysis. Pharmacoeconomics 2008;26(9):781-796.
33. Fenwick E, Claxton K, Sculpher M. Representing uncertainty:the role of cost-effectiveness acceptability curves. Health Econ. 2001;10:779–787.
34. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. Value Health 2009;12(5):739-749.
35. Drummond M, Brown R, Fendrick A, et al. Use of pharmacoeconomics information--report of the ISPOR Task Force on use of pharmacoeconomic/health economic information in health-care decision making. Value Health 2003;6(4):407-416.
36. Hoffmann C, Graf von der Schulenburg J. The influence of economic evaluation studies on decision making. A European survey. The EUROMET group. Health Policy 2000;52(3):179-192.
37. Black W. The CE plane: a graphic representation of cost-effectiveness. Med. Decis. Making 1990;10(3):212-214.
38. Brennan A, Bansback N, Reynolds A, Conway P. Modelling the cost-effectiveness of etanercept in adults with rheumatoid arthritis in the UK. Rheumatology (Oxford). 2004;43(1):62-72.
39. Bansback NJ, Brennan A, Ghatnekar O. Cost effectiveness of adalimumab in the treatment of patients with moderate to severe rheumatoid arthritis in Sweden. Ann. Rheum. Dis. 2005;64(7):995-1002.
40. Chen YF, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W. 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 Technol. Assess. 2006;10(42):1 - 248.
41. National Centre for Pharmacoeconomics. Cost-effectiveness of golimumab (Simponi®) in the treatment of moderate to severe rheumatoid arthritis. Available from: www.ncpe.ie [Accessed 01 July 2010]. 2010.
42. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. New York: Oxford University Press; 2006.
43. Barton GR, Briggs AH, Fenwick EAL. Optimal Cost-Effectiveness Decisions: The Role of the Cost-Effectiveness Acceptability Curve (CEAC), the Cost-Effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfection Information (EVPI). Value Health 2008;11(5):886-897.
44. Health Information and Quality Authority. Health Technology Assessment of a population-based colorectal cancer screening programme March 2009 Available from http://www.hiqa.ie. [Accessed on 05 May 2011].
45. Groot Koerkamp B, Hunink M, Stijnen T, Hammitt J, Kuntz K, Weinstein M. Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. Med. Decis. Making 2007;27:101–111.
46. O'Brien B, Gertsen K, Willan A, al. e. Is there a kink in consumers' threshold value for cost-effectiveness in healthcare? Health Econ. 2002;11:175-180.
47. Severens JL, Brunenberg DEM, Fenwick EAL, O'Brien B, Joore MA. Cost-Effectiveness Acceptability Curves and a Reluctance to Lose. Pharmacoeconomics 2005;23:1207-1214.

48. Fenwick E, Briggs A. Cost-Effectiveness Acceptability Curves in the Dock. Med. Decis. Making 2007;27(2):93-95.
49. Palmer S, Smith P. Incorporating option values into the economic evaluation of health care technologies. J. Health Econ. 2000;19:755–766.
50. Yelin E, Callahan LF. The National Arthritis Data Work Group. Special article the economic cost and social and psychological impact of musculoskeletal conditions. Arthritis Rheum. 1995;38(10):1351-1362.
51. Symmons DPM. Epidemiology of rheumatoid arthritis: determinants of onset, persistence and outcome. Best Practice & Research Clinical Rheumatology 2002;16(5):707-722.
52. Riise T, Jacobsen B, Gran J. Incidence and prevalence of rheumatoid arthritis in the county of Troms, northern Norway. J Rheumatol 2000;27(6):1386-1389.
53. Gabriel S, Crowson S, O'Fallon M. The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955-1985. Arthritis Rheum. 1999;42(3):415-420.
54. Drosos A, Alamanos I, Voulgari P, et al. Epidemiology of adult rheumatoid arthritis in northwest Greece 1987-1995. J. Rheumatol. 1997;24(11):2129-2133.
55. Arnett F, Edworthy S, Bloch D, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988;31(3):315-324.
56. van Gestel A, Anderson J, van Riel P, et al. ACR and EULAR improvement criteria have comparable validity in rheumatoid arthritis trials. American College of Rheumatology European League of Associations for Rheumatology. J. Rheumatol. 1999;26(3):705-711.
57. Aletaha D, Neogi T, AJ. S, Funovits J, Felson D, III BC. Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum. 2010;62:2569-2581.
58. Power D, Codd M, Ivers L, Sant S, Barry M. Prevalence of rheumatoid arthritis in Dublin, Ireland: A population based survey. Ir. J. Med. Sci. 1999;168(3):197-200.
59. Arthritis Ireland. Available from: www.arthritisireland.ie/info/facts.php [Accessed 09 January 2009].
60. Choy EHS, Panayi GS. Cytokine Pathways and Joint Inflammation in Rheumatoid Arthritis. N. Engl. J. Med. 2001;344(12):907-916.
61. Klareskog L, Catrina A, Paget S. Rheumatoid arthritis. The Lancet 2009;373(9664):659-672.
62. Eric-Jan J. A. K, Ben A. W., De Jong, Miek A., Van Leeuwen, Hilde Swinkels Frank H. J. Van Den Hoogen Martin Van 't Hof Leo B. A. Van De Putte Martin H. Van Rijswijk, Walther J. Van Venrooij, Piet L., C. M., Van Riel. The prognostic value of anti-cyclic citrullinated peptide antibody in patients with recent-onset rheumatoid arthritis. Arthritis & Rheumatism 2000;43(8):1831-1835.
63. Alexander. J. M, Harold., Snieder. Alan. S., Rigby,Markku., Koskenvuo, Jaakko., Kaprio, Kimmo., Aho., Alan. J., Silman. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis & Rheumatism 2000;43(1):30-37.
64. Källberg H, Padyukov L, Plenge R, et al. Gene-Gene and Gene-Environment Interactions Involving HLA-DRB1, PTPN22, and Smoking in Two Subsets of Rheumatoid Arthritis. Am. J. Hum. Genet. 2007;80(5):867–875.

65. Stolt P, Bengtsson C, Nordmark B, et al. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Ann. Rheum. Dis. 2003;62(9):835–841.
66. Olsson A, Skogh T, Axelson O, Wingren G. Occupations and exposures in the work environment as determinants for rheumatoid arthritis. Occup. Environ. Med. 2004;61(3):233-238.
67. Klockars M, Koskela R, Järvinen E, Kolari P, Rossi A. Silica exposure and rheumatoid arthritis: a follow up study of granite workers 1940-81. British Medical Journal (Clinical Research Edition) 1987;294(6578):997-1000.
68. Kallberg H, Jacobsen S, Bengtsson C, et al. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. Ann. Rheum. Dis. 2009;68(2):222-227.
69. Xinjun L, Sundquist J, Sundquist K. Socioeconomic and Occupational Risk Factors for Rheumatoid Arthritis: A Nationwide Study Based on Hospitalizations in Sweden. J. Rheumatol. 2008;35(6):986-991.
70. Naz SM, Symmons DPM. Mortality in established rheumatoid arthritis. Best Practice & Research Clinical Rheumatology 2007;21(5):871-883.
71. van Tuyl LHD, Boers M, Lems WF, et al. Survival, comorbidities and joint damage 11 years after the COBRA combination therapy trial in early rheumatoid arthritis. Ann. Rheum. Dis. 2010;69(5):807-812.
72. Kroot E, van Leeuwen M, van Rijswijk M, et al. No increased mortality in patients with rheumatoid arthritis: up to 10 years of follow up from disease onset. Ann. Rheum. Dis. 2000;59:954–958.
73. Ward MM. Recent improvements in survival in patients with rheumatoid arthritis: Better outcomes or different study designs? Arthritis Rheum. 2001;44(6):1467-1469.
74. Goodson NJ, Wiles NJ, Lunt M, Barrett E, Silman A, Symmons D. Mortality in early inflammatory polyarthritis: cardiovascular mortality is increased in seropositive patients. Arthritis Rheum. 2002;46(8):2010–2019.
75. Goodson N, Marks J, Lunt M, Symmons D. Cardiovascular admissions and mortality in an inception cohort of patients with rheumatoid arthritis with onset in the 1980s and 1990s. Ann. Rheum. Dis. 2005;64(11):1595–1601.
76. Gabriel SE, Crowson CS, Kremers HM, et al. Survival in rheumatoid arthritis: A population-based analysis of trends over 40 years. Arthritis Rheum. 2003;48(1):54-58.
77. Chen Y-F, Jobanputra P, Barton P, et al. 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 Technol. Assess. 2006;10(42).
78. Wolfe F, Mitchell DM, Sibley JT, et al. The mortality of rheumatoid arthritis. Arthritis Rheum. 1994;37(4):481-494.
79. Sokka T, Hakkinen A, Krishnan E, Hannonen P. Similar prediction of mortality by the health assessment questionnaire in patients with rheumatoid arthritis and the general population. Ann. Rheum. Dis. 2004;63(5):494-497.
80. Jacobsson L, Knowler WC, Pillemer S, et al. Rheumatoid arthritis and mortality – a longitudinal study in Pima Indians. Arthritis Rheum. 1993;36(8):1045–1053.
81. Wolfe F, Michaud K, DeWitt EM. Why results of clinical trials and observational studies of antitumor necrosis factor (anti-TNF) therapy differ:

methodological and interpretive issues. Ann. Rheum. Dis. 2004;63 (Suppl ii):ii13 - ii17.
82. Myllykangas-Luosujarvi R, Aho K, Kautiainen H, Isomaki H. Shortening of life span and causes of excess mortality in a population-based series of subjects with rheumatoid arthritis. Clin. Exp. Rheumatol. 1995;13(2):149–153.
83. Wallberg-Jonsson S, Ohman M, Dahlqvist S. Cardiovascular morbidity and mortality in patients with seropositive rheumatoid arthritis in Northern Sweden. The Journal of Rheumatology 1997;24(3):445-451.
84. Symmons DP, Jones, M.A., Scott, D.L., Prior, P. Longterm mortality outcome in patients with rheumatoid arthritis: early presenters continue to do well. J. Rheumatol. 1998;25(6):1072-7.
85. Gabriel S, Crowson C, O'Fallon W. Mortality in rheumatoid arthritis: have we made an impact in 4 decades? The Journal of Rheumatology 1999;26(12):2529–2533.
86. Wolfe F, Hawley D, Anderson J. All cause and cause specific mortality in RA: a 25 year study of 1865 RA patients. Arthritis Rheum. 1999;42:S135–S385.
87. Chehata JC, Hassell AB, Clarke SA, et al. Mortality in rheumatoid arthritis: relationship to single and composite measures of disease activity. Rheumatology (Oxford). 2001;40(4):447-452.
88. Bjornadal L, Baecklund E, Yin L, Granath F, Klareskog L, Ekbom A. Decreasing mortality in patients with rheumatoid arthritis: results from a large population based cohort in Sweden, 1964–95. The Journal of Rheumatology 2002;29(5):906–912.
89. Pincus T, Keysor J, Sokka T, Krishnan E, Callahan L. Patient questionnaires and formal education level as prospective predictors of mortality over 10 years in 97% of 1416 patients with rheumatoid arthritis from 15 United States private practices. The Journal of Rheumatology 2004;31(2):229-234.
90. Young A, Koduri G, Batley M, et al. Mortality in rheumatoid arthritis. Increased in the early course of disease, in ischaemic heart disease and in pulmonary fibrosis. Rheumatology (Oxford). 2007;46(2):350-357.
91. National Institute for Clinical Excellence. NICE clinical guideline 79: Rheumatoid arthritis: the management of rheumatoid arthritis in adults. 2009.
92. Visser K, Katchamart W, Loza E, et al. Multinational evidence-based recommendations for the use of methotrexate in rheumatic disorders with a focus on rheumatoid arthritis: integrating systematic literature research and expert opinion of a broad international panel of rheumatologists in the 3E Initiative. Ann. Rheum. Dis. 2009;68:1086-1093.
93. Capell HA, Madhok R, Porter DR, et al. Combination therapy with sulfasalazine and methotrexate is more effective than either drug alone in patients with rheumatoid arthritis with a suboptimal response to sulfasalazine: results from the double-blind placebo-controlled MASCOT study. Ann. Rheum. Dis. 2007;66(2):235-241.
94. Boers M, Verhoeven A, Markusse H, et al. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. The Lancet 1997;350(9074):309-318.
95. Grigor C, Capell H, Stirling A, et al. Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial. The Lancet 2004;364(9430):263-269.

96. Möttönen T, Hannonen P, Leirisalo-Repo M, et al. Comparison of combination therapy with single-drug therapy in early rheumatoid arthritis: a randomised trial. FIN-RACo trial group. The Lancet 1999;353(9164):1568-1573.
97. Goekoop-Ruiterman Y, De Vries-Bouwstra J, Allaart D, et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): A randomized, controlled trial. Arthritis Rheum. 2005;52(11):3381-3390.
98. Katchamart W, Trudeau J, Phumethum V, Bombardier C. Efficacy and toxicity of methotrexate (MTX) monotherapy versus MTX combination therapy with non-biological disease-modifying antirheumatic drugs in rheumatoid arthritis: a systematic review and meta-analysis. Ann. Rheum. Dis. 2009;68(7):1105-1112.
99. Smolen J, Kalden J, Scott D, et al. Efficacy and Safety of leflunomide compared with placebo and sulphasalazine in active rheumatoid arthritis: a double blind randomised, multicentre trial. The Lancet 1999;353(9149):259-266.
100. Strand V, Cohen S, Schiff M, et al. Treatment of active rheumatoid arthritis with leflunomide compared with placebo and methotrexate. Arch. Intern. Med. 1999;159:2542-2550.
101. Kremer JM, Genovese MC, Cannon GW, et al. Concomitant Leflunomide Therapy in Patients with Active Rheumatoid Arthritis despite Stable Doses of Methotrexate. Ann. Intern. Med. 2002;137(9):726-733.
102. Arava: Summary of Product Characteristics updated 21/03/2011. Available from www.medicines.ie. [Accessed 10 June 2011].
103. Weinblatt ME, Kremer JM, Coblyn JS, et al. Pharmacokinetics, safety, and efficacy of combination treatment with methotrexate and leflunomide in patients with active rheumatoid arthritis. Arthritis Rheum. 1999;42(7):1322-1328.
104. Siva C, Eisen S, Shephard R, et al. Leflunomide Use During the First 33 Months After Food and Drug Administration Approval: Experience With a National Cohort of 3,325 Patients. Arthritis & Rheumatism (Arthritis Care & Research) 2003;49(6):745-751.
105. Miller M, Morton G, Anderson, LG, Keroack B, Radis C. One-year community based experience with leflunomide in the treatment of rheumatoid arthritis [abstract]. Arthritis Rheum. 2000;43(Suppl 9):S343.
106. Donahue KE, Gartlehner G, Jonas DE, et al. Systematic Review: Comparative Effectiveness and Harms of Disease-Modifying Medications for Rheumatoid Arthritis. Ann. Intern. Med. 2008;148(2):124-134.
107. Hider S, Silman A, Bunn D, Manning S, Symmons D, Lunt M. Comparing the long-term clinical outcome of treatment with methotrexate or sulfasalazine prescribed as the first disease-modifying antirheumatic drug in patients with inflammatory polyarthritis. Ann. Rheum. Dis. 2006;65(11):1449-1455.
108. Hetland ML, Stengaard-Pedersen K, Junker P, et al. Aggressive combination therapy with intraarticular glucocorticoid injections and conventional DMARDs in early rheumatoid arthritis Two Year Clinical and Radiographic Results From The CIMESTRA Study. Ann. Rheum. Dis. 2007:ard.2007.076307.

109. Dale J, Alcorn N, Capell H, Madhok R. Combination therapy for rheumatoid arthritis: Methotrexate and sulfasalazine together or with other DMARDs. Nature Clinical Practice Rheumatology 2007;3(8):450-458.
110. PBAC. PBAC review of bDMARDs for the treatment of severe active rheumatoid athritis (Public Summary Document). 2009.
111. Barton P, Jobanputra P, Wilson J, Bryan S, Burls A. 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 Technol. Assess. 2004;8(11(iii)):1-91.
112. Scott DL, Smith C, Kingsley G. Joint damage and disability in rheumatoid arthritis: an updated systematic review. Clin. Exp. Rheumatol. 2003;21(Suppl. 31):S20-27.
113. National Institute for Clinical Excellence. Final appraisal determination – Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor (part review of NICE technology appraisal guidance 36, review of NICE technology appraisal guidance 126 and 141. Available from: http://www.nice.org.uk/nicemedia/live/12135/49473/49473.pdf [Accessed 10 August 2010]. 2010.
114. Connock M, Tubeuf S, Malottki K, et al. Certolizumab pegol (CIMZIA®) for the treatment of rheumatoid arthritis. Health Technol. Assess. 2010;14(Suppl. 2):1-10.
115. Bagust A, Boland A, Hockenhull J, et al. Rituximab for the treatment of rheumatoid arthritis. Health Technol. Assess. 2009;13(Suppl. 2):23-29.
116. Canadian Agency for Drugs and Technologies in Health. Clinical and Economic Overview: Biological Response Modifier Agents for Adults with Rheumatoid Arthritis. 2010. Available from: http://www.cadth.ca/media/pdf/TR_RA_Clinical_and_Economic_Overview_e.pdf. [Accessed 01 December 2010].
117. Suarez-Almazor M, Ortiz Z, Lopez-Olivo M, et al. Long-term clinical and cost-effectiveness of infliximab and etanercept for rheumatoid arthritis (Technology Overview no 29) Available from: www.cadth.ca/media/pdf/435_AntiTNF_to_e.pdf [Accessed 12 March 2011]. 2007.
118. Chiou C-F, Choi J, Reyes CM. Cost-effectiveness analysis of biological treatments for rheumatoid arthritis. Expert Review of Pharmacoeconomics and Outcomes Research 2004;4:307-315.
119. Choi HK, Seeger JD, Kuntz KM. A cost-effectiveness analysis of treatment options for methotrexate-naive rheumatoid arthritis. The Journal of Rheumatology, 2002;29(6):1156-1165.
120. Benucci M, Saviola G, Manfredi M, Sarzi-Puttini P, Atzeni F. Cost Effectiveness Analysis of Disease-Modifying Antirheumatic Drugs in Rheumatoid Arthritis. A Systematic Review Literature. International Journal of Rheumatology 2011:Article ID 845496, 6 pages, 2011. doi:10.1155/2011/845496.
121. Schadlich PK, Zeidler H, Zink A, et al. Modelling Cost Effectiveness and Cost Utility of Sequential DMARD Therapy Including Leflunomide for Rheumatoid Arthritis in Germany: II. The Contribution of Leflunomide to Efficiency. Pharmacoeconomics 2005;23(4):395-420.

122. Maetzel A, Strand CV, Tugwell PS, GA. W, Bombardier C. Cost Effectiveness of Adding Leflunomide to a 5-year Strategy of Conventional Disease-Modifying Antirheumatic Drugs in Patients with Rheumatoid Arthritis. Arthritis & Rheumatism (Arthritis Care & Research) 2002;47(6):655-661.
123. Schipper LG, Fransen J, Barrera P, Van Riel PLCM. Methotrexate in combination with sulfasalazine is more effective in rheumatoid arthritis patients who failed sulfasalazine than in patients naive to both drugs. Rheumatology (Oxford). 2009;48(7):828-833.
124. Smolen JS, Breedveld FC, Schiff MH, et al. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology (Oxford). 2003;42(2):244-257.
125. van der Heijde DM, van 't Hof MA, van Riel PL, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Ann. Rheum. Dis. 1990;49(11):916-920.
126. Tugwell P, Boers M. Developing consensus on preliminary core efficacy endpoints for rheumatoid arthritis clinical trials. OMERACT Committee. J. Rheumatol. 1993;20(3):555-556.
127. American College of Rheumatology Committee to Reevaluate Improvement Criteria. A Proposed Revision to the ACR20: The Hybrid Measure of American College of Rheumatology Response. Arthritis & Rheumatism (Arthritis Care & Research) 2007;57(2):193–202.
128. Felson DT, Anderson JJ, Lange MLM, Wells G, LaValley MP. Should improvement in rheumatoid arthritis clinical trials be defined as fifty percent or seventy percent improvement in core set measures, rather than twenty percent? Arthritis Rheum. 1998;41(9):1564-1570.
129. Bathon JM, Martin RW, Fleischmann RM, et al. A Comparison of Etanercept and Methotrexate in Patients with Early Rheumatoid Arthritis. N. Engl. J. Med. 2000;343(22):1586-1593.
130. Siegel JN, Zhen B. Use of the American College of Rheumatology N (ACR-N) index of improvement in rheumatoid arthritis: Argument in favor. Arthritis Rheum. 2005;52(6):1637-1641.
131. Boers M. Use of the American College of Rheumatology N (ACR-N) index of improvement in rheumatoid arthritis: Argument in opposition. Arthritis Rheum. 2005;52(6):1642-1645.
132. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995;38(1):44-48.
133. Wells G, Becker JC, Teng J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. Ann. Rheum. Dis. 2009;68(6):954-960.
134. Sharp JT. Scoring radiographic abnormalities in rheumatoid arthritis. J. Rheumatol. 1989;16:568–569.
135. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. Acta Radiologica Diagnosis (Stockholm) 1977;18:481-491.

136. Van der Heijde D, Dankert T, Nieman F, Rau R, Boers M. Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis. Rheumatology (Oxford). 1999;38:941-947.
137. Bansback NJ, Anis AH, Marra CA. Patient Reported Outcomes for Rheumatoid Arthritis: Where Are We and Where Are We Going? J. Rheumatol. 2008;35(8):1482-1483.
138. Fries J, Spitz P, Young D. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J. Rheumatol. 1982;9(5):789-793.
139. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier D. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J. Rheumatol. 1993;20(3):557-560.
140. Marra C, Esdaile J, Guh D, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. Med. Care 2004;42(11):1125-1131.
141. Hurst N, Jobanputra P, Hunter M, Lambert M, Lochhead A, Brown H. Validity Of EuroQoL--A Generic Health Status Instrument--In Patients With Rheumatoid Arthritis: Economic And Health Outcomes Research Group. Rheumatology (Oxford). 1994;33(7):655-662.
142. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual. Life Res. 2011;Published Online 09 April 2011: DOI: 10.1007/s11136-011-9903-x.
143. Kekow J, Moots RJ, Emery P, et al. Patient-reported outcomes improve with etanercept plus methotrexate in active early rheumatoid arthritis and the improvement is strongly associated with remission: The COMET trial. Ann. Rheum. Dis. 2009 Mar 16. [Epub ahead of print].
144. Lindgren P, Geborek P, Kobelt G. Modelling the cost-effectiveness of treatment of rheumatoid arthritis with rituximab using registry data from Southern Sweden. Int. J. Technol. Assess. Health Care 2009;25(02):181-189.
145. Hurst N, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Rheumatology (Oxford). 1997;36(5):551-559.
146. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ. 1996;5(2):141-154.
147. Dolan P. Modeling Valuations for EuroQol Health States. Med. Care 1997;35(11):1095-1108.
148. Szende A, Williams, A. Eds. Measuring self reported population health: An international perspective based on EQ 5D. Budapest: SpringMed Publishing Ltd.; 2004.
149. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med. Care 1992;30:473 - 483.
150. Brazier J, Roberts, J, Deverill, M. The estimation of a preference-based measure of health from the SF-36. J. Health Econ. 2002;21(2):271-292.
151. Drummond MF, Sculpher MJ, O'Brien B, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes, third edition. Oxford: Oxford University Press; 2005.

152. Kharroubi S, Brazier JE, O'Hagan A. Modelling covariates for the SF-6D standard gamble health state preference data using a nonparametric Bayesian method. Soc. Sci. Med. 2007;64(6):1242-1252.
153. Torrance G, Tugwell P, Amorosi S, Chartash E, Sengupta N. Improvement in health utility among patients with rheumatoid arthritis treated with adalimumab (a human anti-TNF monoclonal antibody) plus methotrexate. Rheumatology (Oxford). 2004;43(6):712-718.
154. Marra CA, Woolcott JC, Kopec JA, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc. Sci. Med. 2005;60(7):1571-1582.
155. van Gestel AM, Prevoo, M.L., van 't Hof, M.A., van Rijswijk, M.H., van de Putte, L.B., van Riel, P.L. 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 Rheum 1999;39(1):34-40.
156. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient Outcomes in arthritis. Arthritis Rheum. 1980;23:137-145.
157. Bryan S, Longworth L. Measuring health-related utility: why the disparity between EQ-5D and SF-6D? Eur J Health Econ 2005;6(3):253-260.
158. Lamers L, Bouwmans C, van Straten A, Donker M, Hakkaart L. Comparison of EQ-5D and SF-6D utilities in mental health patients. Health Econ. 2006;15(11):1229-1236.
159. Gerard K, Nicholson T, Mullee M, Mehta R, Roderick P. EQ-5D versus SF-6D in an Older, Chronically Ill Patient Group. Applied Health Economics and Health Policy 2004;3(2):91-102.
160. Seymour J, McNamee P, Scott A, Tinelli M. Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. Health Econ. 2009;19(6):683-696.
161. Kobelt G, Lindgren P, Lindroth Y, Jacobson L, Eberhardt K. Modelling the effect of function and disease activity on costs and quality of life in rheumatoid arthritis. Rheumatology (Oxford). 2005;44(9):1169-1175.
162. Kobelt G, Eberhardt K, Jonsson L, Jonsson B. Economic consequences of the progression of rheumatoid arthritis in Sweden. Arthritis Rheum. 1999;42:347-356.
163. Tanno M, Nakamura I, Ito K, et al. Modeling and cost-effectiveness analysis of etanercept in adults with rheumatoid arthritis in Japan: a preliminary analysis. Modern Rheumatology 2006;16(2):77-84.
164. Bansback N, Marra C, Tsuchiya A, et al. Using the health assessment questionnaire to estimate preference-based single indices in patients with rheumatoid arthritis. Arthritis Care Res. 2007;57(6):963-971.
165. Harrison MJ, Lunt M, Bansback N, Verstappen S, Watson K, Symmons D. The Validity of Estimating EQ-5D and SF-6D Utility Values from the HAQ. In: American College of Rheumatology Annual Meeting; 2008; San Francisco, October 24-29; 2008.
166. Barton GR, Sach TH, Jenkinson C, Avery AJ, Doherty M, Muir KR. Do estimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? Health & Quality of Life Outcomes 2008;6:51.

167. Craig B, Busschbach J. The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. Population Health Metrics 2009;7:3.
168. Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. J. Health Serv. Res. Policy 1999;4(3):174-84.
169. Sach T, Barton G, Jenkinson C, Doherty M, Avery A, Muir K. Comparing cost-utility estimates: does the choice of EQ-5D or SF-6D matter? Med. Care 2009;47(8):889-894.
170. Harrison M, Davies L, Bansback N, et al. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. Qual. Life Res. 2009;18(9):1195-1205.
171. Williams A. A measurement and valuation of health: a chronicle. Discussion Paper. York, Centre for Health Economics, York Health Economics Consortium, NHS Centre for Reviews & Dissemination, University of York. 1995.
172. Barton G, Sach T, Doherty M, Avery A, Jenkinson C, Muir K. An assessment of the discriminative ability of the EQ-5Dindex, SF-6D, and EQ VAS, using sociodemographic factors and clinical conditions. The European Journal of Health Economics 2008;9(3):237-249.
173. Marra CA, Marion SA, Guh DP, et al. Not all "quality-adjusted life years" are equal. J. Clin. Epidemiol. 2007;60(6):616-624.
174. Harrison MJ, Davies LM, Bansback NJ, et al. Why Do Patients with Inflammatory Arthritis Often Score States "Worse than Death" on the EQ-5D? An Investigation of the EQ-5D Classification System. Value Health 2009;12(6):1026-1034.
175. Adams R, Walsh C, Veale DJ, Bresnihan B, FitzGerald O, Barry M. Understanding the relationship between the EQ-5D, SF-6D, HAQ and disease activity in inflammatory arthritis. Pharmacoeconomics 2010;28(6):477-487.
176. Torrance G, Sackett D, Thomas W. A utility maximization model for evaluation of health care programmes. Health Serv. Res. 1972;7(2):118-133.
177. Torrance GW, Boyle, M. H., Horwood, S. P. Application of Multi-Attribute Utility Theory to Measure Social Preference for Health States. Oper. Res. 1982;30(6):1043 - 1069.
178. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med. Care 2005;43:203 - 220.
179. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. Med. Decis. Making 1994;14:9 - 18.
180. Craig BM, Busschbach JJV. Toward a more universal approach in health valuation. Health Econ. 2010;[Published online 02 Aug 2010].
181. Stolk EA, Oppe M, Scalone L, Krabbe PFM. Discrete Choice Modeling for the Quantification of Health States: The Case of the EQ-5D. Value Health 2010;13(8):1005-1013.
182. Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. Med. Care 2007;45(3):238–244.
183. Craig BM, Oppe M. From a different angle: A novel approach to health valuation. Soc. Sci. Med. 2010;70(2):169-174.

184. Li L, Fu A. Some methodological issues with the analysis of preference-based EQ-5D index score. Health Services and Outcomes Research Methodology 2009;9(3):162–176.
185. Shaw J, Pickard A, Yu S, Chen S, Iannacchione V, Johnson Jea. A median model for predicting United States population-based EQ-5D health state preferences. Value Health 2010;13(2):278-288.
186. Brazier J, Rowen D, Yang Y, Tsuchiya A. Using rank and discrete choice data to estimate health state utility values on the QALY scale. University of Sheffield. HEDS Discussion Paper 09/10. 2009.
187. Craig BM, Busschbach JJV, Salomon JA. Modeling Ranking, Time Trade-Off, and Visual Analog Scale Values for EQ-5D Health States: A Review and Comparison of Methods. Med. Care 2009;47(6):634-641.
188. Solomon D, Karlson E, Rimm E, et al. Cardiovascular Morbidity and Mortality in Women Diagnosed With Rheumatoid Arthritis. Circulation 2003;107(9):1303-1307.
189. McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. J. Health Econ. 2006;25(3):418-431.
190. Coast J. Reprocessing data to form QALYs. Br. Med. J. 1992;305(6845):87-90.
191. Bojke L, Claxton K, Bravo-Vergel Y, Sculpher M, Palmer S, Abrams K. Eliciting Distributions to Populate Decision Analytic Models. Value Health 2010;13(5):557-564.
192. Brazier J, Yang Y, Tsuchiya A, Rowen D. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. The European Journal of Health Economics 2010;11(2):215-225.
193. Wijeysundera HC, Tomlinson G, Norris CM, Ghali WA, Ko DT, Krahn MD. Predicting EQ-5D Utility Scores from the Seattle Angina Questionnaire in Coronary Artery Disease. Med. Decis. Making 2011;31(3):481-493.
194. Walters S, Campbell M. The use of bootstrap methods for analysing health-related quality of life outcomes (particularly the SF-36). Health and Quality of Life Outcomes 2004;2:70 (doi:10.1186/1477-7525-2-70).
195. Lillegraven S, Kristiansen IS, Kvien TK. Comparison of utility measures and their relationship with other health status measures in 1041 patients with rheumatoid arthritis. Ann. Rheum. Dis. 2010;69(10):1762-1767.
196. Rovira J, Antoñanzas F. Economic analysis of health technologies and programmes. A Spanish proposal for methodological standardisation. Pharmacoeconomics 1995;8(3):245-252.
197. U.S. Public Health Service (Panel on Cost-Effectiveness in Health and Medicine). The AMCP Format for Formulary Submissions (Version 2.1, April 2005). Laurenceville (NJ) 2005:Pharmacoeconomic Guidelines Around The World: ISPOR [online] Available from URL: http://www.ispor.org/peguidelines/index.asp [Accessed 30 March 2009].
198. Craig B, Busschbach, J. The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. Popul Health Metr 2009;7(3).
199. Conner-Spady B, Surez-Almazor M. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. Med. Care 2003; 41(7):791-801.

200. Whitehurst DGT, Bryan S. Another Study Showing that Two Preference-Based Measures of Health-Related Quality of Life (EQ-5D and SF-6D) are Not Interchangeable. But Why Should We Expect Them to be? Value Health 2011;14(4):531-538.
201. Prescott-Clarke P, Primatesta P, eds. Health Survey for England 1996. Vol 1: Findings; Vol 11: Survey methodology and documentation. London; The Stationary Office: 1998.
202. Bansback N, Sun H, Guh DP, et al. Impact of the recall period on measuring health utilities for acute events. Health Econ. 2008;17(12):1413-1419.
203. Brazier J, Roberts, J, Tsuchiya, A., Busschbach, J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004;13(9):873-884.
204. Konig H-H, Bernert S, Angermeyer MC, et al. Comparison of Population Health Status in Six European Countries: Results of a Representative Survey Using the EQ-5D Questionnaire. Med. Care 2009;47(2):255-261.
205. Busschbach JJV, Weijnen T, Nieuwenhuizen M, et al. A comparison of EQ-5D time trade-off values obtained in Germany, The United Kingdom and Spain. The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective 2003:143 - 165.
206. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996;276:1253-1258.
207. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. Health Econ. 2011;20(3):348-361.
208. Menzies NA, Salomon JA. Non-monotonicity in the episodic random utility model. Health Econ. 2010;DOI: 10.1002/hec.1683 [First published online 18 November 2010].
209. Harrison MJ, Davies LM, Bansback NJ, et al. Why Do Patients with Inflammatory Arthritis Often Score States "Worse than Death" on the EQ-5D? An Investigation of the EQ-5D Classification System. Value Health 2009;12(6):1026-1034.
210. Bansback NJ, Ara R, Barkham N, et al. Estimating the cost and health status consequences of treatment with TNF antagonists in patients with psoriatic arthritis. Rheumatology (Oxford). 2006;45(8):1029-1038.
211. Kielhorn A, Porter D, Diamantopoulos A, Lewis G. UK cost-utility analysis of rituximab in patients with rheumatoid arthritis that failed to respond adequately to a biologic disease-modifying antirheumatic drug. Curr. Med. Res. Opin. 2008;24(9):2639-2650.
212. Harrison MJ, Davies LM, Bansback NJ, Ingram M, Anis AH, Symmons DPM. The Validity and Responsiveness of Generic Utility Measures in Rheumatoid Arthritis: A Review. J. Rheumatol. 2008;35(4):592-602.
213. National Institute for Clinical Excellence. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor. NICE technology appraisal guidance 195 Available from: www.nice.org.uk/TA195 [Acessed on 11 February 2011]. 2010.
214. Louviere J, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. Health Economics, Policy and Law 2009;4:527-546.

215. Bryan S, Dolan P. Discrete choice experiments in health economics. For better or for worse? European Journal of Health Economics 2004;5:199-202.
216. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. Br. Med. J. 2000;320:1530-1533.
217. PROMIS. Active PROMIS Studies. Available from: www.nihpromis.org [Accessed 01 June 2011].
218. Weinblatt ME, Keystone EC, Furst DE, et al. Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: The ARMADA trial. Arthritis Rheum. 2003;48(1):35-45.
219. Keystone EC, Kavanaugh AF, Sharp JT, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: A randomized, placebo-controlled, 52-week trial. Arthritis Rheum. 2004;50(5):1400-1411.
220. van de Putte LBA, Atkins C, Malaise M, et al. Efficacy and safety of adalimumab as monotherapy in patients with rheumatoid arthritis for whom previous disease modifying antirheumatic drug treatment has failed. Ann. Rheum. Dis. 2004;63(5):508-516.
221. Miyasaka N. Clinical investigation in highly disease-affected rheumatoid arthritis patients in Japan with adalimumab applying standard and general evaluation: the CHANGE study. Modern Rheumatology 2008;18(3):252-262.
222. Kim HY, Lee SK, Song YW, et al. A randomized, double-blind, placebo-controlled, phase III study of the human anti-tumor necrosis factor antibody adalimumab administered as subcutaneous injections in Korean rheumatoid arthritis patients treated with methotrexate. APLAR Journal of Rheumatology 2007;10(1):9-16.
223. Chen DY, Chou SJ, Hsieh TY, et al. Randomized, double-blind, placebo-controlled, comparative study of human anti-TNF antibody adalimumab in combination with methotrexate and methotrexate alone in Taiwanese patients with active rheumatoid arthritis. J. Formos. Med. Assoc. 2009;108(4):310.
224. Weinblatt ME, Keystone EC, Furst DE, Kavanaugh AF, Chartash EK, Segurado OG. Long term efficacy and safety of adalimumab plus methotrexate in patients with rheumatoid arthritis: ARMADA 4 year extended study. Ann. Rheum. Dis. 2006;65(6):753-759.
225. Nesbitt A, Fossati G, Bergin M, et al. Mechanism of action of certolizumab pegol (CDP870): In vitro comparison with other anti-tumor necrosis factor α agents. Inflamm. Bowel Dis. 2007;13(11):1323-1332.
226. Keystone E, Van der Heijde D, Mason JD, et al. Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: Findings of a fifty-two-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. Arthritis Rheum. 2008;58(11):3319-3329.
227. Smolen J, Landewe RB, Mease P, et al. Efficacy and safety of certolizumab pegol plus methotrexate in active rheumatoid arthritis: the RAPID 2 study. A randomised controlled trial. Ann. Rheum. Dis. 2009;68(6):797-804.
228. Fleischmann R, Vencovsky J, van Vollenhoven RF, et al. Efficacy and safety of certolizumab pegol monotherapy every 4 weeks in patients with rheumatoid arthritis failing previous disease-modifying antirheumatic therapy: the FAST4WARD study. Ann. Rheum. Dis. 2009;68(6):805-811.

229. European Medicines Agency. Assessment Report for Cimzia (Doc.Ref.: EMEA/664021/2009) Accessed July 2010. 2009.
230. Weinblatt ME, Kremer JM, Bankhurst AD, et al. A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. N. Engl. J. Med. 1999;340(4):253-259.
231. Moreland LW, Schiff MH, Baumgartner SW, et al. Etanercept therapy in rheumatoid arthritis. A randomized controlled trial. Ann. Intern. Med. 1999;130:478 - 486.
232. Klareskog L, Gaubitz M, Rodriguez-Valverde V, Malaise M, Dougados M, Wajdula J. A long-term, open-label trial of the safety and efficacy of etanercept (Enbrel) in patients with rheumatoid arthritis not treated with other disease-modifying antirheumatic drugs. Ann. Rheum. Dis. 2006;65(12):1578-1584.
233. Klareskog L, van der Heijde D, de Jager JP, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. The Lancet 2004;363(9410):675-681.
234. Emery P, Breedveld F, Hall S, et al. Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. The Lancet 2008;372(9636):375-382.
235. Smolen JS, Kay J, Doyle MK, et al. Golimumab in patients with active rheumatoid arthritis after treatment with tumour necrosis factor [alpha] inhibitors (GO-AFTER study): a multicentre, randomised, double-blind, placebo-controlled, phase III trial. The Lancet 2009;374(9685):210-221.
236. Keystone EC, Genovese MC, Klareskog L, et al. Golimumab, a human antibody to tumour necrosis factor (alpha) given by monthly subcutaneous injections, in active rheumatoid arthritis despite methotrexate therapy: The GO-FORWARD Study. Ann. Rheum. Dis. 2009;68(6):789-796.
237. Emery P, Fleischmann R, Moreland L, et al. Golimumab, a human anti-tumor necrosis factor alpha monoclonal antibody, injected subcutaneously every four weeks in methotrexate-naive patients with active rheumatoid arthritis: Twenty-four-week results of a phase III, multicenter, randomized, double-blind, placebo-controlled study of golimumab before methotrexate as first-line therapy for early-onset rheumatoid arthritis. Arthritis Rheum. 2009;60(8):2272-2283.
238. Denby A. Golimumab. London New Drugs Group 2009;Available from: www.nelm.nhs.uk [Accessed 20 March 2010].
239. Maini R, St Clair EW, Breedveld F, et al. Infliximab (chimeric anti-tumor necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomized phase III trial. ATTRACT Study Group. The Lancet 1999;354:1932 - 1939.
240. Westhovens R, Yocum D, Han J, et al. The safety of infliximab, combined with background treatments, among patients with rheumatoid arthritis and various comorbidities: A large, randomized, placebo-controlled trial. Arthritis Rheum. 2006;54(4):1075-1086.
241. Abe T, Takeuchi T, Miyasaka N, et al. A multicenter, double-blind, randomized, placebo controlled trial of infliximab combined with low dose

methotrexate in Japanese patients with rheumatoid arthritis. J. Rheumatol. 2006;33(1):37-44.
242. Zhang FC, Hou Y, Huang F, et al. Infliximab versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: A preliminary study from China. APLAR Journal of Rheumatology 2006;9(2):122-126.
243. St.Clair EW, Van Der Heijde DMFM, Smolen JS, et al. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: A randomized, controlled trial. Arthritis Rheum. 2004;50(11):3432-3443.
244. Klarenbeek NB, Guler-Yuksel M, van der Kooij SM, et al. The impact of four dynamic, goal-steered treatment strategies on the 5-year outcomes of rheumatoid arthritis patients in the BeSt study. Ann. Rheum. Dis. 2011;70(6):1039-1046.
245. Van Vollenhoven RF, Ashling J. Rheumatoid arthritis registries in Sweden. Clin. Exp. Rheumatol. 2005;23(5 SUPPL. 39):S195-S200.
246. van Vollenhoven R, Ernestam S, Harju1 A, Bratt J, Klareskog L. Etanercept versus etanercept plus methotrexate: a registry based study suggesting that the combination is clinically more efficacious. Arthritis Research and Therapy 2003;5:R347-R351.
247. Hetland ML, Christensen IJ, Tarp U, et al. Direct comparison of treatment responses, remission rates, and drug adherence in patients with rheumatoid arthritis treated with adalimumab, etanercept, or infliximab: Results from eight years of surveillance of clinical practice in the nationwide Danish DANBIO registry. Arthritis Rheum. 2010;62(1):22-32.
248. Kievit W, Adang EM, Fransen J, et al. The effectiveness and medication costs of three anti-tumour necrosis factor {alpha} agents in the treatment of rheumatoid arthritis from prospective clinical practice data. Ann. Rheum. Dis. 2008;67(9):1229-1234.
249. Geborek P, Crnkic M, Petersson IF, Saxne T. Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southern Sweden. Ann. Rheum. Dis. 2002;61:793 - 798.
250. Zink A, Listing JD, Kary S, et al. Treatment continuation in patients receiving biologics or conventional DMARD therapy. Ann. Rheum. Dis. 2005;64:1274 - 1279.
251. Canadian Agency for Drugs and Technologies in Health. Clinical and Economic Overview: Biological Response Modifier Agents for Adults with Rheumatoid Arthritis (July). In: http://www.cadth.ca/index.php/en/cadth/therapeutic-review-pilot/therapeutic-review-reports; 2010.
252. Tran K, Asakawa. K, Cimon K, et al. Pharmacologic-based Strategies for Smoking Cessation: Clinical and Cost-Effectiveness Analyses [Internet]. Ottawa: Canadian Agency for Drugs and Technologies in Health;(Technology Report; no. 130). 2010.
253. Nixon R, Bansback N, Brennan A. The efficacy of inhibiting tumour necrosis factor {alpha} and interleukin 1 in patients with rheumatoid arthritis: a meta-analysis and adjusted indirect comparisons. Rheumatology (Oxford). 2007;46(7):1140-1147.
254. Sutton A, Ades AE, Cooper N, Abrams K. Use of Indirect and Mixed Treatment Comparisons for Technology Assessment. Pharmacoeconomics 2008;26:753-767.

255. Sutton A, Abrams K, Jones D. An illustrated guide to the methods of meta-analysis. J. Eval. Clin. Pract. 2001;7(2):135-148.
256. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat. Med. 2004;23(20):3105-3124.
257. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Stat. Med. 2007;26(6):1237-1254.
258. Schiff M, Keiserman M, Codding C, et al. Efficacy and safety of abatacept or infliximab vs placebo in ATTEST: a phase III, multi-centre, randomised, double-blind, placebo-controlled study in patients with rheumatoid arthritis and an inadequate response to methotrexate. Ann. Rheum. Dis. 2008;67(8):1096-1103.
259. Bucher H, Guyatt G, Griffith L, Walter S. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J. Clin. Epidemiol. 1997;50(6):683-691.
260. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2009;172(1):137–159.
261. Launois R, Avouac B, Berenbaum F, et al. Comparison of Certolizumab Pegol with Other Anticytokine Agents for Treatment of Rheumatoid Arthritis: A Multiple-treatment Bayesian Metaanalysis. The Journal of Rheumatology 2011;38(5):835-845.
262. Lee YH, Woo JH, Rho YH, Choi SJ, Ji JD, Song GG. Meta-analysis of the combination of TNF inhibitors plus MTX compared to MTX monotherapy, and the adjusted indirect comparison of TNF inhibitors in patients suffering from activerheumatoid arthritis. Rheumatol. Int. 2008;28(6):553-559.
263. Alonso-Ruiz A, Pijoan J, Ansuategui E, Urkaregi A, Calabozo M, Quintana A. Tumor necrosis factor alpha drugs in rheumatoid arthritis: systematic review and metaanalysis of efficacy and safety. BMC Musculoskeletal Disorders 2008;9(1):52. (DOI:10.1186/1471-2474-9-52).
264. Venkateshan SP, Sidhu S, Malhotra S, Pandhi P. Efficacy of Biologicals in the Treatment of Rheumatoid Arthritis. Pharmacology 2009;83(1):1-9.
265. Lloyd S, Bujkiewicz S, Wailoo AJ, Sutton AJ, Scott D. The effectiveness of anti-TNF-a therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis. Rheumatology (Oxford). 2010;49(12):2313-2321.
266. Salliot C, Finckh A, Katchamart W, et al. Indirect comparisons of the efficacy of biological antirheumatic agents in rheumatoid arthritis in patients with an inadequate response to conventional disease-modifying antirheumatic drugs or to an anti-tumour necrosis factor agent: a meta-analysis. Ann. Rheum. Dis. 2011;70(2):266-271.
267. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian Meta-Analysis of Multiple Treatment Comparisons: An Introduction to Mixed Treatment Comparisons. Value Health 2008;11(5):956-964.
268. Kristensen LE, Christensen R, Bliddal H, Geborek P, Danneskiold-Samsee B, Saxne T. The number needed to treat for adalimumab, etanercept, and infliximab based on ACR50 response in three randomized controlled trials on established rheumatoid arthritis: a systematic literature review. Scand. J. Rheumatol. 2007;36:411 - 417.

269. Kristensen LE, Jakobsen AK, Bartels EM, et al. The number needed to treat for second-generation biologics when treating established rheumatoid arthritis: a systematic quantitative review of randomized controlled trials. Scand. J. Rheumatol. 2011;40(1):1-7.
270. Devine E, Alfonso-Cristancho R, Sullivan S. Effectiveness of biologic therapies for rheumatoid arthritis: an indirect comparisons approach. Pharmacotherapy 2011;31(1):39-51.
271. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the Standard EQ-5D Three-Level System with a Five-Level Version. Value Health 2008;11(2):275-284.
272. Bergman GJD, Hochberg MC, Boers M, Wintfeld N, Kielhorn A, Jansen JP. Indirect Comparison of Tocilizumab and Other Biologic Agents in Patients with Rheumatoid Arthritis and Inadequate Response to Disease-Modifying Antirheumatic Drugs. Semin. Arthritis Rheum. 2010;39(6):425-441.
273. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 2000;10:325-337.
274. Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. Stat. Med. 2002;21(11):1601-1623.
275. Hyrich K, Watson K, Silman A, Symmons D, Register TBSRB. Predictors of response to anti-TNF-{alpha} therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. Rheumatology (Oxford). 2006;45(12):1558-1565.
276. Oppe M, Al M, Rutten-van Molken M. Comparing Methods of Data Synthesis: Re-Estimating Parameters of an Existing Probabilistic Cost-Effectiveness Model. Pharmacoeconomics 2011;29(3):239-250.
277. Bojke L, Claxton K, Palmer S, Sculpher M. Defining and Characterising Structural Uncertainty in Decision Analytic Models. York, UK: Centre for Health Economics, University of York; 2006 March. CHE Research Paper 9. Available from: http://www.york.ac.uk/media/che/documents/papers/researchpapers/rp9_structural_uncertainty_in_decision_analytic_models.pdf. [Accessed 10 July 2011].
278. Manning W, Fryback D, Weinstein M. Reflecting uncertainty in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p 247–75.
279. Jackson C, Thompson SG, LD. S. Accounting for uncertainty in health economic decision models by using model averaging. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2009;172(2):383–404.
280. Jackson C, Sharples L, Thompson S. Structural and parameter uncertainty in Bayesian cost-effectiveness models. Journal of the Royal Statistical Society: Series C (Applied Statistics), 2009;59(2):233–253.
281. Adams R, Schmitz S, FitzGerald O, Walsh C, Barry M. Does Model Choice in Mixed Treatment Comparisons Affect the Decision in Economic Evaluations? ISPOR Connections 2011;17:2.
282. Kay J, Matteson EL, Dasgupta B, et al. Golimumab in patients with active rheumatoid arthritis despite treatment with methotrexate: A randomized, double-blind, placebo-controlled, dose-ranging study. Arthritis Rheum. 2008;58(4):964-975.

283. Gartlehner G, Hansen RA, Jonas BL, Thieda P, Lohr KN. The comparative efficacy and safety of biologics for the treatment of rheumatoid arthritis: a systematic review and metaanalysis. J. Rheumatol. 2006;33(12):2398-2408.
284. Cooper N, Peters J, Lai M, et al. How Valuable are Multiple Treatment Comparison Methods in Evidence-Based Health-Care Evaluation? Value Health 2011;14(2):371-380.
285. Combe B, Codreanu C, Fiocco U, et al. Efficacy, safety and patient-reported outcomes of combination etanercept and sulfasalazine versus etanercept alone in patients with rheumatoid arthritis: a double-blind randomised 2-year study. Ann. Rheum. Dis. 2009;68(7):1146-1152.
286. Singh JA, Christensen R, Wells GA, et al. Biologics for rheumatoid arthritis: an overview of Cochrane reviews (Review). Cochrane Database of Systematic Reviews 2009(4).
287. Strand V, Sokolove J. Randomized controlled trial design in rheumatoid arthritis: The past decade. Arthritis Research and Therapy 2009;11(1):205 (doi:10.1186/ar2555).
288. Kristensen L, Saxne T, Nilsson J-A, Geborek P. Impact of concomitant DMARD therapy on adherence to treatment with etanercept and infliximab in rheumatoid arthritis. Results from a six-year observational study in southern Sweden. Arthritis Research & Therapy 2006;8(6):R174-R184.
289. Du Pan SM, Dehler S, Ciurea A, Ziswiler H-R, Gabay C, Finckh A. Comparison of drug retention rates and causes of drug discontinuation between anti–tumor necrosis factor agents in rheumatoid arthritis. Arthritis Care Res. 2009;61(5):560.
290. National Centre for Pharmacoeconomics. Cost-effectiveness of golimumab (Simponi®) in the treatment of moderate to severe rheumatoid arthritis. http://www.ncpe.ie/document.php?cid=33&sid=138&docid=189 2010.
291. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 2000;10:325--337.
292. Higgins JPT, Green S, editors. Cochrane Handbook for systematic reviews of interventions, 4.2.5[updated May 2005]. Chichester: John Wiley & Sons, Ltd; 2005.
293. Zink A, Strangfeld A, Schneider M, et al. Effectiveness of tumor necrosis factor inhibitors inrheumatoid arthritisin an observational cohort study: Comparison of patients according to their eligibility for major randomized clinical trials. Arthritis Rheum. 2006;54(11):3399-3407.
294. Finckh A, Dudler J, Wermelinger F, et al. Influence of anti-infliximab antibodies and residual infliximab concentrations on the occurrence of acquired drug resistance to infliximab in rheumatoid arthritis patients. Joint Bone Spine 2010;77(4):313-318.
295. Jobanputra P, Callaghan R. Adverse reactions to and safety of newer disease-modifying antirheumatic drugs for rheumatoid arthritis. Adverse Drug Reaction Bulletin 2003(223):855.
296. Smolen J, Emery P. Infliximab: 12 years of experience. Arthritis Research & Therapy 2011;13(Suppl 1):S2-S20.
297. Kievit W, Fransen J, Oerlemans AJM, et al. The efficacy of anti-TNF in rheumatoid arthritis, a comparison between randomised controlled trials and clinical practice. Ann. Rheum. Dis. 2007;66(11):1473-1478.

298. O'Rourke K, Walsh C, Hutchinson M. Outcome of beta-interferon treatment in relapsing-remitting multiple sclerosis: a Bayesian analysis. J. Neurol. 2007;254(11):1547-1554.
299. Jobanputra P, Barton P, Bryan S, Burls A. The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation. Health Technol. Assess. 2002;6(21).
300. Krishnan E, Sokka T, Hakkinen A, Hubert H, Hannonen P. Normative values for the health assessment questionnaire disability index. Arthritis Rheum. 2004;50:953–60.
301. Breedveld FC, Weisman MH, Kavanaugh AF, et al. The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. Arthritis Rheum. 2006;54(1):26-37.
302. National Institute for Health and Clinical Excellence. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor. NICE technology appraisal guidance 195. 2010. Available from: www.nice.org.uk/TA195 [Accessed on 11 February 2011].
303. Adams R, Craig BM, Walsh CD, et al. The Impact of a Revised EQ-5D Population Scoring on Preference-based Utility Scores in an Inflammatory Arthritis cohort. Value Health 2011;14(6):921-927.
304. Health Services Executive. Nursing Home Support Scheme Cost of Care in Public Nursing Homes. http://www.hse.ie/eng/services/Find_a_Service/Older_People_Services/nhss/costhse.pdf [Accessed 10 Sept 2010]. 2011.
305. Irish Society for Rheumatology. Guidelines for Prescribing TNFalpha Blockers in Adults with Rheumatoid Arthritis. 2006. Available from www.isr.ie. [Accessed 01 September 2009].
306. Madden D, Nolan B, Nolan A. GP reimbursement and visiting behaviour in Ireland. UCD Centre for Economic Research Working Paper Series;WP04/26, 2004.
307. National Centre for Pharmacoeconomics. Cost-effectiveness of certolizumab pegol (Cimzia®) for the treatment of moderate to severe rheumatoid arthritis [Accessed 01 July 2011]. http://www.ncpe.ie 2010.
308. Coyle D JM, Blumenauer B, Cranney A, Maetzel A, Tugwell P, Wells GA. Infliximab and etanercept in patients with rheumatoid arthritis: a systematic review and economic evaluation [Technology report no 64]. Ottawa: Canadian Coordinating Office for Health Technology Assessment; 2006.
309. Schmitz S, Adams R, Walsh C, Barry M, FitzGerald O. A Mixed Treatment Comparison of the efficacy of Anti-TNF agents in rheumatoid arthritis for methotrexate non-responders demonstrates differences between treatments: A Bayesian approach. Published Online First: 2011. Ann. Rheum. Dis. 2011.
310. Hyrich KL, Lunt M, Dixon WG, Watson KD, Symmons DPM, on behalf of the BSRBR. Effects of switching between anti-TNF therapies on HAQ response in patients who do not respond to their first anti-TNF drug. Rheumatology (Oxford). 2008;47(7):1000-1005.
311. Klareskog L, van der Heijde D, de Jager JP, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment

alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. Lancet 2004;363:675 - 681.
312. Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: Systematic review and meta-analysis of rare harmful effects in randomized controlled trials. Journal of the American Medical Association 2006;295(19):2275-2285.
313. Corona R, Bigby M. What are the risks of serious infections and malignancies for patients treated with anti-tumor necrosis factor antibodies? Arch. Dermatol. 2007;143(3):405-406.
314. Wolfe F, Michaud K. The loss of health status in rheumatoid arthritis and the effect of biologic therapy: a longitudinal observational study. Arthritis research & therapy 2010.
315. Cook JP, Vernon JA, Manning R. Pharmaceutical Risk-Sharing Agreements. Pharmacoeconomics 2008;26(7):551-558.
316. Bogglid M, Palace J, Barton P, et al. Multiple Sclerosis risk sharing scheme: two year results of clinical cohort study with historical comparator. British Medical Journal, 2009;b4677 doi: 10.1136/bmj.b4677.
317. Department of Health (UK). A new value-based approach to the pricing of branded medicines - a consultation. Available from [http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH_122760] Accessed 12 November 2011. 2011.
318. Neumann P, Lin P, Greenberg D, et al. Do Drug Formulary Policies Reflect Evidence of Value? Am. J. Manag. Care 2006;12(1):30-36.
319. Fendrick A, Smith D, Chernew M, Shah S. A benefit-based copay for prescription drugs: patient contribution based on total benefits, not drug acquisition cost. Am. J. Manag. Care 2001;7(9):861-867.

Appendices

Appendix 1: Literature Review of Economic Evaluations in RA

Search strategy for literature review.

Literature Search strategies

All strategies were tried and the search revealing the greatest number of results was chosen. Duplicates were removed.

PubMed, EMBASE, Cochrane

Keyword search

Limits Activated: Humans, Randomized Controlled Trial, English, All Adult: 19+ years

1. Rheumatoid arthritis
2. anti-tumor necrosis factor alpha {tw} OR anti-tumour necrosis factor alpha {tw} OR anti-tnf{tw} OR TNFR-Fc{tw} OR Tumor Necrosis Factor-alpha{tw} OR etanercept {tw} OR enbrel {tw} OR infliximab {tw} OR remicade {tw} OR adalimumab{tw} OR humira {tw} OR golimumab{tw} OR simponi{tw} OR certolizumab {tw} OR certolizumab pegol{tw} OR cimzia{tw}
3. Economic evaluation{tw} OR health technology assessment{tw} OR technology appraisal{tw}OR cost-effectiveness{tw} OR cost-utility{tw}
4. #1 AND #2 AND #3

Alternative Search Strategy - without tag

Limits Activated: Humans, Randomized Controlled Trial, English, All Adult: 19+ years

5. Rheumatoid arthritis
6. anti-tumor necrosis factor alpha OR anti-tumour necrosis factor alpha OR anti-tnf OR TNFR-Fc OR Tumor Necrosis Factor-alpha OR etanercept OR

enbrel OR infliximab OR Remicade OR adalimumab OR humira OR golimumab OR simponi OR certolizumab OR certolizumab pegol OR cimzia

7. Economic evaluation OR health technology assessment OR technology appraisal OR cost-effectiveness OR cost-utility
8. #1 AND #2 AND #3

Results 112 records

TUFTS Database (CEA Registry)

Limits: Pharmaceutical, musculoskeletal disease: Results 164 records

Limits: Pharmaceutical, musculoskeletal disease, lifetime: 71 records

Limits: Pharmaceutical, musculoskeletal disease, lifetime,

Relevant Records (Include either a DMARD or biological for RA): 31

Appendix 2: Quality of Life Questionnaires used for data collection

EQ-5D

Note: (VAS is not given here as was not used in the calculation of utility)

Figure 1: EQ-5D (UK English version)

By placing a tick in one box in each group below, please indicate which statements best describe your own health state today.

Mobility

I have no problems in walking about ❑
I have some problems in walking about ❑
I am confined to bed ❑

Self-Care

I have no problems with self-care ❑
I have some problems washing or dressing myself ❑
I am unable to wash or dress myself ❑

Usual Activities (*e.g. work, study, housework, family or leisure activities)*

I have no problems with performing my usual activities ❑
I have some problems with performing my usual activities ❑
I am unable to perform my usual activities ❑

Pain/Discomfort

I have no pain or discomfort ❑
I have moderate pain or discomfort ❑
I have extreme pain or discomfort ❑

Anxiety/Depression

I am not anxious or depressed ❑
I am moderately anxious or depressed ❑
I am extremely anxious or depressed ❑

Short Form-36 (SF-36) Version 1

SHORT FORM 36 HEALTH SURVEY

MRN: [ ][ ][ ][ ][ ][ ][ ] Initials: [ ][ ][ ] Date: [D][D]/[M][M]/[Y][Y]

Instructions: This survey asks for your views about your health. This information will help keep track of how well you fell and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:

	(circle one)
Excellent	1
Very good	2
Good	3
Fair	4
Poor	5

2. Compared to one year ago, how would you rate your health now?

	(circle one)
Much better now than one year ago	1
Somewhat better now than one year ago	2
About the same as one year ago	3
Somewhat worse now than one year ago	4
Much worse now than one year ago	5

SHORT FORM 36 HEALTH SURVEY

MRN: [ ][ ][ ][ ][ ][ ][ ] Initials: [ ][ ][ ] Date: [ ][ ]/[ ][ ]/[ ][ ]

3. The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so how much?

(circle one number on each line)

	ACTIVITIES	Yes, Limited A Lot	Yes, Limited A Little	No, Not Limited At All
a.	**Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports.	1	2	3
b.	**Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.	1	2	3
c.	Lifting or carrying groceries.	1	2	3
d.	Climbing **several** flights of stairs.	1	2	3
e.	Climbing **one** flight of stairs.	1	2	3
f.	Bending, kneeling, or stooping.	1	2	3
g.	Walking **more than half a mile**.	1	2	3
h.	Walking **half a mile**.	1	2	3
i.	Walking **one hundred yards**.	1	2	3
j.	Bathing or dressing yourself.	1	2	3

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

(circle one number on each line)

		YES	NO
a.	Cut down on the **amount of time** you spent on work or other activities.	1	2
b.	**Accomplished less** than you would like.	1	2
c.	Were limited in the **kind** of work or other activities.	1	2
d.	Had **difficulty** performing the work or other activities (for example, it took extra effort)	1	2

SHORT FORM 36 HEALTH SURVEY

MRN: [ ][ ][ ][ ][ ][ ][ ] Initials: [ ][ ][ ] Date: [D][D]/[M][M]/[Y][Y]

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

(circle one number on each line)

		YES	NO
a.	Cut down on the **amount of time** you spent on work or other activities	1	2
b.	**Accomplished less** than you would like.	1	2
c.	Didn't do work or other activities as **carefully** as usual.	1	2

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?

	(circle one)
Not at all	1
Slightly	2
Moderately	3
Quite a bit	4
Extremely	5

7. How much bodily pain have you had during the past 4 weeks?

	(circle one)
None	1
Very mild	2
Mild	3
Moderate	4
Severe	5
Very severe	6

SHORT FORM 36 HEALTH SURVEY

MRN: [ ][ ][ ][ ][ ][ ][ ] Initials: [ ][ ][ ] Date: [ ][ ]/[ ][ ]/[ ][ ]

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

	(circle one)
Not at all	1
A little bit	2
Moderately	3
Quite a bit	4
Extremely	5

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks-

(circle one number on each line)

		All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time
a.	Did you feel full of life?	1	2	3	4	5	6
b.	Have you been a very nervous person?	1	2	3	4	5	6
c.	Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
d.	Have you felt calm and peaceful?	1	2	3	4	5	6
e.	Did you have a lot of energy?	1	2	3	4	5	6
f.	Have you felt downhearted and low?	1	2	3	4	5	6
g.	Did you feel worn out?	1	2	3	4	5	6
h.	Have you been a happy person?	1	2	3	4	5	6
i.	Did you feel tired?	1	2	3	4	5	6

SHORT FORM 36 HEALTH SURVEY

MRN: [ ][ ][ ][ ][ ][ ][ ] Initials: [ ][ ][ ] Date: [D][D]/[M][M]/[Y][Y]

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?

	(circle one)
All of the time	1
Most of the time	2
Some of the time	3
A little of the time	4
None of the time	5

11. How TRUE or FALSE is each of the following statements for you?

(circle one number on each line)

		Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
a.	I seem to get ill more easily than other people.	1	2	3	4	5
b.	I am as healthy as anybody I know.	1	2	3	4	5
c.	I expect my health to get worse.	1	2	3	4	5
d.	My health is excellent.	1	2	3	4	5

Health Assessment Questionnaire (HAQ)

Health Assessment Questionnaire (HAQ)

In this section we are interested in learning how your illness affects your ability to function in daily life. Please feel free to add any comments on the back of this page.

Please tick the response which best describes your usual abilities OVER THE PAST WEEK

	(0) Without ANY Difficulty	(1) With SOME Difficulty	(2) With MUCH Difficulty	(3) UNABLE To Do
1. DRESSING & GROOMING				
Are you able to :				
Dress yourself, including tying shoelaces and doing buttons?				
Shampoo your hair?				
2. ARISING				
Are you able to :				
Stand up from a straight chair?				
Get in & out of bed?				

3. EATING

Are you able to:

Cut your meat? _____ _____ _____ _____

Lift a full cup or
glass to your mouth? _____ _____ _____ _____

Open a new milk carton? _____ _____ _____ _____

4. WALKING

Are you able to :

Walk outdoors on flat
ground? _____ _____ _____ _____

Climb up five steps? _____ _____ _____ _____

Please tick **any AIDS OR DEVICES** that you usually use for any of these activities:

Devices used for dressing _____ Walker _____
(button hook, zipper pull, long-handled shoe horn etc) Crutches
_____

Built up or special utensils _____ Wheelchair
_____

Special or built up chair _____ Cane
_____

Other (specify) _____

Please tick any categories for which you usually need **HELP FROM ANOTHER PERSON.**

Dressing & Grooming _____ Eating _____

Arising _____ Walking _____

5. HYGIENE

Are you able to:

Wash & dry your body? _____ _____ _____ _____

Take a bath? _____ _____ _____ _____

Get on & off the toilet? _____ _____ _____ _____

6. REACH

Are you able to :

Reach & get down a
5 pound object
(such as a bag of sugar)
from just above your head? _____ _____ _____ _____

Bend down to pick up
clothing from the floor? _____ _____ _____ _____

7. GRIP

Are you able to :

Open car doors? _____ _____ _____ _____

Open jars which have
been previously opened? _____ _____ _____ _____

Turn taps on & off? _____ _____ _____ _____

8. ACTIVITIES

Are you able to :

Run errands & shop? _____ _____ _____ _____

Get in & out of a car? _____ _____ _____ _____

Do chores such as
vacuuming or yard work? _____ _____ _____ _____

Please tick any **AIDS or DEVICES** that you usually use for any of these activities:

Raised toilet seat _____ Long-handled appliances in bathroom _____

Bathtub seat _____ Long-handled appliances for reach _____

Bathtub bar _____ Jar opener (for jars previously opened) _____

Other (specify) _____

Please tick any categories for which you usually need **HELP FROM ANOTHER PERSON**

Hygiene _____ Gripping & opening things _____

Reach _____ Errands & Chores _____

Appendix 3: Revised Valuations for EQ-5D Health States

(Combined Rank and TTO)

State	Revised Estimate	State	Revised Estimate	State	Revised Estimate
11111	0.9954	21211	0.837	22322	0.431
11112	0.835	21212	0.765	22323	0.276
11113	0.573	21213	0.503	22331	0.273
11121	0.823	21221	0.753	22332	0.201
11122	0.751	21222	0.682	22333	0.046
11123	0.489	21223	0.42	23132	0.188
11212	0.803	21231	0.417	23222	0.386
11231	0.454	21232	0.345	23232	0.156
11321	0.616	21312	0.59	23313	0.246
11322	0.544	21321	0.578	23321	0.389
11331	0.386	21322	0.507	23332	0.088
12131	0.411	21323	0.352	31222	0.344
12211	0.798	21331	0.349	32211	0.423
12212	0.727	21332	0.277	32222	0.269
12221	0.715	21333	0.122	32223	0.113
12222	0.644	22112	0.721	32231	0.11
12223	0.382	22121	0.71	32232	0.039
12231	0.379	22122	0.638	32313	0.128
12321	0.54	22123	0.376	32331	0.042
13212	0.507	22211	0.761	33212	0.238
13311	0.51	22221	0.678	33232	-0.075
13332	0.125	22222	0.606	33321	0.159
21111	0.869	22223	0.344	33323	-0.068
21121	0.785	22231	0.341	33331	-0.071
21122	0.714	22232	0.269	33332	-0.143
21123	0.452	22233	0.114	33333	-0.298
21131	0.449	22311	0.586		
21133	0.222	22321	0.503		

Appendix 4: Bansback Mapping Model

Model 1 from Bansback et al.

Model 1 regressed the HAQ DI onto the EQ-5D and SF-6D.

	EQ-5D Index			SF-6D Index		
	B	SE	P	B	SE	P
HAQ Index	-0.21	0.01	<0.01	-0.10	0.01	<0.01
Constant	0.86	0.01	<0.01	0.79	0.01	<0.01

Regression Equations:

SF-6D transformation equation:	-0.10 x HAQDI + 0.79
EQ-5D transformation equation:	-0.21 x HAQDI + 0.86

Appendix 5: Regression Equations for mapping from HAQDI and DAS28 to EQ-5D and SF-6D for the Irish Cohort

Mapping from HAQDI in RA

Original EQ-5D transformation equation:	-0.236 x HAQ + 0.792
Revised EQ-5D transformation equation:	-0.168 x HAQ + 0.822
SF-6D transformation equation:	-0.084 x HAQ + 0.669

Mapping from HAQDI in PsA

Original EQ-5D transformation equation:	-0.259 x HAQ + 0.794
Revised EQ-5D transformation equation:	-0.194 x HAQ + 0.837
SF-6D transformation equation:	-0.098 x HAQ + 0.676

Mapping from DAS 28 in RA

Original EQ-5D transformation equation:	-0.084 x DAS28 + 0.887
Revised EQ-5D transformation equation:	-0.059 x DAS28 + 0.888
SF-6D transformation equation:	-0.029 x DAS28 + 0.704

Appendix 6: Systemic Literature Review for Mixed Treatment Comparison

1. Search strategy and inclusion criteria for literature review.

Literature Search strategies

All strategies were tried and the search revealing the greatest number of results was chosen. Duplicates were removed.

PubMed

Keyword search

Limits Activated: Humans, Randomized Controlled Trial, English, All Adult: 19+ years

1. Rheumatoid arthritis{tw}
2. anti-tumor necrosis factor alpha{tw} OR anti-tumour necrosis factor alpha{tw} OR anti-tnf OR TNFR-Fc{tw} OR Tumor Necrosis Factor-alpha{tw} OR etanercept{tw} OR enbrel{tw} OR infliximab{tw} OR remicade{tw} OR adalimumab{tw} OR humira{tw} OR golimumab{tw} OR simponi{tw} OR certolizumab{tw} OR certolizumab pegol{tw} OR Cimzia {tw}
3. #1 AND #2

Alternative Search Strategy - without tag

1. Rheumatoid arthritis
2. Anti-tnf OR anti-tumor necrosis factor alpha OR anti-tumour necrosis factor alpha OR TNFR-Fc OR Tumor Necrosis Factor-alpha OR etanercept OR enbrel OR infliximab OR remicade OR adalimumab OR humira OR golimumab OR simponi OR certolizumab OR certolizumab pegol OR cimzia OR anti tumour necrosis factor alpha
3. #1 AND #2

Results: 243 records

Subject Heading / MeSH Search

Limits Activated: Humans, Randomized Controlled Trial, English, All Adult: 19+ years

1. "Arthritis, Rheumatoid"[Mesh]
2. "Tumor Necrosis Factor-alpha"[Mesh] OR "TNFR-Fc fusion protein" [Supplementary Concept] OR "infliximab" [Supplementary Concept] OR "adalimumab" [Supplementary Concept] OR "golimumab" [Supplementary Concept] OR "CDP870" [Supplementary Concept]
3. #1 AND #2

Results: 238 records

EMBASE

Keyword

Limits activated: Randomized controlled trial, Human, Only in English, Adult: 18 to 64 years, Aged: 65+ years

1. Rheumatoid arthritis
2. Anti-tnf OR anti-tumor necrosis factor alpha OR anti-tumour necrosis factor alpha OR TNFR-Fc OR Tumor Necrosis Factor-alpha OR etanercept OR enbrel OR infliximab OR remicade OR adalimumab OR humira OR golimumab OR simponi OR certolizumab OR certolizumab pegol OR cimzia
3. Combine #1 and #2 using the and operator

Subject Heading / Emtree Search

Limits Activated: Randomized controlled trial, Human, Only in English, Adult: 18 to 64 years, Aged: 65+ years

1. 'rheumatoid arthritis'/exp
2. 'tumor necrosis factor inhibitor'/exp OR 'tumor necrosis factor alpha inhibitor'/exp OR 'tumor necrosis factor alpha antibody'/exp OR 'adalimumab'/exp OR 'golimumab'/exp OR 'certolizumab pegol'/exp
3. Combine #1 and #2 using the and operator

Results: 104 records

The keyword search was repeated for the Cochrane Database.
Clinical trials results were chosen and screened (note it was not possible to specify randomised clinical trial)
Results: 366 records

Inclusion Criteria

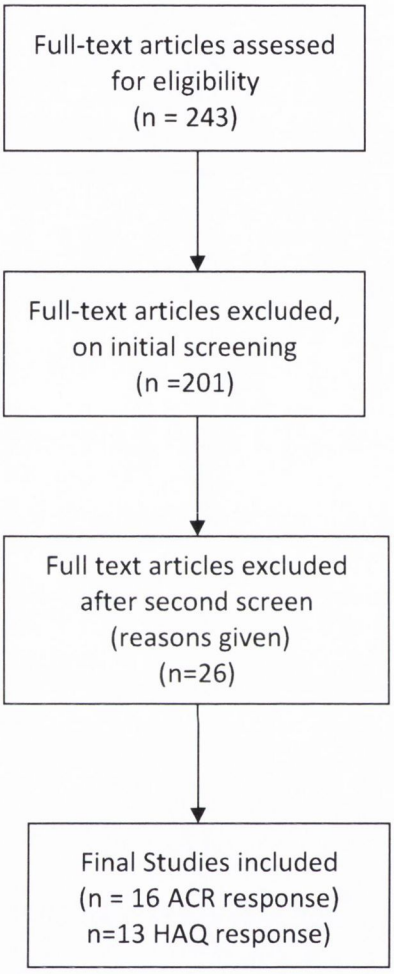
- Study design included randomised controlled trials.
- Population includes patients with established rheumatoid arthritis (in accordance with the American College of Rheumatology) (ACR) criteria (early RA was not included)
- The drugs to be assessed are: infliximab, etanercept, adalimumab, golimumab and certolizumab pegol
- Doses included were those that were used in the trials.
- Comparators: Non-biological Disease Modifying Agents (DMARDs) specifically MTX (primary comparator), against each anti-TNF individually.
- Outcomes: ACR criteria (ACR20 and 50) and quality of life outcome; Health assessment questionnaire (HAQ)

2. Reasons for exclusion of trials from second screen

Year	Drug	Author(s)	Citation/DOI	Reason for exclusion
2006	Adalimumab	Breedeveld et al.	10.1002/art.21519	MTX naïve population
2006	Adalimumab	Schiff et al.	10.1136/ard.2005.043166	Review paper on safety
2004	Adalimumab	Torrance et al.	10.1093/rheumatology/keh153	Duplicate (ARMADA trial)
2003	Adalimumab	Furst et al.	J Rheumatol 2003; 30(12): 2563-2571	Included DMARDs other than and including MTX
2006	Adalimumab	Abe et al.	J Rheumatol 2006; 33(1):37	14 week study
2000	Infliximab	Lipsky et al.	10.1056/nejm200011303432202	54 week data of ATTRACT study
2006	Infliximab	Smolen et al.	10.1002/art.21678	MTX naïve population
2005	Infliximab	Goekoop-Ruiterman et al.	10.1002/art.21405	Early RA
2004	Infliximab Etanercept	Gomez-Puerto et al.	Ann Rheum Dis 2004;63:896	Anti-TNF experienced population
2004	Infliximab Etanercept	Yazici and Erkan	Ann Rheum Dis 2004;63:607-612	Anti-TNF experienced, Observational data
2004	Infliximab	Durez et el.	10.1136/ard.2003.012914	Compared to intravenous methylprednisolone
2004	Infliximab	William St. Clair et al.	10.1002.art.20565	MTX naïve population
2009	Infliximab	Van Vollenhoven et al.	10.1016/S0140-6736(09)60944-2	Included DMARDs other than and including MTX
2008	Etanercept	Lisbona et al.	J Rheum 2008;35(3):394-7	6 week data only
2000	Etanercept	Bathon et al.	NEJM 2000;343(22):1586-1593	MTX naïve population, no baseline HAQ score
2007	Etanercept	Weisman et al.	10.1093/rheumatology/kem033	16 week safety study, efficacy data unavailable
2005	Etanercept	Chen et al.	10.1136/ard.2005.038851	ACR data not reported, Included DMARDs other than and including MTX
2004	Etanercept	Klareskog et al.	Lancet 2004;363:675-681	Included MTX naïve patients
2006	Etanercept	Van der Heijde et al.	10.1002/art.21655	2 year data of Klareskog 2006
2006	Etanercept	Klaresekog et al.	10.1136/ard.2005.038349	Open label study
2006	Etanercept	Combe et al.	10.1136/ard.2005.049650	Included DMARDs other than and including MTX
2004	Etanercept	Lan et al.	J Formos Med Assoc 2004:103(8);618-623	12 week data only
2008	Etanercept	Kekow et al.	10.1136/ard.2008.102509	Early RA population
2008	Etanercept	Emery et al.	10.1016/S0140-6736(08)61000-4	Early RA population
2009	Golimumab	Emery et al.	10.1002/art.24638	MTX naïve population
2009	Golimumab	Smolen et al.	10.1016/S0140-6736(09)60506-7	Anti-TNF experienced population

3. Selection Process

Flow Diagram for anti-TNF therapy in RA



Appendix 7: R Syntax for Analysis on all drugs

(The individual syntax on each drug has not been presented, due to length; however the method is identical)

6 week toxicity

```
> metaprop(c(16,11,97,15),c(1161,2364,2326,301))
  proportion          95%-CI %W(fixed) %W(random)
1     0.0138  [0.0079; 0.0223]    18.88     25.35
2     0.0047  [0.0023; 0.0083]    38.42     25.82
3     0.0417  [0.0339; 0.0506]    37.80     25.82
4     0.0498  [0.0282; 0.0809]     4.91     23.01

Number of trials combined: 4

                    proportion          95%-CI  z  p.value
Fixed effect model      0.0191  [0.0158; 0.0226] NA       --
Random effects model    0.0236  [0.0060; 0.0522] NA       --

Quantifying heterogeneity:
tau^2 = 0.0232; H = 5.74 [4.29; 7.67]; I^2 = 97% [94.6%; 98.3%]

Test of heterogeneity:
     Q d.f.  p.value
 98.76    3 < 0.0001
```

Forest Plot:

```
> sec1<-metaprop(c(16,11,97,15),c(1161,2364,2326,301),c("Kristensen 2006","Du
Pan 2009","Hetland 2010","Geborek 2002"))
> forest(sec1)
```

6-24 week toxicity

```
6-24 week toxicity all drugs
Error: unexpected symbol in "6-24 week"
> metaprop(c(91,120,101,20),c(854,1161,2364,301))
  proportion          95%-CI %W(fixed) %W(random)
1     0.1066  [0.0867; 0.1292]    18.25     25.25
2     0.1034  [0.0864; 0.1223]    24.81     25.62
3     0.0427  [0.0349; 0.0517]    50.49     26.16
4     0.0664  [0.0411; 0.1008]     6.45     22.96

Number of trials combined: 4

                    proportion          95%-CI  z  p.value
Fixed effect model      0.0684  [0.0613; 0.0758] NA       --
```

Random effects model 0.0781 [0.0439; 0.1211] NA --

Quantifying heterogeneity:
tau^2 = 0.0203; H = 4.63 [3.33; 6.44]; I^2 = 95.3% [91%; 97.6%]

Test of heterogeneity:
Q d.f. p.value
64.32 3 < 0.0001

Method: Inverse variance method
Freeman-Tukey double arcsine transformation used for proportions

Forest Plot:

```
> sec1<-metaprop(c(91,120,101,20),c(854,1161,2364,301),c("Zink 2005","Kristensen
2006","Du Pan 2009","Geborek 2002"))
> forest(sec1)
```

6-24 week inefficacy

```
metaprop(c(112,49,24,63,10),c(854,1161,2364,2326,301))
  proportion          95%-CI %W(fixed) %W(random)
1     0.1311 [0.1092; 0.1556]    12.20     20.02
2     0.0422 [0.0314; 0.0554]    16.57     20.19
3     0.0102 [0.0065; 0.0151]    33.73     20.43
4     0.0271 [0.0209; 0.0345]    33.19     20.43
5     0.0332 [0.0160; 0.0602]     4.31     18.93

Number of trials combined: 5

                    proportion          95%-CI  z  p.value
Fixed effect model      0.0311 [0.0272; 0.0353] NA      --
Random effects model    0.0420 [0.0148; 0.0823] NA      --

Quantifying heterogeneity:
tau^2 = 0.0359; H = 6.86 [5.47; 8.59]; I^2 = 97.9% [96.7%; 98.6%]

Test of heterogeneity:
     Q d.f. p.value
 187.97    4 < 0.0001

Method: Inverse variance method
Freeman-Tukey double arcsine transformation used for proportions
```

Forest Plot:

```
> sec1<-metaprop(c(112,49,24,63,10),c(854,1161,2364,2326,301),c("Zink
2005","Kristensen 2006","Du Pan 2009","Hetland 2009","Geborek 2002"))
> forest(sec1)
```

Total withdrawals all drugs at 6 months

```
> metaprop(c(203,170,124,63,31),c(854,1161,2364,2326,301))
  proportion          95%-CI %W(fixed) %W(random)
1     0.2377  [0.2095; 0.2677]    12.20     20.01
2     0.1464  [0.1266; 0.1681]    16.57     20.10
3     0.0525  [0.0438; 0.0622]    33.73     20.21
4     0.0271  [0.0209; 0.0345]    33.19     20.21
5     0.1030  [0.0711; 0.1430]     4.31     19.46

Number of trials combined: 5

                    proportion          95%-CI  z  p.value
Fixed effect model      0.0736  [0.0676; 0.0798] NA       --
Random effects model    0.1022  [0.0414; 0.1859] NA       --

Quantifying heterogeneity:
tau^2 = 0.0743; H = 9.81 [8.21; 11.73]; I^2 = 99% [98.5%; 99.3%]

Test of heterogeneity:
     Q d.f.  p.value
 385.21    4 < 0.0001

Method: Inverse variance method
Freeman-Tukey double arcsine transformation used for proportions
```

Forest plot:

```
> sec1<-metaprop(c(203,170,124,63,31),c(854,1161,2364,2326,301),c("Zink
2005","Kristensen 2006","Du Pan 2009","Hetland 2010","Geborek 2002"))
> forest(sec1)
```

Appendix 8: Costs for RA Model

Item costed			Unit cost (source price year)	Unit Cost (IRL € 2010)	Price year	Source
Healthcare Visits						
Day case (infusion suite 8hrs)			1077.00	1164.00	**2011**	HSE Casemix
Day case			587.00	661.76	2005	HSE Casemix
In-patient stay			4637.00	5227.57	2005	HSE Casemix
O/P visit			150.00	168.60	2010	HSE Casemix
A&E visit			227.00	255.15	2010	HSE Casemix
Gp Visit			25.10	25.10	2010	GP surgery fee in Ireland *
Outpatient Visit				168.60		HSE Casemix
Cost per bed day			689.00	776.75		HSE Casemix
Tests						
FBC			9.26	10.00	2011	SVUH Laboratory
FBC+ESR			17.34	18.73	2011	SVUH Laboratory
ESR			8.32	8.99	2011	SVUH Laboratory
Urea and electrolytes			7.85	8.48	2011	SVUH Laboratory
LFT			12.20	13.18	2011	SVUH Laboratory
CRP			12.74	13.76	2011	SVUH Laboratory
Urinalysis			0.51	0.55	2011	Based on Multistix SG**
Hep B			26.00	28.09	2011	National Virus Reference Lab (UCD)
TB Screen				22.21	2008	Test cost, 6 mins nurse time+ 15min dr. time) Inflated to 2009
Imaging/ Diagnostics						
CT C Spine/ S spine/ L spine/knee				268.00	2008	SJH
MRI C Spine/ S spine/ L spine/knee				268.00	2008	SJH

CXR				112.40	2010	SJH
Xray limbs, hips, shoulder				146.12	2008	SJH
US (hands, ankles, shoulder, wrists)				223.00	2008	SJH
Biopsy				123.00	2007	VHI
Arthroscopy (Day case)				1194.00	2007	DRG 124 Casemix 2009
Medication Costs	**Pack Size**	**Pack Cost**		**Unit Cost**		
Adalimumab pen/syringe	2	1125.18		562.59	2011	HSE reimbursement price April 2011
Etanercept 50mg pen/syringe	4	1031.71		257.93	2011	HSE reimbursement price April 2011
Golimumab 50mg syringe	1	1105.50		1105.50	2011	HSE reimbursement price April 2011
Certolizumab 200mg	2	1031.80		515.90	2011	HSE reimbursement price April 2011
Infliximab 100mg vial	1	690.43		2071.29	2010	MIMs Nov 2009
Rituximab 500mg vial	1	1632.83		1632.83	2011	Cliniscipt 2011
Abatacept 250mg vial	1	359.61		359.61	2010	HSE Price Realignment Nov 2010
Rituximab 500mg vial	1	1632.79		1632.79	2009	MIMs Nov 2009
MTX tablets 2.5mg (Lederle)	28	3.83		0.82	2010	HSE reimbursement price (PCRS Nov 2010)
MTX syringe (Metoject) 10mg/ml	1	18.84		18.84	2011	HSE reimbursement price (PCRS Nov 2010)
MTX syringe (Metoject) 10/1.5mlmg	1	23.12		23.12	2011	HSE reimbursement price (PCRS Nov 2010)
MTX syringe (metoject) 10/2ml	1	25.11		25.11	2011	HSE reimbursement price (PCRS Nov 2010)
Methothrexate Subcut Average	1	22.36		22.36	2011	Average of three doses
Leflunomide 10-20mg tablets	30	67.69		2.26	2011	HSE reimbursement price (PCRS Nov 2010)
Azathioprine 50mg	100	29.20		0.88	2011	HSE reimbursement price (PCRS Nov

						2010)
Hydroxychloroquine 200mg	60	19.80		0.33	2011	HSE reimbursement price (PCRS Nov 2010)
Salazopyrin EN	112	12.28		0.11	2011	HSE reimbursement price (PCRS Nov 2010)
Cyclosporin Caps 100mg (Neoral)	30	89.79		2.99	2011	PCRS April 2011
Gold IM injection (Myocrisin)	10	46.06		4.61	2009	MIMs Nov 2009
Prednisolone 5mg EC (Deltacortril)	100	10.73		0.11	2011	HSE reimbursement price (PCRS Nov 2010)
Triamcinolone (Adcortyl) 10mg/ml	5	6.52		1.30	2011	HSE reimbursement price (PCRS Nov 2010)
Methylprednisolone 500mg Injection (Solu-medrone)	1	10.38		10.38	2011	MIMs April 2011
Diclofenac (Difene) 75mg	56	14.55		0.26	2011	HSE reimbursement price (PCRS Nov 2010)
Other Drug Related Costs				**Unit Cost (or Hourly in case of infusion cost)**		
Compounding		(excl VAT2 1%)		40.00	2008	SVUH 2009(cost of rituximab compunding in pharmacy)
Patient care fee - High Tech Drug Scheme		per month		62.03	2010	2008
Dispensing Fee (Average of 3 fees from sliding scale)				4.30	2010	HSE Pharmacy Fees
Cost of Infusion				134.63	2010	SJH Chemotherapy Infusion Suite Daily Rate €1077/ 8hrs
Palliative Care Costs						
Palliative Care				**Yearly**		

Diclofenac 75mg bd				73.22		HSE reimbursement price (PCRS Nov 2010)
Prednisolone 5mg od				45.38		HSE reimbursement price (PCRS Nov 2010)
1 week Rheumatology Rehabilitation (1 x year)				2518.00		Cost of Care in Public Nursing Homes
Total Palliative Care Cost				**2636.60**		
Staff Costs						

Staff costs (www.dohc	Annual salary (midpoint of scale)	Budget 2009 deductions (5% on first €30,000;7.5% on next €40,000;10% on next €55,000	Salary following deductions	70:30 pay:non-pay rule	Salary per \)	Hours per week (SJH personnel	Salary per h	Source
Senior staff nurse	48,870	2,915	45,955	65650	1258	37.50	€ 33.55	Dohc
Staff Nurse	39,431.81	2,207.39	37,224	53178	1019	37.5	€27.18	DoHC
Public Health Nurse	52,989.66	3,224.22	49,765	71093	1362	37.5	€36.33	
Senior Pharmacist	69,200	4,440	64,760	92514	1773	35.00	€ 50.66	Dohc
Category 1 Consultant	168,844	14,384	154,459	220656	4229	33	€ 128.14	DoHC (avera

*28% of patients are medical card and private patients will be out of pocket (€55-€60) and therefore not a cost to the HSE
**Urinalysis is based on cost of Multistix 10 SG (100 pack) €51.52 (inc. 21% VAT) (MIMS Dec 2009) 1 test = €0.51
*** Nursing Home Support Scheme, Our Lady's Hospice, Harolds Cross Effect from July 2010 to July 2011 €2518 per week
SJH St James Hospital, SVUH St. Vincents University Hospital, PCRS Primary care Reimbursement Service, HSE Health Service Executive

Appendix 9: Comparison of PSAs for BRAM

Table (i): Mean and Standard Deviation of the NMB of each anti-TNF compared to DMARD from the results of five PSA

PSA Run	Fixed Seed	Etanercept NMB	Infliximab NMB	Adalimumab NMB	Golimumab NMB	Certolizumab NMB
1	0.1	-86259.27	-83624.87	-94569.51	-79575.61	-78459.16
2	0.2	-86502.66	-83536.90	-95060.96	-79605.54	-78377.98
3	0.3	-86659.38	-83651.79	-94319.96	-79646.08	-78492.21
4	0.4	-86396.56	-83752.82	-95045.27	-79577.99	-78445.72
5	0.5	-85988.21	-83561.20	-94888.37	-79629.41	-78344.74
Std Dev		288.75	53.22	334.80	30.11	68.03

Appendix 10: Collaborative Work

The technical work in chapter 5 has been carried out as stated throughout the chapter with Susanne Schmitz who is a PhD student in statistics. Her PhD focuses on the methods of evidence synthesis and in particular mixed treatment comparisons. The work required for this thesis required new models to be developed (in particular for the HAQ multiplier) and required the use of WinBUGs. This thesis examines the cost-effectiveness of anti-TNF drugs in RA, which requires decision to be made in relation to the cost-effectiveness model and compiling the inputs for this model. The inputs for clinical efficacy were best calculated by combining data from many different sources. The most appropriate method to use was mixed treatment comparison which uses a Bayesian methodology. While an understanding of the methodology is known by the candidate, the computation required in WinBUGs was seen as outside the expertise required for this thesis. The collaborative work is highlighted in the relevant chapter (Ch. 5).

- The background work, systematic review, data extraction, author contact, and calculation of weighted baseline estimates. 100% Roisin Adams

- Decisions surrounding the assumptions for the model, in particular for the methotrexate effect and inclusion/ exclusion criteria 80% Roisin Adams, 20% Susanne Schmitz.

- Development of MTC model for HAQ multiplier: 100% Susanne Schmitz

- Adaptation of the Nixon model from OR to RR: 100% Susanne Schmitz

- The computation of relative effects in WinBUGS: 100% Susanne Schmitz

- Writing of the paper published in Annals of Rheumatic Disease: Roisin Adams 30% Susanne Schmitz 70%

Survival Analysis

- The background work, systematic review, data extraction 100% Roisin Adams –for completion these were verified by Susanne Schmitz as the data was extracted from graphs.
- The computation of Weibull distribution and fitting of curves in WinBUGS 100% Susanne Schmitz

Appendix 11: Results of One Way Sensivity Analysis

Table 42. PSA Results of Linear utility mapping using original EQ-5D scoring (Irish cohort)

Comparison	ICER	95% Credible Interval	
		Lower	Upper
ETN - Base	78200	65400	98300
CZP - Base	88500	80300	105000
GOL- Base	97600	87000	107000
ADA - Base	110700	90600	138100
IFX - Base	209500	155600	326900

Table 43. Parameter Set 3 Linear Utility mapping using SF-6D measure (Irish cohort)

Comparison	ICER	95% Credible Interval	
		Lower	Upper
ETN - Base	181700	141600	256300
CZP - Base	205000	153900	313300
GOL- Base	225000	163200	363200
ADA - Base	249600	184300	369600
IFX - Base	456200	275000	1369000

Table 44. Parameter Set 4 Linear Utility mapping using original EQ-5D scoring (UK cohort)

Comparison	ICER	95% Credible Interval	
		Lower	Upper
ETN - Base	58300	47600	75500
CZP - Base	66700	62100	73900
GOL- Base	73000	68400	79000
ADA - Base	82900	66800	105700
IFX - Base	158000	116500	244700

Table 45. Parameter Set 5 Quadratic mapping for the SF-6D

Comparison	ICER	95% Credible Interval	
		Lower	Upper
ETN - Base	133200	92600	238300
CZP - Base	151000	107300	245000
GOL- Base	170000	117600	307700
ADA - Base	202700	141200	373800
IFX - Base	411800	254900	1144300

Table 46. Parameter Set 6 Same time on all anti-TNF treatment

Comparison	ICER	95% Credible Interval	
		Lower	Upper
ETN - Base	117300	92600	157000
CZP - Base	144300	133000	163200
GOL- Base	163000	152300	190000
ADA - Base	172600	132300	254500
IFX - Base	347100	229200	707600

Table 47. Parameter Set 7 HAQ Change on anti-TNF treatment

Comparison	ICER	95% Credible Interval	
		Lower	Upper
ETN - Base	405500	192200	-3254500
CZP - Base	484000	231500	4970000
GOL- Base	620000	265000	-2570000
ADA - Base	869400	338600	-1265100
IFX - Base	5630000	597000	-730800

Note: The negative ICER values for the upper CI are higher than the positive values. These values are dominated.

Table 48. Parameter Set 8 Drug price reduction 50%

Comparison	ICER	95% Credible Interval	
		Lower	Upper
ADA - Base	71200	70300	72100
ETN - Base	47400	46900	47900
IFX - Base	170000	165000	175000
GOL- Base	59700	58800	60600
CZP - Base	60000	59100	60900

Table 49. Parameter Set 9 Offset costs included

Comparison	ICER	95% Credible Interval	
		Lower	Upper
ETN - Base	96500	60100	273000
CZP - Base	110000	70000	310000
GOL- Base	121500	72400	299000
ADA - Base	138200	87600	380800
IFX - Base	264200	159600	1076800

Table 50. Parameter Set 10 Utility estimates set to zero

Comparison	ICER	95% Credible Interval	
		Lower	Upper
ETN - Base	102700	85300	131200
CZP - Base	115900	101300	139700
ADA - Base	143700	115400	185100
GOL- Base	151800	111300	138900
IFX - Base	268300	187200	480700

Publications

Peer Reviewed Publications

Adams R, Craig BM, Walsh CD, Veale DJ, Bresnihan B, FitzGerald O, Barry M. The Impact of a Revised EQ-5D Population Scoring on Preference-Based Utility Scores in an Inflammatory Arthritis Cohort. *Value in Health*. 2011;14(6):921-927

Adams R, Walsh C, Veale D, Bresnihan B, FitzGerald O, Barry M. Understanding the Relationship between EQ-5D, SF-6D, HAQ and Disease Activity in Inflammatory Arthritis. *Pharmacoeconomics* 2010;28(6):477-487.

Schmitz S, Adams R, Walsh C, Barry M, FitzGerald O. A Mixed Treatment Comparison of the efficacy of Anti-TNF agents in rheumatoid arthritis for MTX non-responders demonstrates differences between treatments: A Bayesian approach. *Annals of the Rheumatic Diseases* 2011. DOI: 1136/annrheumdis-2011-200228. Available Online first 29 September 2011.

Partially Peer Reviewed Publications

Adams R, Schmitz S, FitzGerald O, Walsh C, Barry M**.** Does Model Choice in Mixed Treatment Comparisons Affect the Decision in Economic Evaluations? *ISPOR Connections* 2011. Vol.17. No.2.

Book Chapter

Adams R. Disease Outcomes in Rheumatoid Arthritis in Scientific Basis of Healthcare: Arthritis. V Preedy (Ed) Science Publishers. 2011 (ISBN 978-1-57808-730-3) In press (March 2012).

Peer Reviewed Conference Proceedings

Oral Presentations

Adams R, Ng CT, Gibbs A, Bresnihan B, Veale D, Tilson L, FitzGerald O, Barry M. Investigating the benefit of anti-TNF therapy on physical and mental functioning via the SF-36 questionnaire in patients with rheumatoid arthritis. *International Society for Pharmacoeconomic Outcomes and Research Annual European Conference*. October 2008, Athens, Greece. (Awarded Best new investigator Podium Presentation).

Adams R, Craig B, Walsh C, Veale D, Bresnihan B, FitzGerald O, Barry M.
A revised scoring system for the EQ-5D produces a change in preference-based utility score, which is closer to disease measures in inflammatory arthritis. *Health Technology Assessment International Conference* June 7-9, 2010, Dublin, Ireland.

Adams R, Schmitz S, Walsh C, FitzGerald O, Barry M.
Does model choice in Mixed Treatment Comparisons affect the outcome for decision makers? *International Society for Pharmacoeconomic Outcomes and Research Annual European Conference,* November 5-7, Prague, Czech Republic.
Schmitz S, Adams R, Walsh C, Barry M. The Use of Continuous Data versus Categorical Data in MTC: the Case of the HAQ Multiplier in Rheumatoid Arthritis. Oral Presentation. *International Society for Pharmacoeconomic Outcomes and Research Annual European Conference*, November 2011, Madrid, Spain.

Poster presentations

Adams R, Walsh C, Schmitz S, Barton P, Barry M. A Cost Utility Analysis of anti-TNF agents for the treatment of Rheumatoid Arthritis. Poster Presentation. *International Society for Pharmacoeconomic Outcomes and Research Annual European Conference*, 6-8 November 2011, Madrid, Spain.

Schmitz S, Adams R, Walsh C, Barry M, FitzGerald O. A Bayesian Mixed Treatment Comparison Demonstrates Differences between Anti-Tumour Necrosis Factor Agents

in Rheumatoid Arthritis. Poster Presentation. *American College of Rheumatology Annual Meeting,* 5-9 November 2011, Chicago, USA.

Adams R, Walsh C, Tilson L, FitzGerald O, Bresnihan B, Veale D, Barry M. A calibration of the relationship between the HAQ, SF-6D, EQ-5D and disease activity in inflammatory arthritis. Poster Presentation. *International Society for Pharmacoeconomic Outcomes and Research Annual European Conference*, October 2009, Paris, France.

Adams R, Mumtaz A, Ng CT, Pontiflex E, Lynch BM, Saber T, Veale D, Bresnihan B, Barry M, FitzGerald O. Improvement in quality of life is demonstrated by improvement in Bath disease activity scores. Poster Presentation. *European League against Rheumatism,* June 2009, Paris, France

Adams R, Turner CP, O'Leary A, Bresnihan B, Veale D, Barry M, Fitzgerald O. Deriving a single index utility from HAQ scores in patients with inflammatory arthritis in an Irish cohort. Poster Presentation. *Irish Society of Rheumatology Annual Meeting*, Sept 2009, Belfast, N. Ireland.

Adams R, Ng CT, Lynch BM, Saber T, Pontiflex E, Molloy M, Grier A, Veale D, Bresnihan B, Barry M, FitzGerald O. QoL scores reflect disease activity both pre and post initiation of biologic therapy in patients with rheumatoid and psoriatic arthritis. Poster Presentation. *British Society of Rheumatology Annual Conference*, April 2009, Glasgow, UK.

Adams R, Craig B, Walsh C, Veale D, Bresnihan B, FitzGerald O, Barry M. Applying a revised scoring system for the EQ-5D in an inflammatory arthritis cohort. *European League against Rheumatism*, June 2010, Rome, Italy.