Monograph:
Brennan, A., Bansback, N. and Nixon, R. (2006) Modelling the cost effectiveness of TNF-α inhibitors in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. Discussion Paper. (Unpublished)

HEDS Discussion Paper 06/12

HEDS Discussion Paper 06/12

ScHARR

SCHOOL OF HEALTH AND

RELATED RESEARCH

The University of Sheffield

ScHARR
School of Health and Related Research

Health Economics and Decision Science

Discussion Paper Series

MODELLING THE COST EFFECTIVENESS OF TNF-α INHIBITORS IN THE MANAGEMENT OF RHEUMATOID ARTHRITIS: RESULTS FROM THE BRITISH SOCIETY FOR RHEUMATOLOGY BIOLOGICS REGISTRY

An Independent Report

Alan Brennan (1,2), Nick Bansback (1,2,3), Richard Nixon (2,4)

1. Health Economics and Decision Sciences, ScHARR, University of Sheffield
2. ARC Epidemiology Unit, University of Manchester (honorary positions)
3. Centre for Health Evaluation and Outcome Sciences, University of British Columbia
4. Medical Research Council Biostatistics Unit, Cambridge

TABLE OF CONTENTS

TABLES AND FIGURES

SUMMARY

In 2001, the National Institute for Clinical Excellence (NICE) recommended TNF-• inhibitors for the treatment of patients with active RA unresponsive to conventional DMARDs. The British Society for Rheumatology Biologics Registry (BSRBR) was established in October 2001 and now has 3 years follow-up, and over 7000 patients.

Our study performs an analysis using a decision analytic model populated by the BSRBR data to evaluate the cost effectiveness of TNF inhibitors over conventional DMARD therapy. In particular, we consider what is the incremental cost per QALY of TNF inhibitors versus traditional DMARDs according to current practice in the UK? What if guidance that only patients achieving moderate or good EULAR Disease Activity Score (DAS28) response are allowed to continue TNF inhibitor therapy after 3 months? Unfortunately, similar analyses based on the ACR20 or ACR50 measures are not possible because the BSRBR does not record the data necessary to produce these composite measures of relative improvement. The analysis examines subgroups based on age, sex, disease duration, number of previous DMARDs, and baseline HAQ disability score. Sensitivity analyses consider alternative assumptions concerning interpretation of the evidence base, including those on HAQ disability progression on traditional DMARD, relationships between HAQ and utility, impact of delayed progression whilst on TNF inhibitors, use of sequential TNF inhibitors and discounting rates.

There are several caveats and limitations detailed in the methods section of this report and re-iterated in the conclusions. In particular, any mortality reduction benefits, which might be attributable to TNF inhibitors, are excluded.

The results are analysed using a number of scenarios. The scenario on current UK practice gives a cost effectiveness estimate of around £24,000 per QALY. If the guidelines set out by NICE in their initial appraisal were strictly adhered to, and non-responders were withdrawn from therapy, this is estimated to reduce to around £22,000. These numbers are within the region that NICE deemed cost effective in the previous appraisal e.g. 'the incremental cost-effectiveness ratio of these therapies (etanercept and infliximab) can be estimated to be in the region of £27,000 to £35,000 per QALY'.

Sensitivity analysis results show several important factors. The assessment of cost effectiveness in the 2001 appraisal was made using rates of discounting set at 6% cost, 1½ QALY. These remain in place for the 2005 appraisal but recent recommendations suggest moving to 3½ % cost, 3½ QALY). If the suggested new discount rates were used cost effectiveness would be estimated at £31,000-32,000 per QALY. Assumptions concerning long-term disease progression on traditional DMARDs (i.e. the control arm) also make a substantial difference to cost-effectiveness. The basecase analysis (Scott et al. data for the average UK progression) may be an under-estimate, and it is clear that treatment of patients with higher rates of progression is more cost-effective. Under our assumptions, sequential therapy with 2 TNF inhibitors appears to have the same order of cost-effectiveness as single therapy.

1. INTRODUCTION

1.1. BACKGROUND

Rheumatoid arthritis (RA) is a chronic, progressive, inflammatory disease that affects approximately 0.8% of the adult population.[Symmons, 2002] RA affects the physical functioning of patients, their psychological and social health, and eventually progresses to substantial disability through the loss of mobility, increased co-morbidity and premature mortality.[Yelin, 1995; Pincus, 1993; Wolfe, 2003; Wong, 2001]. The economic burden of RA to society is substantial and approximates to that of treating coronary heart disease [Callahan, 1998] In 1996, the total economic impact of RA in England was estimated to be £1.256 billion, over half of which was accounted for by loss of earnings while in-patient and long-term institutional care accounted for over 50% of the direct medical cost [McIntosh, 1996].

The addition of new disease-modifying anti-rheumatic drugs (DMARDs) such as the tumour necrosis factor (TNF)-α antagonists has transformed the management of RA. A number of clinical trials have demonstrated the efficacy of anti- TNF- α agents in improving function, significantly inhibiting joint destruction and reducing toxicity in comparison to conventional DMARDs. Costing between £8000 and £15000 per patient year, a recent study in the US demonstrated that the introduction of these new treatments increases the total annual direct cost of a biologic treated patient threefold. [Michaud, 2004]

The additional costs make these agents natural candidates for cost effectiveness analyses (CEA). CEA compares the incremental costs of an intervention over conventional management with its incremental health benefit.[Drummond, 1997] Cost utility ratios are the most popular CEA since health benefit is measured in quality adjusted life years (QALYs) allowing comparisons across other diseases on the most efficient strategy for resource allocation. A number of such analyses have already been developed giving a variety of estimates ranging from cost effective to not cost effective.[Kobelt, 2003; Wong, 2001; Brennan, 2004; Jobanputra, 2002; Kobelt, 2004; Bansback, 2004] Wolfe *et al*, question the validity of most estimates, since estimates of benefit have been derived directly from randomised controlled trials, which they argue are not representative of real clinical practice.[Wolfe, 2004] Kobelt et al used data from a Swedish registry, which does not have a control arm so does not allow a valid incremental estimate of cost utility to be made.[Kobelt, 2004] Therefore no existing evaluation could be considered to be robust.[Bansback, 2005]

In 2001, the National Institute for Clinical Excellence (NICE) recommended TNF- α inhibitors for the treatment of patients with active RA unresponsive to conventional DMARDs based on their assessment of the agents encouraging cost effectiveness.[NICE, 2002] At that time, NICE mandated that all patients with RA exposed to TNF-α inhibitors require follow-up to assess the long-term safety and effectiveness of these drugs. The British Society for Rheumatology Biologics Registry (BSRBR) was

established in October 2001 for this purpose.[http://www.arc.man.ac.uk /webbiologicsreg.htm] With up to 3 years of follow-up data, and over 7000 biologic treated patients recruited by the BSRBR, new questions over the cost effectiveness of these agents are emerging. We performed a cost-utility analysis using a decision analytic model populated by the BSRBR data to evaluate the cost effectiveness of TNF-α inhibitors over conventional DMARD therapy.

1.2. RESEARCH QUESTIONS

The specific research questions focussed on:

1. **What is the incremental cost utility (cost per QALY) of TNF inhibitors according to current practice in the UK versus use of traditional DMARDs only?**
   After the patient has come off the initial TNF antagonist it is assumed that they would switch back to traditional DMARDs. This is the question that was initially reviewed by NICE, and we have focussed this work on re-evaluating whether this was the correct decision.

2. **What would be the cost utility if guidance that only patients achieving moderate or good EULAR Disease Activity Score (DAS28) response are allowed to continue TNF inhibitor therapy after 3 months?**
   An important issue in the treatment of patients with TNF antagonists is the criterion for continuing a patient on treatment. NICE guidance follows the old BSR guidance which stated that patients must at least be a moderate EULAR responder (based on DAS28) at 3 months to continue therapy.[NICE, 2002] The BSRBR shows that this has not necessarily been adhered to. We have explored the impact of using this decision rule in the model.

3. **We also explore the use of a 2nd TNF antagonist in a sequence after the first has failed, based on a small amount of data**

4. **What is the cost utility for subgroups based on;**
   o Age
   o Sex
   o Disease duration
   o Number of previous DMARDs
   o Baseline HAQ disability score.

5. **What is the cost utility if we made alternative assumptions concerning interpretation of the evidence base, including those on**
   o HAQ disability progression on traditional DMARD.
   o Relationships between HAQ and utility
   o Impact of delayed progression whilst on TNF inhibitors
   o Use of sequential TNF inhibitors
   o Discounting rates

1.3. FUNDING FOR THE STUDY

This study was originally proposed by the authors for a research funding bid to the Arthritis and Rheumatism Campaign. The bid was unsuccessful.

The timely requirement for an analysis of the BSRBR for the National Institute for Clinical Excellence (NICE), meant that the project team began work without funding. A small grant of £10,000 from the British Society of Rheumatologists has subsequently met a proportion of the costs.

1.4. CONTRIBUTORS

o Richard Nixon and Nick Bansback performed the statistical modelling of the BSRBR.
o Nick Bansback, Alan Brennan, Richard Nixon and Jason Madan have worked on the cost effectiveness model.
o Deborah Symmons gave clinical advice.
o Mark Harrison aided the statistical tests.
o All authors contributed to the writing of the report.

1.5. CONFLICTS OF INTEREST

The authors have no current specific conflicts of interest. Previous research funding from industry and government is described in detail at Appendix 2 (section 6.2)

2. MATERIAL AND METHODS

2.1. OVERVIEW OF APPROACH

2.1.1. Modelling

We developed a form of micro simulation known as an individual sampling model to describe the natural history of rheumatoid arthritis.

We express the results in terms of quality adjusted life years (QALYs), drug related costs, hospitalisation costs and cost-effectiveness ratios. A stochastic analysis was performed to capture the parameter uncertainty. Results of this analysis are presented in the form of acceptability curves and net benefit distributions.

The simulation tracks patients' health from time of entry to the model until death in 6 monthly cycles. Patients' health is characterised in terms of their health state utility, which treatment they are on, and whether they remain alive. The model simulates a hypothetical patient that follows the course based on the experience of an average cohort. In contrast to a Markov approach, at each decision node a random number decides the route a patient takes based on calculated probability. Therefore each patient represents only one possible route that can be taken. The patients are replicated only 1000 times with different randomly sampled numbers, by which time enough routes have been taken to give the model precision.

2.1.2. Analysis based on TNF inhibitors as a class

This analysis has focussed on patients treated with etanercept (Enbrel® - Wyeth), infliximab (Remicade® - Schering-Plough) and adalimumab (Humira® - Abbott).
Analyses have been performed on TNF inhibitors as a class. We have not looked at the differences in the cost effectiveness between TNF antagonists due to more complex selection bias which would not be sufficiently accounted for in the general case mix adjustment approach [Deeks, 2003] The BSRBR will be able to answer these questions in the future but further methodological work will be required.

2.1.3. Data sources

Overview of BSRSR data

The model uses data from the British Society for Rheumatology Biologics Registry (BSRBR) as its primary source of evidence.

This data is considered to be a valuable source for such an analysis because:

- o This is one of the largest sources of data on the health outcomes of patients using TNF antagonists.
- o The data are well collected.
- o Patients are followed up, and collection is good.
- o A number of economic endpoints are collected.
- o The registry has a control arm with which to make comparisons to traditional DMARDs.

However the challenges in using this registry are:

(i) The registry is not randomised so a number of biases, in particular selection bias might be seen between treatment groups. For instance the availability of anti-TNF α was limited to patients that had failed 3 DMARDs. Therefore a number of patients recruited onto the control arm have yet to fail the 3 DMARDs necessary to attempt anti-TNF α. To control for this problem, we use case-mix adjustments using the treatments given as a dummy variable.

(ii) The timing of measurements may not correspond to clinical events. Measurement of health state utility is made at baseline and then by postal questionnaire at 6 monthly intervals. We are therefore able to derive the improvement in health utility for the first treatment in its first 6 months (although we do not know how it varies through this time). However, for subsequent treatments the postal questionnaire may not coincide with visits to the Rheumatologist or other events when treatment-switching decisions are made. Thus, we do not have a measurement of health utility at the time when the next treatment is attempted so it is difficult to access the exact magnitude of efficacy.

(iii) The BSRBR is still relatively new with data on a substantial number of patients up only to 3 years. For making long term estimates, we need to make assumptions as to whether the 3 year data adequately predict long term effectiveness. Lack of data is particular pertinent in the control arm where recruitment has been slow and is limited mostly to 6 months. We therefore use external data sources for some parameters on the control arm.

(iv) While we try to use a societal perspective, the BSRBR does not incorporate all components of cost and benefit necessary for a full societal approach to be made. Modelling can be used to synthesise such data sources from external sources, but after reviewing the literature reliable estimates are not reported. We have therefore focussed on using the data from the BSRBR.

Other data sources

Where the BSRBR was not able to provide evidence, we used sources from the literature that have been previously reviewed and critiqued. [Barton, 2004]

2.1.4. Subgroups and Covariates Analysis

At the start of the model the baseline characteristics, age, sex, disease duration, number of previous DMARDs used, HAQ -DI, and health state utility is sampled using non parametric bootstrapping from the average characteristics of a sample of distinct patient groups.

These patient groups are based on clinically meaningful characteristics:

- o HAQ [0-0.5, 0.5-1.0, 1.0-1.5, 1.5-2.0, 2.0-2.5, 2.5-3.0]
- o Age [<40, 40-50, 50-60, 60-70, 70+ years]
- o Disease duration [0-5,5-10,10-15,15+ years]
- o Number of previous DMARDs [<2,<3, <4, <5,5+]
- o Gender [Male, Female]
- o And whether the patient is on concomitant DMARD.

These are then used as the baseline characteristics for both intervention and control groups. This is the appropriate methodology given the non linear nature of the analyses and incorporates the correlation between these variables. For example, if it is selected to analyse patients of higher age, then it would be more likely that the disease duration for this patient group would also be higher.

We have used a fixed number of covariates in each of the statistical tests. We did not select parameters based on their statistical significance as we were not interested in why co variables were related to the independent variable, or which was most significant, but rather use all the data we had to predict most accurately. The use of holding all co variables in all analyses is two-fold. Firstly, the BSRBR is an un-randomised registry so making comparisons between treatment groups can be confounded by selection bias. We selected all variables which we deemed would be important prognostic factors for the different outcomes and adjusted for each for these.[Deeks, 2003] Secondly, since we have only ~3 years of data, we want to predict the long time effects. Using time dependent covariates such as disease duration, age, and number of DMARDs, allows for this to become a time dependent model.

2.1.5. Measurement of quality of life

This analysis has focussed on producing a cost utility analysis where the benefits of an intervention are measured using QALYs. Previous models have solely looked at the Health Assessment Questionnaire Disability Index (HAQ -DI), a measurement of disability, and mapped the results directly to utility using a simple linear relationship.[Fries, 1982] A criticism of this approach is that only the impact of the treatment on functional disability is captured, and not the psychological or pain elements associated with the disease. For this analysis, we have a direct measure of health utility since the BSRBR incorporates the SF36, a generic measure of patients' health related quality of life.[Ware, 1993] The SF36 can be translated to a preference based health utility via the SF6D.[Brazier, 2002] The SF6D incorporates domains of physical

functioning, role limitation, social functioning, pain, mental health and energy and vitality. The SF6D has been shown to be a responsive measure in diseases of mild to moderate severity, but since it is based on a measure of general health, it struggles to distinguish between states of severe health.[Brazier, 2004]. This so called 'floor effect' is caused because the levels of ill-health described in the SF36 do not necessarily discriminate between the more severe states. The result is that for increasing severity of HAQ disability, the SF6D will not value below 0.4. (Figure 1)

Figure 1: Scatter plot of HAQ -DI versus SF6D using BSRBR data.
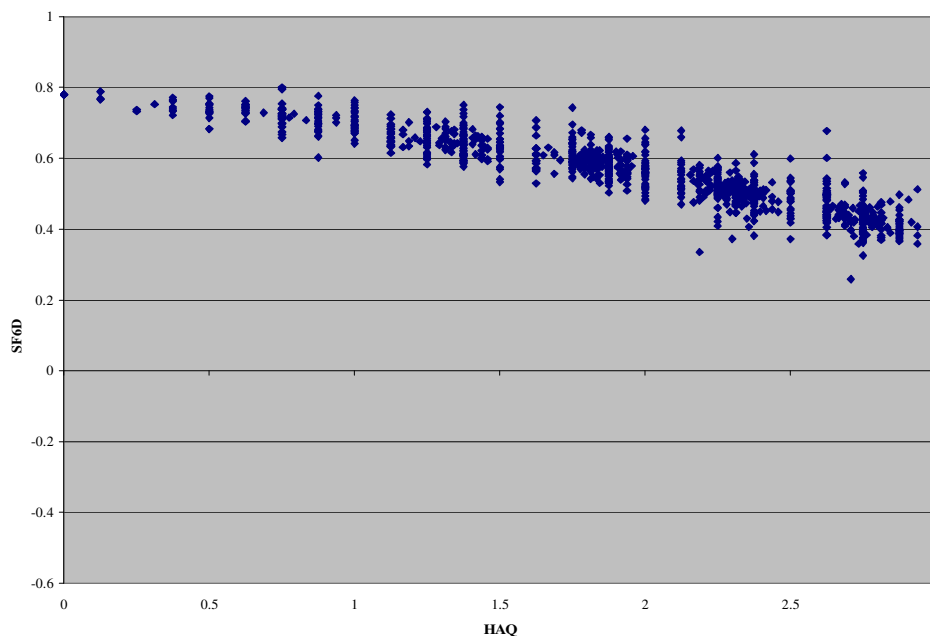
We have therefore attempted to provide an alternative measure of health utility, the EQ5D.[Brooks, 1996; Dolan, 1997] The EQ5D is a popular measure given its ease of application. It measures five domains, mobility, self care, usual activities, pain, anxiety. Since EQ5D was not measured directly in the registry, we used a mapping of the HAQ disability questionnaire to the EQ5D.[Bansback, in submission] This mapping imputes the EQ5D from all the 42 components of the HAQ questionnaire so is more sophisticated than the simple linear relationships previously used. Since this is not a direct measurement of EQ5D, it is inferior to the SF6D in terms of its accuracy of measurement. However, it has been shown that the EQ5D is better at distinguishing between states of severe health in RA [Marra, 2004]

Figure 2shows that it captures a greater range of values, particularly in severe states while Figure 3 shows how the two measures relate.

Figure 2: Scatter plot of HAQ -DI versus imputed EQ5D
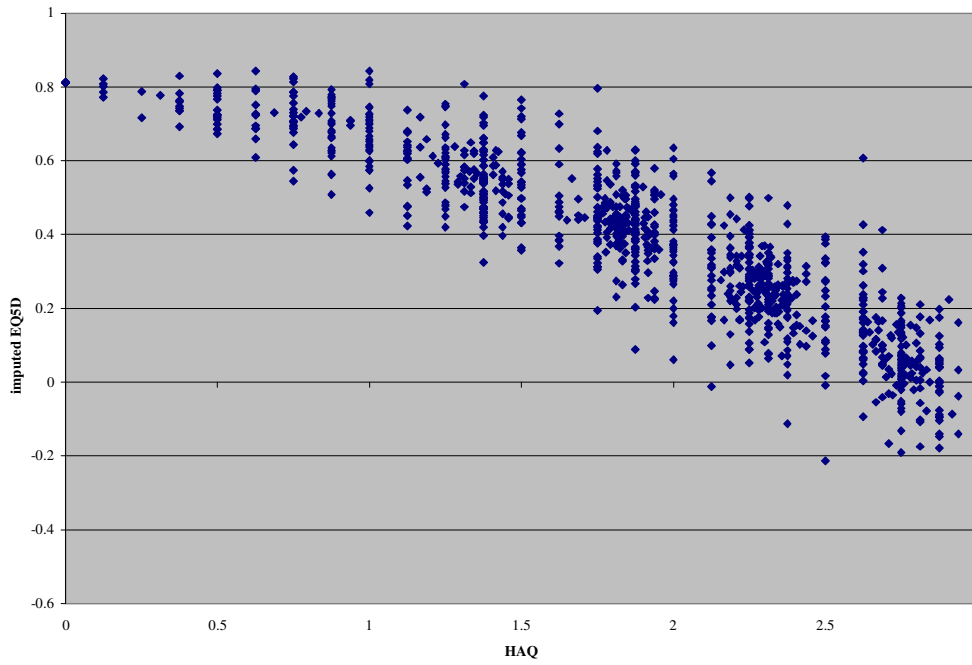
Figure 3: Scatter plot of SF6D versus imputed EQ5D
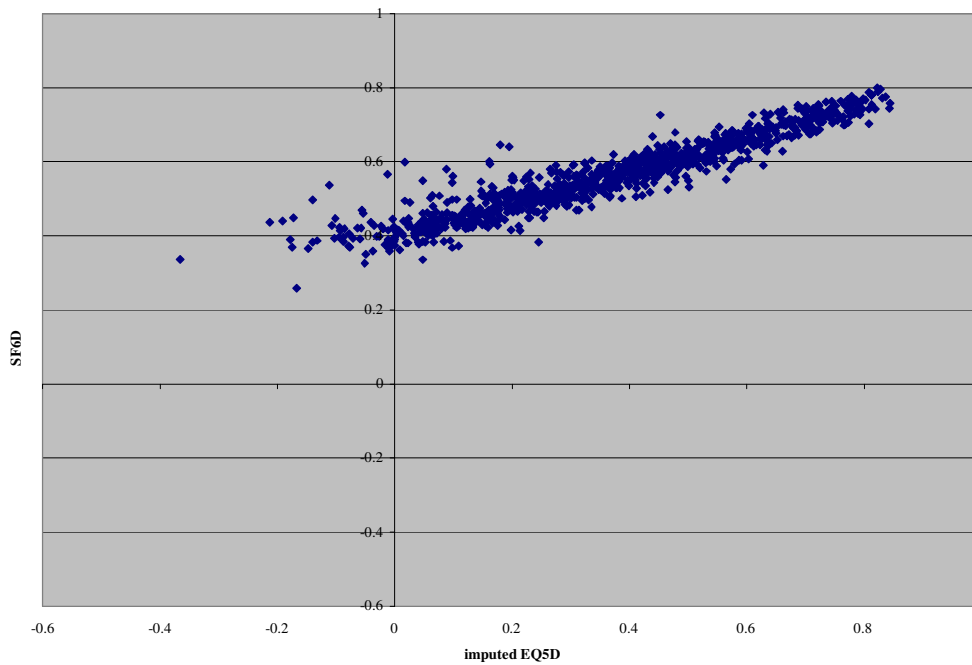
2.1.6. *Probabilistic Sensitivity Analysis Approach*

The uncertainty in model parameters is characterised using probability distributions. Where possible we characterised joint probability distributions for all the uncertain parameters. This was accomplished by using multivariate normal distributions to describe the correlation in uncertainty between the results of the statistical analyses. To do this, the variance covariance matrix is used to capture the joint distributions. Where joint distributions are not described we assume independence between the uncertainty in parameters.

Monte Carlo sampling is used to propagate the parameter uncertainty in the cost effectiveness model. This entails making random draws of the uncertain parameters from their (joint) probability distribution, running the model for each simulated set of parameters and collecting the outputs from each run.[Briggs, 2001] These are then a random sample from the induced probability distribution of model outputs. This process is known as 'probabilistic sensitivity analysis' (PSA). Outputs from the model include mean costs and mean effectiveness. In comparing the cost-effectiveness of two strategies, uncertainty about incremental mean costs and effectiveness can be displayed in the incremental cost-effectiveness plane as a scatter plot of the Monte Carlo output samples. When choosing between two strategies, decision uncertainty is usually expressed graphically through the cost-effectiveness acceptability curve (CEAC), which plots the probability that one treatment is more cost-effective than the other as a function of the societal willingness to pay threshold value of a QALY.

Decisions about whether to reimburse interventions are made with decision uncertainty. The consequences of decision uncertainty, in terms of wasted resources and health gain forgone, can be calculated to inform whether additional evidence should be collected in order that the decision can be reviewed in the future. This is the basis of value of information analysis which can identify those areas where reducing the uncertainty would have the greatest impact upon the decision uncertainty. We present the results of the global expected value of information.

2.2. MODEL PATHWAYS DESCRIPTION

The model pathway is described in Figure 4.

For each modelled patient we examine

o Initial DAS28 response (either a moderate, good, or non DAS28 responder based on their current age, disease duration, number of previous DMARDs, level of health state utility, gender)
o Improvement in health utility over the first 6 months (based on a number of demographic and clinical characteristics, along with the type of treatment they are on, and the type of DAS28 response they are predicted to achieve.)
o Utility progression 6 months onward
o The length of time a patient remains on each treatment (dependent variables such as age, disease duration, number of previous DMARDs and the type of DAS response they are predicted to achieve)

Alongside this we examine

o Drug and monitoring costs
o Costs associated with hospitalisation

When the patient reaches his/her time to withdraw, the model moves on to a 2nd treatment in the sequence, then a 3rd etc…up to 6 treatments. As all the estimates include time dependent variables such as age, disease duration and number of previous DMARDs, at each point the patient withdraws from a treatment and switches to the next, the probabilities of response, magnitude of improvement and time to withdrawal will be different. We do not specify particular DMARDs at different positions in the sequence but rather use the same generalised DMARD in each position based on a weighted average of BSRBR patients DMARD use. After the 6th treatment we assume patients will no longer respond but will still receive some maintenance therapy on DMARDs.

The model runs the same patient through 2 arms i.e. TNF inhibitor therapy versus traditional DMARDs.

The formulation of each of these analyses is given in the next section.

Figure 4: Example of how analyses link together over time to predict health state utility, time on treatment and cost.

Time to withdrawal ~
  DAS response$_t$
  Age$_t$
  Disease duration$_t$
  Number DMARDs$_n$
  Gender
  Concomitant DMARD
  Treatment$_1$

Time to withdrawal ~
  DAS response$_{t+1}$
  Age$_{t+1}$
  Disease duration$_{t+1}$
  Number DMARDs$_{n+1}$
  Gender
  Concomitant DMARD
  Treatment$_2$

DAS response type ~
  Age$_t$
  Disease duration$_t$
  Number DMARDs$_n$
  Gender
  Treatment$_1$

Utility progression

DAS response type ~
  Age$_{t+1}$
  Disease duration$_{t+1}$
  Number DMARDs$_{n+1}$
  Gender
  Treatment$_2$

Utility progression

Repeat until death

Improvement in utility ~
  Utility$_t$
  DAS response$_t$
  Age$_t$
  Disease duration$_t$
  Number DMARDs$_n$
  Gender
  Treatment$_1$

Improvement in utility ~
  Utility$_{t+1}$
  DAS response$_{t+1}$
  Age$_{t+1}$
  Disease duration$_{t+1}$
  Number DMARDs$_{n+1}$
  Gender
  Treatment$_2$

Drug Costs & monitoring and administration ~
  Treatment$_1$

Hospitalisation Cost ~
  Utility$_{t - t+1}$
  Age$_{t - t+1}$
  Disease duration$_{t - t+1}$
  Number DMARDs$_n$
  Gender
  Treatment$_1$

Drug Costs & monitoring and administration ~
  Treatment$_2$

Hospitalisation Cost ~
  Utility$_{t+1 - t+2}$
  Age$_{t+1 - t+2}$
  Disease duration$_{t+1 - t+2}$
  Number DMARDs$_{n+1}$
  Gender
  Treatment$_2$

TREATMENT 1

TREATMENT 2

Disease duration / age

(t)   (t+1)   (t+2)   (t+x)

Number of DMARDs

(n)   (n+1)   (n+2)

18

2.3. PROBABILITY OF RESPONSE TO TNF INHIBITOR AND TRADITIONAL DMARD THERAPY

We estimate the probability of DAS28 (EULAR) response (non/ moderate/ good) using a proportional odds cumulative Logit model (**Table 1**)

The probability of DAS28 response type is sampled at each point when a new treatment is attempted. Therefore, the probability is dependent on how long the patient has been on their existing treatments and the number of exisiting treatments.

Table 1 Statistical modelling of proportional odds cumulative Logit model for predicting type of response

Let π_1 π_2 and π_3 be the probability of a DAS response 0 (poor), 1 (moderate) or 2 (good)

$$L_1 = \log\left(\frac{\pi_1}{1-\pi_1}\right)$$

$$L_2 = \log\left(\frac{\pi_1 + \pi_2}{1-(\pi_1 + \pi_2)}\right)$$

We fit the model

$$L_j = \alpha_j - \gamma^T x$$

To predict the probability of a DAS response we use the equations

$$P(DAS = 0) = \frac{1}{1 + \exp\left(-\left(\alpha_1 - \gamma^T x\right)\right)}$$

$$P(DAS = 2) = 1 - \frac{1}{1 + \exp\left(-\left(\alpha_2 - \gamma^T x\right)\right)}$$

$$P(DAS = 1) = 1 - P(DAS = 0) - P(DAS = 2)$$

where the γ are the coefficients for the covariates.

Table 2: Results of proportional odds cumulative Logit model for predicting type of response

Co variable		SF6D	EQ5D
x_1	Health state utility	2.2691	1.0275
x_2	Age (years)	-0.0209	-0.0182
x_3	Disease duration (years)	0.0097	0.0098
x_4	Previous number of DMARDs	-0.0676	-0.0624
x_5	Gender (0=Male, 1=Female)	-0.3162	-0.2932
x_6	Whether on TNF inhibitor (1=Yes)	0.5608	0.6318
α_1	None I Moderate or Good intercept	-1.1451	-1.6849
α_2	None or Moderate I Good intercept	1.3917	0.8650

A positive coefficient signals an increasing chance of successful response. Thus patients that are older, have had more previous DMARDs or are female, are less likely to have successful response, whilst those who have higher baseline utility, or are on a TNFα inhibitor are more likely to respond.

The coefficient for disease duration is counter intuitive since patients with a greater disease duration appear to be more likely to respond. For an individual tracked through our model, this is counter balanced by the age coefficient.

The variance-covariance matrix for these regression parameters, were recorded and have been used in probabilistic sensitivity analysis initialising a multi-variable normal distribution.

Figure 5shows an example of how age and disease duration and treatment are estimated to affect the probability of type of response for a patient in the intervention and control arms over 10 years in the model. The exact probabilities change dependent on the type of patient.

Figure 5: Example of probability of DAS response type over time for an average patient on the intervention arm (top) and control arm (bottom)
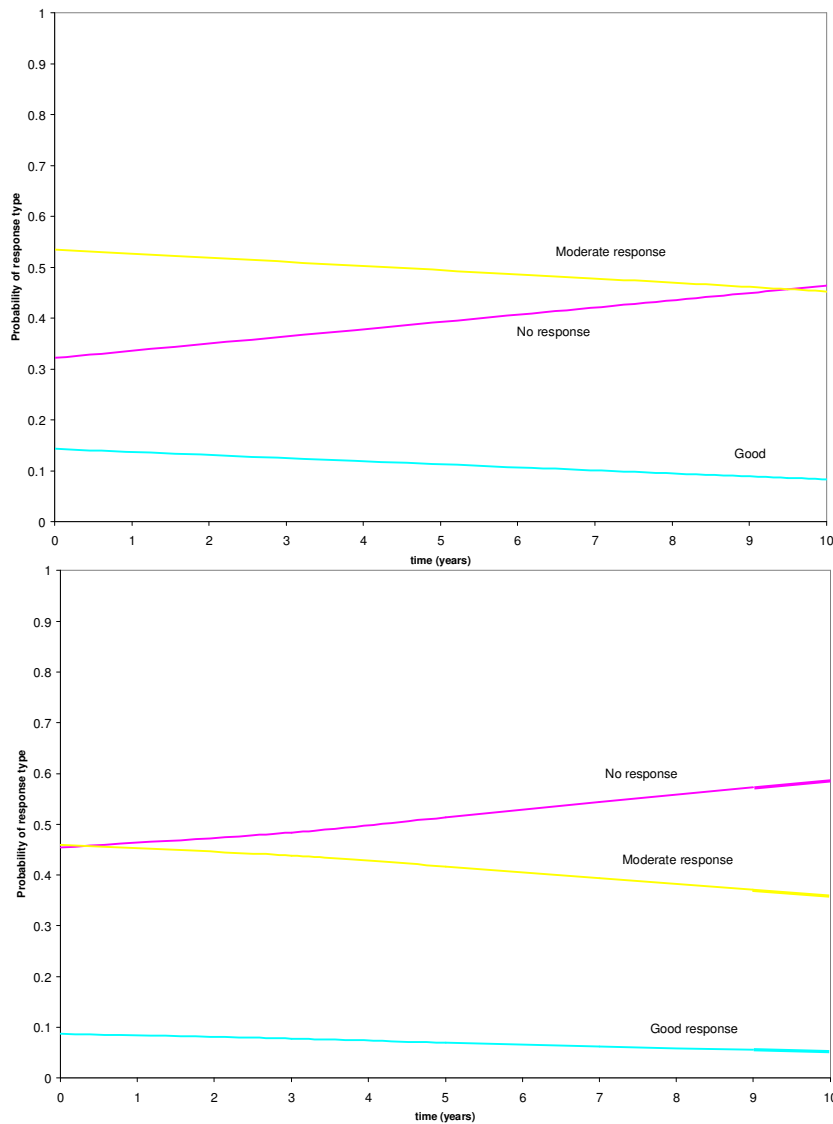
Graphs shown are for an average patient in the BSRBR (female, aged 55, baseline SF6D 0.53, baseline HAQ equal to 2.1, disease duration of 14 years, attempted 5 previous DMARDs)

2.4. INITIAL IMPROVEMENT ON TNF INHIBITOR AND TRADITIONAL DMARD THERAPY

Table 3 Statistical Modelling of Initial Improvement on TNF inhibitor and Traditional DMARD

Predict utility at six month (x+1) from baseline x

We assume utility at 6 months is normally distributed with mean μ

$$u_6 \quad \sim \quad N(\mu, \sigma^2)$$
$$l(\mu) \quad = \quad \alpha + \beta l(u_0) + \gamma^T x$$

Where β is the coefficient for the baseline utility u_o and the γ are the coefficient for the other covariates.

To predict six month utility from baseline use the formula
$$utility = l^{-1}\left(\alpha + \beta l(u_0) + \gamma^T x\right)$$

Utility is transformed to the logit scale and vice versa to maintain utility in a reasonable range by:

Function to transform the range –0.6 :1 to 0.025:0.975

$$t(x) = \frac{0.95(x + 0.6)}{1.6} + 0.025$$

Function to transform the range 0.025:0.975 to –0.6 :1

$$t^{-1}(x) = \frac{1.6(x - 0.025)}{0.95} - 0.6$$

Logit type function for the range –0.6 :1

$$l(x) = \log\left(\frac{t(x)}{1 - t(x)}\right)$$

Inverse logit type function for the range –0.6:1

$$l^{-1}(x) = t^{-1}\left(\frac{1}{1 + e^{-x}}\right)$$

Therapy

The magnitude of improvement (can be worsening) in the first 6 months of a new treatment is then estimated using a multivariate regression model. This used demographic and clinical variables along with the type of DAS28 response. (Table 4)

Table 4: Results of multivariate regression model to predict improvement in utility over 1st 6 months of treatment

Co variable	Description	SF6D	EQ5D
u_0	Health state utility at time x †	0.4972	0.3854
x_1	Age (years)	-0.0026	-0.0049
x_2	Disease duration (years)	-0.0002	-0.0046
x_3	Previous number of DMARDs	-0.0132	-0.0176
x_4	Gender (1=Male?)	0.0256	-0.0419
x_5	Whether on TNF inhibitor (1=Yes)	0.0967	0.1753
x_6	No DAS28 response	reference	reference
x_7	Moderate DAS28 response	0.1482	0.2387
x_8	Good DAS28 response	0.2936	0.4922
α	Intercept	0.5453	0.5629

The results show that patients who are older, have greater disease duration, or higher previous DMARD use achieve slightly lower utility improvements. Patients who receive a TNF inhibitor, or are DAS28 moderate or good responders achieve higher utility improvements. Note, that in each case, the EQ5D utility coefficients are approximately twice the size of those for SF6D

The variance-covariance matrix for these regression parameters, were recorded and have been used in probabilistic sensitivity analysis initialising a multi-variable normal distribution.

Figure 6: Magnitude of utility improvement in the first 6 months based on type of DAS response for average patient on intervention arm (top) and control arm (bottom)
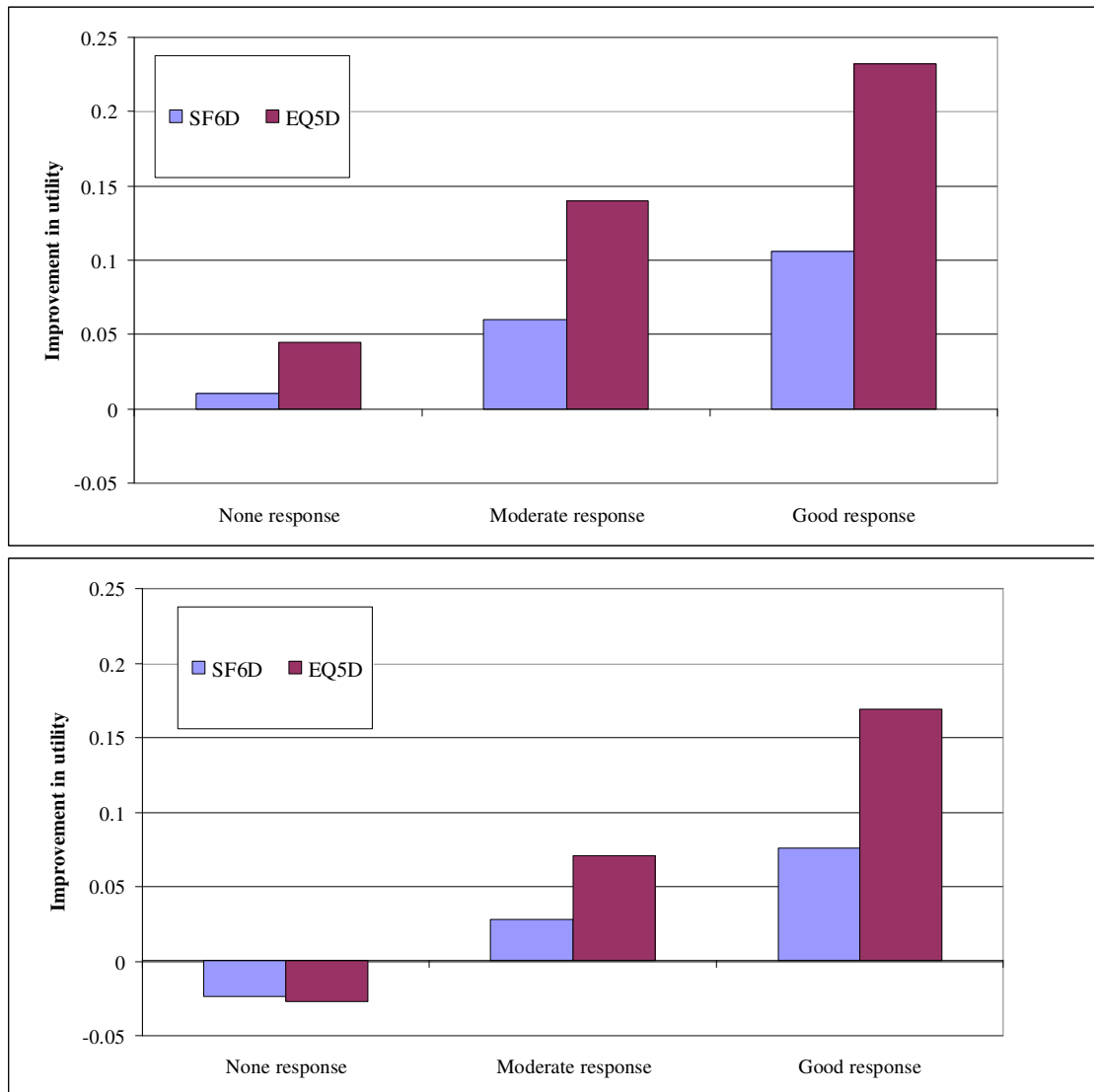
Graphs shown are for an average patient in the BSRBR (female, aged 55, baseline SF6D 0.53 (EQ5D=0.31), baseline HAQ equal to 2.1, disease duration of 14 years, attempted 5 previous DMARDs)

It is apparent from trial evidence on biologics that improvements in disability occur quickly in responders to TNFα treatment. Instead of assuming a straight line between 0 and 6 months, we have adjusted the utility improvement to reflect this. We assume that 80% of a patient's total response will be achieved within the first month. (see figure 7)

Figure 7: Illustration of assumption over utility improvement
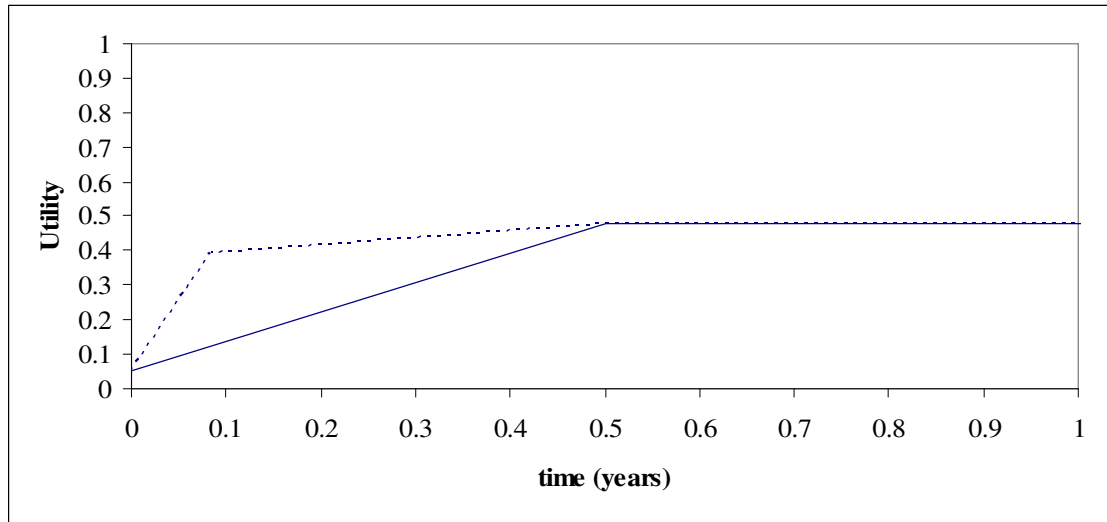
2.5. DURATION OF TNF INHIBITOR TREATMENT – TIME TO WITHDRAWAL

The time on TNF inhibitor treatment is modelled using a Weibull survival analysis. For this analysis, we assumed that patients could only switch treatment at each 6 monthly interval. In the registry, it was apparent that switching sometimes occurred between these clinical assessments. We however made the assumption that the therapy that the patient ended the period on, was the predominant therapy for the entire period.

The type of response a patient is predicted to have achieved is used in estimating the duration on treatment in two ways. First, the type of DAS28 response the patient has achieved is used to predict the time the patient will spend on therapy. We found that patients who achieved a good DAS28 response, would remain on TNF antagonist treatment for longer than moderate or non responders. (Results of a multivariate Weibull survival analysis on treatment are found in Table 6). Secondly, dependent on the rule for deciding whether a patient should remain on therapy, the type of DAS28 response is used to remove patients after 3 months or 6 months. For instance, the current NICE guidelines specify that patients should be at least a moderate responder at 3 months to remain on treatment.[NICE, 2004] If that rule is selected, and the patient is sampled to be a non responder, the Weibull regression will estimate the time on treatment and if this is 3 months, the decision rule is used and withdraws them at this point.

Table 5 Statistical Modelling of Weibull survival analysis

The baseline hazard function is

$$h_0(t) = \frac{\alpha}{\beta^{\alpha}} t^{\alpha-1}$$

Where α is the shape and β the scale parameter and t is the time in months A proportional hazards model is fitted for adjusting the survival for covariates.

$$h(t) = h_0(t)\exp(\gamma^T x)$$

The survival curve is

$$s(t) = \exp\left(-\int_0^t h(u)du\right)$$

$$= \exp\left(-\exp(\gamma^T x)\left(\frac{t}{\beta}\right)^{\alpha}\right)$$

The model is fitted twice. Firstly where the survival time is the time on the first TNF, and secondly where they are the time on any continuous TNF.

Table 6: Multivariate Weibull survival analysis to predict time on 1st and all TNF antagonist treatments

Co variable	Description	First TNF		All TNFs	
		SF6D	EQ5D	SF6D	EQ5D
x_1	Age (years)	-0.003	-0.003	0.0162	0.0154
x_2	Disease duration (years)	0.001	0.002	0.0079	0.0061
x_3	Previous number of DMARDs	0.066	0.066	0.0587	0.0675
x_4	Gender (1=Male?)	-0.750	-0.454	-1.0462	-0.8764
x_5	Whether on TNF inhibitor (1=Yes)	0.137	0.175	0.0918	0.1083
x_6	On concomitant DMARD	0.042	0.078	-0.1630	-0.1166
	No DAS response	reference		reference	
x_7	Moderate DAS response	-1.264	-1.232	-1.2307	-1.2783
x_8	Good DAS response	-1.882	-1.777	-1.7873	-1.6193
$\log(\beta)$	log(scale)	3.764	3.772	4.1826	4.1883
$\log(\alpha)$	log(shape)	0.588	0.582	0.7979	0.8085

The variance-covariance matrix for these regression parameters, were recorded and have been used in probabilistic sensitivity analysis initialising a multi-variable normal distribution.

The existing guidelines on prescribing anti TNF-αs, mandated by NICE recommended that patients should only continue therapy if they were at least a moderate responder in the first 3 months of treatment. More recent guidelines from the BSR state that 6 months is a more appropriate guideline. Since the BSRBR only collects data at 6 monthly intervals, it was difficult to assess the impact of this issue on the cost utility of anti TNFαs directly from the data.

Sometimes, patients who are DAS28 non responders are sampled with a worsening utility. It might be argued that if a patient is showing a disutility due to a treatment it would be stopped. We have looked at this option only in the sensitivity analysis since quality of life is not explicitly used as a decision rule by clinician or patient, particularly if it is a small change which would not necessarily be detectable by a patient (<0.1). This decision rule applies both to TNF inhibitor and traditional DMARD therapy

2.6. DURATION OF DMARD TREATMENT – TIME TO WITHDRAWAL

For estimating the time to withdrawal on a DMARD, since most of the BSRBR data were limited to just the first 6 months, we could not use a survival analysis to estimate long term survival on DMARD treatment. Instead we used figures from Weibull distributions averaged over all DMARDs from the literature. [Barton et al.]

However, unlike the TNF antagonist survival estimates, this is not dependent on the type of DAS28response. Therefore a patient who is sampled to be a DAS28 non responder has the identical survival probability on DMARDs as a DAS28 good responder.

Table 7: Values for Weibull survival on DMARDs [Barton et al.]

Co variable	Description	Coefficient	SE
α	Scale	2.68	0.59
β	Shape	0.80	0.07

2.7. LONGER TERM PROGRESSION OF UTILITY WHILST ON TNF INHIBITOR THERAPY

2.7.1. Assumption of Steady State Based On other Registries

Registries in Sweden and also the US have published data on HAQ trends for 4 years plus suggesting that response to therapy is maintained.[Kobelt 2004] In our basecase analysis we have assumed that utility achieved at 6 months is maintained until treatment withdrawal i.e. that utility progression is zero.

2.7.2. Results from BSRBR data by DAS28 Response

The BSRBR data have the opportunity to examine this but there are few patients with over 18 months follow-up.

We constructed a statistical model to examine the rate of utility progression according to DAS28 response status at 6 months.

The model is based on a similar logistic approach as in Table 3 but with covariates only according to DAS28 response.

Table 8: Results of multivariate regression model to predict improvement in utility over 6 to 18 months TNF inhibitor treatment. (coefficients are monthly change on the Logit scale)

Co variable	Description	SF6D	EQ5D
x_8	Good DAS28 response (reference)	-0.001813	-0.0051
x_6	No DAS28 response (add to reference)	0.008874	0.008164
x_7	Moderate DA28S response (add to reference)	0.001669	0.000909

The results show that there is a slight worsening of utility for patients who achieved good response, whilst moderate responders have progression close to zero and poor responders experience some ongoing but marginal improvement in utility.

2.7.3. Sensitivity Analysis on delayed underlying progression using Age and Disease Duration Parameters.

There is evidence from the trends in radiological data for TNF antagonists that patients on average show at least no radiological progression (some trials show improvements) whereas treatment with MTX and other DMARDs, even when patient respond is associated with radiological progression. Radiological progression has been shown to be strongly linked with worsening disability over the long term.[Scott, 2000].

We therefore assume that patients have the same radiological damage at the end of anti-TNFα treatment as at the start, so their ability to improve is the same at the end as the start. To do this, since radiological damage is not a parameter in our model, we hold patients age and disease duration whilst they remain on the anti-TNFα, so when they do discontinue, they would follow the precise path in progression they would have done had they gone directly to DMARDs at the time they started their TNF antagonist. This concept is well explained in Kirwan et al. [Kirwan, 1999] These assumptions are explored in sensitivity analyses.

Figure 8: Example from Kirwan on delayed progression
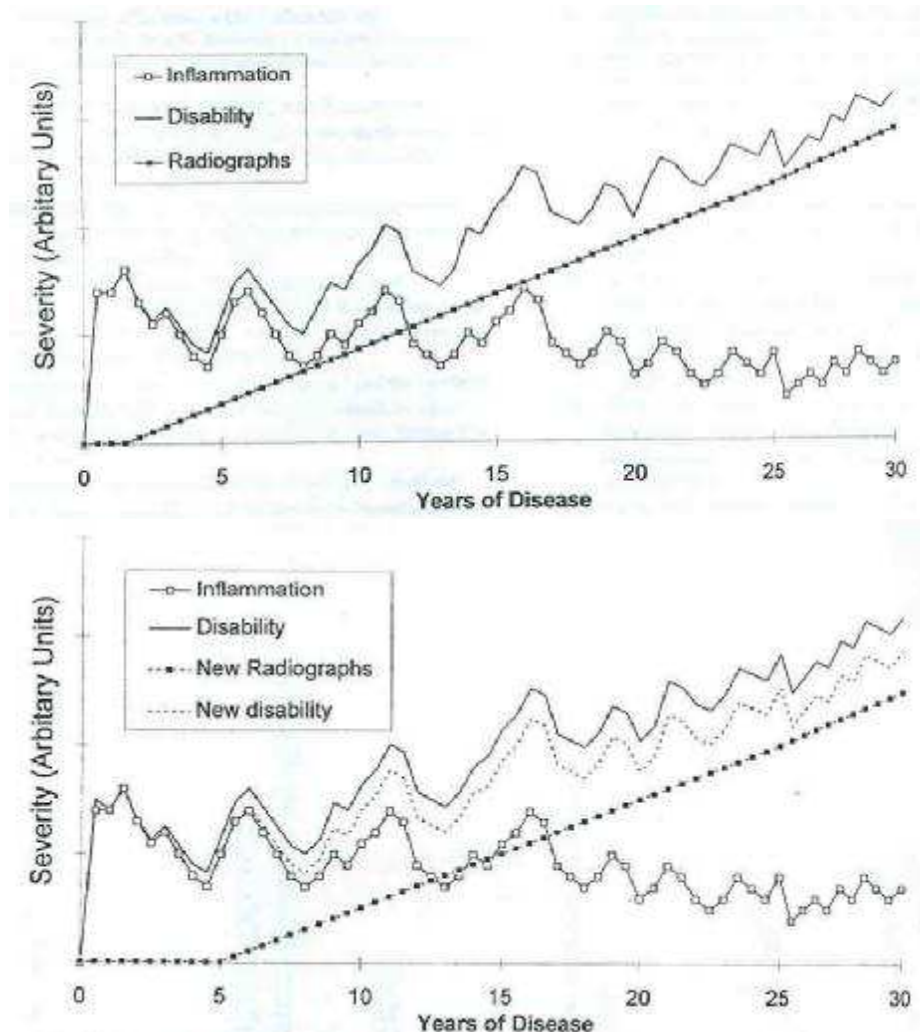
Figure 5. Schematic representation of "progression" in RA after suppression of radiographic change for the first 4 years of disease.

30

2.8. WORSENING AFTER WITHDRAWAL FROM TNF INHIBITOR THERAPY

There are very few data on utility changes after withdrawal from TNF inhibitor.

The analysis undertaken of BSRBR data suggest an overall monthly worsening on the logistic scale of -0.0056. This is a much smaller level of worsening than the initial improvements seen and it would be nonsensical to then apply utility improvements based on Table 3 from this new base because patients would be on a much higher utility than they achieved during their successful response to TNF inhibitor therapy.

Instead we make 2 assumptions.

The first assumption is that when a patient switches treatment, they worsen temporarily until the new treatment becomes effective. This is not seen in the long term data from the registry as the time point of switching treatment is different to the time when the patient answers the questionnaire regarding their disability and quality of life.

The second assumption is the level of utility the patient will return to when they discontinue the treatment they are on. We made the assumption that the level they would reach would be dependent on the progression seen during the previous treatment, and that the worsening would be equal to the initial improvement (Figure 9). This is based on results seen when patients on etanercept were discontinued and their disability quickly rebounded back to near baseline.[Brennan 2004]

Figure 9: Example of how utility changes at end of treatment
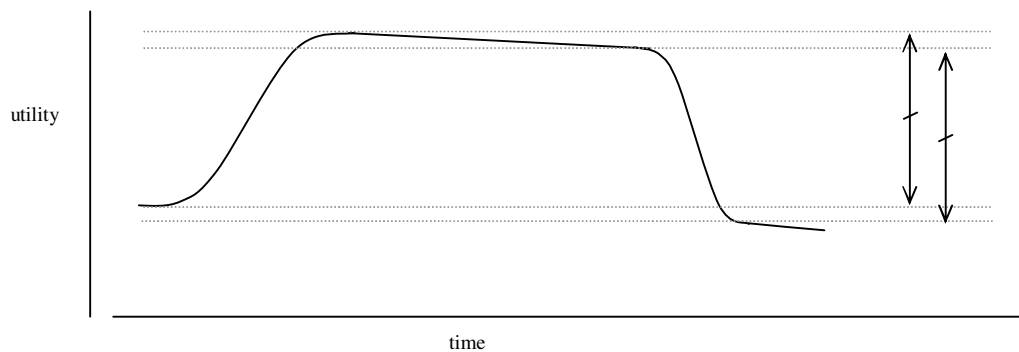
2.9. LONG-TERM PROGRESSION ON TRADITIONAL DMARD THERAPY

2.9.1. HAQ DI score progression on DMARDs - Scott et Al Evidence Re-examined

We again have insufficient data to estimate the progression of utility for a patient whilst on DMARD therapy so we have gone to external sources. Reviewing the literature showed that no studies had looked at health utility progression directly. We therefore used HAQ DI as a proxy for utility as has been done before in economic evaluations of RA therapies. In a paper by Scott et al, the annual progression in HAQ -DI is assessed from 12 cross sectional studies.[Scott, 2000] The weighted average of annual HAQ progression, was calculated to be 0.042 (Table 9).

Table 9: Review of DMARD progression rates from Scott et al

Study	Year	N in study	Mean annual HAQ progression
Wolfe et al	1991	561	0.020
Lassere et al	1995	353	0.045
Sherrer et al	1986	691	0.072
Greenwood et al	1999	701	0.032
Ward et al	1993	282	0.014
Gardiner et al	1993	175	0.030
Callahan et al	1997	100	-0.006
Leymarie et al	1997	370	0.000
Ward et al	1998	182	0.017
Munro et al	1998	440	0.119
Truro cases	1998	33	0.006
Shipps Cross cases	1998	46	0.023
Crude average			0.031
Weighted average			**0.042**

From Scott et al. The links between joint damage and disability in rheumatoid arthritis

To estimate the uncertainty in the average progression on utility we use the figure of 0.58 for the individual variation for a patient with established RA over 4 to 5 years in Scott et al. To calculate the standard error, we first make this an annual variation (0.145) and then divide by the square root of n-1 (=0.0023, where n=3934). This again is converted to utility.

2.9.2. Translating HAQ DI score to Utility

To convert this to health state utility we use results of simple linear regressions ($\Delta SF6D = -0.1008 \Delta HAQ$, $\Delta EQ5D -0.2102 \Delta HAQ$). [Bansback Appendix 3]

2.10. DRUG AND MONITORING COSTS

The BSRBR gave the breakdown of treatment type for each arm. In patients taking TNF antagonists, up to 80% of patients were taking a concomitant DMARD. DMARDs were often taken in combinations of up to 4 treatments. We assumed this breakdown would remain constant over time as patients switch from treatment to treatment.

Table 10 Percentage of patients on each DMARD and TNF inhibitor

	TNF inhibitor Breakdown	DMARD Breakdown
Azathioprine	4%	1%
Hydroxychloroquine	10%	11%
Gold	2%	7%
Leflunomide	11%	16%
Methotrexate	80%	54%
Sulfasalazine	18%	38%
Ciclosporin	3%	1%
Etanercept	43%	0%
Infliximab 2nd year	44%	0%
Adalimumab	13%	0%

We multiplied these proportions by calculated costs for each individual treatment to gain a weighted cost for each strategy (divided into TNF antagonist plus concomitant DMARD, TNF antagonist alone, DMARD).

Table 11 Cost Assumptions

| Treatment | Dose/ schedule | No treatments | | Dose each treatment (mg) | Dose (mg) | | Vial size (mg) | Cost per vial | Cost per | | Total annual cost |
		1st 6 mths	Sub 6 mths		1st 6 months	sub 6 months			1st 6 months	Sub 6 months	
Azathioprine	1-3mg per kg daily	182	182	150	27300	27300	2800	10	97	97	194
Hydroxychloroquine	400mg daily	182	182	400	72800	72800	12000	5	28	28	55
Gold	50mg (first dose 10) weekly	26	26	50	1300	1300	50	9	243	243	487
Leflunomide	100mg for three days then 20 daily	182	182	20	3640	3640	20	2	282	282	564
Methotrexate	15mg weekly	26	26	15.0	390	390	70	3	18	18	36
Sulfasalazine	500mg to 2500mg once daily with dose increasing by 500mg each week to max of 3g	840	910	500	420000	455000	56000	8	63	68	132
Ciclosporin	3.25mg/kg daily	182	182	225	40950	40950	6750	173	1051	1051	2101
Etanercept	25mg twice weekly	52	52	25	1300	1300	25	89	4648	4648	9296
Infliximab 1st year	3mg/kg weeks 0, 2, 6 then every 8	5.5	3.25	210	1155	682.5	100	419.62	4847	2864	7711
Infliximab 2nd Year	3mg/kg every 8 weeks	3.25	3.25	210	682.5	682.5	100	419.62	2864	2864	5728
Adalimumab	40mg every other week	13	13	40	520	520	40	358	4648	4648	9295

[Note most Unit costs from BNF February 2005, Infliximab price has changed from that stated by BNF at that time and has reduced from £451.20 per vial to £419.62. Earlier draft versions of this report used the higher price, the results in this final version use the new lower price of £419.62 per vial.]

Table 12 Monitoring Assumptions

Treatment	Outpatient department visit	GP visit	Half day Hospital attendance	Full blood count	Erythrocyte Sedimentation Rate/ or CRP	Liver function test	Chest x-ray	Urea, electrolytes and creatinine	Protein and glucose	Blood pressure	Total cost
1st 6 Months											
Azathioprine	1	12		13	1	13		1			545
hydroxychloroquine	1	2		3	3	3					207
Gold	1	23		24	1	24		24	24		1,078
Leflunomide	1	12		13	1	7		1	1	13	508
Methotrexate	1	12		13	1	13	1	13			639
Sulfasalazine	1	7		8	1	5					343
Ciclosporin	1	8		7	1	7		9	1	9	418
Etanercept	1	5		6	6	6	1	6			404
Infliximab			5	5	5	5	1	4			1142
Adalimumab	1	5		6	6	6	1	6			404
subsequent 6 months											
Azathioprine	1	5		6		6					280
hydroxychloroquine	1	1		2	2	2					161
Gold	1	8		9		9		9	9		442
Leflunomide	1	3		4		4					209
Methotrexate	1	5		6		6		6			317
Sulfasalazine	1	1		2		2					139
Ciclosporin	1	5				6					213
Etanercept	1	2		3	3	3		3			226
Infliximab			4	4	4	4		4			903
Adalimumab	1	2		3	3	3		3			226

[Unit costs from Barton et al]

2.10.1. Dose Assumptions

The costs for each drug were calculated using the recommended dosages.[BSR guidelines]

The amount of infliximab given to a patient is determined by their weight. The recommended initial dose is 3mg per kg. This is given at week 0, 2, 6 and then subsequent 8 weeks. We assumed 3 initial treatments in the first 6 weeks plus a subsequent eight weekly regimen giving an average use of 20 weeks divided by 8 i.e. 2.5 over the remainder of the first 6 months. The total number of treatments assumed over the first 6 months is therefore 5.5. In the 2nd 6 months and beyond we assumed infusions eight weekly, giving an average of 26/8= 3.25 infusions per 6 month period. Given an person for example of weight 70kg, this would calculate to 1155mg in the first 6 months and 682.5mg in subsequent 6 months.

The actual dose of infliximab is recorded in the registry and showed some differences from this recommended dosing. The possible reasons for differences relate to the provision of infliximab in 100mg vials. The dose calculated based on patient's weight might leave some of a vial unused. What happens to this unused drug is not recorded in the registry. It might be used on another patient, it might be thrown away, or the patient might receive slightly more than the exact calculated dose.

We have used the recommended licensed doses in the model's central estimate and used the BSR data on reported dose in a sensitivity analysis.

[Note: Earlier draft versions of the report and model used a slightly higher number of doses of infliximab in the first 6 months – wrongly assuming 6.25 infusions over the first 6 month period. This has been corrected]

2.10.2. Cost Summary

Table 13: Summary of Drug Costs and monitoring costs at licensed dosages by treatment

	Drug Cost 1st 6 months	Drug Cost subsequent 6 months	Monitoring 1st 6 months	Monitoring subsequent 6 months	Breakdown of use in intervention arm	Breakdown of use in control arm
Azathioprine	97	97	545	280	4%	1%
Hydroxychloroquine	28	28	207	161	10%	11%
Gold	243	243	1078	442	2%	7%
Leflunomide	282	282	508	209	11%	16%
Methotrexate	18	18	639	317	80%	54%
Sulfasalazine	63	68	343	139	18%	38%
Ciclosporin	1051	1051	418	213	3%	1%
Etanercept	4648	4648	404	226	43%	0%
Infliximab	4847	2864	1142	903	44%	0%
Adalimumab	4648	4648	404	226	13%	0%

Table 14: Summary of Total costs by treatment by 6 monthly periods

Monthly period	Licensed dosages			Using reported dosages with dose escalation		
	Anti-TNF in combination with DMARD	Anti-TNF alone	DMARD	Anti-TNF in combination with DMARD	Anti-TNF alone	DMARD
0-6	£6,262	£5,464	£781	£7,265	£6,467	£781
6-12	£4,826	£4,387	£430	£5,955	£5,516	£430
12-18	£4,826	£4,387	£430	£6,211	£5,771	£430
18-24	£4,826	£4,387	£430	£6,641	£6,202	£430

2.11.   COSTS DUE TO HOSPITALISATION

A regression model is used to predict the number of days per 6 months a patient is an inpatient.

We looked at the impact of treatment on adverse events and hospitalisation by building a multivariate regression model looking at the number of days a patient was hospitalised dependent on clinical and demographic factors, along with the treatment type and health state utility value.(Table 15) Previous studies have shown a strong relationship between direct costs (predominantly drug and hospitalisations for joint replacements) and disability.[Michaud 2004]

Table 15: Results of multivariate regression of number of days LOS

Co variable	Description	SF6D	EQ5D
x_1	Health state utility	-1.4690	-0.5467
x_2	Age (years)	0.0080	0.0078
x_3	Disease duration (years)	0.0083	0.0075
x_4	Previous number of DMARDs	0.0690	0.0648
x_5	Gender (1=Male?)	-0.0611	-0.0620
x_6	Whether on TNF inhibitor (1=Yes)	-0.4113	-0.3719
x_7	(Intercept)	0.7883	0.2351

This was multiplied by an average cost for a day in a rheumatology unit (= £287 (range £145 – 368)).[Netten and Curtis, 2002](inflated to 2004).

This methodology underestimates the total non drug related costs of the disease. For instance, it does not look separately at the breakdown of type of adverse event, or type of procedure that may vary between arms. This subject requires further detailed analysis, which the BSRBR provides good quality data to provide.

2.12. OTHER COSTS

A number of important costs are excluded

– costs of institutionalisation due to disability are excluded

– costs of longer-term surgeries unless they are represented in the first 18 months of BSRBR data

– costs to (or quality of life impact on) carers

– lost work productivity due to disability among patients of working age.

2.13. LIFE-TABLES AND MORTALITY ASSUMPTIONS

Life years, the other component of QALYs have not been analysed separately between groups. Instead we have used standard UK life tables (Table 16).[www.gad.gov.uk] We adjust these by standardised mortality ratios for patients with RA (Table 17).[WHO Global Burden of Disease programme]

Whilst evidence is emerging between the relationship between improved disability and increased longevity, given the current short time horizon of the registry, this has not been included.

Table 16: Life tables for the UK population (Probability a person aged x will die before x+1)

	Standard		Adjusted for RA		*(cont)*	Standard		Adjusted for RA	
Age	**Male**	**Female**	**Male**	**Female**	*Age*	**Male**	**Female**	**Male**	**Female**
0	0.0060	0.0048	0.0120	0.0097	51	0.0044	0.0029	0.0070	0.0050
1	0.0004	0.0003	0.0009	0.0007	52	0.0048	0.0032	0.0076	0.0056
2	0.0003	0.0002	0.0005	0.0004	53	0.0052	0.0034	0.0082	0.0059
3	0.0002	0.0002	0.0003	0.0003	54	0.0056	0.0038	0.0090	0.0067
4	0.0002	0.0002	0.0003	0.0003	55	0.0063	0.0041	0.0101	0.0072
5	0.0001	0.0001	0.0003	0.0002	56	0.0072	0.0045	0.0115	0.0079
6	0.0001	0.0001	0.0003	0.0002	57	0.0079	0.0051	0.0126	0.0089
7	0.0001	0.0001	0.0002	0.0002	58	0.0087	0.0053	0.0140	0.0093
8	0.0001	0.0001	0.0002	0.0002	59	0.0098	0.0060	0.0156	0.0104
9	0.0001	0.0001	0.0002	0.0002	60	0.0110	0.0068	0.0177	0.0118
10	0.0001	0.0001	0.0002	0.0002	61	0.0120	0.0074	0.0193	0.0129
11	0.0001	0.0001	0.0003	0.0002	62	0.0133	0.0079	0.0212	0.0139
12	0.0002	0.0001	0.0003	0.0002	63	0.0145	0.0087	0.0232	0.0152
13	0.0002	0.0001	0.0004	0.0002	64	0.0157	0.0097	0.0251	0.0169
14	0.0002	0.0001	0.0004	0.0003	65	0.0176	0.0105	0.0228	0.0158
15	0.0003	0.0002	0.0005	0.0003	66	0.0194	0.0118	0.0252	0.0176
16	0.0004	0.0002	0.0007	0.0005	67	0.0216	0.0130	0.0281	0.0194
17	0.0006	0.0003	0.0011	0.0005	68	0.0237	0.0144	0.0308	0.0216
18	0.0008	0.0003	0.0016	0.0006	69	0.0266	0.0160	0.0346	0.0241
19	0.0008	0.0003	0.0016	0.0006	70	0.0292	0.0178	0.0380	0.0267
20	0.0008	0.0003	0.0017	0.0006	71	0.0328	0.0202	0.0427	0.0303
21	0.0008	0.0003	0.0016	0.0006	72	0.0366	0.0226	0.0475	0.0339
22	0.0009	0.0003	0.0018	0.0006	73	0.0405	0.0253	0.0527	0.0380
23	0.0008	0.0003	0.0017	0.0006	74	0.0456	0.0282	0.0592	0.0423
24	0.0009	0.0003	0.0018	0.0006	75	0.0499	0.0316	0.0649	0.0474
25	0.0009	0.0003	0.0014	0.0006	76	0.0549	0.0351	0.0714	0.0527
26	0.0008	0.0004	0.0013	0.0006	77	0.0606	0.0389	0.0788	0.0584
27	0.0009	0.0004	0.0015	0.0006	78	0.0666	0.0428	0.0866	0.0643
28	0.0009	0.0004	0.0015	0.0006	79	0.0732	0.0472	0.0951	0.0709
29	0.0010	0.0004	0.0016	0.0007	80	0.0797	0.0532	0.1036	0.0798
30	0.0010	0.0004	0.0016	0.0007	81	0.0861	0.0586	0.1119	0.0878
31	0.0010	0.0005	0.0017	0.0009	82	0.0937	0.0657	0.1218	0.0986
32	0.0011	0.0005	0.0018	0.0009	83	0.1042	0.0731	0.1354	0.1096
33	0.0011	0.0005	0.0017	0.0010	84	0.1170	0.0836	0.1521	0.1254
34	0.0012	0.0006	0.0019	0.0011	85	0.1298	0.0931	0.1687	0.1396
35	0.0012	0.0007	0.0019	0.0012	86	0.1408	0.1023	0.1831	0.1535
36	0.0013	0.0007	0.0021	0.0012	87	0.1514	0.1136	0.1968	0.1704
37	0.0014	0.0007	0.0022	0.0013	88	0.1674	0.1267	0.2176	0.1901
38	0.0014	0.0008	0.0023	0.0015	89	0.1821	0.1404	0.2367	0.2106
39	0.0016	0.0009	0.0025	0.0016	90	0.1898	0.1546	0.2468	0.2319
40	0.0017	0.0010	0.0027	0.0017	91	0.2016	0.1710	0.2621	0.2565
41	0.0019	0.0011	0.0030	0.0019	92	0.2235	0.1879	0.2905	0.2818

42	0.0019	0.0012	0.0030	0.0022	93	0.2410	0.2070	0.3134	0.3105
43	0.0021	0.0014	0.0034	0.0025	94	0.2590	0.2215	0.3367	0.3322
44	0.0022	0.0015	0.0035	0.0026	95	0.2807	0.2420	0.3649	0.3631
45	0.0025	0.0016	0.0040	0.0029	96	0.2971	0.2602	0.3863	0.3904
46	0.0028	0.0019	0.0045	0.0034	97	0.3260	0.2766	0.4238	0.4149
47	0.0032	0.0020	0.0051	0.0036	98	0.3408	0.2989	0.4430	0.4484
48	0.0033	0.0022	0.0053	0.0038	99	0.3547	0.3177	0.4610	0.4766
49	0.0039	0.0024	0.0062	0.0043	100	0.3869	0.3435	0.5030	0.5153
50	0.0041	0.0027	0.0066	0.0047					

Table 17: Standardised Mortality Ratios for Rheumatoid Arthritis population

Age	Male	Female
0-24	2	2
25-64	1.6	1.75
65+	1.3	1.5

WHO Global Burden of Disease programme

2.13.1. Exclusion of any Biological Effect on Mortality Risk Reduction

Also excluded from our analysis is any potential effect of TNF inhibitor therapies on mortality. There are studies, which show a significant association between HAQ improvement and reduced mortality risk. [Pincus 2001], [Yelin et al , 2002]

2.14. DISCOUNTING

The discount rates in the basecase are set at 6% for costs and 1.5% for QALYS in line with the NICE protocol (viewed using
http://www.nice.org.uk/pdf/Final_protocol_Anti_TNF.pdf
on 25th May 2005) and as discussed by telephone with the NICE technical lead.

In a sensitivity analysis, we investigate the effects of discounting both benefits and costs at a 3.5% annual rate.

2.15.   NUMBER OF RUNS REQUIRED FOR PROBABILISTIC SENSITIVITY ANALYSIS

The model is a patient level simulation. Therefore running a probabilistic sensitivity analysis requires stability in both 1st and 2nd order uncertainty. We have used methodology recently developed in the University of Sheffield by Prof. Tony O'Hagan and colleagues at the Centre for Bayesian Statistics and health Economics (CHEBS) to quantify how many individual patient runs are necessary for stable results when undertaking probability sensitivity analysis. [O'Hagan et al. 2005]
Early runs of the model suggested a 1000 patient run would give a standard error in cost per QALY of approximately £300 to £400.

The formulae to optimise the number. of individuals sampled per 2nd order parameter sample gave results of 20 if cost variance were used as the metric, and 75 if QALY variance were used.

We therefore decided to run all analyses with 100 Monte Carlo samples for analysing parameter uncertainty. For each sample, we undertook ran 50 hypothetical individual patients. A total of 5000 model runs per evaluation.

3. RESULTS

3.1. BASECASE 1 – CURRENT UK PRACTICE AS PER BSRBR REGISTRY

In the first basecase analysis we make the assumptions on evidence as set out in the figure below.

Figure 10   Summary of Assumptions made in Basecase1

Health state utility estimate	○ SF6D ● EQ5D					
Utility Progression on Biologic after 6m		● Zero ○ BSR 6-18m data				
Allow age/dis dur'n update on Biologic?		○ Yes ● No				
Probabilistic Sensitivity Analysis	● On ○ Off					
Costs to include	☑ Drug ☑ Monitoring ☑ Hospitalisation					
Number of TNFs	● Single use ○ 2 in a sequence					
Discount rates	6.0% ↕ Costs 1.5% ↕ Benefits					
Response Threshold	● Use actual data from BSRBR					
Keep on only Moderate / Good responders for	○ 3m ○ 6m					
Keep on only Good responders for	○ 3m ○ 6m					
Withdrawal	● Do not link to improvement ○ Link to improvement					

Patient population

HAQ (0-3)	○ 0-0.5	○ 0.5-1.0	○ 1.0-1.5	○ 1.5-2.0	○ 2.0-2.5 ○ 2.5-3.0	● All
Age (yrs)	○ <40	○ 40-50	○ 50-60	○ 60-70	○ 70+	● All
Disease duration (yrs)	○ 0-5	○ 5-10	○ 10-15	○ 15+		● All
Number of previous DMARDs	○ <2	○ <3	○ <4	○ <5	○ 5+	● All
Gender	○ Male	○ Female				● All
Combination DMARD	○ Yes	○ No				● All

Treatment cost	● Licenced ○ BSRBR reported dose ○ BSRBR no creep
HAQ progression on DMARD	● Scott et al ○ ERAS II ○ ERAS III / IV
Utility / HAQ relationship	● Bansback ○ High ○ Low

Group Box 227

Table 18 Results Summary Table for BaseCase1 – Current UK Practice as per BSRBR Registry

Results Summary Table

	TNF inhibitor Therapy		Traditional DMARDs		Incremental
	Mean	S.E.	Mean	S.E.	Mean
Lifetime Cost	£ 57,919	£ 4,533	£ 20,706	£ 3,006	£ 37,214
Lifetime QALY	5.1514	0.5657	3.5931	0.6257	1.5583
Cost Per QALY gained					£ 23,882

Probability that Biologic Therapy is Cost-Effective   **Overall EVPI per patient**

at £20,000 per QALY	11%	
at £30,000 per QALY	84%	£ 309

43

The results show an estimated discounted lifetime cost of nearly £58,000 on TNF inhibitor therapy versus around £21,000 on traditional DMARDs. The incremental cost of around £37,000 achieves an estimated discounted QALY gain of 1.5583 over a lifetime. The resulting cost per QALY gained of is £23,882. This is around the range which might be considered acceptable by NICE. The probabilistic sensitivity analyses confirm this, showing an 84% probability of being cost-effective at a £30,000 threshold.

Figure 11 Cost-Effectiveness Plane BaseCase1 – Current UK Practice as per BSRBR Registry (based on 100 models runs)
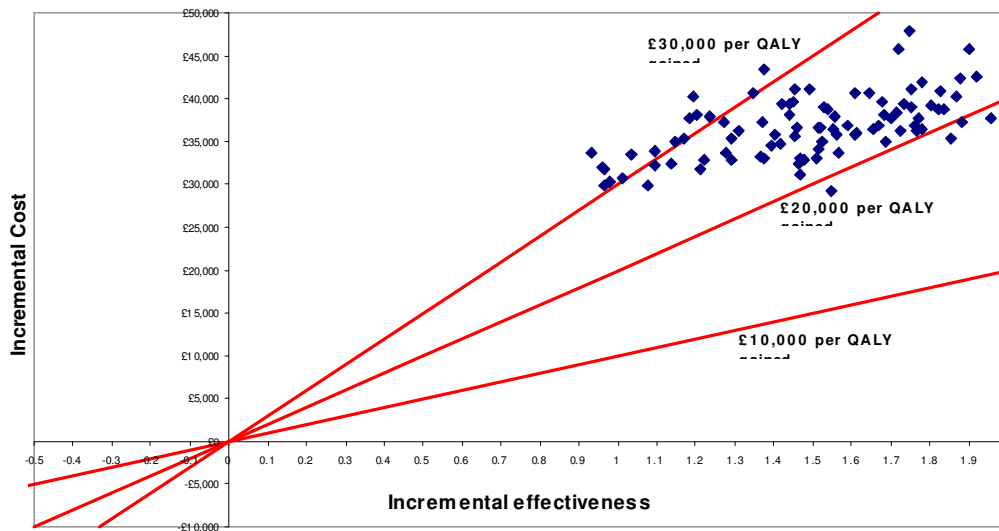
Figure 12nsitivity analyses confirm this, showing an 84% probability of being cost-effective at a £30,000 threshold.

Figure • SEQ Figure * ARABIC • 11• Cost-Effectiveness Plane BaseCase1 – Current UK Practice

3.2. * ARABIC • 11• Cost-Effectiveness Plane BaseCase1 –

0 models runs)

•

Figure • SEQ Figure * ARABIC at the rule to withdraw patients who are not achieving moderate DAS28 response at 3 months is applied. This of course means that the utility gains achieved by the poor responders are forgone, but also that the costs of ongoing treatment for this group are saved.

Table 19 Results Summary Table for BaseCase2 – Modelling the BSR Guidance – Withdrawal at 3 months unless Moderate DAS28 Response Is Achieved

Results Summary Table

	TNF inhibitor Therapy		Traditional DMARDs		Incremental
	Mean	S.E.	Mean	S.E.	Mean
Lifetime Cost	£ 53,884	£ 4,959	£ 20,880	£ 3,058	£ 33,004
Lifetime QALY	4.9634	0.5392	3.4885	0.6018	1.4749
Cost Per QALY gained					£ 22,378

Probability that Biologic Therapy is Cost-Effective		Overall EVPI per patient	
at £20,000 per QALY	20%		
at £30,000 per QALY	95%	£	120

The result is a discounted lifetime cost reduction of around £4,000 on TNF inhibitor therapy, when compared with Basecase1. The incremental cost of around £33,000 achieves an estimated discounted lifetime QALY gain of 1.4749 (0.0834 lower than basecase1). The resulting cost per QALY gained of £22,378 is 7% lower than that for basecase 1.

PSA results show a 95% probability of being below a £30k threshold.

Figure 13 Cost-Effectiveness Plane BaseCase2 – Modelling the BSR Guidance – Withdrawal at 3 months unless Moderate DAS28 Response Is achieved (based on 100 models runs)
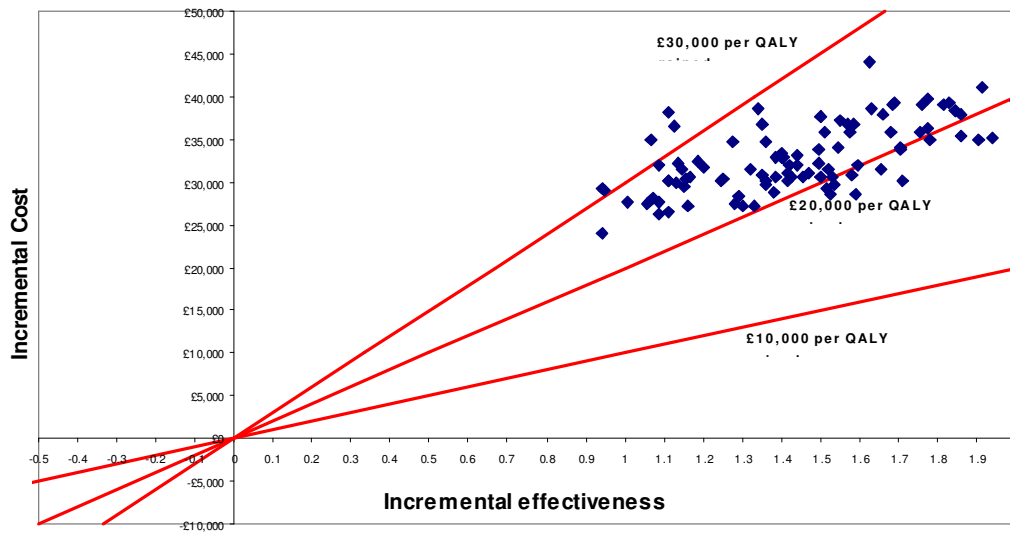
Figure 14 Cost-Effectiveness Acceptability Curve for BaseCase2 – Modelling the BSR Guidance – Withdrawal at 3 months unless Moderate DAS Response Is achieved (based on 100 models runs)
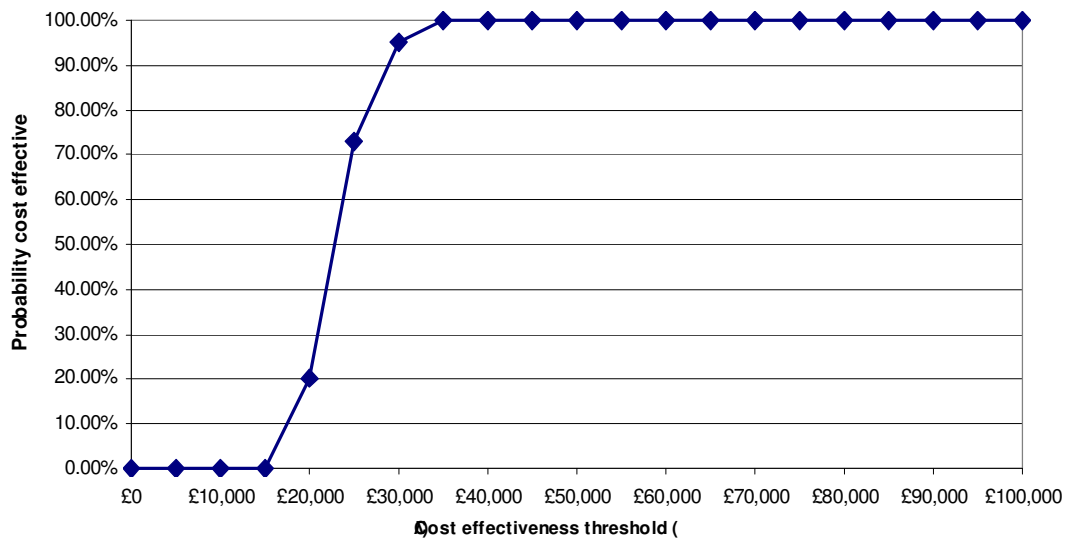
3.3.IMPACT OF DISCOUNTING ASSUMPTIONS (3½% COSTS, 3½% QALYS)

When NICE made the decision to recommend TNF-α treatments for RA, the mechanism for valuing long-term benefits and cost used a discount rate for costs of 6% and for benefits of 1.5%.

It has been agreed for the purposes of the 2005 NICE appraisal that these discount rates will again be used.

However, recent guidance suggests that discount rates of 3.5% should be used for both costs and benefits. This has important implications for treatments of chronic conditions where the system is 'buying' improvements in quality of life several years from now.

For example under the current assumptions (1.5% discounting) a year of full perfect health (i.e. 1 QALY) in 2015 would be 0.8533 discounted QALYs, whereas the suggested new assumptions on discount rates imply its value is just 0.6879 – a 20% reduction in the value of future benefits in 10 years. The effect accelerates over time so that in 20 years the equivalent reduction in the value of long-term health gains is 34%.

3.3.1. Impact of Discounting Assumptions On BaseCase1

Table 20: BaseCase 1 Sensitivity– Impact of Discounting Assumptions (3.5% costs, 3.5% QALYs)

Results Summary Table

	TNF inhibitor Therapy		Traditional DMARDs		Incremental
	Mean	S.E.	Mean	S.E.	Mean
Lifetime Cost	£ 72,545	£ 5,513	£ 31,199	£ 4,054	£ 41,346
Lifetime QALY	4.2843	0.4319	2.9927	0.5685	1.2916
Cost Per QALY gained					£ 32,013

Probability that Biologic Therapy is Cost-Effective		Overall EVPI per patient	
at £20,000 per QALY	0%		
at £30,000 per QALY	36%	£ 1,595	

The results of analysis using 3.5% and 3.5% discount rates, show a cost per QALY gained of around £32,000, suggesting that anti-TNF-α therapies would be considered on the borderline of cost-effective if a £30k willingness to pay for QALYs threshold were applied by decision makers. The probabilistic sensitivity analyses under these assumptions suggest a 36% probability of being below the £30k threshold .

Figure 15 Cost-Effectiveness Plane Sensitivity Analysis– Impact of Discounting Assumptions (based on 100 models runs)
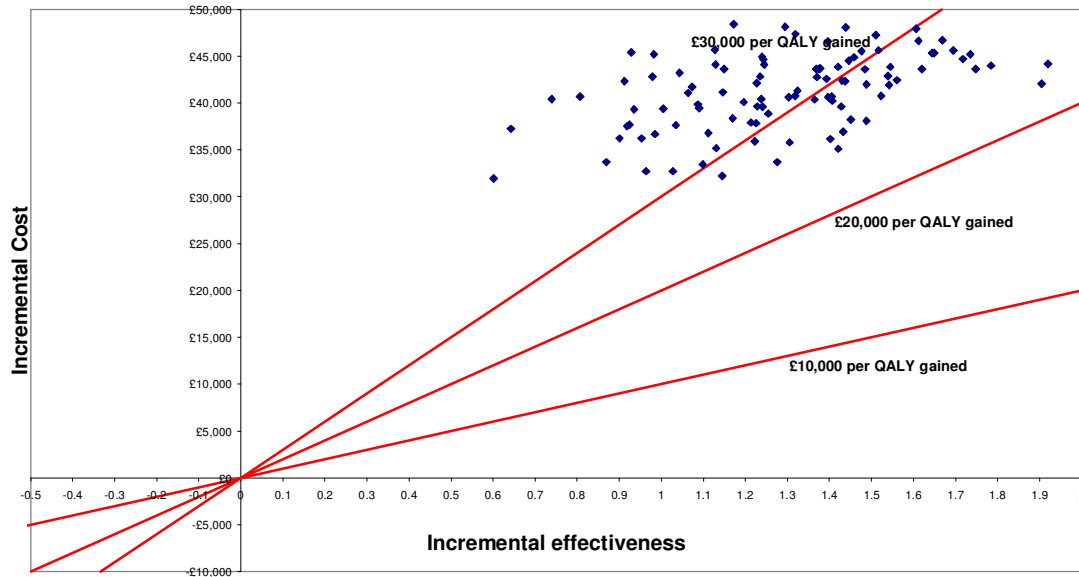
Figure 16 Cost-Effectiveness Acceptability Curve for Sensitivity Analysis– Impact of Discounting Assumptions (based on 100 models runs)
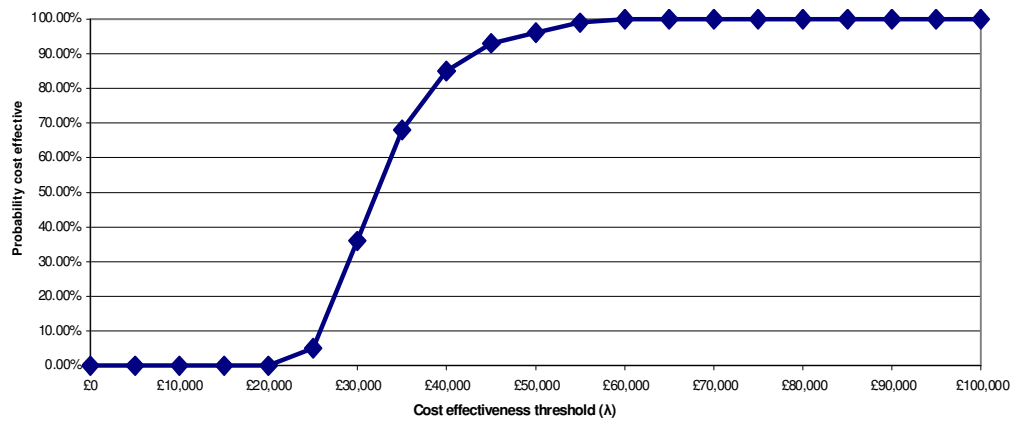
3.3.2. *Impact of Discounting Assumptions On BaseCase2*

The results of using 3.5% discount rates (for both costs and QALYs) in basecase 2 are similar.

Table 21 Basecase2 Sensitivity– Impact of Discounting Assumptions (6% costs, 1.5% QALYs)

Results Summary Table

	TNF inhibitor Therapy		Traditional DMARDs		Incremental
	Mean £	S.E. £	Mean £	S.E. £	Mean £
Lifetime Cost	68,075	6,403	30,443	5,163	37,632
Lifetime QALY	4.2906	0.4914	3.0838	0.5641	1.2068
Cost Per QALY gained					£ 31,184

Probability that Biologic Therapy is Cost-Effective		Overall EVPI per patient
at £20,000 per QALY	0%	
at £30,000 per QALY	43%	£ 1,834

Figure 17 Cost-Effectiveness Plane Sensitivity Analysis– Impact of Discounting Assumptions
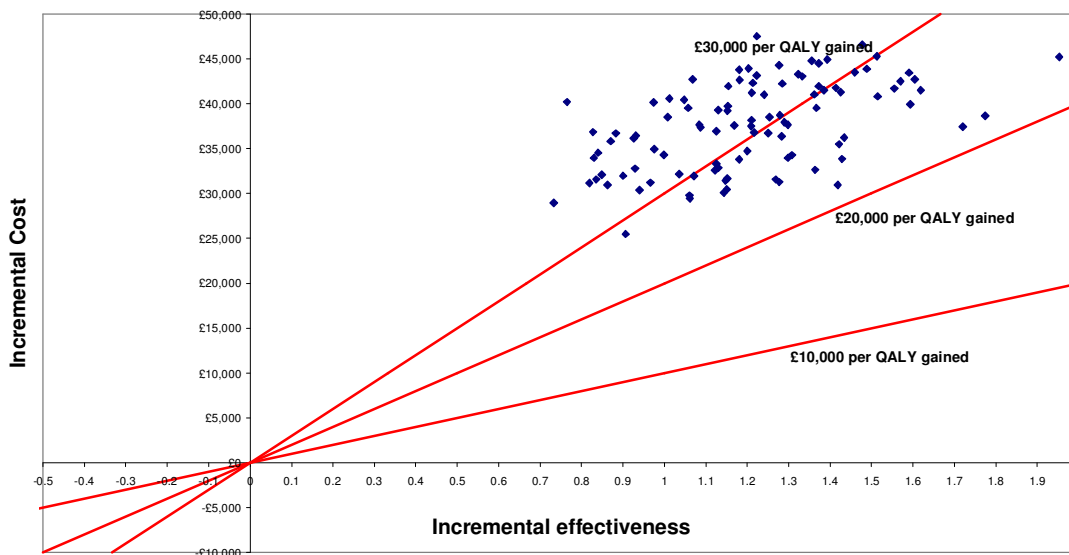
Figure 18 CEAC Analysis– Impact of Discounting Assumptions
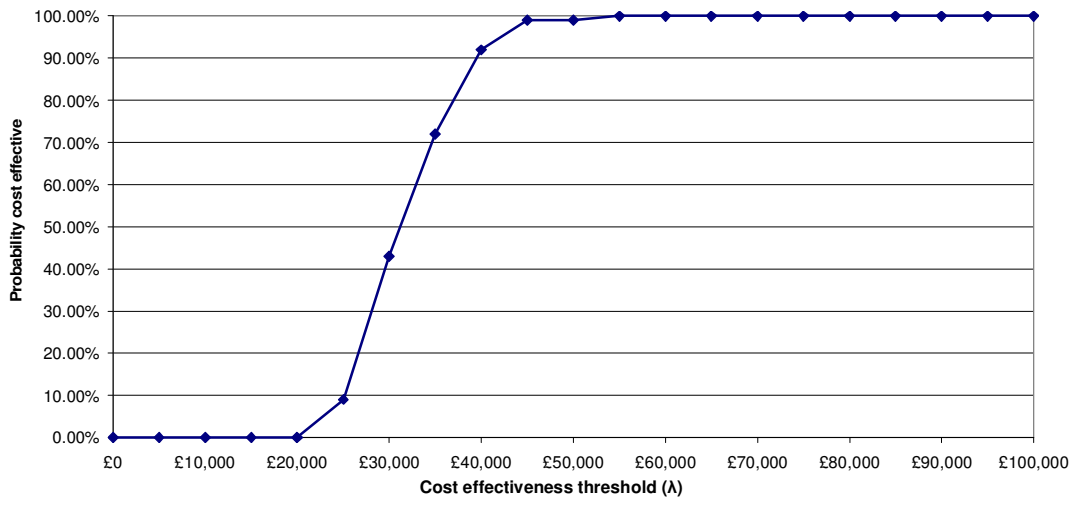
3.4. IMPACT OF USING SF-6D DERIVED UTILITY

The BSRBR data recorded quality of life data using the SF36 questionnaire. This can be translated into health state utilities via the SF6D algorithm as described in 2.1.5. However, as detailed earlier, this instrument suffers from a 'floor-effect' in patients with severe diseases, whereby patients below a certain level of disability are not well discriminated by the SF36 questionnaire. If decision makers were to choose to believe the validity of SF6D based results the effects would be substantial.

Table 22 BaseCase1 Sensitivity Analysis– Impact of Using SF36 Derived Utility

Results Summary Table

	TNF inhibitor Therapy		Traditional DMARDs		Incremental
	Mean	S.E.	Mean	S.E.	Mean
Lifetime Cost	£ 61,382	£ 4,075	£ 22,978	£ 3,162	£ 38,405
Lifetime QALY	8.6342	0.6899	7.8375	0.6637	0.7967
Cost Per QALY gained					£ 48,206

Probability that Biologic Therapy is Cost-Effective		**Overall EVPI per patient**
at £20,000 per QALY	0%	
at £30,000 per QALY	0%	£ -

The incremental QALY gained by TNF inhibitor therapy is almost halved if the SF6D derived utility is used rather than the EQ5D instrument. The result is a cost per QALY that is around double the basecase.

Figure 19 Cost-Effectiveness Plane Sensitivity Analysis– Impact of SF6D Utility on BaseCase1
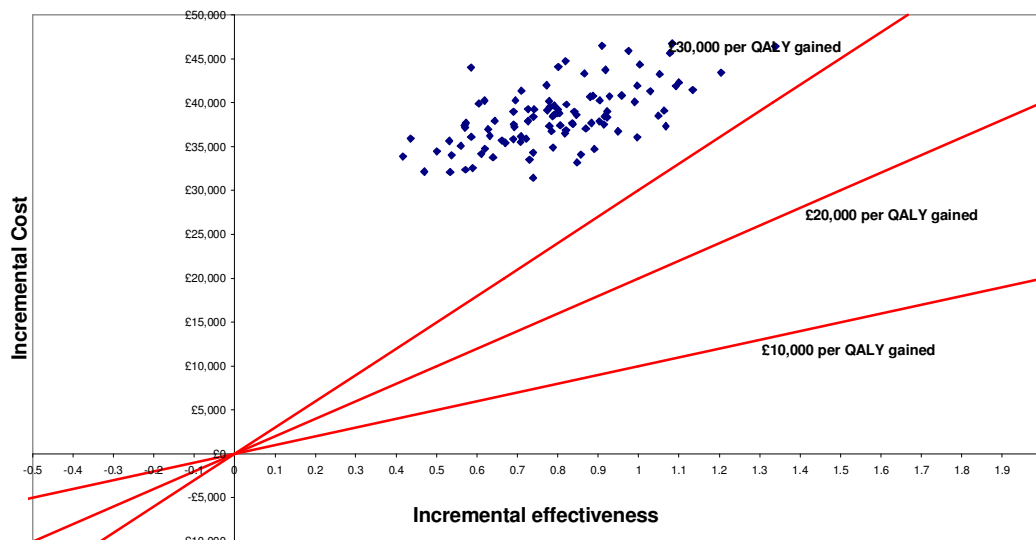
Figure 20 Cost-Effectiveness Acceptability Curve for Sensitivity Analysis– Impact of Using SF36 Derived Utility (based on 100 models runs)
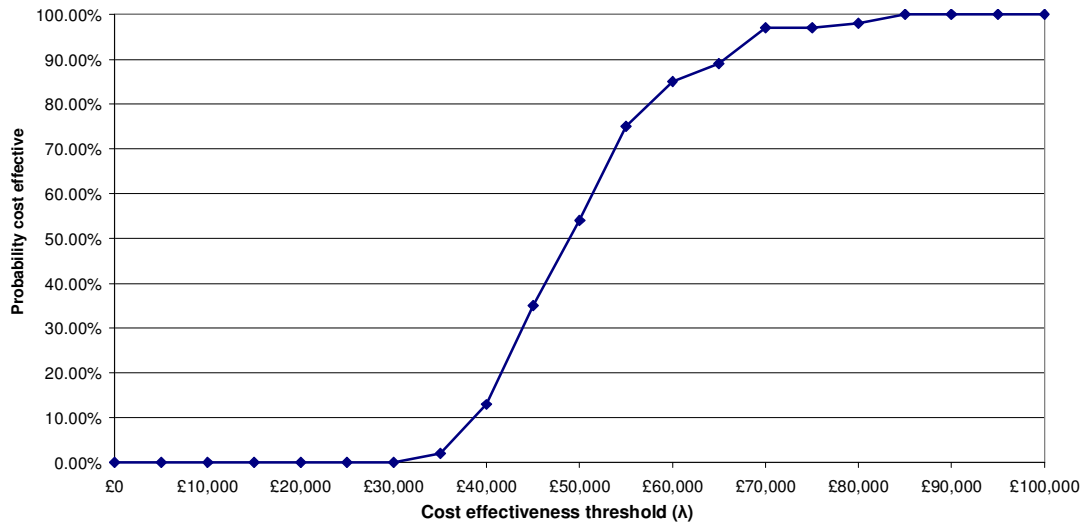
3.5. IMPACT OF DISABILITY PROGRESSION RATE WHILST ON DMARDS

The basecase analysis assumes that disability progression whilst on traditional DMARDs is best estimated by the weighted average of the studies examined by Scott et al. (Table 9). The resulting progression rate assumed was 0.0418 HAQ points per annum. The individual studies examined by Scott et al. show some substantial differences.

The question arises as to what is the relevant rate for patients who currently receive TNF inhibitor therapy in the UK i.e. what would their progression be in the absence of TNF inhibitor therapy. Unfortunately there are too few patients followed for too short a time in the BSRBR control arm to provide any useful evidence for this long-term progression.

Previous analyses examined the Early Rheumatoid Arthritis Study data to make some estimates of this (Brennan et al.) In sensitivity analyses here, we examine the impact of alternative assumptions on HAQ progression for the control arm and the TNF inhibitor arm after TNF inhibitor withdrawal.

We must note too that, in the absence of data to undertake covariate adjustment, the progression rate is assumed equivalent for every patient subgroup within our model. Thus patients who have failed several DMARDs even whilst quite young are assumed to have the same rate of HAQ progression as those who have had only 2 or 3 DMARDs over more than 15 years.

3.5.1. Sensitivity On Disability Progression in Basecase1 - ERAS data on Patients who have failed 2 DMARDs (worsening = 0.07 HAQ points per annum)

Table 23 Basecase1 Sensitivity Analysis– Impact of Using ERAS data on Patients who have failed 2 DMARDs (worsening = 0.07 per annum)

Results Summary Table

	TNF inhibitor Therapy		Traditional DMARDs		Incremental
	Mean	S.E.	Mean	S.E.	Mean
Lifetime Cost	£ 57,695	£ 3,954	£ 20,898	£ 2,927	£ 36,797
Lifetime QALY	4.4602	0.5731	2.4752	0.6452	1.9850
Cost Per QALY gained					£ 18,537

Probability that Biologic Therapy is Cost-Effective		Overall EVPI per patient	
at £20,000 per QALY	68%		
at £30,000 per QALY	100%	£	-

If a HAQ progression rate on traditional DMARDs were 0.07 points per annum then the cost per QALY gained by TNF inhibitor therapies would be around £18,500.

The PSA analyses would then suggest a 100% probability of being under the £30k threshold)

Figure 21 Cost-Effectiveness Plane Sensitivity Analysis– Impact of 0.07 pa HAQ progression assumption on Basecase1
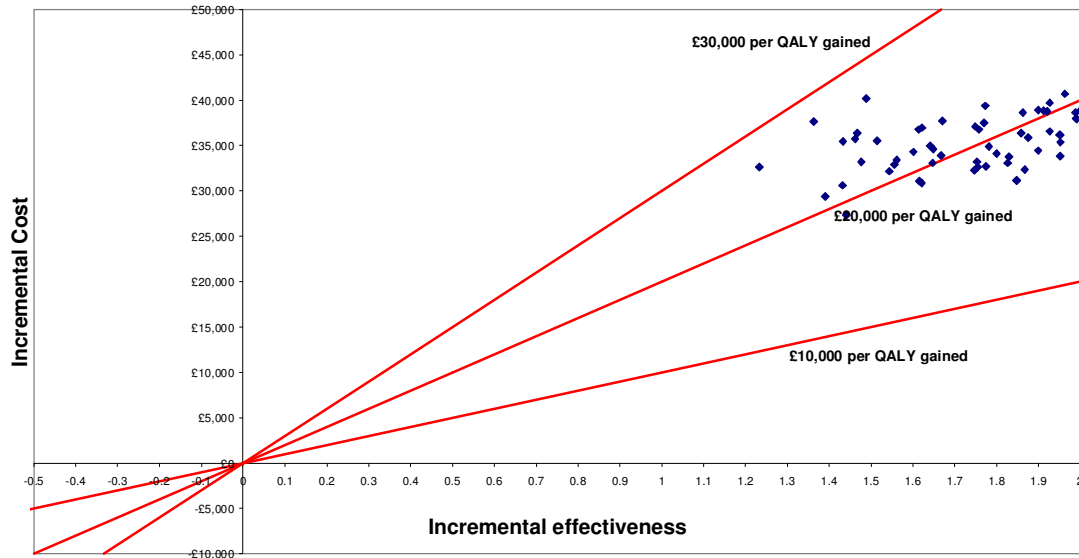
Figure 22 Cost-Effectiveness Acceptability Curve for Sensitivity Analysis– Impact of Using ERAS data on Patients who have failed 2 DMARDs (worsening = 0.07 per annum)
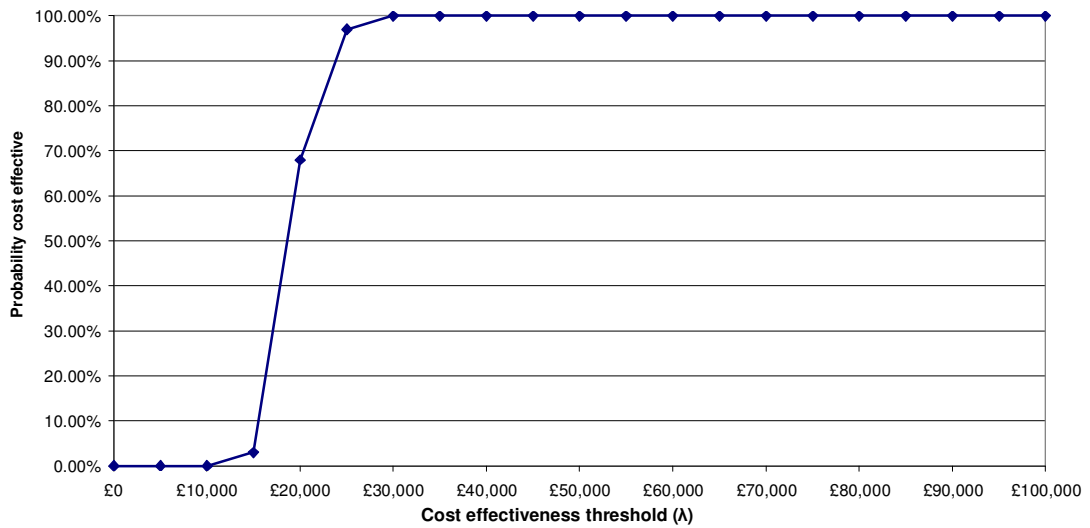
3.5.2. Sensitivity On Disability Progression in Basecase1 - ERAS data on Patients in Functional Class III/IV (worsening = 0.13 HAQ points per annum)

Table 24 Basecase 1 Sensitivity– Impact of ERAS Functional Class III/IV (0.13 per annum)

Results Summary Table

	TNF inhibitor Therapy		Traditional DMARDs		Incremental
	Mean	S.E.	Mean	S.E.	Mean
Lifetime Cost	£ 58,118	£ 4,243	£ 21,384	£ 2,990	£ 36,734
Lifetime QALY	3.2364	0.6397	0.3320	0.7059	2.9044
Cost Per QALY gained					£ 12,648

Probability that Biologic Therapy is Cost-Effective		Overall EVPI per patient
at £20,000 per QALY	100%	
at £30,000 per QALY	100%	£ -

The results if patients were progressing at a very high rate of 0.13 HAQ points per annum would be around £12,500 per QALY.

In such circumstances, the PSA suggests 100% probability of cost-effectiveness at both £30,000 and £20,000 thresholds.

Figure 23 C-E Plane– Impact of 0.13 pa HAQ progression on BaseCase1 (note rescaled axes)
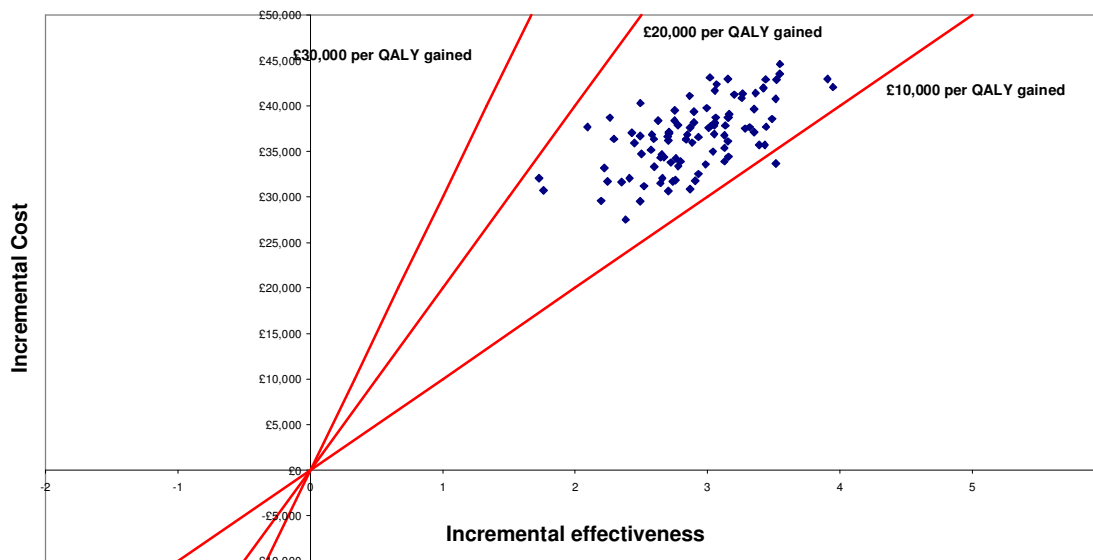
Figure 24 CEAC– Impact of Using ERAS Functional Class III/IV (worsening = 0.13 per annum)
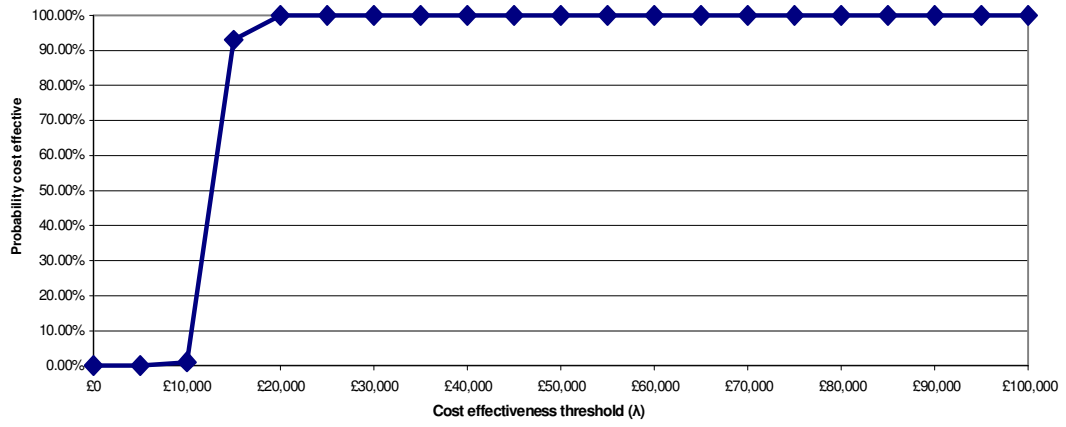
3.5.3. *Sensitivity On Disability Progression in Basecase2*

The equivalent sensitivity analyses using alternative data for disability progression on DMARDs from the ERAS database give slightly lower cost per QALY when applied to Basecase2.

Table 25 Basecase 2 Sensitivity– Impact of Using ERAS data on (a) Patients who have failed 2 DMARDs (worsening = 0.07 per annum) and (b) Functional Class III/IV (0.13 per annum)

Part a - Patients who have failed 2 DMARDs (worsening = 0.07 per annum)

Results Summary Table

	TNF inhibitor Therapy		Traditional DMARDs		Incremental
	Mean	S.E.	Mean	S.E.	Mean
	£	£	£	£	£
Lifetime Cost	53,131	4,686	20,971	2,708	32,160
Lifetime QALY	4.3928	0.6572	2.6122	0.6414	1.7807
Cost Per QALY gained					£ 18,061

Probability that Biologic Therapy is Cost-Effective		**Overall EVPI per patient**	
at £20,000 per QALY	76%		
at £30,000 per QALY	100%	£ -	

Part b - Functional Class III/IV (0.13 per annum)

Results Summary Table

	TNF inhibitor Therapy		Traditional DMARDs		Incremental
	Mean	S.E.	Mean	S.E.	Mean
	£	£	£	£	£
Lifetime Cost	54,426	4,974	21,463	3,269	32,963
Lifetime QALY	2.9544	0.7025	0.3086	0.8421	2.6458
Cost Per QALY gained					£ 12,459

Probability that Biologic Therapy is Cost-Effective		**Overall EVPI per patient**	
at £20,000 per QALY	100%		
at £30,000 per QALY	100%	£ -	

3.6. SUMMARY OF SENSITIVITY ANALYSES UNDERTAKEN ON BASECASE1

Table 26 sets out the results for all of the sensitivity analyses undertaken on basecase 1. The analyses discussed earlier on SF6D, HAQ progression and discounting are presented in the table as analyses 1b, 1c, 3 and 6.

3.6.1. Utility Relationship with HAQ (Analysis 2a)

Previous work undertaken by Birmingham assessing cost effectiveness of TNF inhibitors used an assumption that utility was related to HAQ disability score with a linear slope of –0,3. If this assumption is used the cost per QALY is around £20,000 (a 16% reduction as compared with the basecase).

3.6.2. Two TNF inhibitors in sequence (Analysis 5)

When we have analysed the use of 2 TNF inhibitors in sequence, making the assumption that the probability of response and utility gained following a switch to a second TNF inhibitor are able to be modelled using the same relationships as seen from the BSRBR data, then we find the results only marginally different from the basecase (cost per QALY increased by just 2%).

3.6.3. Withdrawal assumptions based on BSR guidance (Analyses 7,8)

Current guidance suggests that patients now achieving moderate response should withdraw at 3 months. If this policy (7a) were implemented it would reduce the overall costs of the overall lifetime costs by around 7%, whilst reducing the overall QALY gain by around 4%. This results in a cost per QALY of around £22,000 (7% lower than the basecase).

Withdrawal at 3 months if good response is not achieved (7b) has much more substantial effect on cost reductions but also reduces the QALY gain and produces a cost per QALY of the same order of around £23,000 per QALY.

The implementation of the stopping rules at 3 or 6 months as set out above, does not have as big an impact on cost effectiveness as the authors expected *a priori*. The main reason for this is that there is some utility gain obtained for those persons who do not achieve moderate response i.e., even patients achieving poor response on average improved their utility (see table 12d in appendix 1, section 6.1).

Sensitivity Analysis 8, which examines a rule to withdraw any patient whose utility actually worsens make little difference to the results.

Table 26 Summary of All Sensitivity Analyses Undertaken On BaseCase1

		TNF inhibitor Cost	DMARD Cost	TNF inhibitor QALY	DMARD QALY	Incremental Cost	Incremental QALY	Cost Per QALY gained	%Diff in CPerQ from Basecase	%<20k	%<30k	Overall EVPI per person at £30k
	BaseCase1	£ 57,919	£ 20,706	5.1514	3.5931	£ 37,214	1.5583	£ 23,882	0%	11%	84%	£ 309
	Sensitivity Analysis	<-------------% Difference from BaseCase1------------------>						<-------------Result Summary--------------->				
1c	HAQ 0.13pa	0%	3%	-37%	-91%	-1%	86%	£ 12,648	-47%	100%	100%	£-
1b	HAQ 0.07pa	-0%	1%	-13%	-31%	-1%	27%	£ 18,537	-22%	68%	100%	£-
2a	Utility per HAQ -0.3	1%	2%	-7%	-18%	0%	19%	£ 20,016	-16%	48%	99%	£ 8
2b	Utility per HAQ -0.1	-0%	-3%	8%	19%	1%	-18%	£ 29,503	24%	0%	59%	£2,032
3	SF36	6%	11%	68%	118%	3%	-49%	£ 48,206	102%	0%	0%	£-
4	Age/DisDur Increase	1%	2%	-3%	0%	0%	-9%	£ 26,428	11%	2%	75%	£ 818
5	2 TNF inhibitors In Sequence	11%	3%	5%	1%	16%	14%	£ 24,328	2%	9%	88%	£ 462
6	Discount (3.5%,3.5%)	25%	51%	-17%	-17%	11%	-17%	£ 32,013	34%	0%	36%	£1,595
7a	Withdraw at 3m if not moderate	-7%	1%	-3%	-3%	-11%	-4%	£ 22,203	-7%	20%	97%	£ 106
7b	Withdraw at 6m if not moderate	-7%	1%	-2%	0%	-11%	-8%	£ 23,084	-3%	14%	93%	£ 148
7c	Withdraw at 3m if not good	-39%	1%	-19%	-1%	-61%	-59%	£ 22,476	-6%	26%	87%	£ 333
7d	Withdraw at 6m if not good	-39%	2%	-19%	-2%	-61%	-59%	£ 22,316	-7%	30%	85%	£ 348
8	Withdraw if Utility Worsens	-2%	2%	3%	6%	-5%	-4%	£ 23,824	-0%	10%	88%	£ 347
9a	Baseline HAQ = [0.0 to 0.5]	2%	-14%	-35%	-99%	11%	112%	£ 12,524	-48%	100%	100%	£-
9b	Baseline HAQ = [0.5 to 1.0]	3%	-6%	82%	121%	8%	-7%	£ 27,863	17%	0%	66%	£1,391
9c	Baseline HAQ = [1.0 to 1.5]	4%	2%	62%	90%	6%	-4%	£ 26,306	10%	3%	74%	£ 816
9d	Baseline HAQ = [1.5 to 2.0]	1%	3%	-6%	-11%	-1%	3%	£ 22,962	-4%	22%	93%	£ 317
9e	Baseline HAQ = [2.0 to 1.5]	3%	1%	41%	56%	4%	6%	£ 23,433	-2%	14%	93%	£ 208
9f	Baseline HAQ = [2.5 to 3.0]	-4%	1%	-47%	-65%	-7%	-6%	£ 23,695	-1%	8%	92%	£ 148
10a	Age <40	9%	5%	68%	71%	11%	63%	£ 16,252	-32%	94%	100%	£-
10b	Age 40 to 50	8%	6%	23%	18%	9%	36%	£ 19,057	-20%	67%	99%	£ 6
10c	Age 50 to 60	2%	1%	3%	3%	3%	4%	£ 23,444	-2%	18%	91%	£ 350
10d	Age 60 to 70	-6%	-2%	-32%	-35%	-9%	-25%	£ 29,147	22%	0%	56%	£1,702
10e	Age 70+	-19%	-9%	-58%	-61%	-25%	-51%	£ 36,312	52%	0%	19%	£ 598

61

11a	Disease Duration 0 to 5 yrs	0%	-7%	29%	33%	4%	21%	£ 20,581	-14%	42%	100%	£-
11b	Disease Duration 5 to 10 yrs	-1%	-1%	13%	15%	-0%	7%	£ 22,207	-7%	20%	95%	£ 167
11c	Disease Duration 10 to 15 yrs	2%	2%	3%	1%	2%	8%	£ 22,633	-5%	15%	93%	£ 224
11d	Disease Duration 15+ yrs	-1%	-0%	-15%	-19%	-1%	-6%	£ 25,310	6%	9%	79%	£ 758
12a	Previous DMARDs <2	-0%	-8%	27%	26%	4%	31%	£ 18,949	-21%	67%	100%	£-
12b	Previous DMARDs <3	1%	-9%	30%	35%	6%	18%	£ 21,585	-10%	29%	99%	£ 26
12c	Previous DMARDs <4	-0%	-6%	17%	20%	2%	8%	£ 22,585	-5%	17%	95%	£ 113
12d	Previous DMARDs <5	-3%	-4%	-6%	-7%	-2%	-4%	£ 24,438	2%	8%	86%	£ 634
12e	Previous DMARDs 5+	2%	9%	-10%	-13%	-2%	-5%	£ 24,628	3%	5%	90%	£ 300
13a	Gender = Male	-5%	4%	-16%	-16%	-10%	-17%	£ 25,998	9%	1%	80%	£ 603
13b	Gender = Female	-0%	-0%	-1%	-3%	-0%	3%	£ 23,193	-3%	16%	89%	£ 439
14a	TNF inhibitor + Combination DMARD	6%	4%	-1%	1%	7%	-5%	£ 27,111	14%	1%	76%	£1,051
14b	TNF inhibitor Monotherapy	-0%	3%	1%	2%	-2%	1%	£ 23,280	-3%	12%	91%	£ 291
15a	Dose as per BSRBR reported	18%	-2%	-1%	-3%	28%	2%	£ 29,999	26%	0%	53%	£3,215
15b	Dose without Dose Creep	18%	2%	0%	1%	27%	-0%	£ 30,263	27%	0%	47%	£2,712
16a	Use BSR Utility prog'n long-tem	0%	1%	-23%	-3%	-1%	-71%	£ 82,471	245%	0%	3%	£ 141
16b	Use BSR Utility prog'n till 18m	1%	2%	-7%	-5%	1%	-12%	£ 27,477	15%	3%	69%	£1,352

3.6.4. Subgroup analysis by baseline HAQ score (Analysis 9)

Analyses 9a to 9f show that patients with a higher baseline HAQ score on average achieve slightly better cost effectiveness (i.e. lower cost per QALY gain) . For example the baseline HAQ subgroup [2.5 to 3.0] has a cost per QALY of approximately £24,000 whilst the group baseline HAQ = [0.5 to 1.0] has a cost per QALY of around£27,000. NOTE the results for baseline HAQ [0.0 to 0.5] (9a) should be ignored as it is based on less than 5 individual patients in the database.

3.6.5. Age subgroups (Analysis 10)

The age subgroup analysis suggests that it is significantly more cost effective to treat patients at a younger age. For example the cost per QALY for patients aged 70+ is over £36,000 whereas the cost per QALY for patients under 40 is around£16,000. This is because patients at a younger age live for a longer time and therefore have more time to accrue the long-term benefits of TNF inhibitor therapy in terms of delayed disease progression.

3.6.6. Disease duration sub groups. (Analysis 11)

Treating patients with shorter disease duration appears slightly more cost effective than treating those with longer disease duration. The cost per QALY for patients in the subgroup disease duration 0 to 5 years is around£20,000. Whereas for a disease duration of over 15 years, the cost per QALY is around£25,000. This is probably because disease duration is partly correlated with age.

3.6.7. Sub group analysis based on previous DMARD use (Analysis 12).

The analyses show that the number of previous DMARDS makes a marginal difference and that treating patients with fewer DMARDS is slightly more cost effective than those who have previously used 5+ DMARDS. Again this is probably because the probability of improvement and the level of initial improvement is slightly lower as the number of previous DMARDS increases. Again patients with fewer DMARDS will also on average have slightly lower age.

3.6.8. Male versus female (Analysis 13)

Subgroups based on gender alone make very little different to the results with males marginally less cost effective than females.

NOTE, one caution on this analysis is that we have assumed, at this stage, an average weight of 70kg for each patient. For infliximab based therapy costs are proportional to

weight and therefore larger differences may occur between males and females were this to be taken into account.

3.6.9.   Combination or monotherapy (Analysis 14)

TNF inhibitor plus combination DMARD therapy patients has a slightly higher cost per QALY (around£27,000) as compared with those on monotherapy (around£23,000).

As discussed in section 2.10.1 on methods, there are some important caveats around the reported dose data in the BSRBR data. If we make a worse case assumption that these doses are all correct then the cost per QALY would increase substantially to around£30,000.

3.6.10. Utility progression from 6 months onward (Analysis 16)

In the basecase analysis we have assumed that successful responders to therapy maintain their level of utility until such time as they withdraw from therapy, i.e. the rate of utility progression from 6 months onwards is 0.   Some BSRBR data does exist on utility progression. There are several caveats around it, including in particular, that it is relatively short term with not many patients followed beyond 18 months.

At this stage we have analysed the data using one simple covariate, (DAS28 response level achieved at 6 months) rather that the more complex set of covariates set out in other statistical analyses. The results suggest that patients achieving good response by 6 months have a marginal worsening of utility over 6 to 18 months. Moderate DAS28 responders have a utility progression of approximately 0 and the poor DAS28 responders actually have a continued utility gain. If we utilise these data over 6 to 18 months and then apply an assumption that patients have zero utility progression from then onwards, the cost per QALY would be approximately 15% higher at around £27,500 (analysis 16b).

If we were to assume that these data actually applied long term then the cost effectiveness would be around £80,000. This latter analysis is nonsensical because it soon becomes apparent that, after a number of years, patients who have had initially poor response have had a much higher level of utility (because they are assumed to continue with a linear upwards progression) than those who initially had a very good response who are progressing marginally downwards.

3.6.11. Cautions upon all the sensitivity analyses in basecase 1

It must be remembered in interpreting these analyses that the absolute result is less important than the change from baseline.   For example, if the true level of HAQ progression on DMARD therapy were 0.07 per annum, then we would re-base the

basecase to around £18,500 per QALY and all of the sensitivity analyses would occur around this basecase.

It should also be noted that:
  a. we have modelled TNF inhibitors as a class and not disentangled Etanercept from Infliximab or Adalimumab.
  b. We have assumed an average weight per patient of 70kg.
  c. All of the assumptions set out in the methods section apply.

3.7. SENSITIVITY ANALYSES ON BASECASE2

A similar set of sensitivity analyses have been undertaken on basecase 2 and are reported below.

The resulting effects are all of the same order as those shown for basecase 1.

Table 27 Summary of All Sensitivity Analyses Undertaken On Basecase 2

		TNF inhibitor Cost	DMARD Cost	TNF inhibitor QALY	DMARD QALY	Incremental Cost	Incremental QALY	Cost Per QALY gained	%Diff in Cost Per QALY from BaseCase	%<20k	%<30k	Overall EVPI per person at £30k
	BaseCase2	£ 53,884	£ 20,880	4.9634	3.4885	£ 33,004	1.4749	£ 22,378	0%	20%	95%	£ 120
	Sensitivity Analysis	<------------------Difference from BaseCase1---------------------->						<------------------Result Summary------------------>				
1c	HAQ 0.13pa	1%	3%	-40%	-91%	-0%	79%	£ 12,459	-44%	100%	100%	£-
1b	HAQ 0.07pa	-1%	0%	-11%	-25%	-3%	21%	£ 18,061	-19%	76%	100%	£-
2a	Utility per HAQ -0.3	-0%	4%	-8%	-18%	-2%	15%	£ 19,026	-15%	62%	99%	£ 11
2b	Utility per HAQ -0.1	1%	-1%	11%	24%	2%	-17%	£ 27,680	24%	0%	70%	£1,370
3	SF36	7%	13%	73%	125%	3%	-48%	£ 44,174	97%	0%	0%	£-
4	Age/DisDur Increase	0%	-0%	-2%	1%	1%	-8%	£ 24,507	10%	9%	84%	£ 512
5	2 TNF inhibitors In Sequence	17%	1%	6%	-1%	27%	23%	£ 23,036	3%	18%	94%	£ 459
6	Discount (3.5%,3.5%)	26%	46%	-14%	-12%	14%	-18%	£ 31,184	39%	0%	43%	£1,834
7a	Withdraw at 3m if not moderate	0%	0%	0%	0%	0%	0%	£ 22,378	0%	20%	95%	£ 120
7b	Withdraw at 6m if not moderate	1%	-1%	3%	4%	2%	0%	£ 22,733	2%	23%	91%	£ 386
7c	Withdraw at 3m if not good	-32%	1%	-12%	5%	-53%	-52%	£ 21,610	-3%	34%	86%	£ 275
7d	Withdraw at 6m if not good	-34%	-0%	-18%	-4%	-56%	-52%	£ 20,670	-8%	43%	93%	£ 114
8	Withdraw if Utility Worsens	-2%	-1%	3%	7%	-3%	-6%	£ 23,266	4%	13%	94%	£ 212
9a	Baseline HAQ = [0.0 to 0.5]	5%	-18%	-38%	-103%	19%	117%	£ 12,283	-45%	100%	100%	£-
9b	Baseline HAQ = [0.5 to 1.0]	6%	-7%	87%	125%	13%	-3%	£ 26,174	17%	1%	77%	£ 887
9c	Baseline HAQ = [1.0 to 1.5]	6%	2%	66%	96%	9%	-6%	£ 25,872	16%	5%	77%	£ 894
9d	Baseline HAQ = [1.5 to 2.0]	2%	2%	-4%	-6%	1%	0%	£ 22,657	1%	21%	91%	£ 191
9e	Baseline HAQ = [2.0 to 1.5]	4%	-0%	45%	60%	7%	8%	£ 22,199	-1%	18%	98%	£ 83
9f	Baseline HAQ = [2.5 to 3.0]	-6%	-1%	-47%	-62%	-9%	-9%	£ 22,470	0%	17%	93%	£ 253
10a	Age <40	13%	2%	78%	82%	21%	68%	£ 16,056	-28%	95%	100%	£-
10b	Age 40 to 50	8%	6%	24%	22%	8%	29%	£ 18,839	-16%	72%	100%	£-
10c	Age 50 to 60	1%	1%	4%	7%	1%	-2%	£ 22,933	2%	20%	97%	£ 133
10d	Age 60 to 70	-6%	-4%	-27%	-28%	-7%	-26%	£ 28,133	26%	1%	69%	£1,259
10e	Age 70+	-23%	-10%	-58%	-60%	-31%	-54%	£ 33,589	50%	1%	26%	£ 632

11a	Disease Duration 0 to 5 yrs	2%	-6%	30%	34%	6%	21%	£ 19,613	-12%	54%	100%	£-
11b	Disease Duration 5 to 10 yrs	-0%	-2%	15%	20%	1%	4%	£ 21,672	-3%	32%	97%	£ 119
11c	Disease Duration 10 to 15 yrs	3%	5%	4%	4%	1%	5%	£ 21,566	-4%	34%	96%	£ 120
11d	Disease Duration 15+ yrs	0%	-0%	-12%	-14%	0%	-8%	£ 24,408	9%	11%	86%	£ 624
12a	Previous DMARDs <2	1%	-10%	32%	34%	8%	28%	£ 18,811	-16%	64%	100%	£-
12b	Previous DMARDs <3	1%	-9%	32%	42%	7%	11%	£ 21,750	-3%	28%	98%	£ 58
12c	Previous DMARDs <4	0%	-7%	18%	22%	5%	10%	£ 21,261	-5%	29%	99%	£ 2
12d	Previous DMARDs <5	-2%	-2%	-3%	-2%	-1%	-5%	£ 23,183	4%	17%	90%	£ 244
12e	Previous DMARDs 5+	1%	6%	-12%	-15%	-2%	-4%	£ 22,935	2%	12%	94%	£ 146
13a	Gender = Male	-7%	0%	-16%	-15%	-11%	-17%	£ 24,063	8%	10%	91%	£ 312
13b	Gender = Female	1%	-0%	3%	2%	2%	3%	£ 22,139	-1%	24%	97%	£ 23
14a	TNF inhibitor + Combination DMARD	6%	3%	0%	2%	8%	-3%	£ 24,879	11%	3%	88%	£ 191
14b	TNF inhibitor Monotherapy	-0%	-1%	3%	3%	0%	3%	£ 21,685	-3%	27%	96%	£ 127
15a	Dose as per BSRBR reported	18%	1%	0%	0%	28%	-0%	£ 28,838	29%	0%	66%	£1,926
15b	Dose without Dose Creep	20%	1%	-1%	-3%	32%	5%	£ 27,973	25%	1%	67%	£1,301
16a	Use BSR Utility prog'n long-tem	0%	1%	-25%	5%	-0%	-93%	£ 335,680	1400%	0%	0%	£-
16b	Use BSR Utility prog'n till 18m	2%	3%	-2%	5%	1%	-20%	£ 28,143	26%	3%	65%	£1,605

4. DISCUSSION

4.1. CONCLUSIONS

A cost-effectiveness model based on data from the British Society for Rheumatology Biologics Registry Modelling has been constructed.

The objective of the study was to assess the cost effectiveness of anti-TNF-α inhibitors in the management of RA in the UK and in particular analyse:

- What is the incremental cost utility (cost per QALY) of TNF-α inhibitors according to current practice in the UK versus use of traditional DMARDs only?
- What would be the cost utility if guidance that only patients achieving moderate or good EULAR Disease Activity Score (DAS28) response are allowed to continue TNF inhibitor therapy after 3 months?
- We also explore the use of a 2nd TNF antagonist in a sequence after the first has failed, based on a small amount of data
- What is the cost utility for subgroups based on Age, Sex, Disease duration, Number of previous DMARDs, Baseline HAQ disability score?
- What is the cost utility if we made alternative assumptions concerning interpretation of the evidence base, including those on HAQ disability progression on traditional DMARD, relationships between HAQ and utility, impact of delayed progression whilst on TNF inhibitors, use of sequential TNF inhibitors, and discounting rates?

It should be remembered that the conclusions presented here are dependent on the methodology and assumptions described in section 2 of this report.

1. The results of our analysis suggest that the cost-effectiveness of current practice appears around £24,000 per QALY. It is difficult to apply one basecase, so instead we present a number of scenarios. If the guidelines set out by NICE in their initial appraisal were strictly adhered to, and non-responders were withdrawn from therapy, this would reduce to around £22,000.

2. These numbers are within the region that NICE deemed cost effective in the previous appraisal ('the incremental cost-effectiveness ratio of these therapies (etanercept and infliximab) can be estimated to be in the region of £27,000 to £35,000 per QALY'. [NICE FAD, 4.2.5]

3. The assessment of cost effectiveness in the 2001 appraisal was made using rates of discounting set at 6% cost, 1½ QALY. These still apply in the 2005 appraisal but new recommendations are coming (3½ % cost, 3½ QALY). The impact of this on the longer-term benefits make a substantial difference to the discounted cost per QALY gained. If the suggested new discount rates were used then the cost effectiveness of the TNF inhibitors are estimated to be £31,000-32,000 per QALY.

4. Assumptions concerning the long-term disease progression on traditional DMARDs (i.e. the control arm) also make a substantial difference to cost-effectiveness. The basecase analysis (Scott et al. data for the average UK progression) may be under-estimate. Treatment of patients with higher rates of progression is much more cost-effective.

5. If SF6D data were held to be a valid measure of quality of life improvements in severely disabled RA patients, then TNF inhibitors are estimated to have a much higher cost-effectiveness (almost £50,000 per QALY). The authors view is that EQ5D (used in the basecase) is a more sensitive and reliable measure for patients with severe RA.

6. Other sensitivity analyses show that patients who are younger (and have more lifetime in which to benefit from improved disability), have higher baseline disability, and fewer previous DMARDs appear slightly more cost-effective.

7. Sequential therapy with 2 TNF inhibitors appears to have the same order of cost-effectiveness as single therapy but the analysis undertaken assumes (in the absence of evidence on correlation) that response to a 2nd TNF inhibitor is independent of response to the first.

4.2. LIMITATIONS, FURTHER ANALYSES AND POSSIBLE RESEARCH IMPLICATIONS

There are several caveats, limitations, and implications for possible further research, the most important of which are:

a) The analysis has attempted to account for any selection bias between the TNF inhibitor and control arms in the BSRBR data by adjusting many of the parameters based on the main covariates. There are sometimes limitations to this approach . Further analyses using propensity methods might be valuable .

b) The analysis views the TNF inhibitor therapies (infliximab, etanercept and adalimumab) as a class rather than disentangling each individual therapy. A number of potential selection biases, mainly down to availability of treatment over the first 4 years of use make fair adjustments difficult. With additional data, this may become possible.

c) The data on doses used in the BSRBR database require some caution in interpretation and our basecase analysis assumed standard recommended doses. A more detailed analysis of some fields collected in the BSRBR, in particular regarding the dosages of treatment would be useful. Sensitivity analysis demonstrated that this is a potentially sensitive variable to the cost effectiveness of treatment

d) Costings for the 'class' of TNF inhibitors were based on a weighted average of TNF inhibitor use in the BSRBR data. The standard cost for infliximab assumed a patient weight of 70kg. We did not include the cost of wastage. A survey of what happens to surplus infliximab might also be useful.

e) The worsening in quality of life utility over time, caused by the progression in radiographic damage and subsequent disability on patients taking both TNF inhibitors and traditional DMARDs are important parameters in the model. While the parameter in the model is applied equally to all patients in the control arm, different rates might effect the delayed progression and change the incremental cost effectiveness. An investigation of data sources, which might enable a covariate adjustment model for this parameter, would probably be worthwhile to populate the parameter for patients in DMARD therapy. This would require long-term data on patients using conventional DMARD therapy, with fields such as disease duration HAQ, age and sex recorded.

f) A similar covariate adjustment for duration of traditional DMARD therapy before withdrawal might also be useful. This would require similar data, but also the response to treatment.

g) To understand the long-term progression of HAQ and utility whilst on TNF inhibitor treatment, ongoing collection of HAQ-DI score will be vital for registry patients. This opposes current plans to stop collection after 5 years.

h) While the SF36 as collected by the BSRBR will be of some use for the evaluation of quality of life in patients taking TNF inhibitors, it is not the most appropriate instrument for obtaining health state utility values in severe RA. The introduction of the EQ5D (5 questions) might be of value from a health economic perspective.

i) Excluded from our analysis is any potential effect of TNF inhibitor therapies on mortality. There are studies, which show a significant association between HAQ improvement and reduced mortality risk. Again, evidence in the next few years from the BSRBR will allow for this examination.

j) We have not examined the costs, or the quality of life impact, of individual adverse events. More detailed analysis, particularly on chronic events would be helpful in future evaluations.
–

k) Detailed costing of hospitalisations are probably under-estimated because patients about to receive joint surgery for example may not have started TNF inhibitor therapy. Ongoing data collection on hospitalisations would also be valuable.

l) Any mortality reduction benefits, which might be attributable to TNF inhibitors, are excluded.

m) A number of important costs are excluded
    –   costs of institutionalisation due to disability are excluded
    –   costs of longer-term surgeries unless they are represented in the first 18 months BSRBR data
    –   costs to (or quality of life impact on) carers
    –   lost work productivity due to disability among patients of working age

n) The probabilistic sensitivity analysis approach is valid for most of the variables based on BSRBR data but for some other, particularly on costs of drug and monitoring, we have assumed standard cost. This may under-estimate the overall uncertainty somewhat.

5. REFERENCES

Bansback N, Brennan A, Ghatnekar O. The cost effectiveness of adalimumab in the treatment of moderate to severe rheumatoid arthritis patients in Sweden Ann Rheum Dis Published Online First November 18th 2004 doi: 10.1136/ard.2004.027565

Bansback N, Marra C, Tsuchiya A, Anis A, Guh D, Hammond T, Brazier J. Using the Health Assessment Questionnaire to estimate preference-based single indices in patients with Rheumatoid Arthritis. Submitted for publication. (appendix 3, section 6.3)

Bansback NJ, Regier DA, Ara R, Brennan A, Shojania K, Esdaile JM, Anis AH, Marra CA. An Overview of Economic Evaluations for Drugs Used in Rheumatoid Arthritis Focus on Tumour Necrosis Factor-α Antagonists. Drugs 2005; 65 (4): 473-496

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)

Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics 2002. 21(2), 271-292

Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Economics. 2004. 13(9) 873 – 884

Brennan A, Bansback NJ, Reynolds A, et al. Modelling the cost effectiveness of etanercept in adults with rheumatoid arthritis in the UK. Rheumatology 2004; 43: 62-72

Briggs A. Handling uncertainty in economic evaluation and presenting the results. pp172-214 in Drummond M. McGuire A. (Eds.) Economic evaluation in health care: merging theory with practice. OHE / Oxford University Press, Oxford 2001

British National Formulary. (www,bnf.org/bnf/) accessed February 2005.

British Society for Rheumatology .Guidelines for prescribing TNF-a blockers in adults with rheumatoid arthritis. 2001

British Society for Rheumatology. National Guidelines For the Monitoring of Second Line Drugs. 2000

Brooks R. EuroQol: the current state of play. Health Policy 1996;37(1):53-72.

Callahan LF. The burden of rheumatoid arthritis: facts and figures. J Rheumatol Suppl 1998; 53: 8-12

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7(27):iii-x, 1-173.

Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35(11):1095-108.

Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: The Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:789-793.

Government Actuary's Department. (www,gad.gov.uk) accessed December 2004

Kobelt G, Eberhardt K, Geborek P. TNF inhibitors in the treatment of rheumatoid arthritis in clinical practice: costs and outcomes in a follow up study of patients with RA treated with etanercept or infliximab in southern Sweden. Ann Rheum Dis 2004; 63: 4-10

Kobelt G, Jonsson L, Young A, et al. The cost-effectiveness of infliximab (Remicade•) in the treatment of rheumatoid arthritis in Sweden and the United Kingdom based on the ATTRACT study. Rheumatology 2003; 42: 326-35

Maetzel A, Tugwell P, Boers M, et al., on behalf of the OMERACT 6 Economics Research group. Economic evaluation of programs or interventions in the management of rheumatoid arthritis: defining a reference case. J Rheumatol 2003; 30: 891-6

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):1571-82

Marra CA, Esdaille JM, Guh D, Kopec JA, Brazier JE, Koehler BE, Chalmers A, Anis AH. (2004) A comparison of four indirect methods of assessing values in rheumatoid arthritis. Medical Care (in press).

McIntosh E. The cost of rheumatoid arthritis. Br J Rheumatol 1996;35:781–90

Michaud K, Messer J, Choi HK, et al. Direct medical costs and their predictors in patients with rheumatoid arthritis: a three-year study of 7527 patients. Arthritis Rheum 2003; 48:2750-62

National Institute for Clinical Excellence. Guidance on the use of etanercept and infliximab for the treatment of rheumatoid arthritis. Technology Appraisal Guidance – No. 36. March 2002

National Institute for Clinical Excellence. Guide to the Methods of Technology Appraisal. (2004). Available at www.nice.org.uk

Netten A, Curtis L, Unit Costs of Health and Social Care. Personal Social Services Research Unit, University of Kent, 2002

O'Hagan, A and Stevenson, MD (2005). Monte Carlo probabilistic sensitivity analysis for patient level simulation models. In preparation

Pincus T, Callahan LF. The 'side effects' of rheumatoid arthritis: joint destruction, disability, and early mortality. Br J Rheumatol 1993; 32 Suppl. 1: 28-37

Pincus T, Sokka T, Wolfe F. Premature mortality in patients with rheumatoid arthritis: evolving concepts. Arthritis Rheum 2001;44:1234–6.

Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, Scott D, Silman A. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology 2002; 41: 793

Ware J, Snow K, Kosinski M, Gandek B. SF-36 Health Survey: Manual and Interpretation Guide. Boston: The Health Institute, New England Medical Center, 1993.

Wolfe F, Michaud K, DeWitt EM. Why results of clinical trials and observational studies of antitumour necrosis factor (anti-TNF) therapy differ: methodological and interpretive issues. Ann. Rheum. Dis, Nov 2004; 63: ii13 - ii17

Wolfe F, Michaud K, Gefeller O, et al. Predicting mortality in patients with rheumatoid arthritis. Arthritis Rheum 2003; 48: 1530-42

Wong JB, Ramey DR, Singh G. Long-term morbidity, mortality, and economics of rheumatoid arthritis. Arthritis Rheum 2001; 44: 2746-9

Wong JB, Singh G, Kavanaugh A. Estimating the cost-effectiveness of 54 weeks of infliximab for rheumatoid arthritis. Am J Med 2002; 113: 400-8

Yelin E, Callahan LF, for the National Arthritis Data Work Group. The economic cost and social psychological impact of musculoskeletal conditions. Arthritis Rheum 1995; 38: 1351-6

Yelin E, Trupin L, Wong B, Rush S. The impact of functional status and change in functional status on mortality over 18 years among persons with rheumatoid arthritis. J Rheumatol. 2002 Sep;29(9):1851-7.

Yelin E, Wanke LA. An assessment of the annual and long-term direct costs of rheumatoid arthritis: the impact of poor function and functional decline. Arthritis Rheum 1999; 42: 1209-18

6. APPENDICES

6.1. APPENDIX 1 – BASIC DATA FOR VALIDATION

1. Baseline HAQ

	Baseline HAQ		Distribution of HAQ					
	n	Mean HAQ	0-.5	.5 to 1	1 to 1.5	1.5 to 2	2 to 2.5	2.5 to 3
DMARD	482	1.63	11.62	12.24	19.71	22.41	22.61	11.41
Biologics (all)	6165	2.1	1.41	3.44	10.82	26.44	36.79	21.1

2. Baseline EQ5D

	Baseline EQ5D		Distribution of EQ5D															
	n	Mean EQ5D	<-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
DMARD	432	0.46	0.93	0	0.23	0	0.93	2.31	6.02	6.94	9.72	7.41	14.35	16.9	13.89	13.66	6.71	0
Biologics (all)	5981	0.293	1.14	0.02	0.47	1.15	2.22	4.9	10.87	12.52	15.63	16.62	15.28	9.55	5.78	3.09	0.7	0.05

3. Baseline SF6D

	Baseline SF6D		Distribution of SF6D															
	n	Mean SF6D	<-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
DMARD	427	0.565	0	0	0	0	0	0	0	0	0	9.13	18.74	38.64	21.78	9.13	6.71	0
Biologics (all)	5827	0.486	0	0	0	0	0	0	0	0	0	22.45	33.33	34.58	7.6	1.49	0.7	0.05

4. Mean HAQ change from baseline to 6 months

	All patients		DAS good responders		DAS moderate responders		Not good or moderate		DAS good and mod combined	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
DMARD	30	-0.079	10	-0.425	12	-0.104	8	0.391	22	-0.25
Biologics (all)	1681	-0.33	367	-0.533	958	-0.327	356	-0.128	1325	-0.384

5. Mean EQ5D change from baseline to 6 months

	All patients		DAS good responders		DAS moderate responders		Not good or moderate		DAS good and mod combined	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
DMARD	31	-0.001	10	0.116	12	-0.002	9	-0.128	22	0.052
Biologics (all)	1617	0.144	352	0.192	922	0.157	343	0.062	1274	0.166

6. Mean SF6D change from baseline to 6 months

	All patients		DAS good responders		DAS moderate responders		Not good or moderate		DAS good and mod combined	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
DMARD	23	0.017	7	0.106	10	-0.016	6	-0.03	17	0.034
Biologics (all)	1599	0.082	351	0.117	910	0.086	338	0.036	1261	0.095

7a. Numbers remaining on first therapy

Time on treatment	0	6	12	18	24	30	36	42	48	54	60
DMARD arm *											
Biologics (all)	0	41.72	27.41	16.05	8.52	6.29	0	0	0	0	0

* Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.
Only looked at non-censored patients

7b. Kaplan Meier survival on first therapy

Time on treatment	0	6	12	18	24
DMARD arm *					
Biologics (all)	1	0.952	0.839	0.779	0.722

* Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.

7c. Mean HAQ for survivors on first therapy

Time on treatment	0	6	12	18	24
Biologics (all)	2.099	1.763	1.701	1.725	1.654

* Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.
  Only look at non-censored patients
  On therapy for at least 6 months

7d. Mean EQ5D for survivors on first therapy

Time on treatment	0	6	12	18	24
Biologics (all)	0.293	0.446	0.456	0.454	0.47

7e. Mean SF6D for survivors on first therapy

Time on treatment	0	6	12	18	24
Biologics (all)	0.486	0.574	0.583	0.581	0.594

8a. Numbers remaining on first therapy for good and moderate responders at 6 months

Time on treatment	0	6	12	18	24	30	36	42	48	54	60
DMARD arm *											
Biologics (all)	0	4.3	40.46	28.57	15.6	11.06	0	0	0	0	0

* Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.
  Only look at non-censored patients

8b. Kaplan Meier survival on first therapy for good and moderate responders at 6 months

Time on treatment	0	6	12	18	24
DMARD arm *					
Biologics (all)	1	0.957	0.895	0.841	0.786

8c Mean HAQ for survivors on first therapy for Good and Moderate responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	2.088	1.688	1.664	1.684	1.664

* Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.
  Only look at non-censored patients

8d Mean EQ5D for survivors on first therapy for Good and Moderate responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	0.3	0.475	0.471	0.463	0.463

8e Mean SF6D for survivors on first therapy for Good and Moderate responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	0.487	0.588	0.588	0.584	0.592

9a Numbers remaining on first therapy for poor responders at 6 months

Time on treatment	0	6	12	18	24	30	36	42	48	54	60
DMARD arm *											
Biologics (all)	0	17.28	49.54	19.45	8.7	5.03	0	0	0	0	0

* Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.
   Only look at non-censored patients

9b Kaplan Meier survival on first therapy for poor responders at 6 months

Time on treatment	0	6	12	18	24
DMARD arm *					
Biologics (all)	1	0.827	0.596	0.516	0.456

* Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.

9c Mean HAQ for survivors on first therapy for poor responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	2.22	2.081	1.979	1.991	1.854

* Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.
   Only look at non-censored patients

9d Mean EQ5D for survivors on first therapy for poor responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	0.26	0.335	0.357	0.382	0.364

9e Mean SF6D for survivors on first therapy for poor responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	0.474	0.514	0.542	0.552	0.537

10b Kaplan Meier survival on any biologic therapy

Time on treatment	0	6	12	18	24
DMARD arm					
Biologics (all)	1	0.998	0.933	0.907	0.882

10c Mean HAQ for survivors on any biologic therapy

Time on treatment	0	6	12	18	24
Biologics (all)	2.099	1.774	1.742	1.766	1.717

Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.
Only look at non-censored patients
On therapy for at least 6 months

10d Mean EQ5D for survivors on any biologic therapy

Time on treatment	0	6	12	18	24
Biologics (all)	0.293	0.44	0.442	0.442	0.449

10e Mean SF6D for survivors on any biologic therapy

Time on treatment	0	6	12	18	24
Biologics (all)	0.486	0.57	0.576	0.574	0.586

11b Kaplan Meier survival on any biologic therapy for Good and Moderate responders
at 6 months

Time on treatment	0	6	12	18	24
DMARD arm					
Biologics (all)	1	0.998	0.969	0.948	0.927

11c Mean HAQ for survivors on any biologic therapy for Good and Moderate responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	2.088	1.694	1.684	1.703	1.683

Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.
Only look at non-censored patients

11d Mean EQ5D for survivors on any biologic therapy for Good and Moderate responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	0.3	0.473	0.461	0.461	0.459

11e Mean SF6D for survivors on any biologic therapy for Good and Moderate responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	0.487	0.586	0.585	0.582	0.587

12b Kaplan Meier survival on any biologic therapy for poor responders at 6 months

Time on treatment	0	6	12	18	24
DMARD arm *					
Biologics (all)	1	0.998	0.855	0.813	0.769

* Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.

12c Mean HAQ for survivors on any biologic therapy for poor responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	2.22	2.067	2.017	2.047	1.961

Difficult to find for DMARDS as a patient can be on up to three DMARDS at a time.
Only look at non-censored patients

12d Mean EQ5D for survivors on any biologic therapy for poor responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	0.26	0.335	0.353	0.35	0.328

12e Mean SF6D for survivors on any biologic therapy for poor responders at 6 months

Time on treatment	0	6	12	18	24
Biologics (all)	0.474	0.513	0.532	0.531	0.556

13a Mean HAQ worsening after withdrawal

Time on treatment	0	6	12	18
Biologics (all)	2.148	2.128	2.092	NA

13a Mean HAQ worsening after withdrawal (number)

Time on treatment	0	6	12	18
Biologics (all)	137	74	15	0

13b Mean EQ5D worsening after withdrawal

Time on treatment	0	6	12	18
Biologics (all)	0.29	0.282	0.301	NA

13b Mean EQ5D worsening after withdrawal (number)

Time on treatment	0	6	12	18
Biologics (all)	141	76	15	0

13c Mean SF6D worsening after withdrawal

Time on treatment	0	6	12	18
Biologics (all)	0.515	0.51	0.486	NA

13c Mean SF6D worsening after withdrawal (number)

Time on treatment	0	6	12	18
Biologics (all)	135	69	15	0

14a Mean HAQ worsening after withdrawal for moderate and good responders

Time on treatment	0	6	12	18
Biologics (all)	2.089	2.106	2.25	NA

14a Mean HAQ worsening after withdrawal for moderate and good responders (number)

Time on treatment	0	6	12	18
Biologics (all)	52	27	4	0

14b Mean EQ5D worsening after withdrawal for moderate and good responders

Time on treatment	0	6	12	18
Biologics (all)	0.338	0.324	0.301	NA

14b Mean EQ5D worsening after withdrawal for moderate and good responders (number)

Time on treatment	0	6	12	18
Biologics (all)	53	28	4	0

14c Mean SF6D worsening after withdrawal for moderate and good responders

Time on treatment	0	6	12	18
Biologics (all)	0.543	0.521	0.425	NA

14c Mean SF6D worsening after withdrawal for moderate and good responders (number)

Time on treatment	0	6	12	18
Biologics (all)	51	27	4	0

15a Mean HAQ worsening after withdrawal for poor responders

Time on treatment	0	6	12	18
Biologics (all)	2.182	2.188	2.125	NA

15a Mean HAQ worsening after withdrawal for poor responders (number)

Time on treatment	0	6	12	18
Biologics (all)	59	32	6	0

15b Mean EQ5D worsening after withdrawal for poor responders

Time on treatment	0	6	12	18

| Biologics (all) | 0.272 | 0.257 | 0.334 | NA |

15b Mean EQ5D worsening after withdrawal for poor responders (number)

Time on treatment	0	6	12	18
Biologics (all)	62	33	6	0

15c Mean SF6D worsening after withdrawal for poor responders

Time on treatment	0	6	12	18
Biologics (all)	0.5	0.501	0.533	NA

15c Mean SF6D worsening after withdrawal for poor responders (number)

Time on treatment	0	6	12	18
Biologics (all)	0.272	0.257	0.334	NA

6.2. APPENDIX 2 PREVIOUS RESEARCH FUNDING CONFLICTS OF INTEREST STATEMENT

At the time of the analysis, Alan Brennan (AB) and Nick Bansback (NB) had received previous research funding from 3 companies for work in rheumatoid arthritis (see below). This work was completed and declared prior to applying with the ARC Epidemiology Unit to undertake analysis of the BSRBR data. AB and NB have also previously received research funding from 1 company for work on biologics in other indications (see below). Again, this work was completed and declared prior to applying with the ARC Epidemiology Unit to undertake analysis of the BSRBR data. AB and NB have received sponsorship to academic conferences from 2 companies

Colleagues in ScHARR are completing a separate analysis for 1 company.
Other ongoing work does not represent a conflict of interest.

AB and NB have completed the following projects in the area of biologics in RA:
1. Modelling cost effectiveness of etanercept in the UK. Funded by Wyeth. Project completed 2001.
2. Modelling cost effectiveness of adalimumab in 10 countries including the US. Funded by Abbott. Project completed June 2004
3. Cost effectiveness of a genetic test to detect responders to anakinra. funded by Interleukin Genetics. Completed 2002.

AB and NB has also been involved in projects concerning biologics in other indications:
4. Cost effectiveness of etanercept in the treatment for Psoriatic Arthritis. Funded by Wyeth. Completed

AB , NB and RN have two further projects related to RA.
5. Funded by United States Government (Agency for Healthcare Research and Quality) researching biologics in RA (to complete June 2005).
6. A methodology project examining methods for optimising clinical trial development decisions, using RA therapies as one case study. This is funded by a company, which does not have a biologic product in the RA market.

Other University of Sheffield staff
7. Cost effectiveness of etanercept in the treatment of Ankylosing spondylitis. Funded by Wyeth, completing 2005. (This project has not involved the staff working on the BSR registry analysis).

Richard Nixon:- declares no conflict of interest.

6.3. APPENDIX 3 – EQ5D MAPPING

In submission – not to be quoted, cited or distributed without the authors prior consent

Using the Health Assessment Questionnaire to estimate preference-based single indices in patients with Rheumatoid Arthritis

Nick Bansback [1,3]
Carlo Marra, [2]
Aki Tsuchiya [3]
Aslam Anis [1]
Daphne Guh [1]
Tony Hammond [4]
John Brazier [3]

Abstract

Objective. To estimate the preference-based measures, EQ-5D and SF-6D from the Health Assessment Questionnaire Disability Index (HAQ-DI) in patients with Rheumatoid Arthritis (RA), and to characterise components that are predictors of health state utility.

Methods. Patients participating in two studies in the UK (n=132) and Canada (n=310) with RA, were administered the HAQ, EQ-5D and the SF-36. The SF-36, a generic measure of quality of life was converted into the preference-based SF-6D. From these results we developed models of the relationship between the HAQ and SF-6D and EQ-5D using various regression analyses.

Results. The optimal model developed for the EQ-5D entered levels for each item as independent variables. A Root Mean Squared Error(RMSE) of 0.18 suggested a relatively good fit. For the SF-6D, RMSE were lower (equal to 0.09) suggesting better predictions than for EQ-5D, but models with more explanatory variables did not improve the results. For both measures, components of the reach dimension appeared the least predictive variables whilst the components of arising, eating, hygiene and activities appeared to be most important.

Conclusions. Our approach enabled calculations of quality adjusted life years (QALYs) from existing trials where only the HAQ has been measured. All of the aspects of the HAQ may not be reflected in the preference-based measures, and this method is suboptimal to direct measurement of health state utility in clinical trials. Given this limitation, our approach provides an alternative for researchers who need health-state utility values, but had not included a preference-based measure in their clinical study because of resource constraints or a desire to limit the patient burden.

Key Words: economics, utility theory, rheumatoid arthritis, Quality-Adjusted Life Years

[1] Centre for Health Evaluation and Outcome Sciences, St. Paul's Hospital, Vancouver, BC, Canada

[2] Centre for Clinical Epidemiology and Evaluation, Vancouver Coastal Research Institute, Vancouver, BC, Canada

[3] Health Economics and Decision Science, School of Health and Related Research

(ScHARR), The University of Sheffield, UK

[4] Rheumatology Department, Maidstone Hospital, Kent, UK.

Nothing to declare

NJ Bansback, MSc., Health Economist, Centre for Health Evaluation and Outcome Sciences,
Honorary Research Fellow, Health Economics and Decision Sciences, University of Sheffield,
UK
CA Marra, Pharm.D., Ph.D., Assistant Professor, Faculty of Pharmaceutical Sciences,
University of British Columbia, Head, Health Economics Program, Centre for Clinical
Epidemiology and Evaluation, Vancouver Coastal Research Institute, CANADA
A Tsuchiya, Ph.D., Lecturer in Health Economics, Health Economics and Decision Sciences,
University of Sheffield, UK
AH Anis, Ph.D., Associate Professor and Director, Master of Health Administration,
Department of Health Care and Epidemiology, University of British Columbia, Health
Economist, Centre for Health Evaluation and Outcome Sciences, CANADA
Daphne Guh, Statistician, Centre for Health Evaluation and Outcome Sciences, CANADA
A Hammond, MD, Consultant Rheumatologist, Maidstone Hospital, UK
JE Brazier, Ph.D., Professor of Health Economics, Health Economics and Decision Sciences,
University of Sheffield, UK

Corresponding author and address for reprints:

Nick Bansback
Centre for Health Evaluation and Outcome Sciences
St. Paul's Hospital
570-24 1081 Burrard Street
Vancouver, B.C. V6Z 1Y6

Email: nbansback@cheos.ubc.ca
Tel: (001) 604-682-2344 (63561)
Fax: (001) 604-806-8778

Running title: Estimating a preference-based single index from the HAQ

Introduction

Given the scarcity of health care resources, public and private agencies have become interested in both the effectiveness and cost-effectiveness of health care interventions.[1] The preferred approach towards measuring benefits in cost effectiveness analyses is to value health status in a single unit of measurement known as utilities which are used to derive "quality adjusted life years" (QALYs). Such cost utility analyses (CUA) are particularly informative for health policy decisions because they allow direct comparison of the efficiency of healthcare resource expenditure across a wide variety of conditions and treatments.[2] Utilities are obtained by asking patients to make judgments about the value of particular health states or outcomes. Preference-based instruments are formal methods for quantifying these judgments and can be obtained directly from patients, or from one of a number of generic measures valued by general population samples, such as the Health Utilities Index or EQ-5D.

Many clinical studies do not use a preference-based measure due to lack of resources or time, or because the commonly used generic preference-based measures are regarded as unsuitable for the condition.[3] In a majority of Rheumatoid Arthritis (RA) clinical trials the Health Assessment Questionnaire is the primary and often sole measure of quality of life.[4] While the HAQ was primarily designed to measure only aspects of physical function and pain, it has been shown to be highly correlated with many generic and disease-specific measures of health related quality of life.[5] Subsequently, linear transformations between HAQ and utility have previously been used in CUA.[6,7]. While other disease specific measures such as the RAQol are being developed, only more recent clinical trials have measured a preference based measure.[8]

As a result, there are many clinical trials whose results are not amenable to populating CUA. Estimating a relationship between the HAQ and a preference-based measure would make it possible to estimate QALY scores from existing clinical data where the HAQ has been measured but preference based instruments have not.[3,9] Moreover, in trials where one such

preference based instrument had been measured, it could also be possible to evaluate another. Such analyses have previously been attempted for outcomes in asthma and obesity.[9,10] This paper uses data from the UK and Canada to map two such preference based instruments, the EQ-5D and the SF-6D from the HAQ questionnaire.

The instruments

Health Assessment Questionnaire

The HAQ is a self-completed questionnaire, developed as a comprehensive measure of outcome in patients with a wide variety of rheumatic diseases, including rheumatoid arthritis, osteoarthritis, juvenile rheumatoid arthritis, lupus, scleroderma, ankylosing spondylitis, fibromyalgia, and psoriatic arthritis. Although its complete form includes an assessment of mortality, disability, pain and symptom levels, drug side-effects and resource utilization, most studies in practice only use the physical disability scale. This scale assesses upper and lower limb function in relation to the degree of difficulty encountered in performing daily living tasks. These tasks include walking, dressing, bathing and shopping. The HAQ contains 20 items distributed across eight components. The scores range from 0 (without any difficulty) to 3 (unable to do). The highest score on any item within one component represents the dimension score. The respondent also indicates whether he or she uses aids or devices (14 items) or help from other people (8 items). The scores for each dimension are corrected for the use of aids or devices, summated and transformed to give an overall disability index (HAQ-DI) between 0-3. A score of 0 represents no disability and 3 very severe, high-dependency, disability.[4]

EQ-5D

The EQ5D has five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has one item, and each item has three levels with 1 denoting no problems and 3 denoting extreme problems.[11] The number of theoretically

possible health states is $3^5 = 243$. EQ5D can be reported in terms of a 5-digit profile indicating the level on each dimension, or in terms of a preference-based single index number. The latter is obtained by applying algorithms that link the 5-digit health state description with average valuations obtained from members of the public using the time trade-off method, or the visual analogue scale. In this study, EQ5D indices are obtained using the so-called MVH A1 value set, derived from a population survey in the UK using 10-year time trade-offs.[12]

SF-6D

The SF-6D has been derived from the SF-36.[13] The SF-36 is a generic measure of health that generates scores across eight dimensions of health.[14] It has become one of the most widely used generic measures of health through out the world, but was it not originally designed for use in economic evaluation. A research team at the University of Sheffield in collaboration with Dr. Ware has estimated a preference-based single index measure of health from the SF-36.[13] The index is estimated via a health state classification called the SF-6D derived from the SF-36 and is composed of six multi-level dimensions of health. It was constructed from a sample of 11 items selected from the SF-36 to minimise the loss of descriptive information and defines 18,000 health states. A selection of 249 states defined by the SF-6D have been valued by a representative sample of the UK general population (n=611) using the standard gamble (SG) valuation technique. Like the EQ-5D, regression models were estimated to predict single index scores for all health states defined by the SF-6D. The resultant algorithm can be used to convert SF-36 data at the individual level to a preference-based index.

Table 1: Characteristics of quality of life instruments

Materials and Methods

Study populations

Participants from two locations were recruited. In Vancouver, Canada, 319 patients with a clinical diagnosis of Rheumatoid Arthritis were followed up quarterly over 3 time periods from eight private rheumatology offices. In Maidstone Hospital, UK, a single observation from 151 patients with a clinical diagnosis of Rheumatoid Arthritis was measured under routine management in a district hospital department of Rheumatology. All patients were administered the HAQ, SF-36 and EQ-5D. The two samples were pooled to create a single source (total of 925 observations) which we hoped would provide estimates more generalisable to North American and European populations(Table 2). The predictive ability of the estimates on the two individual cohorts is also studied.

Table 2: Summary Statistics of the two cohorts

Statistical analysis

For the primary analysis, the relationships between scores on the EQ-5D, SF-6D and the HAQ-DI were examined by fitting linear regression models estimated by generalised estimating equation (GEE) algorithms where the correlation matrix takes the structure of an auto regressive of order 1. The effect of pooling populations from two different countries was explored. The generalisability of the models is examined using 3-fold cross validation where the data is split into 3 subsets stratified by country. The following regression models were evaluated:

Model 1: EQ-5D & SF-6D indices regressed on the HAQ-DI score. This assumes that the 8 dimensions of the HAQ-DI carry equal weight; the 42 items within a given domain carry equal weight; and the intervals between response choices for each item are equal;

Model 2: EQ-5D & SF-6D indices regressed on the 8 HAQ-DI dimension scores, where the dimension is treated as continuous variable. This assumes that the 42 items of HAQ-DI carry equal weight within a given domain and the intervals between response choices for each item are equal;

Model 3: EQ-5D & SF-6D indices regressed on the 42 HAQ-DI item scores, where the responses to each item are treated as a continuous scale. This assumes that the intervals between response choices for each item are equal;

Model 4: EQ-5D & SF-6D indices regressed on the individual levels of the HAQ-DI dimension scores, where each level is entered as a dummy variable with level one as the baseline (i.e., 3 x 8 dummy codes representing the 4 possible responses for each dimension). This assumes that the 42 items have equal weight within a given domain but does assume the dimensions have cardinal properties.

Model 5: EQ-5D & SF-6D indices regressed on the individual levels of the HAQ-DI item scores, where each level is entered as a dummy variable with level one as the baseline (i.e., 3 x 20 dummy codes representing the 4 possible responses for each item of the 8 domains, and 1 x 22 dummy codes representing the dichotomous parameters). This makes the least stringent assumptions and does not assume that the response choices have cardinal properties.

The significance or sign of the beta coefficients was not of primary interest of this exercise given we are interested in predictive ability rather than explanatory power of the variables. Since most datasets will collect all items of the HAQ, all coefficients were included in the final models of 1 to 3. Due to the large number of dummy variables in models 4 and 5 this was not practical. Instead, they were developed using a backwards stepwise selection procedure, systematically removing the least significant variable until only significant variables remain (p<0.05). Residual plots were examined for nonlinear patterns and non-constant error variance. Final regression models were then assessed by 3-fold cross validation and applied to the UK and Canadian samples.

The criterion for judging the performance of the model is the goodness of fit between observed and predicted outcomes as reported in terms of the root of the mean square error (RMSE).[**Error! Bookmark not defined.**] This is the most indicative measure given the

objective of the analyses is not to explain the relationship between the HAQ-DI and the EQ-5D and the SF-6D indices, but to predict the EQ-5D and SF-6D indices from the HAQ-DI. Any reduction in performance between the development and cross-validation sample RMSE is reported.

Results

At baseline, patients in the Canadian cohort were slightly older (61 versus 56 p<0.001) and a greater percentage of patients were female (78% versus 67%, p<0.01). The mean HAQ in the UK patients was substantially higher (1.41 versus 1.11, p<0.01). This was reflected in both the EQ-5D scores where UK patients had a statistically significant different mean score of 0.51 versus 0.63 in the Canadian patients, and the SF-6D where UK patients had a mean score of 0.62 versus 0.68 in the Canadian sample. In total, there were 16 missing HAQ-DI responses, 4 EQ-5D, and 17 SF-6D (all less than 2% of the study population). (Table 2)

The Canadian population had a small but significantly higher estimated utility score, above what was explained by HAQ-DI (B=0.06 for EQ-5D and B=0.04 for SF-6D, p<0.05). However, as no country effect was found on the interaction with HAQ components (i.e. the gradient), an estimated utility gain using these algorithms would not be changed by this difference between countries. We have therefore presented the mean estimates from an amalgamated data source of patients from both countries. Examination of plots of predicted versus actual utility for the models suggest a relatively linear relationship between HAQ-DI and utility(Figure 1).

Table 2: Summary statistics of the two cohorts

Figure 1: Predicted versus actual SF-6D (top) and EQ-5D (bottom) scores.

In model 1, regressions resulted in a RMSE of 0.207 and 0.092 in the EQ-5D and SF-6D respectively.(Table 3) As expected there was a negative relationship between the HAQ-DI and the EQ-5D (Beta = -0.210) and the SF-6D (Beta = -0.101), which were statistically significant (p<0.001).(Table 4)

Table 3: Performance for models 1 to 5

Table 4: Final regression equations for models 1 and 2

The RMSE slightly improved when model 2 was analysed (equal to 0.202 (EQ-5D) and 0.089 (SF-6D)) (Table 4). All coefficients were negative except for hygiene (i.e. a worsening in hygiene predicts an improvement in utility), but neither coefficients were statistically significant. Other non significant terms were the grip dimension for the EQ-5D (p=0.671) and the dressing and grooming dimension for the SF-6D (p=0.093).

Again, in model 3, the RMSE improved in the development sample as compared to the previous model (RMSE = 0.189 and 0.086 in the development set EQ-5D and SF-6D respectively). However the cross validation demonstrated that the RMSE could be as high as 0.206 and 0.095 for the EQ-5D and SF-6D respectively, figures that are greater than the RMSE seen in the cross validation of model 1. On their own, getting in and out of bed (H4), using crutches (H12), help from another person with walking (H21), bending down to pick something off the floor (H26) and doing chores such as vacuuming or yardwork (H32) were the only statistically significant variables in the EQ-5D prediction. In the SF-6D prediction, shampooing your hair (H2) getting in and out of bed (H4), washing and drying your body (H22), opening jars (H28), doing chores such as vacuuming or yardwork (H32), other aids or devices (H38), and help in errands and chores from another person (H42) were statistically significant variables. (Table 5)

Table 5: Final regression equations for model 3

In model 4, the RMSE worsened from model 3 (to 0.199 and 0.088). Cross validation demonstrated again that the model 4 performed no better than models 1 and 2. Some 11 of the 24 variables were included in the EQ-5D, while 13 of the 24 variables were included in the SF-6D.(Table 6) Elements of dressing and grooming did not feature in either the EQ-5D and SF-6D estimates. Six components were predictors of both the EQ-5D and the SF-6D.

Table 6: Final regression equations for model 4

Lastly, model 5 gave the best performance for the EQ-5D with a RMSE's of 0.183 and 0.178 in the developmental and cross validation set respectively. The results for the SF-6D were no better than model 3 with RMSE's of 0.086 and 0.087. In the SF-6D, no components of the reach dimension were included in the final model. More items in the components of arising, eating, hygiene and activities were included than in the other components.(Table 7)

Table 7: Final regression equations for model 5

Discussion

We anticipated that models with more available predictors would account for a higher proportion of the variance and would therefore perform better as measured by the RMSE. While this hypothesis was accurate for the EQ-5D where model 5 proved to be the best performing, it was not the case for the SF-6D or models 2 to 4 for the EQ-5D. The results for the cross validation are perhaps the most important as they predict how generalisable the models will be to external populations. This would indicate that model 5 is the most appropriate model for estimating the EQ-5D, whilst model 2 or model 4 the most appropriate for the SF-6D. The performance of models for the SF-6D always outperformed models for the EQ-5D due to the smaller range in scale of the SF-6D. While the benefits of using the later models versus the simple estimate in model 1 would seem small in terms of the change in RMSE's, these models on the whole will provide more accurate estimates, partly due to their ability to account for the small non linearity seen in the relationship between HAQ and utility, particularly at severe states of disability. (Figure 2)

There are a number of important issues that need further consideration. Firstly, whether these results would be generalisable to external populations. To address this we developed the models using a combination of data from the UK and Canada. Patients in the Canadian dataset were older but had less severe RA. The mean HAQ-DI for both patient groups is similar to many studies recently published analysing the effectiveness of biologic therapies. Comparisons between Canadian and UK RA patients' quality of life have previously been explored and the differences between the two cohorts is not be unexpected.[15] This heterogeneity is important as it means the models can be used for estimation across a wider range of patients. The country effect discovered would appear not to be due to age since the Canadian population was older, but could have been due to another clinical or demographic characteristic not measured in our study. The models were tested on both the UK and Canadian samples. RMSE were always higher for the UK population since there was a smaller sample in which to test. Further external validation would always add assurance to the results, but we consider the results as they stand as good predictions.

Secondly, this exercise provides a method that will always be suboptimal in comparison to a trial that uses a preference-based questionnaire directly. Given the objectives of the study, there are other approaches that could be employed to derive a single index from the HAQ-DI. A survey of the general population could be used to value a sample of states defined by the HAQ-DI using a preference elicitation technique such as standard gamble (SG) or time trade-off (TTO). This would not only generate an enormous number of health states but more importantly each state would contain 42 pieces of information, which most respondents would find impossible to process. Instead, a selection of the most important items of the HAQ-DI could be selected, similar to how the SF-6D uses only 15 questions from the SF-36. Another approach is to administer the HAQ alongside a preference-elicitation technique such as TTO and SG. Regression techniques could then estimate preference weights for each of the items of the HAQ-DI using the SG or TTO response as the dependent variable. However, results from such a study would not meet the reference case for either NICE or the Washington Panel on Cost Effectiveness in Medicine who prefer social preferences elicited using a choice-based

method.[16,2] This exercise could act as a precursor to such studies, but given limited resources, we have undertaken a more pragmatic approach.

Lastly, there has been an argument that the HAQ-DI does not adequately measure aspects of quality of life, measured by the preference based instruments such as mental health and pain.[17] Given this, the models show that HAQ-DI does explain much of the preference based measures we have studied, with small RMSE's (<0.2). Perhaps such aspects of quality of life such as pain are highly correlated to domains so are indirectly covered. Explaining why there is a relationship between the two measures was not the purpose of this study, but rather exploring if there is a translation between the two measures. Importantly, the method described in this paper is not designed and would not predict accurately the utility of an individual but rather predict the average utility of a cohort. Figure 2 demonstrates that across the range of the HAQ-DI, the model prediction for both the EQ-5D and SF-6D is close. Only in the first group (HAQ 0 - .5) is the prediction significantly different to the actual utility (p= <0.01). Even in the higher HAQ groups where we have less patients, the predictions appear good.

Figure 2: Predicted and Actual EQ-5D scores and confidence intervals across HAQ groups within the sample

Another argument concerns whether the generic utility based measures are accurate representations of patients' preferences in RA. Aspects of the condition captured by the HAQ-DI might not be covered in the preference-based measures. Concerns about the EQ-5D and SF-6D in RA patients have previously been demonstrated.[18] This paper is not aimed to make claims on the superiority or defects of different preference-based measures, but to give researchers a method of estimating these what are now frequently used instruments.

Much of the paper has concentrated on studies where no preference-based measure has been administered. Given the SF-6D does not perform well in patients with severe RA due to a floor effect, there is potentially a use when only one preference-based questionnaire is administered.[18,19] This is the case in the British Society for Rheumatology Biologics

Registry (BSRBR) that measured only the SF-36

(http://www.arc.man.ac.uk/webbiologicsreg.htm). This approach would allow an estimate of

EQ-5D utility to additionally be calculated.

The approach examined in this article is to empirically map the relationship between a non-preference-based HRQOL instrument and a preference-based measure. The approach has the advantage of being able to utilise existing valuation data and offers a shortcut for researchers who need health-state utility values, but have not used a preference-based measure in their clinical study because of resource constraints or a desire to limit the patient burden. This could be used to estimate the improvement in utility in important trials such as the ATTRACT trial of infliximab or the TEMPO etanercept trial.[20,21] The results presented here suggest that such a model can be useful in predicting preference-based values and that the models achieve a reasonable goodness of fit.

References

1 Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford Medical Publications, 1987.

2 Gold, M.R., Siegel J.E., Russell L.B., Weinstein M.C. (1996) Cost-Effectiveness in Health and Medicine. Oxford University Press, Oxford.

3 Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use health status measures in economic evaluation. Health Technol Assess 1999;3(9)

4 Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: The Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:789-793.

5 Scott DL, Garrood T. Quality of life measures: use and abuse. Best Pract Res Clin Rheumatol. 2000;14(4):663-687

6 Brennan A Bansback N Reynolds A Conway P. Modelling the cost-effectiveness of etanercept in adults with rheumatoid arthritis in the UK. Rheumatology 2004; 43: 62-72

7 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)

8 Torrance GW, 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 2004; 43: 712-718

9 Tsuchiya A, Brazier J, McColl E, Parkin D. Deriving preference-based single indices from non-preference based condition specific instruments:converting the AQLQ into EQ-5D. [www document]. 2002 Available from http://www.shef.ac.uk/~sheg/discussion/discussion.htm. 2002.

10 Brazier JE, Kolotkin RL, Crosby RD, Williams GR. Estimating a Preference-Based Single Index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) Instrument from the SF-6D. Value Health. 2004; 7(4) 490-498

11 Brooks R. EuroQol: the current state of play. Health Policy 1996;37(1):53-72.

12 Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35(11):1095-108.

13 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics 2002. 21(2), 271-292

14 Ware J, Snow K, Kosinski M, Gandek B. SF-36 Health Survey: Manual and Interpretation Guide. Boston: The Health Institute, New England Medical Center, 1993.

15 Neville C. A comparison between UK and Canadian disease characteristics and valuation of quality of life explored. J Rheum 2001

16 National Institute for Clinical Excellence. Guide to the Methods of Technology Appraisal. (2004). Available at www.nice.org.uk

17 Wolfe F, Michaud K. HAQ-based Utilities and SF6D Systematically Overvalue Quality of Life (QOL) in RA Patients with Severe RA, Pain and Psychological Distress (ACR 2003)

18 Marra CA, Esdaille JM, Guh D, Kopec JA, Brazier JE, Koehler BE, Chalmers A, Anis AH. (2004) A comparison of four indirect methods of assessing values in rheumatoid arthritis. Medical Care (in press).

19 Brazier J, Roberts J. Tsyuchiya A. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Economics. 2004 In press

20 Maini R, St Clair EW, Breedveld F, Furst D, Kalden J, Weisman M, et al. Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group Lancet 1999;354:1932-9

21 Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, 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;363:670-1.

Table 1: Characteristics of quality of life instruments

	Domains/ Categories	Design	Number of individual health state valuations	Method of Scoring	Sample	Country	Boundaries
HAQ	Dressing and grooming, arising, eating, walking, hygiene, reach, grip, activities	20 questions 4 levels each + 22, 2 level questions	24	Scalar	NA	NA	0.00-3.00
SF-6D	Physical function, role limitation, social function, pain, mental health, vitality	6 questions between 4 and 6 levels.	18,000	Preference weighted	611 (representative)	UK	0.30-1.00
EQ-5D	Mobility, usual activities, self-care, pain, anxiety	5 questions 3 levels.	243	Preference weighted	3395 (representative)	UK	-0.59-1.00

Table 2: Summary statistics of baseline characteristics in the two cohorts

	UK (n=132)		Canada (n=310)		Total (n=442)		
	Mean	Range	Mean	Range	Mean	Range	P[†]
Sex (% Female)	67		78		76		0.009
Mean age (SD)	56.01 (13.62)	(17-82)	61.43 (13.6)	(19-90)	60.76 (13.6)	(17-90)	<0.001
HAQ Disability							
N	132		308		440		
Mean Index (SD)	1.41 (0.8)	(0-3)	1.11 (0.77)	(0-3)	1.15 (0.78)	(0-3)	0.004
HAQ Domains, modal level (% of total)							
Dressing & Grooming	2 (35)	(0-3)	0 (46)	(0-3)	0 (39)	(0-3)	<0.001
Rising	1 (41)	(0-3)	0 (54)	(0-3)	0 (44)	(0-3)	<0.001
Eating	1 (35)	(0-3)	0 (40)	(0-3)	0 (35)	(0-3)	0.001
Walking	2 (41)	(0-3)	0 (45)	(0-3)	0 (41)	(0-3)	0.022
Hygiene	2 (43)	(0-3)	3 (30)	(0-3)	0 (31)	(0-3)	<0.001
Reach	3 (30)	(0-3)	0 (31)	(0-3)	2 (30)	(0-3)	0.001
Grip	2 (57)	(0-3)	2 (61)	(0-3)	2 (59)	(0-3)	0.006
Activities	2 (35)	(0-3)	2 (28)	(0-3)	2 (30)	(0-3)	0.469
SF-6D							
N	129		302		431		
Mean Index (SD	0.62 (0.11)	(0.27-0.92)	0.68 (0.13)	(0.26-1)	0.68 (0.13)	(0.26-1)	<0.001
SF-6D Domains, modal level (% of total)							
Physical functioning	4 (31)	(1-6)	5 (30)	(1-6)	5 (28)	(1-6)	0.003
Role limitation	4 (46)	(1-4)	2 (63)	(1-4)	2 (54)	(1-4)	<0.001
Social functioning	3 (36)	(1-5)	3 (43)	(1-5)	3 (40)	(1-5)	<0.001
Pain	5 (33)	(1-6)	4 (27)	(1-6)	4 (27)	(1-6)	0.001
Mental health	3 (36)	(1-5)	2 (41)	(1-5)	2 (38)	(1-5)	0.010
Energy and vitality	5 (34)	(1-5)	3 (35)	(1-5)	3 (33)	(1-5)	<0.001
EQ-5D							
N	131		308		439		
Mean Index (SD)	0.51 (0.31)	(-0.35-1)	0.63 (0.25)	(-0.48-1)	0.62 (0.27)	(-0.48-1)	<0.001
EQ-5D Domains, modal level (% of total)							
Mobility	2 (78)	(1-2)	2 (62)	(1-3)	2 (66)	(1-3)	0.004
Self Care	1 (52)	(1-3)	1 (71)	(1-3)	1 (65)	(1-3)	<0.001
Usual Activities	2 (71)	(1-3)	2 (66)	(1-3)	2 (63)	(1-3)	0.003
Pain	2 (77)	(1-3)	2 (79)	(1-3)	2 (79)	(1-3)	0.008
Anxiety	1 (52)	(1-3)	1 (64)	(1-3)	1 (60)	(1-3)	0.001

[†] ordinal data compared using independent sample t-tests, categorical data compared using chi-squared test

Table 3: Performance for models 1 to 5

Model	EQ-5D Index				SF-6D Index			
	RMSE Develop	RMSE Cross Validation	RMSE Canada	RMSE UK	RMSE Develop	RMSE Cross Validation	RMSE Canada	RMSE UK
1	0.2070	0.1896	0.1763	0.2558	0.0916	0.0870	0.0849	0.0986
2	0.2021	0.1920	0.1771	0.2649	0.0886	0.0845	0.0819	0.0989
3	0.1885	0.2056	0.1865	0.2772	0.0858	0.0955	0.0938	0.1032
4	0.1991	0.1946	0.1776	0.2758	0.0884	0.0841	0.0814	0.0993
5	0.1829	0.1780	0.1610	0.2410	0.0863	0.0866	0.0839	0.0983

Table 4: Final regression equations for models 1 and 2

Model		EQ-5D Index				SF-6D Index			
		B	**SE**	**Z**	**P**	**B**	**SE**	**Z**	**P**
1	HAQ index	-0.2102	0.0116	-18.07	<0.001	-0.1008	0.0072	-13.96	<0.001
	Constant	0.8553	0.0120	71.15	<0.001	0.7893	0.0087	90.36	<0.001
2	Dressing & Grooming	-0.0300	0.0111	-2.70	0.007	-0.0078	0.0046	-1.68	0.093
	Arising	-0.0522	0.0129	-4.05	<0.001	-0.0262	0.0046	-5.69	<0.001
	Eating	-0.0461	0.0113	-4.07	<0.001	-0.0178	0.0045	-3.93	<0.001
	Walking	-0.0282	0.0098	-2.89	0.004	-0.0115	0.0038	-3.08	0.002
	Hygiene	0.0100	0.0074	1.34	0.179	0.0056	0.0031	1.82	0.069
	Reach	-0.0199	0.0082	-2.42	0.016	-0.0100	0.0037	-2.67	0.008
	Grip	-0.0035	0.0082	-0.43	0.671	-0.0142	0.0039	-3.68	0.002
	Activities	-0.0487	0.0097	-5.01	<0.001	-0.0238	0.0043	-5.55	<0.001
	Constant	0.8371	0.0118	70.83	<0.001	0.7858	0.0091	86.00	<0.001

Table 5: Final regression equations for model 3

		EQ-5D Index				SF-6D Index			
		B	**SE**	**Z**	**P**	**B**	**SE**	**Z**	**P**
DRESSING & GROOMING									
-Dress yourself, including tying shoelaces and doing buttons?	H_1	0.0112	0.0192	0.59	0.558	0.0125	0.0081	1.54	0.124
-Shampoo your hair?	H_2	-0.0295	0.0167	-1.77	0.077	-0.0192	0.0077	-2.52	0.012
ARISING									
-Stand up from a straight chair?	H_3	0.0049	0.0180	0.27	0.786	-0.0102	0.0081	-1.26	0.208
-Get in and out of bed?	H_4	-0.0738	0.0221	-3.34	0.001	-0.0298	0.0083	-3.57	<0.001
EATING									
-Cut your meat?	H_5	0.0054	0.0169	0.32	0.751	-0.0037	0.0086	-0.43	0.666
-Lift a full cup or glass to your mouth?	H_6	-0.0247	0.0176	-1.40	0.161	-0.0097	0.0085	-1.14	0.253
-Open a new milk carton?	H_7	-0.0188	0.0155	-1.21	0.227	-0.0088	0.0076	-1.17	0.243
WALKING									
-Walk outdoors on flat ground?	H_8	-0.0288	0.0228	-1.26	0.207	0.0008	0.0096	0.08	0.937
-Climb up five steps?	H_9	0.0166	0.0197	0.84	0.399	-0.0025	0.0075	-0.34	0.736
AIDS OR DEVICES									
-Cane	H_{10}	-0.0022	0.0239	-0.09	0.927	0.0032	0.0114	0.28	0.779
-Walker	H_{11}	0.0345	0.0355	0.97	0.332	-0.0297	0.0185	-1.60	0.109
-Crutches	H_{12}	-0.2028	0.0768	-2.64	0.008	0.0130	0.0362	0.36	0.719
-Wheelchair	H_{13}	-0.0571	0.0527	-1.08	0.280	0.0084	0.0269	0.31	0.754
-Dressing	H_{14}	0.0230	0.0266	0.87	0.387	0.0136	0.0112	1.21	0.225
-Utensils	H_{15}	-0.0544	0.0328	-1.66	0.098	0.0049	0.0123	0.40	0.693
-Chair	H_{16}	0.0518	0.0286	1.81	0.070	0.0011	0.0113	0.09	0.925
-Other?	H_{17}	-0.0510	0.0475	-1.07	0.283	0.0072	0.014	0.51	0.607
HELP FROM ANOTHER PERSON									
-Dressing and grooming	H_{18}	-0.0379	0.0291	-1.30	0.193	-0.0065	0.0105	-0.62	0.536
-Arising	H_{19}	-0.0202	0.0323	-0.62	0.533	-0.0091	0.0107	-0.85	0.396
-Eating	H_{20}	0.0580	0.0322	1.80	0.072	-0.0142	0.0135	-1.05	0.292
-Walking	H_{21}	0.0453	0.0204	2.22	0.027	-0.0021	0.0098	-0.22	0.828
HYGIENE									
-Wash and dry your body?	H_{22}	-0.0131	0.0150	-0.87	0.382	-0.0228	0.0052	-4.41	<0.001
-Take a tub bath?	H_{23}	0.0027	0.0079	0.34	0.733	0.0017	0.0039	0.42	0.673
-Get on and off the toilet?	H_{24}	-0.0047	0.0178	-0.26	0.791	0.0067	0.0066	1.01	0.311
REACH									
-Reach and get down a 5-pound object from just above your head?	H_{25}	-0.0224	0.0118	-1.90	0.058	-0.0074	0.0049	-1.52	0.129
-Bend down to pick up clothing from the floor?	H_{26}	-0.0343	0.0160	-2.14	0.033	-0.0044	0.0064	-0.69	0.491
GRIP									
-Open car doors?	H_{27}	-0.0208	0.0168	-1.24	0.217	0.0092	0.0077	1.20	0.231
-Open jars, which have been previously opened?	H_{28}	-0.0186	0.0150	-1.24	0.216	-0.0129	0.0064	-2.01	0.044
-Turn faucets on and off?	H_{29}	-0.0227	0.0179	-1.27	0.203	-0.0028	0.0063	-0.45	0.653
ACTIVITIES									
-Run errands and shop?	H_{30}	-0.0167	0.0165	-1.02	0.309	-0.0022	0.0066	-0.33	0.739
-Get in and out of a car?	H_{31}	-0.0287	0.0181	-1.59	0.112	-0.0051	0.0062	-0.82	0.412
-Do chores such as vacuuming or yardwork?	H_{32}	-0.0345	0.0112	-3.07	0.002	-0.0157	0.0048	-3.29	0.001
AIDS OR DEVICES									
-Raised toilet seat	H_{33}	-0.0316	0.0256	-1.23	0.218	0.0019	0.0112	0.17	0.862
-Bathtub seat	H_{34}	-0.0175	0.0277	-0.63	0.528	-0.0157	0.0118	-1.33	0.184
-Jar opener	H_{35}	0.0163	0.0185	0.88	0.378	-0.0092	0.0068	-1.35	0.179
-Bathtub bar	H_{36}	0.0137	0.0207	0.66	0.510	0.0104	0.0086	1.21	0.226
-Long handles appliances for reach	H_{37}	0.0126	0.0245	0.52	0.606	-0.0003	0.0105	-0.03	0.977
-Other	H_{38}	0.0331	0.0322	1.03	0.303	-0.0278	0.0135	-2.06	0.040
HELP FROM ANOTHER PERSON									
-Hygiene	H_{39}	0.0420	0.0217	1.94	0.053	0.0131	0.0081	1.63	0.103
-Reach	H_{40}	0.0052	0.0211	0.25	0.806	0.0164	0.008	2.04	0.042
-Gripping and opening	H_{41}	0.0363	0.0211	1.72	0.085	-0.0124	0.0081	-1.54	0.124
-Errands and chores	H_{42}	-0.0401	0.0224	-1.79	0.074	-0.0210	0.0085	-2.48	0.013
Constant		0.8041	0.0122	65.68	<0.001	0.7648	0.0086	89.4	<0.001

Table 6: Final regression equations for model 4

	EQ-5D Index				SF-6D Index			
	B	SE	Z	P	B	SE	Z	P
Arising = 2	-0.0545	0.0189	-2.89	0.004				
Arising = 3	-0.1164	0.0531	-2.19	0.028				
Eating = 1	-0.0726	0.0189	-3.85	<0.001	-0.0247	0.0063	-3.95	<0.001
Eating = 2	-0.085	0.0255	-3.33	<0.001	-0.0521	0.0092	-5.64	<0.001
Eating = 3					-0.1144	0.0353	-3.24	0.001
Walking = 1	-0.0585	0.0154	-3.80	<0.001	-0.0212	0.007	-3.02	0.002
Walking = 2	-0.0919	0.0236	-3.90	<0.001	-0.0377	0.0102	-3.68	<0.001
Walking = 3	-0.1888	0.0382	-4.94	<0.001	-0.0558	0.0155	-3.61	<0.001
Hygiene = 1	-0.0578	0.0213	-2.72	0.007				
Hygiene = 3	-0.0693	0.0193	-3.59	<0.001	-0.0194	0.0066	-2.95	0.003
Reach = 1					-0.0703	0.024	-2.93	0.003
Reach = 2					-0.0151	0.0067	-2.24	0.025
Reach = 3	0.0501	0.0213	2.35	0.019				
Grip = 2					-0.0199	0.008	-2.48	0.013
Grip = 3					-0.0202	0.0086	-2.35	0.019
Activities = 1					-0.0402	0.0122	-3.29	0.001
Activities = 2					-0.0203	0.0061	-3.33	<0.001
Activities = 3	-0.1418	0.0527	-2.69	0.007				
Constant	-0.0606	0.017	-3.56	<0.001	-0.0331	0.0088	-3.77	<0.001

Table 7: Final regression equations for model 5

	EQ-5D Index				SF-6D Index			
	B	**SE**	**Z**	**P**	**B**	**SE**	**Z**	**P**
DRESSING & GROOMING								
$H_1=2$	-0.1514	0.0394	-3.85	<0.001	.	.	.	.
$H_1=3$	.	.	.	.	0.0458	0.0202	2.26	0.024
$H_2=3$	.	.	.	.	-0.0551	0.0166	-3.32	<0.001
ARISING								
$H_4=1$	-0.0762	0.0171	-4.47	<0.001	-0.0382	0.0077	-4.93	<0.001
$H_4=2$	-0.1150	0.0480	-2.40	0.017	-0.0881	0.0167	-5.27	<0.001
$H_4=3$	-0.5847	0.0813	-7.19	<0.001	-0.3443	0.0341	-10.08	<0.001
EATING								
$H_5=2$	.	.	.	.	-0.0520	0.0147	-3.53	0.004
$H_6=2$	-0.1371	0.048	-2.86	0.004	.	.	.	.
$H_7=1$	-0.0410	0.0168	-2.44	0.015	.	.	.	.
$H_7=2$	-0.0783	0.0272	-2.88	0.004	.	.	.	.
WALKING								
$H_8=2$	-0.0955	0.0427	-2.24	0.025	.	.	.	.
$H_9=2$	.	.	.	.	-0.0358	0.0180	-1.99	0.046
$H_9=3$	0.1176	0.0499	2.36	0.018	.	.	.	.
AIDS OR DEVICES								
$H_{11}=1$	.	.	.	.	-0.0443	0.0190	-2.33	0.020
$H_{13}=2$	-0.1367	0.0411	-3.33	<0.001	.	.	.	.
$H_{16}=1$	0.0664	0.0255	2.61	0.009	.	.	.	.
HELP FROM ANOTHER PERSON								
$H_{21}=1$	.	.	.	.	0.0181	0.0087	2.09	0.037
HYGIENE								
$H_{22}=1$	.	.	.	.	-0.0284	0.0076	-3.73	0.002
$H_{22}=2$	.	.	.	.	-0.0579	0.0192	-3.02	0.003
$H_{22}=3$	.	.	.	.	-0.0507	0.0164	-3.09	0.002
$H_{23}=1$	-0.0487	0.0163	-2.99	0.003	-0.0337	0.0072	-4.66	<0.001
$H_{24}=1$	-0.0520	0.0204	-2.54	0.011	.	.	.	.
$H_{24}=2$	-0.1131	0.0431	-2.62	0.009	.	.	.	.
REACH								
$H_{26}=2$	-0.1375	0.0374	-3.68	0.002	.	.	.	.
$H_{26}=3$	-0.1344	0.0625	-2.15	0.032	.	.	.	.
GRIP								
$H_{27}=2$	-0.0756	0.0373	-2.02	0.043	.	.	.	.
$H_{27}=3$	-0.2002	0.0706	-2.84	0.005	0.0540	0.0202	2.67	0.008
$H_{28}=3$	.	.	.	.	-0.0823	0.0135	-6.08	<0.001
ACTIVITIES								
$H_{30}=1$	-0.0505	0.0189	-2.67	0.008	.	.	.	.
$H_{31}=1$	-0.0684	0.0194	-3.52	<0.001	.	.	.	.
$H_{31}=2$	-0.0819	0.0376	-2.18	0.030	.	.	.	.
$H_{32}=1$	.	.	.	.	-0.0378	0.0107	-3.55	<0.001
$H_{32}=2$	.	.	.	.	-0.0613	0.0120	-5.13	<0.001
$H_{32}=3$	-0.0903	0.0276	-3.27	0.001	-0.0841	0.0176	-4.76	<0.001
AIDS OR DEVICES								
$H_{35}=1$	.	.	.	.	-0.0141	0.0066	-2.15	0.032
$H_{36}=1$	.	.	.	.	0.0166	0.0072	2.31	0.021
HELP FROM ANOTHER PERSON								
$H_{41}=1$	.	.	.	.	-0.0157	0.007	-2.26	0.024
Constant	0.8016	0.0107	74.67	<0.001	0.7709	0.0098	78.44	<0.001

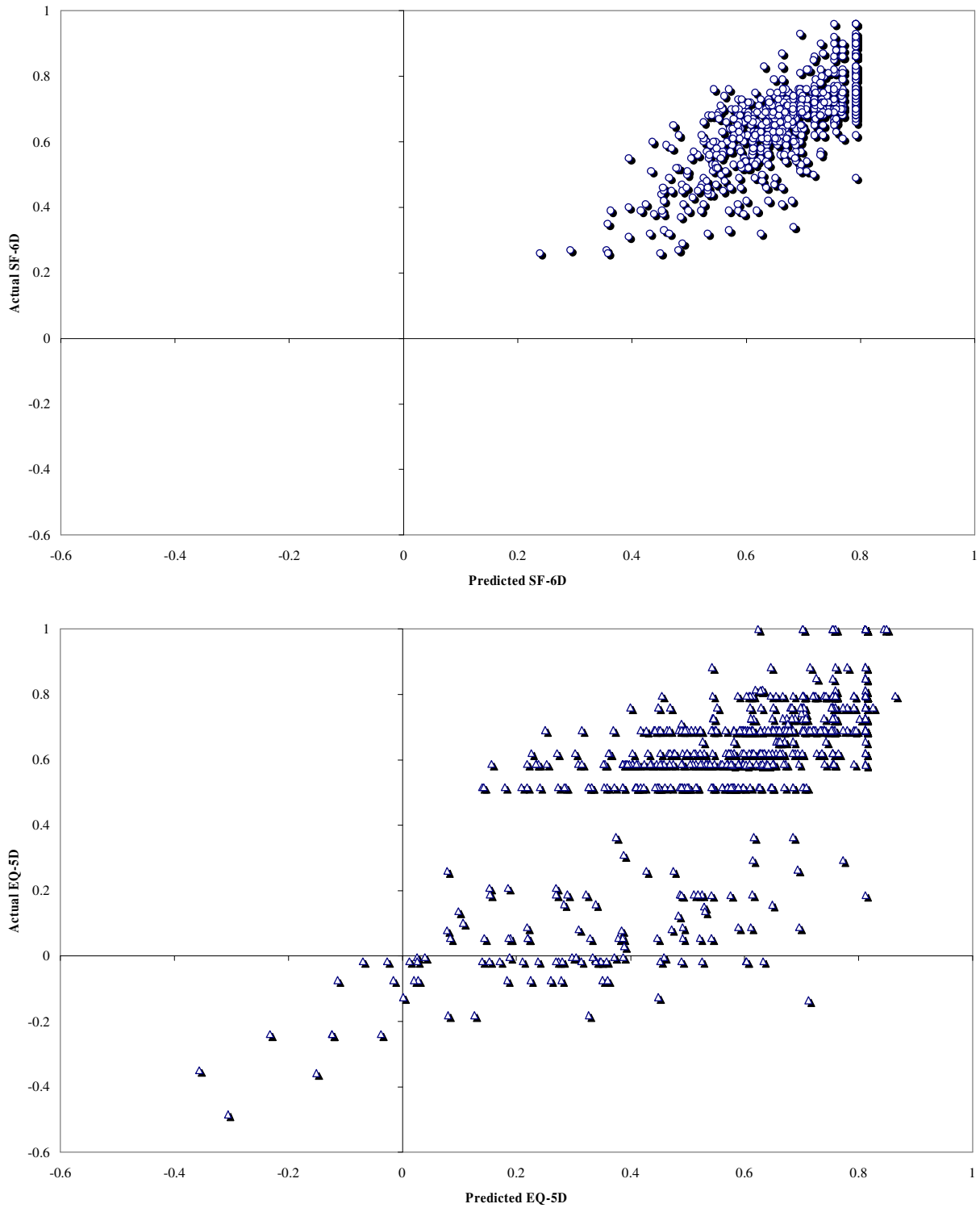**Figure 1: Predicted versus actual SF-6D (top) and EQ-5D (bottom) scores.**

Figure 2: Predicted and Actual EQ-5D (model 5) and SF-6D (model 4) scores and confidence intervals across HAQ groups within the sample