**Technology Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Clinical Excellence**

# Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor

## Addendum report


**Produced by** West Midlands Health Technology Assessment Collaboration

**Authors** Kinga Malottki, Pelham Barton, Angelos Tsourapas, Paresh Jobanputra, David Moore, Yen-Fu Chen

**Correspondence to** Yen-Fu Chen
y.chen.3@bham.ac.uk


**Date completed** 28.01.2010.


**Source of funding**: This report was commissioned by the NIHR HTA Programme as project number 04/26/02


**Rider on responsibility for report**
The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

**This report should be referenced as follows:**
TBC

Commercial in confidence information has been [redacted]

This page is intended blank.

**Contents**

# 1 Introduction

This addendum report contains additional information from the assessment group for the technology appraisal of adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor. Specifically this addendum contains:

- The critique of manufacturers' submissions from the original assessment group report (included for context),
- Critique of the mixed treatment comparisons and indirect comparisons contained within the submissions of the manufacturers of the technologies,
- Further critique of the economic models submitted by the manufacturers of the technologies,
- A revised economic section from the original assessment group report which includes further details of the BRAM economic model with additional scenario analyses,
- Additional sensitivity analysis to assess impact of differences in assumptions between models.

# 2 Critique of manufacturers' submissions

A submission was received from each company with each submission containing an economic analysis. However, only four manufactures included a model. Table 1 provides a brief summary of the five economic analyses provided.

## 2.1 Abbott submission (Adalimumab)

A discrete event simulation model was built to evaluate the cost-effectiveness of adalimumab. The type of evaluation undertaken was a cost-utility analysis with outcomes measured in QALYs.

Adalimumab was compared to all interventions included in the scope; etanercept, infliximab, rituximab and abatacept, all combined with methotrexate. In each of these five strategies, each drug was followed by gold, then leflunomide, then cyclosporin, then rescue therapy. A comparison was also made with a strategy of traditional DMARDs only (gold, then leflunomide, then cyclosporin, then rescue therapy) and also a strategy where adalimumab (or etanercept) is followed by rituximab, then gold, then leflunomide, then cyclosporin, then rescue therapy.

It is assumed that the population has already had an inadequate response to at least two traditional DMARDs, since these are patients who had had an inadequate response to a TNF inhibitor. Therefore, methotrexate, sulfasalazine and hydroxychloroquine are not considered as comparators in the economic evaluation.

Response rates are assumed to be equal across all TNF inhibitors. In addition, drug, administration and monitoring costs of adalimumab and etanercept are assumed to be equal. Therefore, adalimumab and etanercept are evaluated in the same treatment sequence and results for these two drugs are considered similar throughout the submission.

New biologic agents (tocilizumab, golimumab and certolizumab pegol) were excluded from the analysis since these drugs were considered not yet available in the UK.

**Table 1 Summary of methods used in industry economic analyses**

| **Submission features** | **Adalimumab (Abbott)** | **Etanercept (Wyeth)** | **Infliximab (Schering-Plough Limited)** | **Rituximab (Roche)** | **Abatacept (Bristol-Myers Squibb)** |
|---|---|---|---|---|---|
| Population | Adult patients with active RA who have had an inadequate response to methotrexate, sulfasalazine, hydroxychloroquine and one TNF inhibitor | Adult patients with active RA who have had an inadequate response to etanercept | Adult patients with active RA who have had an inadequate response to two non-biologic DMARDs and one TNF inhibitor | Adult patients with active RA who have had an inadequate response to a TNF inhibitor | Adult patients with moderate to severe RA who have had an inadequate response to at least one TNF inhibitor |
| Interventions and comparators | **Gold**→Leflunomide→ Ciclosporin→rescue vs. **ADA/ETN**→Gold→ Leflunomide→Ciclosporin →rescue vs. **IFX**→Gold→ Leflunomide→Ciclosporine →rescue vs. **RTX**→Gold→ Leflunomide→Ciclosporin →rescue vs. **ABT**→Gold→ Leflunomide→Ciclosporin →rescue vs. **ADA/ETN→RTX** → Gold→Leflunomide→ Ciclosporin→rescue | **ETN/IFX/ADA**→ DMARDs→ 'salvage therapy' vs. **DMARDs**→ DMARDs→ 'salvage therapy' vs. **RTX** → DMARDs→ 'salvage therapy' | **ADA**→DMARDs vs. **ETN**→DMARDs vs. **IFX**→DMARDs vs. **ABT**→DMARDs vs. **RTX**→DMARDs vs. **ADA→RTX**→DMARDs vs. **ETN→RTX**→DMARDs vs. **IFX→RTX**→DMARDs vs. **DMARDs** | **RTX**→Leflunomide→ Gold Cyclosporin→ palliative care vs. **ETN**→Leflunomide→ Gold→Cyclosporin→ palliative care vs. **ADA**→Leflunomide→ Gold→Cyclosporin→ palliative care vs. **IFX**→Leflunomide→ Gold→Cyclosporin→ palliative care vs. **ABT**→Leflunomide→ Gold→Cyclosporin→ palliative care vs. **Leflunomide**→Gold→ Cyclosporin→ palliative care | **ABT→IFX→** Leflunomide→Gold→ Azathioprine→Ciclosporin→ Penicillamine→Palliative care vs. **RTX→IFX→** Leflunomide→Gold→ Azathioprine→Ciclosporin→ Penicillamine→Palliative care **ABT→ TNF inhibitors→** Leflunomide→Gold→ Azathioprine→Ciclosporin→ Penicillamine→Palliative care vs. **TNF inhibitors→TNF inhibitors**→Leflunomide→ Gold→Azathioprine→ Ciclosporin→Penicillamine→ Palliative care |
| Form of analysis | Cost-utility analysis | Cost-utility analysis | Cost-utility analysis | Cost-utility analysis | Cost-utility analysis |
| Model used | Discrete event simulation | Markov model | Patient-simulation | Patient-level simulation | Patient-level simulation |

| **Submission features** | **Adalimumab (Abbott)** | **Etanercept (Wyeth)** | **Infliximab (Schering-Plough Limited)** | **Rituximab (Roche)** | **Abatacept (Bristol-Myers Squibb)** |
|---|---|---|---|---|---|
| | model | | | | |
| Cycle length | Continuous | 6-month | 1-month | 6-month | Continuous |
| Time horizon | Lifetime | Lifetime | Lifetime | Lifetime | Lifetime |
| Base case results presented - ICERs (£/QALY) | ADA/ETN vs. DMARDs: £15,962<br>IFX vs. DMARDs: £21,529<br>RTX (9-months) vs. DMARDs: £10,986<br>ABT vs. DMARDs: £30,104<br><br>ADA/ETN→RTX vs. DMARDs: £13,797 | TNF inhibitors vs. DMARDs: £14,501<br>TNF inhibitors vs. Rituximab: £16,225 | ADA vs. DMARDs: £35,138<br>ETN vs. DMARDs: £35,898<br>IFX vs. DMARDs: £28,661<br>ABT vs. DMARDs: £44,769<br>RTX vs. DMARDs: £17,422 (9-month dose of RTX),<br>£27,161 (6-month dose of RTX)<br><br>ADA+RTX vs. DMARDs: £27,998 (9-month dose of RTX),<br>£32,345 (6-month dose of RTX)<br>ETN+RTX vs. DMARDs: £27,936 (9-month dose of RTX),<br>£32,412 (6-month dose of RTX)<br>IFX+RTX vs. DMARDs: £24,236 (9-month dose of RTX),<br>£28,617 (6-month dose of RTX) | RTX vs. ETN: RTX dominates<br>RTX vs. IFX: RTX dominates<br>RTX vs. ABT: RTX dominates<br>RTX vs. ADA: £310,771<br>RTX vs. DMARDs: £5,311 | ABT→IFX vs. RTX→IFX: £20,438<br><br>ABT→TNF inhibitors vs. TNF inhibitors→TNF inhibitors: £23,019 |

| **Submission features** | **Adalimumab (Abbott)** | **Etanercept (Wyeth)** | **Infliximab (Schering-Plough Limited)** | **Rituximab (Roche)** | **Abatacept (Bristol-Myers Squibb)** |
|---|---|---|---|---|---|
| PSA results | ADA/ETN→RTX vs. DMARDs: 100% cost-effective at £30,000 per QALY<br><br>RTX vs. DMARDs probability of being cost-effective ~60% at £20,000/QALY<br><br>ADA/ETN vs. DMARDs probability of being cost-effective ~40% at £20,000/QALY | Not presented | RTX (9-month dose) vs. DMARDs: probability of being cost-effective >90%<br>IFX vs. DMARDs: probability of being cost-effective ~60% at £30,000/QALY<br>IFX+RTX vs. RTX: probability of being cost-effective >40% at £30,000/QALY | RTX vs. ETN: RTX is 100% cost-effective, dominating 74% of iterations<br>RTX vs. IFX: RTX is 100% cost-effective, dominating 70% of iterations<br>RTX vs. ADA: RTX is 100% cost-effective, dominating 37% of iterations<br>RTX vs. ABT: RTX is 100% cost-effective, dominating 70% of iterations<br>RTX vs. DMARDs: RTX is 100% cost-effective | PSA<br>Probability of Abatacept being cost-effective at £30,000 per QALY:<br>99% when compared with Rituximab<br>97% when compared with TNF inhibitors |
| HAQ→QoL | EQ-5D=0.82-0.11*HAQ-0.07*HAQ$^2$ | EQ-5D=0.76-0.28*HAQ | NA | EQ-5D=0.82-0.11*HAQ-0.07* HAQ$^2$ | HUI3=0.76-0.28*HAQ +0.05*Female |
| Adverse events | Included.<br>Rates of tuberculosis (for TNF inhibitors) from BSRBR.<br>Rates of mild, moderate and serious adverse events of etanercept, infliximab and leflunomide from an observational study.<br><br>Leflunomide was used as a | Included.<br>Serious adverse events were modelled at £1,181<br>Adverse events of conventional DMARDs assumed to be more frequent that those of TNF inhibitors. | Although the submission provides background evidence on adverse events, they have not been included in the model. | Not included.<br>The clinical section of the submission indicates that the incidence of adverse events is very similar across all treatments in the appraisal. | Included.<br>Sources were mainly published sources.<br>Abatacept has the lowest rates in all adverse events apart from sinusitis. |

| Submission features | Adalimumab (Abbott) | Etanercept (Wyeth) | Infliximab (Schering-Plough Limited) | Rituximab (Roche) | Abatacept (Bristol-Myers Squibb) |
|---|---|---|---|---|---|
| | proxy for all traditional DMARDs<br>Etanercept was used as a proxy for adalimumab, abatacept and rituximab. | | | | |
| Mortality | Applied a treatment-specific mortality effect. Produced a parametric version of the mortality risk, with adjustments for treatment and HAQ. | Used a baseline mortality of 1.63 times general population mortality, with an adjustment for change in HAQ (not clear how they have implemented this as they did not supply their model electronically). | Used mortality ratios dependent on age and gender but no variation by HAQ or treatment. | Started from general population mortality and applied a multiplier of 1.33 to the power of the HAQ score, with the parameter 1.33 varied in sensitivity analysis. | Started from general population mortality and applied a multiplier of 1.33 to the power of the HAQ score, with the parameter 1.33 varied in sensitivity analysis. |

**Adverse events**

Adverse events were included in the economic analysis. Rates of tuberculosis associated with each of the TNF inhibitors (adalimumab, etanercept, infliximab) were based on data from the BSRBR. [1] Rates of mild, moderate and serious adverse events were estimated from an observational study in Sweden, which evaluated the safety of patients receiving etanercept, infliximab or leflunomide.[2] Values for these drugs were used as proxies for other drugs. The effect of this was that the rate of adverse events was higher for conventional DMARDs than for biologics.

**HAQ to Utility**

A quadratic mapping mechanism was used in order to convert HAQ scores to EQ-5D scores (EQ-5D=0.82-0.11*HAQ-0.07* $HAQ^2$). This equation was estimated through EQ-5D data collected in tocilizumab trials (OPTION and LITHE).[3] The linear mapping mechanism reported in the same study (EQ-5D=0.89-0.28*HAQ) was explored in a sensitivity analysis.

**Results**

The base case results show that all drugs [adalimumab/etanercept, infliximab, rituximab and abatacept (all followed by traditional DMARDs)] may represent cost-effective treatment options when compared with a sequence of traditional DMARDs. Rituximab had the lowest ICER (£10,986) while abatacept the highest (£30,104). The strategy of introducing rituximab after adalimumab/etanercept (i.e. as a third line biologic) had an ICER of £13,797 per QALY, when compared to traditional DMARDs. The ICERs are as follows:

- Adalimumab/Etanercept vs. DMARDs: £15,962 per QALY
- Infliximab vs. DMARDs: £21,529 per QALY
- Rituximab (9-month dose) vs. DMARDs: £10,986 per QALY
- Abatacept vs. DMARDs: £30,104 per QALY
- Adalimumab/Etanercept + Rituximab vs. DMARDs: £13,797 per QALY

ICERs of Adalimumab/Etanercept (followed by DMARDs) vs. DMARDs presented in the sensitivity analyses varied from £11,191 per QALY to £26,456 per QALY, with adalimumab/etanercept being cost-effective in the vast majority of the scenarios explored.

The PSA results for 100 replications (for a cohort of 20,000 patients per replication) showed that at a WTP of £30,000 per QALY, adalimumab/etanercept followed by rituximab is the most cost-effective strategy, with the probability of being cost-effective being close to 1. At a WTP of £20,000 per QALY, rituximab followed by conventional DMARDs is cost-effective, with a probability of being cost-effective at around 60%, while there is a 40% (approx) chance of adalimumab/etanercept followed by conventional DMARDs being cost-effective. The submission, however, states: 'although the CEAC shows the probability that a treatment sequence is the most cost-effective option at various willingness to pay thresholds, it does not show all treatment strategies which can be considered cost-effective at these threshold(s)'. Therefore, the submission concludes that although the strategy of adalimumab/etanercept followed by conventional DMARDs is never shown to be cost-effective (submission Figure 3.3.2.1), the deterministic results showed that it is cost-effective, with an ICER of under £16,000 per QALY. The MS fails to point out though that both rituximab followed by conventional DMARDs and adalimumab/etanercept followed by rituximab had lower ICERs (£10,986 and £13,797 respectively).

## 2.2 Wyeth submission (Etanercept)

A Markov model (6-month cycle) was built to evaluate the cost-effectiveness of etanercept. The type of evaluation undertaken was a cost-utility analysis with outcomes measured in QALYs. However, Wyeth did not provide the model that produced the results presented in the submission.

Patients in the model were assumed to receive initial treatment with methotrexate, then switch to sulfasalazine, then switch to a '1st TNF inhibitor'. It is unclear in text which TNF inhibitor this was. However, cost data suggests that it is etanercept in all strategies compared. Therefore it is assumed that the population modelled were patients whose first failed TNF inhibitor was etanercept.

The three strategies compared are: second TNF inhibitor, DMARDs and 'Rituximab', all followed by traditional DMARDs and then the 'best supportive care' (salvage therapy). It is

unclear though which TNF inhibitor is compared in the 'second TNF inhibitor' strategy. Cost data suggests that it was an average of etanercept, adalimumab and infliximab combined with methotrexate. Similarly, in the 'DMARDs' strategy, it was unclear which DMARD was compared: cost data suggests that it was methotrexate. Finally, the DMARD following a TNF inhibitor seems to be sulfasalazine (again based on cost data).

Cost-effectiveness results were presented for a range of assumed HAQ changes of both the TNF inhibitor (etanercept/infliximab/adalimumab) and the conventional DMARDs.

**Adverse events**

Adverse events were included in the economic analysis. For simplicity, only serious adverse events were modelled, assuming that they last for one cycle (6 months) only. The cost of a serious adverse event was estimated at £1,181, which included 2 GP visits and 7 inpatient days. Text (submission page 33) suggests that various published sources were used for the rates of adverse events for each drug. Adverse events rates for all TNF inhibitors were assumed to be the same as etanercept. Data in the table suggest that rates of adverse events are more frequent in traditional DMARDs than in biologics.

**HAQ to Utility**

A linear mapping mechanism was used in order to convert HAQ scores to EQ-5D scores during each model cycle (EQ-5D = 0.76 - 0.28*HAQ).[4] It was assumed that patients experiencing serious adverse events would lose 0.05 units of utility (or 10% of a QALY) over one year.

**Results**

Results were presented for a range of assumed HAQ changes of both TNF inhibitor (etanercept/infliximab/adalimumab) and conventional DMARDs. The ICER for TNF inhibitors vs. conventional DMARDs was £14,501, when a HAQ drop of 0.55 was assumed for the TNF inhibitors and no change was assumed for the conventional DMARDs. The ICER for TNF inhibitors vs. Rituximab was £16,225, when a HAQ drop of 0.55 was assumed for the TNF inhibitors and a HAQ drop of 0.40 was assumed for Rituximab.

PSA results were not presented in the submission.

## 2.3 Schering-Plough Limited submission (Infliximab)

A patient-simulation/individual sampling model was used to evaluate the cost-effectiveness of infliximab. The type of evaluation undertaken was a cost-utility analysis with outcomes measured in QALYs.

Nine treatment sequences were compared in the cost-effectiveness analysis:

- Adalimumab/Etanercept/Infliximab/Rituximab/Abatacept each followed by a sequence of traditional DMARDs
- Adalimumab/Etanercept/Infliximab each followed by Rituximab and then a sequence of traditional DMARDs
- Sequence of traditional DMARDs

Patients in the model could receive a maximum of two biologic DMARDs followed by a maximum of three non-biologic DMARDs and were limited to a maximum of five treatments within each of the nine sequences. New biologic agents (tocilizumab, golimumab and certolizumab pegol) are excluded from the analysis since these drugs were considered not yet available in the UK.

The baseline characteristics of patients in the GO-AFTER trial where treatment with a TNF inhibitor (golimumab) following withdrawal from one or more previous TNF inhibitors (adalimumab, etanercept or infliximab) was investigated, were considered for the start of the model.

**Adverse events**

Adverse events were not included in the model although evidence on adverse events was included in the efficiency part of the submission.

**HAQ to Utility**

There was no mapping mechanism applied on EQ-5D scores. Utility gains or losses were modelled directly using a QoL measure. Each treatment was associated with an initial utility gain, which was estimated from BSRBR data.

**Results**

The base case results showed that adalimumab, etanercept, infliximab and rituximab (followed by traditional DMARDs) may represent cost-effective treatment options whereas abatacept (followed by traditional DMARDs) did not represent a cost-effective treatment option, when all strategies are compared with a sequence of traditional DMARDs. The ICERs were as follows:

- Adalimumab vs. DMARDs: £35,138 per QALY
- Etanercept vs. DMARDs: £35,898 per QALY
- Infliximab vs. DMARDs: £28,661 per QALY
- Abatacept vs. DMARDs: £44,769 per QALY
- Rituximab (9-month dose) vs. DMARDs: £17,422 per QALY
- Rituximab (6-month dose) vs. DMARDs: £27,161 per QALY

Further analysis, adding rituximab after the TNF inhibitors (adalimumab, etanercept, infliximab) was performed. Infliximab had the lowest ICER for both doses of rituximab explored (6-month/9-month) when compared to both traditional DMARDs and rituximab (both followed by traditional DMARDs). The ICERs were as follows:

vs. DMARDs

- Adalimumab+Rituximab (9-month dose): £27,998 per QALY
- Adalimumab+Rituximab (6-month dose): £32,345 per QALY
- Etanercept+ Rituximab (9-month dose): £27,936 per QALY
- Etanercept +Rituximab (6-month dose): £32,412 per QALY
- Infliximab Rituximab (9-month dose): £24,236 per QALY
- Infliximab Rituximab (6-month dose): £28,617 per QALY

vs. Rituximab

- Adalimumab+Rituximab (9-month dose): £41,747 per QALY
- Adalimumab+Rituximab (6-month dose): £39,084 per QALY
- Etanercept+ Rituximab (9-month dose): £42,477 per QALY
- Etanercept +Rituximab (6-month dose): £39,673 per QALY
- Infliximab Rituximab (9-month dose): £33,274 per QALY
- Infliximab Rituximab (6-month dose): £30,549 per QALY

Overall, when compared to DMARDs, rituximab had the lowest ICER for both 9-month (£17,422 per QALY) and 6-month doses (£27,161 per QALY). Among TNF inhibitors (etanercept, infliximab, adalimumab), infliximab had the lowest ICER (£28,661 per QALY).

ICERs in the sensitivity analyses varied from £16,752 per QALY (Rituximab vs. DMARDS, when a HAQ improvement of 0.01 per annum was assumed for all biologic DMARDS) to £58,850 per QALY (Infliximab+Rituximab vs. Rituximab, when the weight of the patient was assumed to be 120kg).

The PSA results showed that, when compared to traditional DMARDs, the probability of rituximab (9-month dose) being cost-effective was greater than 90% at a range of WTP thresholds greater than £20,000 per QALY. When a 6-month dose was assumed for rituximab, the probability of rituximab being cost-effective was marginally greater than the probability of infliximab being cost-effective, at WTP>£20,000 per QALY. The probability of infliximab (vs. DMARDs) being cost-effective was ~60% at £30,000 per QALY. When compared to rituximab, the probability of infliximab followed by rituximab being cost-effective was greater than 40% at £30,000 per QALY.

## 2.4 Roche submission (Rituximab)

A patient-level simulation was built to evaluate the cost-effectiveness of rituximab. The type of evaluation undertaken was a cost-utility analysis with outcomes measured in QALYs.

Rituximab was compared to all interventions included in the scope; adalimumab, etanercept, infliximab and abatacept. In addition, rituximab was compared to a strategy of traditional

DMARDs. In all strategies compared, the first active treatment was followed by salvage therapy consisting of leflunomide, gold and cyclosporin followed by palliative care. Response rates of leflunomide, gold and cyclosporin were assumed to be equivalent to MTX for this population. Comparison of rituximab against the new biological agents (tocilizumab, golimumab and certolizumab pegol) was not performed as these treatments were considered not used in routine clinical practice in the NHS.

**Adverse events**

Adverse events were not included in the economic analysis. The clinical section of the submission indicates that the incidence of adverse events was very similar across all treatments in the appraisal. Given that rituximab was compared head to head with each of the interventions in the scope, it was assumed that the costs of treating an adverse event would be the same in all strategies compared and therefore the cost-effectiveness ratios would not be affected by these costs.

**HAQ to Utility**

A quadratic mapping mechanism was used in order to convert HAQ scores to EQ-5D scores during each model cycle (EQ-5D=0.82-0.11*HAQ-0.07* $HAQ^2$). This equation was estimated through EQ-5D data collected in two Roche phase III trials (DMARD-IR) for tocilizumab. The linear mapping mechanism used by Bansback[5] (HUI3=0.76-0.28*HAQ+0.05*Female) was explored in a scenario analysis.

The model also assumed that the relationship of HAQ score to patient reported utility was independent of the number of previous biologics used. Moreover, for the base-case analysis, the model allowed for estimates of QALYs being less than one, when patients progress to very high HAQ scores. However, this relationship was not explored in the sensitivity analysis by adding a restriction to the negative QALY values.

**Results**

The base case results showed that rituximab dominates etanercept (Incremental Costs: -£13,246, Incremental QALYs: 0.0168), infliximab (Incremental Costs: -£10,490, Incremental QALYs:

0.0699) and abatacept (Incremental Costs: -£16,075, Incremental QALYs: 0.0606). When compared to adalimumab, rituximab was less costly (Incremental Costs: -£13,551) but also less effective (Incremental QALYs:-0.0436) with an ICER of £310,771 per QALY. When compared to the traditional DMARDs strategy, rituximab was more costly (Incremental Costs: £6,323) but also more effective (Incremental QALYs: 1.0705), with an ICER of £5,311 per QALY.

Overall, TNF inhibitors (etanercept, infliximab, adalimumab) were dominated by rituximab, i.e. rituximab was more effective and less costly. Adalimumab was marginally more effective but also more costly than rituximab, resulting in an ICER of £310,771 per QALY. When compared to traditional DMARDs, rituximab was cost effective at £5,311 per QALY.

ICERs in the sensitivity analyses varied from £4,898 per QALY (vs. traditional DMARDs when a 9-month time to retreatment was assumed for rituximab) to £326,397 per QALY (vs. Adalimumab when a linear mapping mechanism was assumed for the HAQ to QoL conversion), while in most of the scenarios rituximab dominated the other strategies (i.e. rituximab was less costly and more effective).

The PSA results for 1,000 Monte Carlo simulations showed that the probability of rituximab being cost-effective is 100% at a wide range of WTP thresholds (5,000 - £400,000 per QALY).

## 2.5 Bristol-Myers Squibb Pharmaceuticals LTD submission (Abatacept)

A patient-level simulation model was built to evaluate the cost-effectiveness of abatacept. The type of evaluation undertaken was a cost-utility analysis with outcomes measured in QALYs. Baseline patients characteristics were from the ATTAIN trial. Data from ATTAIN, REFLEX and BSRBR were used for the treatment efficacy of the drugs modelled.

Abatacept was compared to all interventions included in the scope; adalimumab, etanercept, infliximab and rituximab. However, TNF inhibitors were also grouped under a 'basket' of TNF inhibitors and these were the base case comparator. The rationale was reported as based on the conclusions from the NICE appraisal of the sequential use of TNF inhibitors.[6] In addition, the

submission argued that TNF inhibitors were grouped because there were no data available to conclude on the efficacy of different TNF inhibitors, after a failure of a first TNF inhibitor.

The 'basket' labelled TNF inhibitors was defined through use of market share data estimated through survey data (BMS data on file). These were: 22% etanercept, 52% adalimumab, 24% infliximab and 2% rituximab for the second line treatment, and 15% etanercept, 9% adalimumab, 37% infliximab and 38% rituximab for the third line, as presented on p. 134 of the submission. Patients in the model were randomly assigned to one of the three 'basket' treatments, based on these data, after excluding rituximab. Efficacy, costs and other parameters related to that therapy were applied to the proportion of patients receiving that therapy. Total costs and outcomes of the 'basket' treatment are the sum of the three 'basket' therapies.

There were two main comparisons. In the first comparison abatacept was compared to rituximab, both followed by infliximab, then traditional DMARDs, then palliative care. In the second comparison, abatacept was compared to a 'basket' of TNF inhibitors, both followed by another 'basket' of TNF inhibitors, then traditional DMARDs, then palliative care.

Traditional DMARDs were not considered as comparators in the economic analysis on the basis that this target population (RA patients with an inadequate response to TNF inhibitors) should have tried multiple traditional DMARDs, and so it was assumed that clinicians were unlikely to revert to these therapies. DMARDs were only included as part of the sequence of treatments after an insufficient response or intolerance to multiple biological therapies (after failure of three biologic DMARDs). After failing DMARDs, patients received NSAIDs only (palliative care).

Other new biologic agents were not considered as comparators for two reasons. Firstly, price information for the new biological therapies was not available at the time of writing. Secondly, new biological therapies were considered not routinely used in the NHS.

In summary, this submission did not consider a 'non-biologic' strategy. All strategies compared included at least two biologic DMARDs (patients with an inadequate response to one TNF inhibitor).

**Adverse events**

Adverse events were assumed to reduce quality of life as well as reducing costs. The following adverse events were included in the economic analysis: infusion related reaction, injection site

reactions, upper respiratory tract infection and urinary tract infection, rash, nausea, neutropenia, hypotension, leucopenia, severe allergic reaction and sinusitis. The sources for the rates of the adverse events were mainly published data.[7,8] Abatacept had the lowest rates of all adverse events apart from sinusitis.

**HAQ to Utility**

A linear mapping mechanism was used in order to convert HAQ scores to HUI3 scores during each model cycle (HUI3 = 0.76 - 0.28*HAQ + 0.05*Female).{8908/id] The submission discussed the available sources for conversion of HAQ to utility, and selected the formula above for the base case analysis, on the basis that this formula was used in previous RA appraisals and models[5,9,10] and was preferred over other algorithms{8902,8909/id} by the ERG in the original abatacept appraisal. The submission acknowledged that the average baseline HAQ score of 1.5 from the formula selected might not be appropriate for a population with an inadequate response to one TNF inhibitor, and therefore explored the EQ-5D approach[11] in sensitivity analysis

**Results**

The base case results showed that abatacept was cost-effective when compared to rituximab (both followed by infliximab as the third biologic) with an ICER of £20,438 per QALY. Abatacept was also cost-effective when compared to a 'basket' of TNF inhibitors (both followed by another 'basket' of TNF inhibitors) with an ICER of £23,019 per QALY. Overall, results showed the ICERs for abatacept were all below £30,000 whether compared with single or a 'basket' of TNF inhibitors, or rituximab.

ICERs for abatacept in the sensitivity analyses varied from £14,145 per QALY (vs. rituximab, when a 1.5% discount rate was assumed for QALYs) to £40,534 (vs. rituximab, when the abatacept HAQ progression rate was assumed to be 0.012 than -0.013 in the base case).

The PSA results showed that the probability of abatacept being cost-effective was 99% at £30,000 per QALY when compared to rituximab. When compared to a 'basket' of TNF inhibitors, the probability of abatacept being cost-effective was 97% at £30,000 per QALY. However, the submission failed to report any other PSA results (particularly below the £30,000 per QALY threshold). From the presented figures it seems that at £20,000 per QALY, both

rituximab and the 'basket' of TNF inhibitors were cost-effective when compared to abatacept, with the probabilities being >50% and >95% respectively.

## 2.6 Summary

A key issue is the appropriate comparator to be used. All but one submissions choose conventional DMARDs as their base case comparator. One submission has not considered a strategy of conventional DMARDs at all, assuming a switch to a third biologic in all strategies compared.

All submissions used the same type of economic evaluation, with cost per QALY being offered as efficiency measure.

There is some variation in the methods used and sources of data for important model inputs such as quality of life scores or baseline population characteristics. Three submissions considered adverse events in their model; however, methods and sources of rates and costs of adverse events varied.

# 3 Critique of indirect comparisons and mixed treatment comparisons (MTC) included in manufacturer submissions

Four of the manufacturers (Abbott, Schering-Plough, Roche and Bristol-Myers Squibb) used results from indirect comparisons and/or MTC to inform their model. This section provides a critical appraisal of these analyses and highlights issues that may impact on the validity of their results.

Before commencing on the critique of indirect comparison/MTC, it is pertinent to clarify the definition of these terms. NICE's Methods guide (2008) states that '*a mixed treatment comparison (MTC) includes trials that compare the interventions head-to-head and indirectly'* whereas an indirect comparison is a '*synthesis of data from a network of trials'*. These two terms have been used inconsistently and sometimes inter-changeably in some of the manufacturer submissions. In this section of the assessment report, all the syntheses of data from a network of trials without incorporating evidence from head-to-head trials are referred to as indirect comparisons in line with the Methods Guidance. This also avoids creating a false impression that direct evidence from head-to-head trials was included in these analyses. Only analyses that incorporate both direct and indirect evidence were referred to as MTC.

For the RA population defined in the scope of this appraisal (patients who had inadequate response to a TNF inhibitor), no head-to-head trial between the five technologies under assessment was identified by the assessment group and the manufacturers, and thus it was not possible to carry out an MTC. Indirect comparison was possible between rituximab and abatacept through placebo-controlled trials of respective drugs. This was conducted by Roche and Bristol-Myers Squibb.

Due to lack of head-to-head trials and a complete absence of placebo controlled trials for the three TNF inhibitors under assessment in the population defined by the scope, three manufacturers have attempted to carry out indirect comparisons/MTCs by extending inclusion criteria to RA population outside the scope (e.g. patients who had not been treated with a TNF inhibitor and/or patient who had not been treated with MTX). One head-to-head trial exists in this broader population and thus MTC combining direct and indirect evidence is possible. The key issue for this approach is whether basic assumptions with regard to clinical and methodological homogeneity and exchangeability of estimated treatment effects between trials held.

**Indirect comparisons in patient population specified in the scope**

Roche and Bristol-Myers Squibb performed indirect comparisons for the RA population defined in the scope using network meta-analyses / Bayesian methods (see Table 2, Roche TNF-IR IC and Bristol-Myers Squibb IC). Both indirect comparisons were based on the same placebo-controlled RCTs for rituximab (REFLEX trial) and abatacept (ATTAIN trial), and additionally included a placebo-controlled RCT for tocilizumab (RADIATE). A further golimumab RCT (GO-AFTER) was also in the Bristol-Myers Squibb analysis. No placebo-controlled trial for the patient population defined in the scope was identified for adalimumab, etanercept and infliximab, and thus it was not possible to include the three TNF inhibitors. The selection and inclusion of tocilizumab and golimumab trials in the indirect comparison seemed arbitrary as they provided no evidence regarding the relative effectiveness of rituximab versus abatacept. The inclusion of these trials had little impact on estimates of relative effectiveness (expressed as response rates to ACR response criteria and relative risks/odds ratios) between rituximab and abatacept compared to results from a pair-wise indirect comparison conducted by the assessment group based on the same rituximab and abatacept trials (see bottom of Table 3).

Roche and Bristol-Myers Squibb used results from indirect comparisons described above to inform their model (ACR responses for Roche; HAQ changes for Bristol-Myers Squibb). However this was restricted to the estimates of effectiveness for rituximab and abatacept and was not applicable for the estimates of effectiveness for TNF inhibitors. For TNF inhibitors, Roche used results from a separate MTC based on different patient populations outside the scope (described below) whereas Bristol-Myers Squibb used observational data from BSRBR registry. The comparisons of effectiveness between TNF inhibitors and rituximab/abatacept in their models were therefore *not* based on an indirect comparison or MTC.

**MTCs in patient population outside the scope**

Three manufacturers have carried out MTCs based on RCTs of RA population outside the scope (e.g. patients who had not been treated with a TNF inhibitor and/or patient who had not been treated with MTX; see Table 2: Abbott MTC, Schering-Plough MTC and Roche DMARD-IR MTC).

Due to the broad inclusion criteria beyond the scope of the appraisal, substantial clinical and statistical heterogeneity exists between the RCTs included in the MTCs. The basic requirement for indirect comparisons/MTCs regarding the exchangeability of relative treatment effects between the included studies could not be assumed and thus the validity of the results was questionable. The violation of the basic requirement was particularly

prominent in the MTCs conducted by Abbott and Schering-Plough in which RCTs of early RA patients who were naïve to MTX treatment were included in the analyses along with RCTs of late RA patients who had inadequate response to MTX and/or TNF inhibitors.

Despite the broad inclusion criteria for the MTCs, clinical and methodological similarity/ difference between included studies was only briefly described or not mentioned at all. Statistical heterogeneity between included studies were either not assessed or (where assessed) only dealt with by using random effects model without further exploration of potential source of heterogeneity. All the MTCs included a head-to-head trial (ATTEST, comparing infliximab to abatacept) but did not examine the direct evidence separately from indirect evidence. Consistency between direct and indirect evidence was not examined.

There is an appreciable difference between the results obtained from the three MTCs (which were based on population outside the scope) and the actual results (where available) observed in RCTs conducted in relevant populations defined in the scope (see Table 3). For ACR response criteria, results from these MTCs tend to overestimate the response rates (for both intervention and control arms but to a different extent) compared to the response rates observed in relevant RCTs.

The substantial heterogeneity among studies included in these MTCs and the discrepancy between the results from these analyses and those actually observed in RCTs raise serious concern with regard to the validity of the MTCs as well as the validity of economic evaluations that utilised data from them.

**Table 2. Summary of indirect comparisons (ICs) / mixed treatment comparisons (MTCs) reported in manufacturer submissions**

| | **Abbott (adalimumab) MTC** | **Schering-Plough (infliximab) MTC** | **Roche (rituximab): TNF-IR IC & DMARD-IR MTC** | **Bristol-Myers Squibb (abatacept) IC** |
|---|---|---|---|---|
| **Literature search** | Based on a number of previous studies/reports (Nixon et al. 2007, Wailoo 2008, Bristol-Myers Squibb submission and Evidence Review Group report for TA141) plus an updated search of PUBMED from January 1 2005 to May 31 2009. | Search of EMBASE, MEDLINE, MEDLINE in process and Cochrane Library from inception to April 2009; bibliographies of identified studies. | Search of Medline and EMBASE from 1990 through 2007. | Search of multiple databases from 1 January 1990 to 8 May 2009, conference abstracts, manufacturers and NICE web sites, bibliographies of identified studies. |
| **Inclusion criteria** | **Design**: clinical trial<br>**Population**: broader than scope (including patients not previous treated with TNF inhibitors and/or MTX)<br>**Intervention**: broader than scope (including anakinra, certolizumab pegol, golimumab, and tocilizumab)<br>**Outcome**: need to report ACR response<br>**Other**: at least 6 month follow-up time | **Design**: double-blind RCTs; >24 week (except rituximab trials)<br>**Population**: broader than scope (including RA patients of any stage)<br>**Intervention**: broader than scope (including certolizumab pegol, golimumab and tocilizumab)<br>**Outcome**: need to report ACR response criteria or mean change in HAQ score<br>**Other**: published as full papers in English | **Design**: RCTs of duration ≥ 6 months<br>**Population**: two analyses were performed:<br><br>***TNF inadequate responder (TNF-IR) indirect comparison***: same as scope;<br><br>***DMARD inadequate responder (DMARD-IR) MTC***:<br>**Population**: outside scope (including patients who had inadequate response to DMARD but predominantly not previously treated with a TNF inhibitor).<br>**Intervention**: broader than scope (including tocilizumab) | **Design**: RCTs<br>**Population**: same as scope<br>**Intervention**: broader than scope (including certolizumab pegol, golimumab and tocilizumab)<br>**Outcome**: clinically relevant outcomes<br>**Other**: published as full papers in English and conducted in Europe or America |

| | Abbott (adalimumab) MTC | Schering-Plough (infliximab) MTC | Roche (rituximab): TNF-IR IC & DMARD-IR MTC | Bristol-Myers Squibb (abatacept) IC |
|---|---|---|---|---|
| | | | **Outcome**: need to report ACR response criteria / ACR core disease parameters<br>**Other**: published as full papers in English, German, French, and Dutch. | |
| **Included studies** | **29 RCTs**, plus 1 'open label randomised study', 3 prospective cohort study, 1 study based on registry<br><br>**Within scope (2)**: abatacept (1), rituximab (1), plus 5 other studies of TNF inhibitors.<br>**Outside scope (27)**[c]: abatacept (4), adalimumab (5) etanercept (5), infliximab (2), anakinra (3), certolizumab pegol (1), golimumab (3), tocilizumab (5) | **34 RCTs**<br><br>**Within scope (2)**: rituximab (1), abatacept (1)<br>**Outside scope (32)**[c]: adalimumab (7), etanercept (5), infliximab (4), rituximab (2), abatacept (4), certolizumab pegol (3), golimumab (3), tocilizumab (5) | ***TNF-IR indirect comparison:* 3 RCTs**<br>**Within scope (2)**: rituximab (1), abatacept (1)<br>**Outside scope (1)**: tocilizumab (1)<br><br>***DMARD-IR MTC:* 18 RCTs**<br>**Within scope (0)**: none.<br>**Outside scope (18)** [a, c]: adalimumab (4), etanercept (4), infliximab (3), rituximab (2)[b], abatacept (3), tocilizumab (3) | **4 RCTs**<br><br>**Within scope (2)**: rituximab (1), abatacept (1)<br>**Outside scope (2)**: tocilizumab (1) and golimumab (1) |
| **Assessment of homogeneity and similarity* between included studies** | Not stated. | Not stated. Plots of the treatment effect on ACR response against baseline HAQ and disease duration were used to selected covariables into the analyses. | Homogeneity at each ACR response level was assessed using Q-statistics. Stated that 'baseline characteristics across the trials were comparable with respect to ACR core parameters'. | Not stated. |
| **Outcome analysed** | ACR20, 50 and 70. | ACR20, 50 and 70. | ACR20, 50 and 70. | Multiple outcomes including ACR responses; response criteria derived from DAS HAQ |

| | Abbott (adalimumab) MTC | Schering-Plough (infliximab) MTC | Roche (rituximab): TNF-IR IC & DMARD-IR MTC | Bristol-Myers Squibb (abatacept) IC |
|---|---|---|---|---|
| | | | | scores; withdrawal, DAS and HAQ change from baseline; various outcomes on adverse events, component outcomes of ACR criteria; SF-36 component summary scores |
| **Analytical methods** | Bayesian hierarchical models estimated with WinBUGS. ACR responses were modelled on a log-odds ratio scale. Log-odds ratios of responses were adjusted for addition of MTX, disease duration and baseline HAQ among other variables. Also used 'Fully-conditional predictive mean matching' to impute data. | Network meta-analyses conducted on an ordered logit scale. Analyses were performed both with and without adjustment of disease duration. | Analyses were performed with WinBUGS and conducted with non-informative priors. Results for TNF inhibitors were pooled. | Models were fitted using WinBUGS, employing Markov chain Monte Carlo (MCMC) simulation. Both fixed-effects and random-effects estimation was conducted for all analyses. |
| **Input into the manufacturer model** | Using Bayesian hierarchical models, posterior mean predicted treatment response rates (predicted for a patient with a disease duration of 11 years and an average HAQ score of 2.1). | Odds ratios (adjusted for disease duration) for ACR responses derived from indirect comparison were used in the model. | For rituximab and abatacept, ACR response rates from TNF-IR indirect comparisons were used. For TNF inhibitors, ACR response rates from DMARD-IR MTC were firstly discounted by 30% and then used in the model. | Results from indirect comparison for HAQ change were used in the model, but only for rituximab and abatacept. Data from registry (BSRBR) on HAQ change were used for TNF inhibitors. |
| **Comments** | Included trials of both early and late RA populations with very different treatment history (e.g. patients who had inadequate | Included trials of both early and late RA populations with very different treatment history (e.g. patients who had inadequate | Patient populations included in TNF-IR indirect comparison were in line with the scope. The major limitation of the analysis | Patient populations included in the indirect comparison were in line with the scope. The major limitation of the analysis was |

| | Abbott (adalimumab) MTC | Schering-Plough (infliximab) MTC | Roche (rituximab): TNF-IR IC & DMARD-IR MTC | Bristol-Myers Squibb (abatacept) IC |
|---|---|---|---|---|
| | response to a TNF inhibitor vs. patients who were naïve to TNF inhibitors vs. patients who were naïve to MTX). The basic requirement for indirect comparisons with regard to exchangeability of relative treatment effect between trials cannot be assumed and thus the validity of the results is questionable.<br><br>Also the indirect comparison included evidence from multiple study design (i.e. RCTs and observational studies). RCT evidence did not appear to have been analysed separately from evidence from observational studies. The nature of randomised comparison therefore may not have been preserved. In addition different search strategies and inclusion criteria were applied for different technologies. | response to a TNF inhibitor vs. patients who were naïve to TNF inhibitors vs. patients who were naïve to MTX). The basic requirement for indirect comparisons with regard to exchangeability of relative treatment effect between trials cannot be assumed. The validity of the results is questionable particularly because the indirect comparison used MTX as the reference standard (i.e. the hub of the evidence network) for comparison.<br><br>The proportional odds assumption of the ordered logit model (i.e. treatment effect was constant across ACR20, 50 and 70) did not seem to be consistent with observations from REFLEX and ATTAIN trials. | was that only one trial each was available for rituximab and abatacept and no trial was available for the three TNF inhibitors.<br><br>The inclusion of the tocilizumab trial appeared arbitrary as it provided no information regarding relative effectiveness of rituximab and abatacept. The inclusion of the trial had little impact on the estimates of relative effectiveness (in terms of ACR responses) between rituximab and abatacept compared to a pair-wise adjusted indirect comparison conducted by the assessment group based on the same trials (see bottom of Table 3).<br><br>Relative risks were translated into response rates using the pooled placebo response as baseline. Given the substantial heterogeneity between studies (e.g. placebo response rates for ACR20 ranged from 15% to | that only one trial each was available for rituximab and abatacept and no trial was available for the three TNF inhibitors.<br><br>The inclusion of the tocilizumab and golimumab trials appeared arbitrary as they provided no information regarding relative effectiveness of rituximab and abatacept. The inclusion of these trials had little impact on the estimates of relative effectiveness (in terms of ACR responses) between rituximab and abatacept compared to a pair-wise adjusted indirect comparison conducted by the assessment group based on the same trials (see bottom of Table 3). |

| | Abbott (adalimumab) MTC | Schering-Plough (infliximab) MTC | Roche (rituximab): TNF-IR IC & DMARD-IR MTC | Bristol-Myers Squibb (abatacept) IC |
|---|---|---|---|---|
| | | | 72% according to Figure 35 of Roche submission), the validity of pooling placebo response across studies and consequently the relative risks derived from it was questionable. | |

*As described in Song F et al. (2009) BMJ:338:b1147.

[a] Four studies were excluded from main analyses (but included in sensitivity analyses) because the 'treatment arms in these trials were fundamentally different from the remaining trials': no DMARD background treatment was provided in three studies; the other study evaluated combination therapy with a biologic agent and sulfasalazine.

[b] Approximately one-third of patients in this study had previously been treated with a TNF inhibitor.

[c] One trial included both abatacept and infliximab.

**Table 3 Comparison of ACR responses between data observed in RCTs and results of indirect comparisons (ICs) and mixed treatment comparisons (MTCs)**

| | Interventions/comparators* | ACR20 | ACR50 | ACR70 |
|---|---|---|---|---|
| **ACR responses** | | | | |
| **Control (traditional DMARD/placebo/none)** | | | | |
| **Data from RCTs** | **GO-AFTER (wk 14)** | 18% | 6% | 2% |
| | **REFLEX (wk 24)** | 18% | 5% | 1% |
| | **ATTAIN (wk 24)** | 20% | 4% | 2% |
| **Results from IC/MTC** | **Abbott MTC (model input)** | 25% | 10% | 4% |
| | **Roche DMARD-IR MTC** | 32% | 12% | 4% |
| | **Roche TNF-IR IC** | 15% | 4% | 1% |
| | **TNF inhibitors** | | | |
| **Data from RCT** | **GO-AFTER (golimumab 50 mg) wk 24** | 34% | 18% | 12% |
| | **GO-AFTER (golimumab 100 mg) wk 24** | 44% | 20% | 10% |
| **Results from IC/MTC** | **Abbott MTC (model input)** | 64% | 40% | 21% |
| | **Roche DMARD-IR MTC** | | | |
| | **Adalimumab** | 66% | 44% | 18% |
| | **Etanercept** | 64% | 36% | 14% |
| | **Infliximab** | 60% | 33% | 14% |
| | **30% degradation of Roche DMARD-IR MTC (model input)** | | | |
| | **Adalimumab** | 46% | 31% | 13% |
| | **Etanercept** | 45% | 25% | 10% |
| | **Infliximab** | 42% | 23% | 10% |
| | **Rituximab** | | | |
| **Data from RCT** | **REFLEX (rituximab) wk 24** | 51% | 27% | 12% |
| **Results from IC/MTC** | **Abbott MTC (model input)** | 62% | 38% | 20% |
| | **Roche DMARD-IR MTC** | 60% | 35% | 18% |
| | **Roche TNF-IR IC (model input)** | 46% | 23% | 14% |
| | **Abatacept** | | | |
| **Data from RCT** | **ATTAIN (abatacept) wk 24** | 50% | 20% | 10% |
| **Results from IC/MTC** | **Abbott MTC (model input)** | 55% | 31% | 15% |
| | **Roche DMARD-IR MTC** | 59% | 33% | 15% |
| | **Roche TNF-IR IC (model input)** | 43% | 22% | 8% |
| **Estimates of relative effectiveness** | | | | |
| | **TNF inhibitors vs. control (odds ratios)** | | | |
| **Data from RCT** | **GO-AFTER (golimumab 50 mg) wk 24** | 2.55 | 4.12 | 4.0 |
| | **GO-AFTER (golimumab 100 mg) wk 24** | 3.87 | 4.67 | 3.5 |
| **Results from IC/MTC** | **Schering-Plough MTC, adalimumab** | [redacted] | [redacted] | [redacted] |
| | **Schering-Plough MTC, etanercept** | | | |
| | **Schering-Plough MTC, infliximab** | | | |
| | **Bristol-Myers Squibb IC, golimumab 50 mg** | 2.55 | 4.30 | n/a |
| | **Bristol-Myers Squibb IC, golimumab 100 mg** | 3.90 | 4.89 | n/a |
| | **Rituximab vs. control (odds ratios)** | | | |
| **Data from RCT** | **REFLEX (rituximab) wk 24** | 4.77 | 7.00 | 13.67 |
| **Results from IC/MTC** | **Schering-Plough MTC** | [redacted] | [redacted] | [redacted] |
| | **Bristol-Myers Squibb IC** | 4.84 | 7.27 | 16.38 |

| | | | | |
|---|---|---|---|---|
| | **Abatacept vs. control (odds ratios)** | | | |
| **Data from RCT** | **ATTAIN (abatacept) wk 24** | 4.18 | 6.53 | 7.40 |
| **Results from IC/MTC** | **Schering-Plough MTC** | [redacted] | [redacted] | [redacted] |
| | **Bristol-Myers Squibb IC** | 4.20 | 6.98 | 9.28 |
| | **Rituximab vs. abatacept (relative risks)** | | | |
| **Results from ICs** | **Assessment group IC** | 1.12 | 1.00 | 1.80 |
| | **Roche TNF-IR IC** | 1.06 | 1.05 | 1.75 |
| | **Bristol-Myers Squibb IC (ratio of odds ratios)** | 1.14 | 1.07 | 1.85 |

*All interventions and comparators were assumed to be used with ongoing MTX
Commercial in confidence information is [redacted]

# 4 Further critique of manufacturers' models

A description of the models included in each of the manufacturer's submissions and a summary of results from this modelling is provided in section 2. A critique of indirect and/or mixed treatment comparisons that were used to inform the models is given in section 3. Building upon these two prior sections, this section aims to provide further critique of the manufacturers' models by highlighting issues and uncertainties related to data input and assumptions used..

Data input and assumptions used in manufacturer models are summarised in Table 4 at the end of this section. Key issues relating to characteristics of starting population, estimates of clinical effectiveness (short-term and long-term), mapping of effectiveness data to utility, discontinuation rule(s) and treatment duration, handling of adverse events and mortality, estimates of costs and other relevant factors are discussed below for each of the model.

## 4.1 Abbott (adalimumab)

**Characteristics of starting population**

The characteristics of the starting population were based on data from BSRBR[12] which is appropriate. These published data were collected in 2006 and are slightly dated. The starting population in Abbott model had a slightly higher HAQ score at baseline compared to the equivalent population described in the current BSR submission (2.1 vs. 2.0). The current BSR submission to NICE[13] (Section 4, Table 4-1) highlights a trend over the past 8 years that patients treated more recently have shorter disease durations, lower disease activity scores, lower HAQ scores and have tried fewer conventional DMARDs before starting a TNF inhibitor.

**Treatment sequence**

The stated assumptions that patients will have tried methotrexate, sulfasalazine and hydroxychloroquine (and thus these drugs are not evaluated) are clinically appropriate. The evaluated sequences include gold as the comparator or first traditional DMARD after failing biologics (see Table 1 in section 2). Sequences that consider gold early are increasingly

unlikely. Gold is now likely to be used much later during treatment (for example see the West Midlands survey, Appendix 10.11 of the main report). In addition, whilst azathioprine has limited efficacy, this drug would still be tried in patients with resistant disease. This drug should therefore be used late in the sequence.

**Estimates of clinical effectiveness - short-term**

Clinical effectiveness was estimated according to ACR response rates obtained from the manufacturer's MTC which included RCTs of very heterogeneous patient populations outside the scope of this appraisal as well as a few selected observational studies of relevant populations within the scope. As described in section 6.3, the validity of the MTC was questionable. The ACR responses estimated from the MTC for control groups (i.e. placebo or DMARDs for which patients had had inadequate response) were used for conventional DMARDs in the model. These response rates, if estimated correctedly, would not have reflected the response rates for a conventional DMARD which patients have not previously tried.

Mapping of ACR responses to HAQ change was based on an RCT (DE019) of adalimumab used as the *first* biologic therapy. Mapping using alternative data from PREMIER (an RCT of adalimumab in early RA, MTX naïve patients) suggested the relationship between ACR response to treatment and changes in HAQ score will differ depending on the population being treated. Therefore mapping based on data from a subgroup of patients in DE019 with a HAQ score greater than 2 was used by the manufacturer in a sensitivity analysis.

**Estimates of clinical effectiveness - long-term**

The base case assumed HAQ progression on biologics is the same as that of the general population (0.03 per year). An annual increase of 0.045 for conventional DMARD and 0.06 for non-responders was assumed. Zero HAQ progression on biologic treatment was explored in sensitivity analysis. Whilst previous analyses have considered the possibility that HAQ does not progress at all in a population of patients treated with a TNF inhibitor this assumption lacks face validity. Remission was achieved by 7% of patients in a large cohort of RA patients and minimum disease activity was achieved by around 20% including those on a TNF inhibitor.[14] On the basis that a majority of RA patients treated with a TNF inhibitor have continued disease activity it is not credible that HAQ does not change with time in this population.

The model assumed that following treatment withdrawal, the HAQ score would immediately worsen by an exactly equivalent amount to the initial improvement. A sensitivity analysis was conducted in which the HAQ score worsens by 75% of the initial gain. It seems appropriate to explore several possible scenarios. Patients experiencing a severe flare of disease are unlikely to left in this state and unlikely to suffer a prolonged worsening of function because of the short term use of corticosteroids combined with other DMARDs and/or a biologic as appropriate.

**Mapping of effectiveness data to utility**

HAQ scores were converted to EQ-5D according to equations developed by Ducournau 2009[15] using data from tocilizumab trials (OPTION and LITHE) in patients who had had inadequate response to MTX. Two equations (linear and non-linear) were available. The non-linear equation was used for the base case analysis while the linear equation was examined in sensitivity analyses.

**Discontinuation rule and treatment duration**

The model demands for an ACR50 response at six months in order that patients are eligible to continue treatment. This threshold appears too high compared to clinical practice. It is clear from BSRBR and other data that patients continue treatment with a TNF inhibitor despite not meeting NICE stipulated DAS28 criteria (so called 'stayers' in BSRBR analyses). This suggests that there is worthwhile clinical benefit despite a failure to meet thresholds (which are derived from populations and have limitations when applied to individual patients; see main report section 3.1.8.2 – DAS response criteria).

Withdrawal rates used in the base case analysis for TNF inhibitors are based on a shared frailty model previously developed by the Decision Support Unit using BSRBR data for patients receiving 2nd TNF inhibitor. Withdrawal rates for abatacept and rituximab was assumed to be the same as for TNF inhibitors.

**Handling of adverse events and mortality**

A reduction in mortality (independent of age, HAQ and co-morbidity) for patients on anti-TNF treatment was assumed based on Jacobsson et al.[16] This assumption was also applied for rituximab and abatacept. A hazard ratio of 0.92 for male and 0.52 for female was used. The reported mortality advantages for patients on TNF inhibitor treatment compared with conventional DMARDs, need great care in interpretation because of selection biases involved in treating patients with a TNF inhibitor which may not be sufficiently adjusted for. Sicker

individuals, those with cardiac failure and those with previous malignancies are much less likely to be treated.

**Estimates of costs**

Abbott state that the drug costs of adalimumab and etanercept are similar but fail to acknowledge that this only applies to adalimumab used every other week. The license for adalimumab permits dose increases so the drug may be administered every week (potentially doubling drug costs). European data, including the UK, suggests that around 8% of patients need an increase in their dose of adalimumab. This figure may be an underestimate as many investigators reported that financial constraints inhibited dose increases.[17]

The dose of leflunomide is 20 mg per day not 25 mg as stated. The stated dose for cyclosporin was 2.5 mg per kg. In practice this can range from 2.5 to 4 mg per kg.

The stated six outpatient visits *and* 11 nurse visits during the first 6 months for patients starting a TNF inhibitor are excessive for etanercept and adalimumab. For infliximab the necessary assessments can be done on the day a patient receives an infusion though it may be appropriate to include a nurse visit at other times to ensure that MTX safety is maintained. So, there will be 5 visits for infusions during the first 6 months. Blood and other monitoring can be done at these and additional 2 nurse visits would be needed to ensure a minimum of monthly checks were made.

Two outpatient visits and six nurse visits were assumed for monitoring after first 6 months. An outpatient visit every 3 months is appropriate say for a period of around 18 months but after this, in stable patients with well controlled disease, monitoring by a rheumatologist can be reduced to every 6 months. Frequency of blood testing for concomitant MTX can be done at nurse visits or GP practices where there are shared care agreements.

Disease related hospital costs (inpatient days and joint replacement procedures) were estimated based on HAQ band using data from the Norfolk Arthritis Register (NOAR) database.[18] Higher costs are more likely with higher HAQ scores but for items such as joint replacement this is only likely to apply to those with persistently raised HAQ scores (i.e. those with more fixed damage) rather than in whom HAQ scores rise due to flares of inflammatory disease. The latter group have a higher risk of hospitalisation because of this but rates in contemporary practice are low because of the use of corticosteroids.

## 4.2 Wyeth (etanercept)

Wyeth did not submit an electronic version of the model. Overall the description with regard to methods for identifying data and justification for the selection of data was very limited.

**Characteristics of starting population**

The mean age of the starting population was 53 years and was based on TEMPO trial. This is an RCT of TNF naïve patients (mean disease duration 6.6 years) who had *not* experienced treatment failure with MTX. The rationale for choosing this trial is not described. The modelling appears to start when patients first receive RA treatment (MTX) so it is not clear why a starting cohort of early RA patients was not chosen. The starting population in TEMPO was younger than the BSRBR cohort at study entry (mean age 56 years), but it is difficult to ascertain whether patients' age would be similar to BSRBR data (i.e. reflecting UK population and practice) when the patients reached the point of failing a TNF inhibitor. Other characteristics of the starting population were not described, including baseline HAQ score.

**Treatment sequence**

The identity of drugs in the treatment sequence was not clearly described. For example, the terms '1st TNF-α inhibitor', '2nd TNF-α inhibitor' and 'DMARD after TNF' were used without further clarification. The costs for the 2nd TNF inhibitor (the intervention under evaluation) were assumed to be the average of adalimumab, etanercept and infliximab + MTX. The assumed costs for the 2nd TNF inhibitor (£4,159.68) therefore do not reflect the (higher) costs for etanercept + MTX (£4,687.83) according to the table of unit costs provide in Wyeth submission.

**Estimates of clinical effectiveness - short-term**

Short-term HAQ improvement for TNF inhibitor (average -0.48; varied between -0.55 to -0.41 depending on reasons for withdrawal of previous TNF inhibitor) was based on data from the ReAct study, an observational study of switching to *adalimumab* after failing a TNF inhibitor. Short-term HAQ improvement for conventional DMARD was assumed to be zero according to BSRBR data. In contrast with the -0.48 observed in ReAct study, short-term HAQ improvement for TNF inhibitor observed in BSRBR was only -0.11 but this data was not used in the model. The estimates of effectiveness for the model were therefore taken from studies using different methods of data collection and thus inappropriate for comparison.

Various sources have been cited for HAQ improvements on other treatments but the citations may be incorrect (e.g. the cited references for DMARDs before first TNF inhibitor appears to be uncontrolled studies of second line biologics).

**Estimates of clinical effectiveness - long-term**

Long-term HAQ progression for patients on TNF inhibitors (and rituximab) was assumed to be zero according to Wick et al.[19] Various levels of HAQ progression were applied for patients on conventional DMARDs based on assumption.

**Mapping of effectiveness data to utility**

HAQ score was converted to EQ-5D using the equation reported by Brennan et al.[4]

**Discontinuation rule and treatment duration**

This was not described.

**Handling of adverse events and mortality**

Various probabilities of experiencing a serious adverse event was assigned for each treatment. The cited references included: a systematic review including probably first line biologic use, narrative reviews, methodological papers discussing HAQ and quality of life (incorrectly cited?). Mortality rates were adjusted according to change in HAQ using an equation but the source of the equation was not cited.

**Estimates of costs**

Resource use was based on HAQ score according to Taylor et al.[20]

## 4.3 Schering-Plough (infliximab)

**Characteristics of starting population**

The characteristics of starting population were based on GO-AFTER (a golimumab trial in patients who had inadequate response to TNF inhibitors): mean age 54, female 79%, baseline HAQ 1.61. The starting population was younger and had much lower baseline HAQ score compared to corresponding patients in BSRBR. Baseline utility (EQ-5D and SF-6D) was imputed from baseline HAQ using simple linear regression (lower HAQ corresponding to

higher utility). The consequence is that the estimated baseline utility may have been higher than it should be.

**Treatment sequence**

The model compared the five technologies against conventional DMARDs. It also compared each of the three TNF inhibitors followed by rituximab against conventional DMARDs.

**Estimates of clinical effectiveness - short-term**

Effectiveness for biologics was measures according ACR response, which were then mapped to EULAR response using an algorithm derived from GO-AFTER data.

Effectiveness data for biologics were obtained from a network meta-analysis of RCTs largely outside the scope. The validity of the network meta-analysis was questionable (see critiques on MTC). Effectiveness data for conventional DMARDs were obtained from EULAR response estimated by Brennan et al.[21] using regression analysis based on BSRBR data. It appears that EULAR response for corresponding patients who switched to a second TNF inhibitor (rather than conventional DMARDs) was available from the same analysis but this data was not used in the model. Instead estimates of effectiveness for TNF inhibitors were taken from the MTC and thus the data for comparative effectiveness were obtained from different sources that may not be comparable.

**Estimates of clinical effectiveness - long-term**

For patients receiving biologics, the base case analysis assumed zero utility progression. A sensitivity analysis was carried out assuming utility progression was equal to that observed in the BSRBR (by EULAR response), which suggest that utility worsens for EULAR good responders, is close to zero for moderate responders and improves marginally for none responders.[22] This seems counter-intuitive.

A further assumption was made that patients have the same radiological damage at the end of biologic treatment as at the start and therefore their ability to improve on further treatment was also retained. This was implemented in the model by holding age and disease duration constant for the time on biologic. The impact of this assumption is unclear and does not seem to have been explored in sensitivity analysis.

**Mapping of effectiveness data to utility**

For the base case, utility was estimated to be a function of EULAR response, treatment (on biologic treatment or not), health state utility at time of treatment initiation, age, disease duration, number of previous DMARDs and gender according to an analysis of BSRBR data.[23]

**Discontinuation rule and treatment duration**

Withdrawal data for TNF inhibitors was taken from the BSRBR analysis of patients receiving a second TNF inhibitor.[24] All patients receiving biologics who do not achieve a moderate or good EULAR response were discontinued from treatment at 6 months. Treatment withdrawals were assumed to be the same for rituximab and abatacept. This assumption may over-estimate the proportions of people who continue with these therapies although data are limited. For rituximab, in the German registry (RABBIT; Strangfeld et al[25]), 39% of people had no response after 6 months. However at 12 months 68% of patients had gone on to receive a second infusion. What proportion of the remaining 32% goes on to receive a further infusion is not yet known. Further attrition with subsequent courses is likely but difficult to estimate. Withdrawal data for conventional DMARDs was taken from Barton et al.[26]

**Handling of adverse events and mortality**

No impact of treatment on mortality was assumed in the model.

**Estimates of costs**

It was assumed that where possible the monitoring and administration for biologics and methotrexate was carried out concurrently. This seems appropriate. Two cost assumptions are presented for rituximab based on a 6 month or 9 month dosing frequency. The 6 month dosing frequency was based on market research rather than actual data from systematically collected data and may not be appropriate.

Vial optimisation was assumed in the base case. The assumptions are based on a questionnaire survey of rheumatology units (33% response rate). In many institutions vial sharing is achieved by central (pharmacy) preparation of infusions in advance of patient arrival. This can lead to drug wastage where patients are deemed not be fit for infusion or fail to turn up. In any case all any savings from vial sharing are dwarfed by dose escalation.[27] In the cited systematic review 44% of patients treated with infliximab had the drug dose increased.

## 4.4 Roche (rituximab)

**Characteristics of starting population**

The starting population was based on REFLEX trial: mean age 52.4 years, 81% female, disease duration 11.9 years, prior DMARD 2.5 (excluding MTX). Over half of the patients in REFLEX were recruited from USA and thus the cohort does not reflect UK population/practice, as exemplified in the much younger age compared to the BSRBR cohort.

**Treatment sequence**

The treatment sequence did not contain azathioprine.

As mentioned before, whilst azathioprine has limited efficacy, this drug would still be tried in patients with resistant disease an thus should therefore be used late in the sequence.

**Estimates of clinical effectiveness - short-term**

For rituximab and abatacept, ACR response rates from TNF-IR indirect comparisons (based on trials of patients who had failed one or more TNF inhibitor) were used. For TNF inhibitors, ACR response rates from DMARD-IR MTC (based on trials of patients naïve to TNF inhibitors) were firstly discounted by 30% and then used in the model. The validity for the DMARD-IR MTC was questionable (see critiques on MTC). The estimates of effectiveness for TNF inhibitors and rituximab/abatacept were therefore taken from different set of analyses which are not comparable.

**Estimates of clinical effectiveness - long-term**

Long-term HAQ progression for patients staying on treatment was set at zero (and also assumed to be zero for other biologics) according to observation from long-term extension of the REFLEX trial. An annual progression of 0.0225 was assumed for conventional DMARDs and 0.03 for palliative care. These were slightly lower than figures used in other manufacturer models.

**Mapping of effectiveness data to utility**

HAQ scores were converted to EQ-5D according to the non-linear equation developed by Ducournau 2009[15] using data from tocilizumab trials. An additional analysis that included age as a covariate in the non-linear model was also performed.

**Discontinuation rule and treatment duration**

Continuation of treatment (for all drugs) was subject to achieving an ACR 20 or higher at the end of the first 6-month cycle. Subsequently the same annual withdrawal rate (9.5%) for all biologics was assumed. This was based on Geborek 2002;[2] an average of two estimates for etanercept (8%) and infliximab (12%) used as the first biologic therapy. The same annual withdrawal rate (27%) was assumed for all traditional DMARDs. This was based on Bansback et al. 2005,[5] which cited Wolfe 1995[28] as the source. The data is likely to be outdated for some of the DMARDs.

**Handling of adverse events and mortality**

Adverse events were not included in the model.

**Estimates of costs**

Drug acquisition, administration and monitoring costs were estimated based on 5 year average. This may not accurately reflect the costs of drugs with higher start-up costs.

## 4.5 Bristol-Myers Squibb (abatacept)

**Characteristics of starting population**

The characteristics of the starting population were based on the ATTAIN RCT.[7] Using data from a recent UK cohort (BSRBR[12]) might have been a more appropriate approach. Compared to the BSRBR data, patients in the ATTAIN trial were on average slightly younger (58 vs. 53.4 years), had a longer disease duration (9 vs. 12.2 years) and more patients were receiving glucocorticoids (44-52% vs. 70.2%). The mean HAQ score was slightly lower in the ATTAIN trial than in BSRBR data (1.8 vs. 2.0) and the DAS28 score was slightly higher (6.5 vs. 6.4).

**Treatment sequence**

It was assumed that a conventional DMARD is not likely to be used after a failure of the 1st TNF inhibitor. This is arguable and it is likely that at least a proportion of rheumatologists may seek to try drugs such as leflunomide, gold or cyclosporin in this circumstance.

Penicillamine is included although it is used rarely today. The treatment sequences described, which were based on Barton 2004,[26] are credible.

**Estimates of clinical effectiveness - short-term**

Clinical effectiveness in the first six months was estimated using HAQ scores. For rituximab and abatacept these were obtained from an MTC (see MTC critique). For TNF inhibitors the estimate was based on a BSRBR data analysis by the Decision Support Unit for NICE[29] and it used the adjusted result for switchers with long duration of second treatment (the report concluded that this is a good estimate for a year of treatment). For conventional DMARDs data from early RA patients was used.[30-32] This data does not come from the population relevant to the scope (patients who failed a TNF inhibitor), it was probably not possible to identify more relevant data.

**Estimates of clinical effectiveness - long-term**

For long term HAQ progression there were two sets of data: one vs. rituximab and one vs. TNF inhibitors. For abatacept there was a further HAQ reduction on treatment based on an analysis of ATTAIN and an extension of rituximab trials[7,33] (-0.0729 and -0.013 respectively). For all other treatments (biologic drugs and conventional DMARDs) an annual increase in HAQ of 0.012 was assumed based on an Evidence Review Group STA report on rituximab (calculation was actually based on non-biologic data).[34] It is unclear why only patients on abatacept were assumed to further improve after the initial effect of the treatment, while all the other treatments are associated with deterioration.

**Mapping of effectiveness data to utility**

The algorithm mapping HAQ to utility was based on a conference abstract (Boggs 2002[35]). A linear equation (intercept 0.76, slope -0.28, female 0,05) was used for that purpose.

**Discontinuation rule and treatment duration**

The treatment duration was based on data from ATTAIN LTE for abatacept (clinical study report 029). For all other treatments data for first biologic use from Barton 2004[26] was utilised. As there was no data for adalimumab and rituximab, an average for all biologics was assumed. These may not be directly applicable to the present decision problem.

The data used in the model differs from that in BSRBR but it is unclear if these parameters affect results.

Discontinuation rates due to adverse events in the first six months for abatacept and rituximab were based on a mixed treatment comparison (see MTC critique). For all other treatments data from studies and reviews in TNF inhibitor naïve patients was used.[30,36-40] The applicability of their results might be limited, although for conventional DMARDs probably no data in the relevant population was available. The proportion of patients discontinuing due to adverse events was the lowest for abatacept (2.3%) and adalimumab (2.8%) and was the highest for conventional DMARDs (12-20%).

**Handling of adverse events and mortality**

The submission states that *"The event rates for abatacept and rituximab were derived from the mixed treatment comparison* [please see comments]. *The event rates for etanercept, adalimumab and infliximab were derived from individual trials and the event rates for conventional DMARDs were based on the literature (as used in Chen et al, 2006)."*

The utility loss due to adverse events was based on data from an Evidence Review Group STA report on erlotinib for relapsed non-small cell lung cancer.[41] Neutropenia and leucopaenia were associated with a utility loss of 0.15 and all other adverse events with a utility loss of 0.05. The applicability of these estimates to RA patients might be limited.

For mortality a HAQ mortality hazard ratio of 1.33 (95% CI: 1.10, 1.61) was used based on Wolfe 1994.[42]

**Estimates of costs**

The submission states that drug costs were based on the doses recommended in the drugs' SmPCs. Drug treatment costs were taken from MIMS. The number of abatacept vials used is assumed to be 2.85. This implies vial sharing. Currently fewer than 200 patients have been treated with abatacept in the UK. Presently it is unlikely that significant vial sharing can occur unless many more patients are treated. Since dose wastage for infliximab is assumed it would also be appropriate to model dose wastage with abatacept.

Drug administration costs were based on Chen 2005[43] and an Evidence Review Group STA on rituximab.[34] Monitoring costs were based on Barton 2004[26] and Curtis 2008.[44] These sources seem to be credible.

Hospitalisation resource use was based mainly on data from the Norfolk Arthritis Register Database (which included joint replacement).[45] Joint replacement surgery was included in the model separately and therefore it was deduced from the Norfolk data assuming that two thirds of RA hospitalisations are due to joint replacement (as stated in Pugner 2000[46]).

Time to joint replacement was assumed to be the same as in Barton 2004[26] and its impact on HAQ was based on Wolfe 1998.[47] The cost of joint replacement was assumed to be around £6,000.[48]

NHS Reference costs for 2007-08 were used for adverse events.

**Table 4 Data input and assumptions used in manufacturer models**

| | BRAM | Abbott | Wyeth | Schering-Plough | Roche | Bristol-Myers Squibb |
|---|---|---|---|---|---|---|
| Baseline characteristics | Based on BSRBR:[13] mean age 58 IQR 15, 81% female, baseline mean HAQ 2 | Based on BSRBR:[12] median age 58 years, 79% female, baseline HAQ 2.1 | Based on TEMPO trial (a trial in TNF inhibitor naïve patients): mean age 53 | Based on GO-AFTER (golimumab trial in patients who had inadequate response to TNF inhibitors): mean age 54, female 79%, baseline HAQ 1.61 | Based on REFLEX trial: mean age 52.4 years, 81% female, disease duration 11.9 years; prior DMARD 2.5 (excluding MTX) | Based on the ATTAIN trial. |
| Treatment sequence (after the failure of one TNF inhibitor) | Compared each of the five technologies against conventional DMARDs | Compared each of the five technologies against conventional DMARDs and a strategy of TNF inhibitor followed by rituximab. | Compared TNF inhibitor (as a class) to conventional DMARD and rituximab. | Compared each of the five technologies against conventional DMARDs. Also compared each of the three TNF inhibitors followed by rituximab against conventional DMARDs. | Compared each of the five technologies against conventional DMARDs. | Compared various strategies of sequential use of two biologics.<br><br>Although penicillamine is used rarely today the sequences described, which were based on Barton 2004,[26] are credible. |
| Estimates of clinical effectiveness – short-term | HAQ improvement based on data from the systematic review – where available; for | Based on ACR response rates obtained from MTC of trials outside the scope (see critiques | Short-term HAQ improvement was based on data from ReAct study, an observational study | ACR response was then mapped to EULAR response using an algorithm derived from GO- | Effectiveness of treatments (ACR responses)<br><br>Mapping HAQ to | Mean HAQ change in the first 6 months based on MTC for ABT and RTX and on BSRBR |

| | BRAM | Abbott | Wyeth | Schering-Plough | Roche | Bristol-Myers Squibb |
|---|---|---|---|---|---|---|
| | DMARDs halved effectiveness in early RA was used due to lack of data for the relevant population | on MTC).<br><br>Response rates were assumed to be equal for the three TNF inhibitors.<br>ACR response rates were mapped to HAQ changes using data from DE019 trial (adalimumab as first biologic therapy, outside scope). | of switching to adalimumab after failing a TNF inhibitor.<br><br>Assumed zero HAQ improvement for patients switched to conventional DMARDs according to BSRBR data.[12] | AFTER data.<br><br>For biologics, data were obtained from a network meta-analysis of RCTs largely outside the scope. The validity of the network meta-analysis was questionable (see critiques on MTC).<br><br>For conventional DMARDs, data were taken from a regression analysis of BSRBR data.[49] | utility | analysis[29] for TNF inhibitors. For DMARDs sources in early RA were used.[30-32] |
| Estimates of clinical effectiveness – long term | HAQ progression:<br>• 0 – on biologic treatments,<br>• 0.045/year on conventional DMARD,<br>• 0.06/year on palliation. | Base case assumed HAQ progression on biologics is the same as general population (0.03 per year[18]) and an annual increase of 0.045 for conventional DMARD and 0.06 | Different HAQ changes for medium and long term. Sources include published literature (however references seem to be incorrect) | Base case assumed utility progression on biologics is zero.<br><br>Patients' age and disease duration were held constant for the time on biologic based on | Long-term HAQ progression for patients staying on treatment was set at 0 (also assumed to be 0 for other biologics) according to observation from long-term extension | For abatacept only there was a further HAQ reduction on treatment based on an analysis of ATTAIN and an extension of rituximab trials.[7,33] For all other |

| | BRAM | Abbott | Wyeth | Schering-Plough | Roche | Bristol-Myers Squibb |
|---|---|---|---|---|---|---|
| | | for non-responders. Zero HAQ progression on biologic treatment was explored in sensitivity analysis.<br><br>Assumed that following treatment withdrawal, the HAQ score would immediately worsen by an exactly equivalent amount to the initial improvement. A sensitivity analysis was conducted in which the HAQ score worsens by 75% of the initial gain. | | the assumption that no radiologic progression occurs during treatment.<br><br>An annual HAQ progression of 0.042 was assumed according to Scott et al[50] for patients receiving conventional DMARDs. | of REFLEX trial.<br><br>An annual progression of 0.0225 was assumed for conventional DMARDs and 0.03 for palliative care. | treatments (biologic drugs and conventional DMARDs) an annual increase in HAQ of 0.012 was assumed based on an Evidence Review Group report on rituximab.[34] |
| Discontinuation rule and treatment duration | No formal quitting rule, but based on available data.<br>For long term survival on treatment | The minimal response required for continuation of treatment after the initial 6 month period is ACR50. | | Withdrawal data for TNF inhibitors (assumed the same for rituximab and abatacept) was taken from a BSRBR | Continuation of treatment (for all drugs) were subject to achieving an ACR 20 or higher at the end of first 6- | The treatment duration was based on data from ATTAIN LTE for abatacept (clinical study report 029). |

| | BRAM | Abbott | Wyeth | Schering-Plough | Roche | Bristol-Myers Squibb |
|---|---|---|---|---|---|---|
| | Weibull curves were fitted to the available data:<br>• TNF inhibitors – from BSR submission<br>• RTX – from REFLEX LTE<br>• ABT – from BMS submission<br>DMARDS – General Practice Research Database (GPRD) data | Withdrawal rates used in the base case analysis for TNF inhibitors are based on a shared frailty model previously developed by the Decision Support Unit using BSRBR data for patients receiving 2nd TNF inhibitor.<br>Withdrawal rates for abatacept and rituximab was assumed to be the same as for TNF inhibitors. | | analysis[51] inhibitor. Patients receiving biologics who do not achieve a moderate or good EULAR response were discontinued from treatment at 6 months.<br><br>Withdrawal data for conventional DMARDs was taken from Barton et al.[26]<br><br>The utility rebound following treatment discontinuation was equal to the initial utility gain. | months. Subsequently the same annual withdrawal rate (9.5%) for all biologics was assumed. This was based on Geborek 2002;[2] an average of two estimates for etanercept (8%) and infliximab (12%) used as the first biologic therapy.<br><br>Assumed the same annual withdrawal rate (27%) for all traditional DMARDs. This was based on Bansback et al. 2005,[5] which cited Wolfe 1995[28] as the source. | For all other treatments data for first biologic use from Barton 2004[26] was utilised.<br><br>Discontinuation rates due to adverse events in the first six months for abatacept and rituximab were based on a mixed treatment comparison (please see comments). For all other treatments data from studies and reviews in TNF inhibitor naïve patients was used.[30,36-40] |
| Mapping of effectiveness data to utility | Quadratic equation using dataset supplied by Hurst and reported in | HAQ scores were converted to EQ-5D according to equations (linear and | | Utility was estimated to be a function of EULAR response, treatment (on | HAQ scores were converted to EQ-5D according to the non-linear equation | The algorithm mapping HAQ to utility was based on a conference abstract |

| | BRAM | Abbott | Wyeth | Schering-Plough | Roche | Bristol-Myers Squibb |
|---|---|---|---|---|---|---|
| | Hurst *et al* (1997) in the absence of any more recent data. | non-linear) developed by Ducournau 2009[15] using data from tocilizumab trials. The non-linear equation was used for the base case analysis while the linear equation was examined in sensitivity analyses. | | biologic treatment or not), health state utility at time of treatment initiation, age, disease duration, number of previous DMARDs and gender. | developed by Ducournau 2009[15] using data from tocilizumab trials. | (Boggs 2002[35]) |
| Adverse events | Not incorporated into the model | Data on the occurrence of mild, moderate and serious adverse events for etanercept, infliximab and leflunomide were estimated from Gebroek 2002.[2] Adverse events for adalimumab, rituximab and abatacept were assumed to be the same as etanercept. | Data was obtained from various literature sources (however references seem to be wrong) | Not included in the model. | Not included in the model. | Occurrence of adverse events was based on MTC (please see comments). Utility decrements for AEs based on ERG report on erlotinib for relapsed non-small cell lung cancer STA |

| | BRAM | Abbott | Wyeth | Schering-Plough | Roche | Bristol-Myers Squibb |
|---|---|---|---|---|---|---|
| | | Rates of tuberculosis associated with each of the TNF inhibitors were based on data from the BSRBR[52] | | | | |
| Mortality | Basic mortality was taken from standard life tables. A relative risk (1.33) per unit HAQ was applied. For PSA a lognormal distribution was assumed (95% CI: 1.10, 1.61). | Assumed a reduction in mortality for patients receiving TNF inhibitors based on Jacobsson et al.[16] The reduction also applies to rituximab or abatacept. | At baseline 1.63 times standard mortality from UK life tables. Adjusted based on HAQ Δ(mortality) = Current mortality x [0.375 x Δ(HAQ)] | No impact of treatment on mortality was assumed in the model. | Mortality was adjusted according HAQ score based on Barton et al.[26] | HAQ mortality hazard ratio based on Wolfe 1994 |
| Drug costs and other costs | Costs are made up of drug and monitoring costs. A "start-up" cost reflects higher dosage and additional monitoring, as appropriate for each treatment.<br>Unit costs were | Based on Monthly Index of Medical Specialties (July 2009) assuming an average patient weight of 70kg.<br><br>Disease related hospital costs (inpatient days and joint replacement | Unit drug costs from BNF. Other costs from Curtis 2007 (Table 10 MS). | It was assumed that where possible the monitoring and administration for biologics and methotrexate was carried out concurrently. This seems appropriate. Two cost assumptions are | Drug acquisition, administration and monitoring costs were estimated based on 5 year average. This may not accurately reflect the costs of drugs with higher start-up costs. | Drug costs were based on MIMS. Drug administration costs were based on Chen 2005[43] and an Evidence Review Group on rituximab.[34] Monitoring costs were based on Barton 2004[26] and |

| | BRAM | Abbott | Wyeth | Schering-Plough | Roche | Bristol-Myers Squibb |
|---|---|---|---|---|---|---|
| | based on:<br>• For tests and visits – values from Chen 2006 inflated to 2008 and from Curtis 2008<br>For drugs – BNF 58 accessed online | procedures) were estimated based on HAQ band using data from the Norfolk Arthritis Register (NOAR) database[53] | | presented for rituximab based on a 6 month or 9 month dosing frequency. The 6 month dosing frequency was based on market research. | | Curtis 2008.[44] Hospitalisation resource use was based mainly on data from the Norfolk Arthritis Register Database (which included joint replacement).[45] The cost of joint replacement was assumed to be around £6,000.[48] NHS Reference costs for 2007-08 were used for adverse events. |
| Vial sharing | No vial sharing assumed | - | - | Dosing frequency for RTX; Vial sharing | The latest UK market research data indicates that rituximab is given every 8.7 months on average (GfK HealthCare, January 200; Roche data on file) (p.200) | The number of abatacept vials used is assumed to be 2.85. This implies vial sharing. Currently fewer than 200 patients have been treated with abatacept in the UK. Presently it is unlikely that |

| | BRAM | Abbott | Wyeth | Schering-Plough | Roche | Bristol-Myers Squibb |
|---|---|---|---|---|---|---|
| | | | | | | significant vial sharing can occur unless many more patients are treated. Since dose wastage for infliximab is assumed it would also be appropriate to model dose wastage with abatacept. |

## 4.6 Discussion

A few common issues were identified in the critique of manufacturer models:

- Starting population might not reflect UK population and practice
- Validity and uncertainty in translating effectiveness measures into utility
- Validity of indirect comparisons/MTCs carried out in trials of heterogeneous population
- Uncertainty in the relative effectiveness between individual TNF inhibitors and between these drugs and rituximab/abatacept
- Uncertainty related to the effectiveness of conventional DMARD
- Uncertainty in long-term disease progression on various treatments
- Different discontinuation rules
- Different assumptions with regard to dosing interval or vial optimisation

One particular challenge for this technology assessment/appraisal was an absence of RCTs for the three TNF inhibitors. It is the assessment group's view that evidence for technologies other than abatacept and rituximab is not appropriate for mixed treatment comparison (MTC) or indirect comparison. Different approaches have been used by the assessment group and the manufacturers under this circumstance. The assessment group evaluated evidence from observational studies in detail in the absence of relevant RCTs for adalimumab, etanercept and infliximab which is an unusual situation. The most appropriate data from either RCTs or observational studies for each of the technology under assessment were then selected for economic modelling.

In order to conduct a valid indirect comparison, a network of RCTs which are comparable with respect to patient population and study design is needed. As stated above, no RCT conducted in relevant patient population was found for the three TNF inhibitors. In order

to perform indirect comparisons beyond abatacept and rituximab, one or more assumptions have to be made (as the manufacturers did):

Assumption (1) - The effectiveness and safety of different TNF inhibitors are the same (e.g. evidence from trials of golimumab is applicable to the three TNF inhibitors under assessment);

Assumption (2) - Treatment effects are comparable between trials conducted in patients with different treatment history (DMARDs and biologics) and duration of RA, among other characteristics.

No evidence currently allows verification of assumption (1). To confirm or refute assumption (2) requires a systematic and comprehensive review far beyond the scope of this technology assessment / appraisal. Based on limited information provided in the MTCs included in the manufacturer submissions, it appears substantial clinical, methodological and statistical heterogeneity exists between trials conducted in populations beyond the scope of this appraisal. The validity of analyses based on this assumption is thus questionable. It should therefore be born in mind that potential uncertainties relating to these assumptions may not have been adequately reflected in the results of indirect comparisons/MTCs and the economic evaluations based on them.

# 5 Independent economic assessment

The Assessment Group's own independent analysis was carried out using the Birmingham Rheumatoid Arthritis Model (BRAM), which has been further updated to allow for a non-linear relationship between HAQ and utility. Additional coding has been added to the model to facilitate the use of probabilistic sensitivity analysis (PSA). This means putting a distribution around all parameters in the model. Unless there is a good reason to treat a parameter as fixed, some distribution has been used.

The BRAM is an individual sampling model. A large number of virtual patient histories is simulated with the accumulation of costs and QALYs. The basic model structure is shown in Figure 1. A complete description of the model follows here. A list of the assumptions in the model is given in Appendix 3.

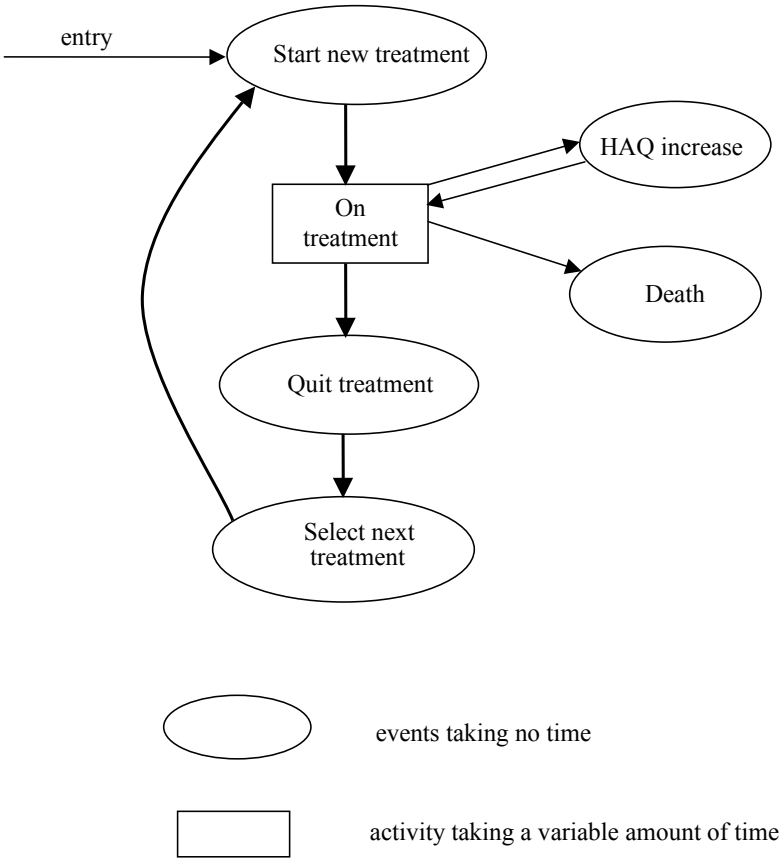
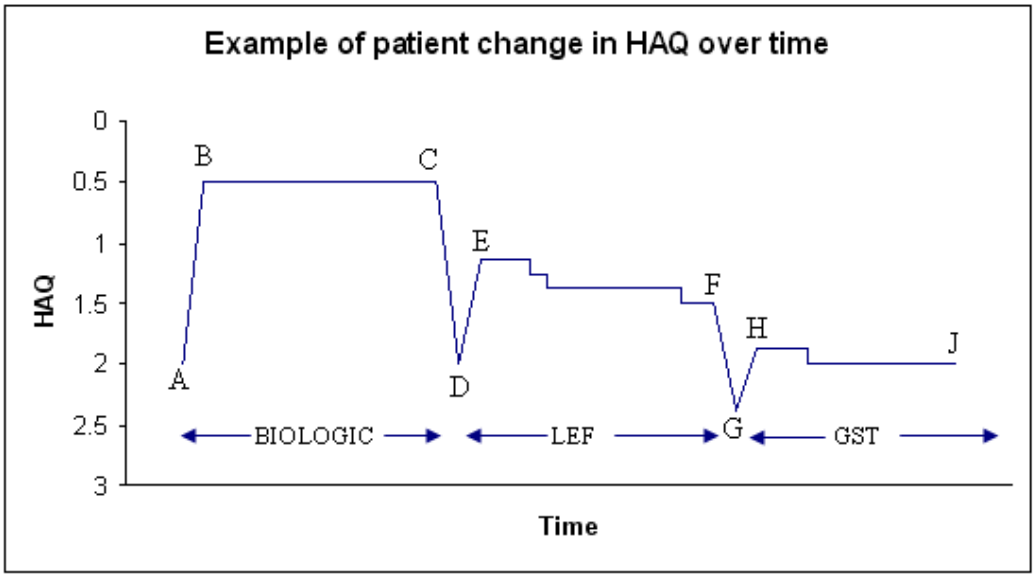



**Figure 1 Basic structure of the model**

## 5.1 Methods

Patients are assumed to follow a sequence of treatments. This involves: starting a treatment, spending some time on that treatment, quitting a treatment if it is toxic or ineffective, and starting the next treatment. The pattern is then repeated as long as active treatments are available. The final treatment in any strategy is palliation.

The HAQ disability index (see assessment report Appendix 10.1) is used as the marker for disease severity. Scores on this scale range from 0 (best) to 3 (worst) in multiples of 0.125. Patients' HAQ scores are assumed to improve (decrease) on starting a treatment and this improvement is lost on quitting the treatment regardless of reason for quitting. While on treatment, a patient's condition is assumed to decline slowly over time. This is modelled by occasional increases of 0.125 in HAQ score. The mean time between such increases in HAQ is allowed to vary by treatment; see Figure 2 for a possible HAQ trajectory. In the reference case analysis, HAQ is assumed to remain constant while a patient is successfully treated with a biological agent: this is modelled by a very large mean time to increase in HAQ.




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

**Figure 2 Possible trajectory of HAQ over time**

### 5.1.1 Strategies to be compared

The current appraisal is concerned solely with the decision to be made at the point of failure of a first TNF inhibitor. Accordingly, the starting population consists of patients who have reached that point in a sequence of treatments. Table 5 shows the treatment sequences compared in this appraisal.

**Table 5 Treatment sequences compared in the BRAM for this appraisal**

| Strategy name | ADA | ETN | IFX | RTX | ABT | DMARDs |
|---|---|---|---|---|---|---|
| 1st | ADA | ETN | IFX | RTX | ABT | LEF |
| 2nd | LEF | LEF | LEF | LEF | LEF | GST |
| 3rd | GST | GST | GST | GST | GST | CyA |
| 4th | CyA | CyA | CyA | CyA | CyA | AZA |
| 5th | AZA | AZA | AZA | AZA | AZA | Pall |
| 6th | Pall | Pall | Pall | Pall | Pall | |

All biologics are assumed to be taken in combination with methotrexate.

Note that previous versions of the BRAM used a starting population of DMARD-naïve patients, and generated a range of different decision populations within the model. Strategies compared also allowed different choices of treatment options depending on toxicity of previous treatments. While the coding to allow this flexibility remains within the model, such flexibility is not required within the present appraisal.

The choice of DMARDs following biologic therapy has been made in line with expected practice and excludes any DMARDs that are likely to have been used before biologic therapy.

### 5.1.2 Data used in the BRAM

What follows is a detailed description of the data and sources thereof. Updated literature reviews have been used wherever possible.

#### Initial patient data

Table 6 and Table 7 show the information about the initial population. As stated earlier, the initial population is a population immediately following failure of a first TNF inhibitor. The values are based on the BSRBR submission to NICE.[13]

**Table 6 Initial age and gender distribution**

| | Age (years) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 15-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65-74 | 75-84 | Total |
| Male | 0.0 | 0.4 | 1.9 | 5.2 | 6.5 | 3.8 | 1.2 | 19 |
| Female | 0.1 | 1.5 | 8.2 | 22.1 | 27.7 | 16.3 | 5.1 | 81 |

**Table 7 Starting distribution of HAQ scores**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| HAQ | 0.125 | 0.25 | 0.375 | 0.5 | 0.625 | 0.75 | 0.875 | 1 |
| % | 0.0 | 0.1 | 0.2 | 0.5 | 0.7 | 1.2 | 1.7 | 2.2 |
| HAQ | 1.125 | 1.25 | 1.375 | 1.5 | 1.625 | 1.75 | 1.875 | 2 |
| % | 2.9 | 3.6 | 4.3 | 5.1 | 5.8 | 6.6 | 7.2 | 7.7 |
| HAQ | 2.125 | 2.25 | 2.375 | 2.5 | 2.625 | 2.75 | 2.875 | 3 |
| % | 8.1 | 8.4 | 8.3 | 8.0 | 7.1 | 5.9 | 3.7 | 0.7 |

## Starting treatments

As in the previous version of the BRAM, the change in HAQ on starting a new DMARD is sampled on an individual basis and takes the form of a multiplier applied to the HAQ score on starting treatment. This multiplier is sampled from a beta distribution. The method used to estimate the parameters of the beta distribution is the same as in a previous report.[30]

To illustrate the method, consider the calculations used in the previous report for leflunomide. The data available were baseline HAQ mean 1.03 (SD 0.62) and HAQ improvement mean 0.48 (SD 0.5).[A1] An Excel spreadsheet was set up to create a starting population of 10,000 virtual patients with HAQ scores drawn from a normal distribution with mean and standard deviation supplied by the user. Each generated HAQ score was converted to the nearest legitimate value (multiples of 0.125 in the range 0-3). The parameters supplied were adjusted to compensate for the effect of this conversion, so that the mean and standard deviation of the population generated corresponded to the data. In this case, this involved adjusting the mean of the underlying distribution to 1.01 and the standard deviation to 0.66. The sample mean and standard deviation then agreed with the data.

A beta distribution was found to match the given mean and standard deviation for HAQ improvement. In this case the parameters were $a = 0.57$ and $b = 0.65$. Figure A1 shows the simulated population in this case. Each square within the graph represents a possible pair of

values of starting HAQ and HAQ on treatment: the darker the square, the larger the number of simulated patients with that pair of HAQ values. It can be seen that there was a high proportion of patients with equal HAQ on treatment compared with before treatment. In this example, the sampled population contained a large number of zero initial HAQ values. These are omitted from the graphs, but included in the calculations relating to HAQ improvement.

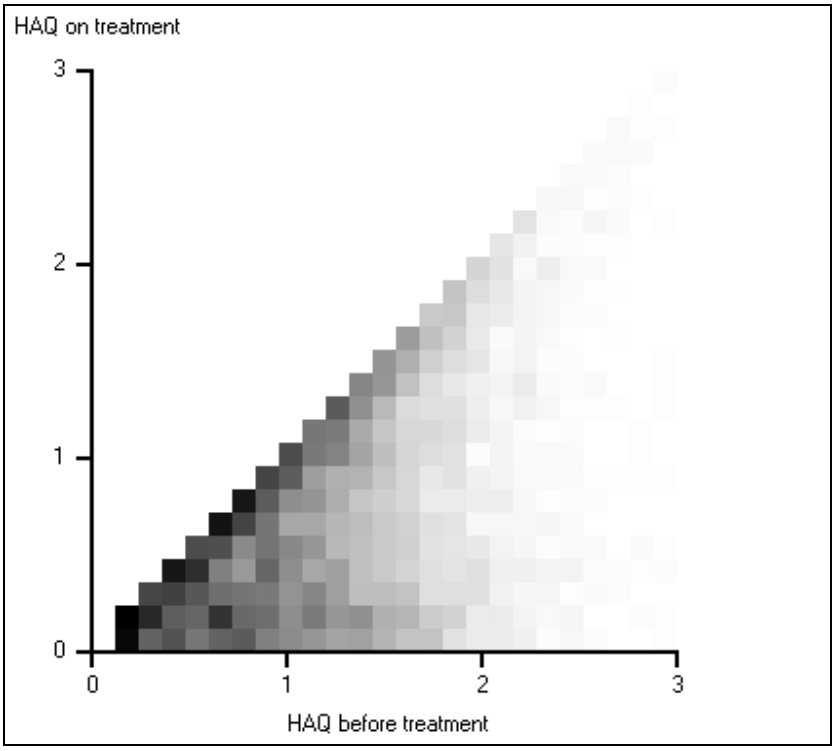



Figure A1. Modelled distribution of HAQ change on starting leflunomide (from previous report).

In the current report, for biologic DMARDs, the parameters have been re-estimated using the best available data for use immediately after a first TNF inhibitor. For conventional DMARDs to be used after biologics, the only available data was from trials in early RA. The effectiveness was halved for use in late RA.

When a patient starts a new treatment in the model, a random number is drawn to determine the HAQ improvement for that patient. Consider for example a patient about to start leflunomide with a HAQ score of 2 and suppose that the random number drawn is 0.5. The value of 0.5 indicates that the improvement multiplier should be at the median of the relevant distribution. In the case of leflunomide, using the values from Table 8, the median is 0.358 so the HAQ should improve by

$0.358 \times 2 = 0.716$. However, because HAQ is measured on a discrete scale, the improvement must be rounded to the nearest multiple of 0.125 which in this case is 0.75. The HAQ on treatment would then be $2 - 0.75 = 1.25$, and the 0.75 improvement (reduction) would be lost on quitting treatment. Had the starting HAQ score been 1, the improvement would have been 0.375 to give a HAQ on treatment of 0.625.

Table 8 shows the point estimates for the parameters of the beta distributions used. However, these values are not known with certainty, so some variation must be included in the probabilistic sensitivity analysis. In the absence of any obvious way of measuring the uncertainty around the parameters, an assumption was made that each could be independently sampled from a Normal distribution with a standard deviation equal to 0.1 times the point estimate. This is still likely to underestimate the uncertainty in these parameters, but is preferable to using fixed values. Note that although the same point estimates have been used for etanercept and infliximab, separate and independent samples have been used for the two drugs in the PSA. This principle has been applied throughout the model. In such cases, it is not known in which direction the difference between the treatments should be, but it is not a reasonable assumption that the treatments should take identical values.

**Table 8 Beta distributions for HAQ multipliers (point estimates)**

| Treatment | *A* | *B* | Mean | Source |
|---|---|---|---|---|
| ADA | 0.32 | 0.92 | 0.26 | Bombardieri 2007[54,55] |
| ETN | 0.21 | 0.75 | 0.22 | Bingham 2009[56] |
| IFX | 0.21 | 0.75 | 0.22 | Assume same as ETN |
| RTX | 0.20 | 0.75 | 0.21 | REFLEX[8,57,58] |
| ABT | 0.33 | 0.85 | 0.28 | ATTAIN[7,59-63] |
| LEF | 0.285 | 0.935 | 0.23 | Effectiveness halved from values used in previous report[30] |
| GST | 0.225 | 0.925 | 0.20 | |
| CyA | 0.065 | 0.325 | 0.17 | |
| AZA | 0.10 | 0.90 | 0.10 | |

For probabilistic sensitivity analysis, the values a and b are drawn from Normal distributions with standard deviation 0.1 times the point estimate (see text).

**Added in response to consultees' comments:** The values here give leflunomide a higher immediate effectiveness than any of the biologics. This is offset in part by the assumption described below about changes in HAQ while on treatment. However, it is stressed that these values are not being used for a comparison in which the biologic treatments replace leflunomide

in a sequence of treatments. Additional scenario analyses have been added to consider alternative assumptions.

**Time on treatments**

The model allows for two stages of early quitting of treatment. For conventional DMARDs, this facility has been used with parameters preserved from Chen *et al* (2006).[30] For TNF inhibitors and abatacept, a single stage of early quitting has been included in line with available data, while for rituximab no early quitting can be allowed, because it is necessary to model the full costs of each cycle of treatment. The values used are in Table 9. For long term survival on treatment, Weibull curves were fitted to the available data.

In the form used, a random variable *X* has a Weibull distribution with shape parameter *a* and scale parameter *b* if $\left(\frac{X}{b}\right)^a$ has an exponential distribution with unit mean. If $a = 1,$ the Weibull reduces to the exponential distribution with mean *b*; in any case *b* is the time until $\frac{1}{e} \approx 37\%$ of the original population remains. If $a < 1,$ then the hazard decreases with time; if $a > 1,$ the hazard increases. The values used are shown in Table 10. For convenience, the mean of the distribution is also shown for the point estimates of the parameters.

For TNF inhibitors, the same principle as for initial effectiveness has been applied: independent samples were drawn each time from the same distribution. For rituximab, the time sampled is then taken up to the nearest multiple of the assumed time between treatment cycles.

**Table 9 Probability of early quitting of treatment**

| Treatment | Parameter | Point estimate | Distribution | Source |
|---|---|---|---|---|
| ADA | Quit at 12 weeks | 9.9% | Beta(89,810) | Bombardieri (2007)[54,55] |
| | Toxicity if above | 56.2% | Beta(50,39) | |
| ETN | Quit at 13 weeks | 5.2% | Beta(21,385) | Bingham (2009)[64] and Buch (2005)[65] |
| | Toxicity if above | 16.7% | Beta(2,10) | Bingham (2009)[66] |
| IFX | Quit at 16 weeks | 23% | Beta(3,10) | OPPOSITE[67] |
| | Toxicity if above | 66.7% | Beta(2,1) | |
| RTX | No early withdrawal (see text) | | | |
| ABT | Quit at 6 months | 13.6% | Beta(35,223) | ATTAIN[7,59-62,68] |
| | Toxicity if above | 25.7% | Beta(9,26) | |

| Treatment | Parameter | Point estimate | Distribution | Source |
|---|---|---|---|---|
| LEF | Quit at 6 weeks | 13% | Beta(13,87) | Geborek (2002)[2] |
| | Quit 6-24 weeks | 30% | Beta(30,70) | |
| | Toxicity if above | 33.3% | Beta(10,20) | |
| GST | Quit at 6 weeks | 14% | Beta(10,62) | Hamilton (2001)[69] |
| | Quit 6-24 weeks | 27.1% | Beta(19.5,52.5) | |
| | Toxicity if above | 66.7% | Beta(6.5,13) | |
| CyA | Quit at 6 weeks | 8% | Beta(16,184) | Yocum (2000)[70] |
| | Quit 6-24 weeks | 24% | Beta(48,152) | |
| | Toxicity if above | 50% | Beta(24,24) | Marra (2001)[71] |
| AZA | Quit at 6 weeks | 15% | Beta(15,85) | Willkens (1995)[72] |
| | Quit 6-24 weeks | 25% | Beta(25,75) | |
| | Toxicity if above | 50% | Beta(12.5,12.5) | |

**Table 10 Times to quitting treatments**

| Treatment | *A* | 95%CI | *b* (years) | 95%CI | Mean (years) | Source |
|---|---|---|---|---|---|---|
| TNF inhibitors | 0.701 | (0.634,0.768) | 3.211 | (3.022,3.412) | 4.06 | BSRBR submission[73] |
| RTX | 0.474 | (0.403,0.545) | 5.1 | (3.742,6.951) | 11.31 | REFLEX long-term extension[74] |
| ABT | 0.81 | (0.734,0.886) | 5.49 | (5.166,5.834) | 6.17 | BMS submission[75] |
| LEF | 1 | (0.905,1.095) | 5.98 | (5.627,6.355) | 5.98 | GRPD database[76] |
| GST | 0.48 | (0.434,0.526) | 1.81 | (1.703,1.923) | 3.91 | |
| CyA | 0.5 | (0.452,0.548) | 4.35 | (4.094,4.623) | 8.70 | |
| AZA | 0.39 | (0.353,0.427) | 4.35 | (4.094,4.623) | 15.53 | |

Normal distributions used for parameter a; lognormal for parameter b. Standard errors for TNF inhibitors and RTX estimated from data. For other treatments, the same proportional variability as for TNF inhibitors has been assumed. Mean time on treatment based on the point estimate of the parameters.

Details of the implementation are as follows. For conventional DMARDs, the survival time is assumed to follow a distribution of the type shown in Figure A2, which is based on the data for leflunomide. The first step represents cessation of treatment after 6 weeks, which is assumed to be for toxicity. The second step represents cessation between 6 and 24 weeks after starting treatment, which could be for toxicity or inefficacy. At each appropriate stage in the running of the model, two variables $u1$ and $u2$ are each drawn from a uniform distribution between 0 and 1. Figure A3

shows how these numbers are used. The value of $u1$ is first used in the beta distribution to determine the HAQ improvement described earlier. Then $u2$ is used to determine the time on treatment.

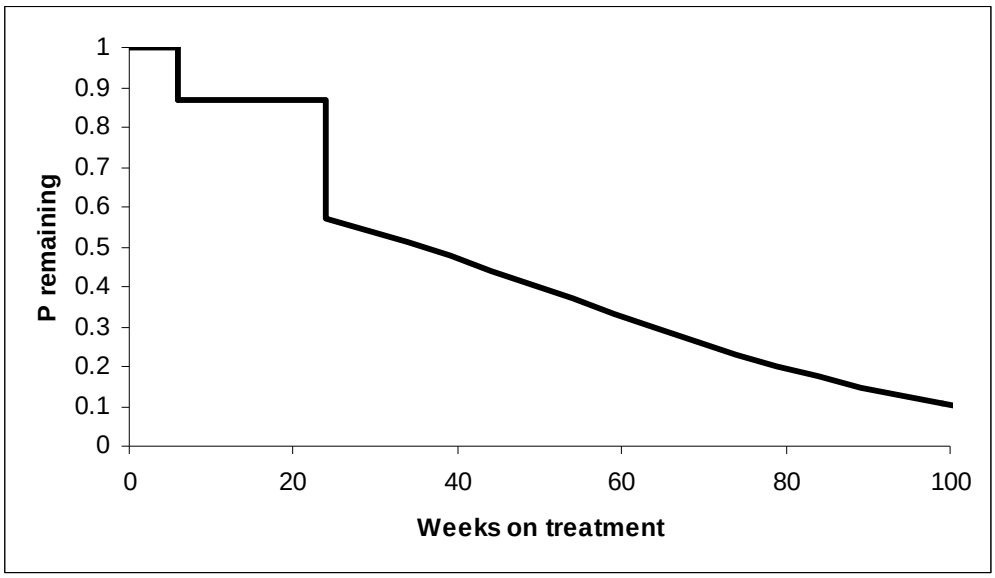



Figure A2. Survival time on a treatment (based on leflunomide data)

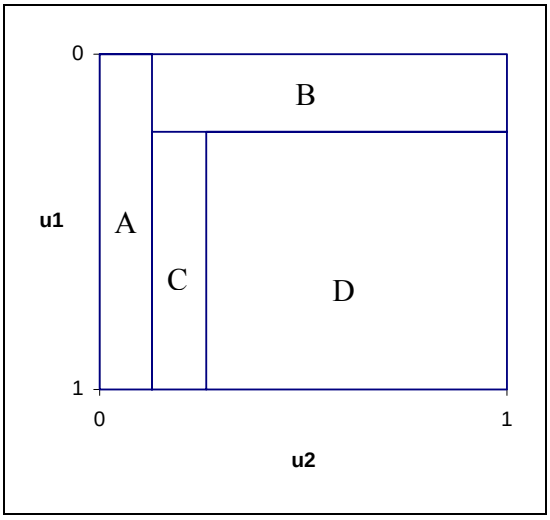



Figure A3. Early cessation of treatment

The four zones in figure A3 represent the following:

A withdrawal within 6 weeks (assumed due to toxicity);

B withdrawal between 6 and 24 weeks for inefficacy;

C withdrawal between 6 and 24 weeks for toxicity;

D remaining on the treatment after 24 weeks.

In implementation, critical values are calculated each time the population parameters are sampled for each treatment, so that the areas of the four zones in Figure A3 correspond to the probabilities sampled from the distributions indicated in Table 76. Then, for each individual, the values of $u1$ and $u2$ are compared to those critical values in the following way:

If $u2$ is below its lower critical value, then the individual is in Zone A, and quits through toxicity after 6 weeks.

Otherwise, $u1$ is compared to its critical value. If $u1$ is below the critical value, then the individual is in Zone B, and quits through ineffectiveness after 24 weeks.

Otherwise, $u2$ is compared to its higher critical value. If $u2$ is below this value, then the individual is in Zone C, and quits through toxicity after 24 weeks.

Otherwise, the individual is in Zone D, and remains on treatment beyond 24 weeks. The value of $u2$ is converted to a value from the appropriate Weibull distribution to determine the time on treatment.

For TNF inhibitors and abatacept, the 6 week quitting was not used, and the time shown in Table 76 was used in place of the 24 week limit used for conventional DMARDs. The implication of this is that for all modelled treatments except rituximab, those individuals with the lowest HAQ improvement on starting treatment all quit early.

**HAQ changes on treatment**

In the reference case analysis, it is assumed that HAQ remains constant while on any biologic treatment. Mean rates of HAQ increase of 0.045/year on conventional DMARDs and 0.06/year on palliation are modelled as mean times to increase (by 0.125) of 2.7 years and 2 years respectively. In the PSA these times are sampled from normal distributions with standard deviations 0.27 years

and 0.2 years respectively. Again, the times for the conventional DMARDs are sampled independently each time.

**Costs**

Costs are made up of drug costs plus monitoring costs. As in previous versions, the model includes an annual usage cost for each treatment, together with a "start-up" cost reflecting higher dosage and additional monitoring early in treatment, as appropriate for each treatment. Table 11 shows the unit costs for tests and visits and Table 12 the unit costs for drugs.

An administration cost of £141.83 is assumed for each dose of IFX, RTX, and ABT. This figure is inflated from the figure of £124 used in earlier versions of the BRAM. Monitoring assumptions for conventional DMARDs are shown in Table 13. It is assumed that monitoring for biologic therapies is included within the monitoring for methotrexate or administration costs, so no additional monitoring cost is included for these. Combining the monitoring assumptions with the unit costs then leads to start-up and annual usage costs as shown in Table 10.1 Test costs – first and subsequent years

| Treatment | Cost (pretreatment) | Cost (steady state yearly) | Cost (additional in first year) |
|---|---|---|---|
| MTX | £30.27 | £107.28 | £35.76 |
| LEF | £12.54 | £53.64 | £54.12 |
| GST | £12.54 | £108.36 | £180.60 |
| CyA | £16.93 | £52.68 | £53.96 |
| AZA | £12.45 | £107.28 | £53.64 |

Table 10.2 Test administration costs – first and subsequent years

| Treatment | Cost (pretreatment) | Cost (steady state yearly) | Cost (additional in first year) |
|---|---|---|---|
| MTX | £71 | £852 | £284 |
| LEF | £71 | £426 | £639 |
| GST | £71 | £852 | £1420 |
| CyA | £71 | £852 | £142 |
| AZA | £71 | £852 | £426 |

Table 14. Note that since these costings are based on fixed prices and monitoring rules, rather than measured resource use, the prices are not varied in the probabilistic sensitivity analysis. All costs were discounted at 3.5% per annum from the start of the model.

**Table 11 Unit costs for tests and visits**

| Test | Cost (£) | Source |
|---|---|---|
| FBC | 4.55 | Values from Chen *et al* (2006)[30] inflated to 2008 prices using the Hospital and Community Health Services inflation index (Curtis, 2008)[77] |
| ESR | 3.51 | |
| BCP | 4.39 | |
| CXR | 17.82 | |
| Urinalysis | 0.09 | |
| Visit | | |
| GP | 36 | Curtis (2008)[77] |
| Hospital outpatient | 71 | |
| Specialist nurse visit | 35.50 | Assumed half of outpatient visit |
| Administration of infusion | 141.83 | Chen *et al* (2006)[30] inflated to 2008 prices |

**Table 12 Unit costs for drugs**

| Treatment | Cost | Assumptions |
|---|---|---|
| ADA | £357.50 per dose | 26 doses per year |
| ETN | £178.75 per dose | 52 doses of 50 mg per year |
| INF | £419.62 per vial | 70 kg patient; drug wastage |
| RTX | £873.15 per 500mg vial | Dosage of 2×1000 mg every 8.7 months in base case |
| ABT | £242.17 per 250 mg | 750 mg every 4 weeks |
| MTX | 11.7p per tablet | 15 mg per week |
| LEF | £1.70 per day | 20 mg per day |
| GST | £11.23 per dose | 50 mg ampoule administered at GP visit |
| CyA | £5.37 per day | 225 mg per day |
| AZA | 40.3p per day | 150 mg per day |

Source: BNF 58 accessed online

Table 9.1 Drug costs – first and subsequent years

| Treatment | Cost (steady state yearly) | Cost (additional in first year) | Assumptions |
|---|---|---|---|
| ADA | £9295 | 0 | 26 doses per year |
| ETN | £9295 | 0 | 52 doses of 50 mg per year |
| INF | £7553.16 | £1258.86 | 6 doses per year; 1 additional dose in first year; 3 vials per dose |
| RTX | £4817.38 | 0 | Each course is 4 500mg vials; multiply by 12/8.7 for annual cost |

| | | | |
|---|---|---|---|
| ABT | £9444.63 | £726.51 | 13 doses of 750mg = 39 times unit cost; 1 additional dose in first year |
| MTX | £36.50 | 0 | 6 tablets per week for 52 weeks |
| LEF | £620.50 | 0 | 365 times daily cost |
| GST | £134.76 | £224.60 | Steady state 12 doses per year; additional 20 doses in first year |
| CyA | £1960.05 | 0 | 365 times daily cost |
| AZA | £147.10 | 0 | 365 times daily cost |

Table 9.2 Administration costs – first and subsequent years

| Treatment | Cost (steady state yearly) | Cost (additional in first year) | Assumptions |
|---|---|---|---|
| ADA | £0 | £106.50 | 3 visits to nurse specialist |
| ETN | £0 | £106.50 | 3 visits to nurse specialist |
| INF | £850.98 | £141.83 | 6 doses per year; 1 additional dose in first year |
| RTX | £391.26 | 0 | Two infusions per course; multiply by 12/8.7 for annual cost |
| ABT | £1843.79 | £141.83 | 13 infusions per year; 1 additional infusion in first year |
| MTX | 0 | 0 | |
| LEF | 0 | 0 | |
| GST | £432 | £720 | Steady state 12 doses per year; additional 20 doses in first year; GP visit for each dose |
| CyA | 0 | 0 | |
| AZA | 0 | 0 | |

**Table 13 Monitoring assumptions**

| Treatment | Pretreatment | On treatment |
|---|---|---|
| MTX | FBC, ESR, BCP, CXR | FBC, BCP every 2 weeks for 4 months then monthly |
| LEF | FBC, ESR, BCP, urinalysis | FBC every 2 weeks for 6 months, every 8 weeks thereafter. BCP monthly for 6 months, every 8 weeks thereafter. |
| GST | FBC, ESR, BCP, urinalysis | FBC, BCP, urinalysis every week for up to 21 injections, then every 2 weeks for 3 months, then every 3 weeks for 3 months, then monthly. Treatment given by i.m. injections |
| CyA | FBC, 2×BCP, ESR, urinalysis | FBC, BCP every 2 weeks for 4 months, then BCP monthly |
| AZA | FBC, ESR, BCP | FBC, BCP weekly for 6 weeks, then every 2 weeks for 3 visits, then monthly |
| Pall | | Outpatient visit every 3 months |

Table 10.1 Test costs – first and subsequent years

| Treatment | Cost (pretreatment) | Cost (steady state yearly) | Cost (additional in first year) |
|---|---|---|---|
| MTX | £30.27 | £107.28 | £35.76 |
| LEF | £12.54 | £53.64 | £54.12 |
| GST | £12.54 | £108.36 | £180.60 |
| CyA | £16.93 | £52.68 | £53.96 |
| AZA | £12.45 | £107.28 | £53.64 |

Table 10.2 Test administration costs – first and subsequent years

| Treatment | Cost (pretreatment) | Cost (steady state yearly) | Cost (additional in first year) |
|---|---|---|---|
| MTX | £71 | £852 | £284 |
| LEF | £71 | £426 | £639 |
| GST | £71 | £852 | £1420 |
| CyA | £71 | £852 | £142 |
| AZA | £71 | £852 | £426 |

**Table 14 Treatment costs**

| Treatment | Start-up (£) | Annual use (£) |
|---|---|---|
| ADA | 527.53 | 10290.78 |
| ETN | 527.53 | 10290.78 |
| IFX | 1821.72 | 9399.92 |
| RTX | 421.03 | 6204.42 |
| ABT | 1289.37 | 12284.20 |
| LEF | 776.66 | 1100.14 |
| GST | 2628.74 | 1527.12 |
| CyA | 283.89 | 2864.73 |
| AZA | 563.09 | 1106.38 |
| Pall | 0.00 | 284.00 |

Costs for hospitalisation and joint replacement are estimated by a cost per unit HAQ score. In the base case analysis, this was set at £1120 per unit HAQ. This was inflated from the previous figure of £860 per unit included in previous versions of the BRAM. Scenario analysis includes various alternative costings here based on industry submissions.

### Mortality

Basic mortality was taken from standard life tables. A relative risk per unit HAQ was applied. The point estimate for this relative risk was set to 1.33, sampling in the PSA from a lognormal distribution with 95% confidence interval (1.10,1.61).

### Quality of life (QoL) scores

In the reference case analysis, a quadratic equation was used to relate HAQ score to QoL score. This was of the form $QoL = a - b_1 HAQ - b_2 HAQ^2$, where the coefficients are shown in Table 15. It is noted that this equation gives negative values (indicating a state worse than death) for high HAQ scores. While this reflects the fact that individual patients in the dataset used to generate the equation gave EQ-5D responses which map to scores below zero on the standard UK tariff, it is acknowledged that the use of negative QoL scores is controversial. Accordingly, coding was added to allow such scores to be adjusted to zero in the model. This coding was used in scenario analysis.

**Table 15 Coefficients in HAQ to QoL equation**

| Coefficient | Point estimate | 95% confidence interval |
|---|---|---|

| | | |
|---|---|---|
| $A$ | 0.804 | (0.711,0.897) |
| $b_1$ | 0.203 | (0.054,0.351) |
| $b_2$ | 0.045 | (-0.007,0.096) |

Source: Birmingham analysis of dataset from Hurst. Note that the coefficient b2 takes a negative value in approximately 9 per cent of model replications. However, the positive value of b1 ensures that QoL decreases with increasing HAQ.

It was assumed that start and end effects could be modelled as one-off deductions proportional to the change in QoL score. The multiplier was set to a base case value of 0.2 (years), sampled from a Normal distribution with standard deviation 0.02 (separately for start and end).

Accumulated QALYs were discounted at 3.5% per annum from the starting point of the model.

## 5.2 Results

When an individual sampling model is run with a fixed parameter set, it must be run with a large number of patients to produce a precise estimate of the population mean cost and QALY differences between strategies. When such a model is run using probabilistic sensitivity analysis, the aim is to produce a distribution for the population outcomes which reflects the parameter uncertainty. This is done by sampling repeatedly from the joint distribution of parameters, and then for any parameter set, sampling a sufficient number of individuals.

Figure 3 shows the overall design of such a model run.

Note that a new set of patients is sampled for each parameter set, but the same patients are run through each of the possible strategies. Preliminary exploration suggested that 5000 patients per parameter set would be appropriate. For the reference case analysis, 2000 parameter sets were sampled from the parameter distributions as described in the previous section. For each parameter set, 5000 individual patient attributes were sampled and these patients were run through each of the six strategies defined in Table 5.

| | |
|---|---|
| Parameter set 1: $QoL = 0.7688 - 0.1723HAQ - 0.0506HAQ^2, etc$ | |
| | Patient 1.1: Female, starting age 45.0947, starting HAQ 2.875 |
| | Patient 1.2: Female, starting age 51.2780, starting HAQ 2.75 |
| | Repeat up to patient 1.*M* |
| Parameter set 2: $QoL = 0.8209 - 0.2087HAQ - 0.0359HAQ^2, etc$ | |
| | Patient 2.1: Female, starting age 50.6852, starting HAQ 2.625 |
| | Patient 2.2: Female, starting age 59.4641, starting HAQ 1.625 |
| | Repeat up to patient 2.*M* |
| Repeat up to parameter set *N*. | |

**Figure 3 Running an individual sampling model under probabilistic sensitivity analysis**

### 5.2.1 Reference case

The discounted lifetime costs and QALYs for each patient were calculated and the mean results for each parameter set output. The overall mean of these results forms the reference case estimate for the mean cost and QALY of each strategy: the 2.5 and 97.5 percentiles give the limits of the 95% credible interval. Note that these percentiles are likely to come from different parameter sets not just between strategies, but also for costs and QALYs for any particular strategy. These results are shown in Table 16. In each case, the lower credible limit for QALYs is negative, reflecting the use of an equation which allowed negative quality of life scores; see the scenario analysis for the effect of changing this assumption.

**Table 16 Results for single strategies in reference case analysis**

| Treatment | Mean Cost | 95% Credible Interval | | Mean QALY | 95% Credible Interval | |
|---|---|---|---|---|---|---|
| ADA | 74800 | 68800 | 81000 | 2.89 | -2.12 | 7.87 |
| ETN | 75100 | 68700 | 81500 | 2.80 | -2.21 | 7.84 |
| IFX | 73000 | 66100 | 79700 | 2.80 | -2.24 | 7.82 |
| RTX | 69400 | 62700 | 76400 | 3.10 | -1.78 | 7.95 |
| ABT | 93000 | 86200 | 100100 | 3.28 | -1.46 | 8.05 |
| DMARDs | 49000 | 43300 | 54900 | 2.13 | -3.27 | 7.46 |

Incremental results were obtained by subtraction for each parameter set, thus producing a sample of 2000 points from the incremental cost-effectiveness distribution between any pair of strategies. Again, the 95% credible interval can be found for cost and QALY differences: note that although the mean results can be inferred from Table 13 (subject to rounding effects), the relevant percentiles cannot. The results are shown in Table 14, which shows all pairwise comparisons. Scatterplots for the comparisons between the biologic strategies and conventional DMARDs

alone are shown in Figure 4, together with the cost-effectiveness acceptability curves for these five comparisons: the remaining scatterplots are in Appendix 1.

**Table 17 Differences between strategies in reference case analysis**

| Comparison | Diff Cost | 95% Credible Interval | | Diff QALY | 95% Credible Interval | |
|---|---|---|---|---|---|---|
| ADA – DMARDs | 25800 | 24100 | 27500 | 0.75 | 0.33 | 1.23 |
| ETN – DMARDs | 26100 | 24200 | 27900 | 0.67 | 0.30 | 1.10 |
| IFX – DMARDs | 24000 | 19500 | 26800 | 0.67 | 0.29 | 1.12 |
| RTX - DMARDs | 20400 | 17500 | 23200 | 0.96 | 0.41 | 1.61 |
| ABT - DMARDs | 44000 | 41300 | 46700 | 1.15 | 0.52 | 1.88 |
| ADA - RTX | 5400 | 2200 | 8700 | -0.21 | -0.52 | 0.03 |
| ETN - RTX | 5700 | 2400 | 9100 | -0.29 | -0.63 | -0.04 |
| IFX - RTX | 3600 | -1600 | 7600 | -0.30 | -0.62 | -0.05 |
| ABT - RTX | 23600 | 19800 | 27400 | 0.18 | -0.10 | 0.50 |
| ADA - ABT | -18200 | -21300 | -15200 | -0.39 | -0.77 | -0.12 |
| ETN - ABT | -18000 | -21200 | -14600 | -0.47 | -0.88 | -0.17 |
| IFX - ABT | -20000 | -25100 | -16200 | -0.48 | -0.88 | -0.17 |
| ADA - ETN | -300 | -2800 | 2100 | 0.08 | -0.09 | 0.29 |
| ADA - IFX | 1800 | -1400 | 6500 | 0.09 | -0.10 | 0.29 |
| ETN - IFX | 2000 | -1200 | 6800 | 0.00 | -0.17 | 0.19 |

Diff = difference, calculated by subtracting the value for the strategy named second from the value for the strategy named first.

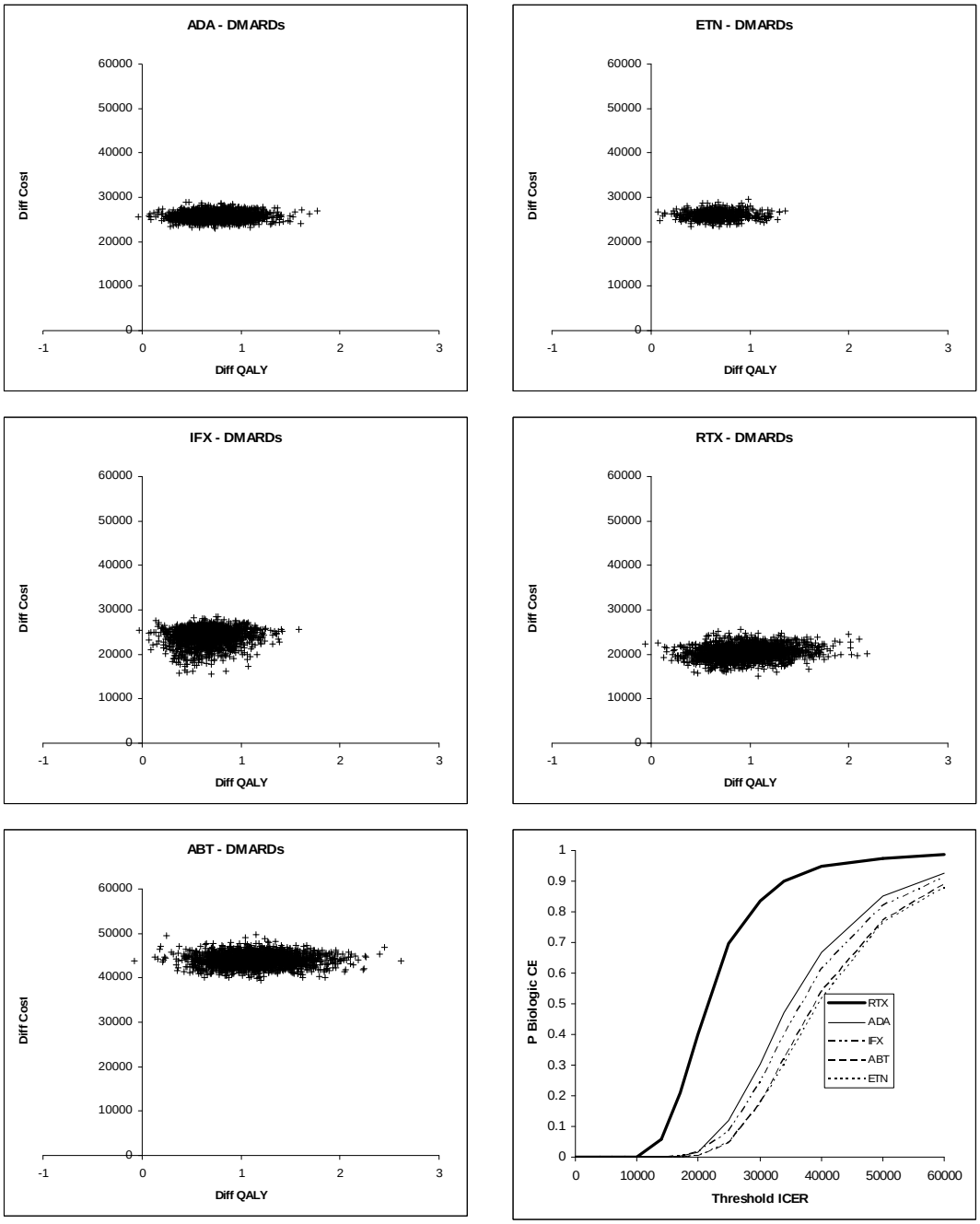



**Figure 4 Cost-effectiveness scatterplots for main comparisons in the reference case**

Similar remarks apply to the incremental cost-effectiveness ratio (ICER), which is found by dividing the difference in mean cost by the difference in mean QALY. Finally, the proportion of model replications for each biologic strategy appears cost-effective compared to any other is

shown, using a threshold ICER of £20,000/QALY and £30,000/QALY. These results are shown in Table 18.

**Table 18 ICERs for reference case analysis**

| Comparison | ICER | 95% Credible Interval | | Proportion of cases CE at £20,000/QALY | Proportion of cases CE at £30,000/QALY |
|---|---|---|---|---|---|
| ADA - DMARDs | 34300 | 20900 | 79100 | 0.02 | 0.30 |
| ETN - DMARDs | 38900 | 23500 | 89000 | 0.00 | 0.17 |
| IFX - DMARDs | 36100 | 21200 | 82000 | 0.02 | 0.24 |
| RTX - DMARDs | 21100 | 12800 | 49700 | 0.40 | 0.84 |
| ABT - DMARDs | 38400 | 23000 | 84700 | 0.00 | 0.17 |
| ADA - RTX | RTX | Not meaningful | | 0.00 | 0.00 |
| ETN - RTX | RTX | Not meaningful | | 0.00 | 0.00 |
| IFX - RTX | RTX | Not meaningful | | 0.00 | 0.00 |
| ABT - RTX | 130600 | 47900 | 0.00 | 0.00 | 0.00 |
| ADA - ABT | *46400* | *23100* | 0.99 | 0.90 | 0.90 |
| ETN - ABT | *37800* | *20100* | 0.98 | 0.77 | 0.79 |
| IFX - ABT | *41700* | *22000* | 0.99 | 0.84 | 0.84 |
| ADA – ETN | ADA | Not meaningful | | 0.84 | 0.84 |
| ADA – IFX | 20500 | Not meaningful | | 0.50 | 0.61 |
| ETN – IFX | 456700 | Not meaningful | | 0.20 | 0.24 |

CE = cost-effective. The proportion of cases cost-effective relates to the strategy given first on each line. ICER in italics means that the strategy named second is more costly and more effective. Where a strategy name is given in place of an ICER, the named strategy dominates its comparator (less costly and more effective). A 95% credible interval for the ICER is not meaningful in cases where the cost-effectiveness scatterplot is not confined to one half of the plane.

### 5.2.2 Scenario analysis

A number of different scenarios have been run. Details of each scenario and the results are in Appendix 2), and a summary is in Table 19, Table 20 and Table 21. It should be noted that although it is always possible to give a result based on the mean of the probabilistic analysis, the results for comparison between TNF inhibitors almost invariably are from a distribution covering all four quadrants of the cost-effectiveness plane, and thus the mean results are subject to enormous uncertainty in that case. The sole exception to this is the scenario "Vary time on TNF inhibitors".

**Table 19 Results from scenario analysis: Comparisons against DMARDs strategy (ICER in £/QALY)**

| Scenario | ADA – DMARDs | ETN – DMARDs | IFX - DMARDs | RTX - DMARDs | ABT - DMARDs |
|---|---|---|---|---|---|
| Reference | 34300 | 38900 | 36100 | 21100 | 38400 |
| Vary time on TNF inhibitors | 34300 | 38400 | 37700 | 21200 | 38500 |
| Same time on all biologics | 34400 | 38700 | 35900 | 21100 | 39500 |
| RTX cycle time 6 months | 34300 | 38900 | 35900 | 32600 | 38400 |
| RTX cycle time 11.6 months | 34200 | 38800 | 35900 | 11400 | 38400 |
| Poor late DMARDs | 28100 | 31100 | 28800 | 16300 | 32100 |
| HAQ change on biologics | 61300 | 76300 | 68900 | 46000 | 63300 |
| Adverse event costs included | 34700 | 39900 | 36800 | 22500 | 38800 |
| No offset costs | 36900 | 41400 | 38600 | 23600 | 41000 |
| Extra cost for palliation | 33400 | 37800 | 35000 | 20100 | 37600 |
| No negative QoL scores | 48600 | 56500 | 52100 | 30700 | 52800 |
| Linear equation HAQ to QoL | 38600 | 43800 | 40600 | 23700 | 42300 |

Small variations in results where neither strategy had changed parameters reflect the first and second order sampling in the model.

**Table 20 Results from scenario analysis: Comparisons of other biologics against RTX (ICER in £/QALY)**

| Scenario | ADA - RTX | ETN – RTX | IFX - RTX | ABT - RTX |
|---|---|---|---|---|
| Reference | RTX | RTX | RTX | 130600 |
| Vary time on TNF inhibitors | RTX | RTX | *4100* | 131800 |
| Same time on all biologics | 206000 | RTX | RTX | 131200 |
| RTX cycle time 6 months | *430* | RTX | *14700* | 51500 |
| RTX cycle time 11.6 months | RTX | RTX | RTX | 861100 |
| Poor late DMARDs | RTX | RTX | RTX | 158600 |
| HAQ change on biologics | RTX | RTX | RTX | 96400 |
| Adverse event costs included | RTX | RTX | RTX | 126100 |
| No offset costs | RTX | RTX | RTX | 134100 |
| Extra cost for palliation | RTX | RTX | RTX | 131000 |
| No negative QoL scores | RTX | RTX | RTX | 140700 |
| Linear equation HAQ to QoL | RTX | RTX | RTX | 130900 |

ICER in italics means that the strategy named second is more costly and more effective. Where a strategy name is given in place of an ICER, the named strategy dominates its comparator (less costly and more effective).

**Table 21 Comparisons between biologics other than RTX (ICER in £/QALY)**

| Scenario | ADA - ABT | ETN - ABT | IFX - ABT | ADA - ETN | ADA - IFX | ETN - IFX |
|---|---|---|---|---|---|---|
| Reference | *46400* | *37800* | *41700* | ADA | 20500 | 456700 |
| Vary time on TNF inhibitors | *47700* | *38900* | *39100* | *72800* | 28700 | 39300 |
| Same time on all biologics | *84100* | *42700* | *53700* | ADA | 21600 | 351500 |
| RTX cycle time 6 months | *46300* | *37800* | *42000* | ADA | 21700 | 1325400 |
| RTX cycle time 11.6 months | *46400* | *37800* | *41800* | ADA | 20700 | 591000 |
| Poor late DMARDs | *40100* | *33500* | *36900* | ADA | 20600 | 316000 |
| HAQ change on biologics | *66500* | *50600* | *57600* | ADA | 24300 | IFX |
| Adverse event costs included | *46700* | *37400* | *41700* | ADA | 19000 | 502600 |
| No offset costs | *49000* | *40500* | *44400* | ADA | 23500 | 460000 |
| Extra cost for palliation | *45800* | *37300* | *41200* | ADA | 20300 | 452000 |
| No negative QoL scores | *60300* | *48300* | *53700* | ADA | 25300 | 7430000 |
| Linear equation HAQ to QoL | *49100* | *40300* | *44600* | ADA | 23100 | 667000 |

ICER in italics means that the strategy named second is more costly and more effective. Where a strategy name is given in place of an ICER, the named strategy dominates its comparator (less costly and more effective). Small variations in results where neither strategy had changed parameters reflect the first and second order sampling in the model. It should be stressed that the comparisons between TNF inhibitors are based in each case (except "Vary time on TNF inhibitors") on the mean values from a distribution which covers all four quadrants of the cost-effectiveness plane.

### 5.2.3 Summary of model results

The reference case model results show similar costs and QALYs for the TNF inhibitors, with somewhat lower costs and QALYs for rituximab and higher costs and QALYs for abatacept. Compared to conventional DMARDs alone, the incremental cost-effectiveness ratio for rituximab is somewhat lower than for the other biologics. Rituximab dominates the TNF inhibitors (lower cost and more QALYs). The ICER for abatacept compared to rituximab is over £100,000/QALY.