Evidence Review Group Report

Tocilizumab for the treatment of rheumatoid arthritis

Produced by: West Midlands Health Technology Assessment Collaboration

[redacted]

Authors: Catherine Meads, Lecturer[1]
Mark Jit, Senior Lecturer[2]
Angelos Tsourapas, Research Fellow[2]
Khalid Ashfaq, Systematic Reviewer[1]
Martin Connock, Systematic Reviewer[1]
Anne Fry-Smith, Information Specialist[1]
Paresh Jobanputra, Consultant and Clinical Service Lead[3]

[1] Unit of Public Health, Epidemiology and Biostatistics, University of Birmingham
[2] Unit of Health Economics, University of Birmingham
[3] Department of Rheumatology, Selly Oak Hospital, University Hospital Birmingham NHS Foundation Trust

Correspondence to: Catherine Meads, [redacted]

Date completed: April 2009

Source of Funding: This report was commissioned by the NIHR HTA Programme as project number 08/90/01.

Declared competing interests of the authors: No competing interests - Meads, Jit, Tsourapas, Ashfaq, Connock, Fry-Smith. Competing interest – Jobanputra (one day's paid consultancy for Roche in 2008 for a preparatory meeting for their submission to NICE).

Acknowledgements: Linda Briscoe and Susan Pritchard for their administrative assistance with this project.

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.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

This report should be referenced as follows: Meads C, Jit M, Tsourapas A, Ashfaq K, Connock M, Fry-Smith A, Jobanputra P. Tocilizumab for the treatment of rheumatoid arthritis. A single technology appraisal. West Midlands, Health Technology Assessment Collaboration. 2009.

Contribution of authors:
Meads wrote the clinical effectiveness sections, edited the report and managed the team. Jit analysed the economic model and wrote the cost effectiveness sections. Tsourapas analysed the mixed treatment comparison and wrote the MTC section. Ashfaq conducted background research and wrote most of Appendix 4. Connock re-analysed the meta-analyses and wrote the meta-analysis part. Fry-Smith conducted additional searches and wrote sections 4.1.1 and 5.0.0. Jobanputra advised on clinical aspects of rheumatoid arthritis and its treatment and made numerous manuscript suggestions.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Evidence Review Group Report 1
1 SUMMARY 6
1.1 Scope of the submission 6
1.2 Summary of submitted clinical effectiveness evidence 6
1.3 Summary of submitted cost effectiveness evidence 6
1.4 Commentary on the robustness of submitted evidence 7
1.4.1 Strengths 7
1.4.2 Weaknesses 7
1.4.3 Areas of uncertainty 8
1.5 Key issues 8
2 BACKGROUND 10
2.1 Critique of manufacturer's description of underlying health problem 10
2.2 Critique of manufacturer's overview of current service provision 11
3 Critique of manufacturer's definition of decision problem 12
Population 12
Intervention 12
3.1 Population 14
3.2 Intervention 15
3.3 Comparators 16
3.4 Outcomes 17
3.4.1 ACR outcome and its components 18
3.4.2 Disease Activity Score 20
3.4.3 Other outcome measures 20
3.5 Time frame 20
3.6 Other relevant factors 21
4 CLINICAL EFFECTIVENESS 21
4.1 Critique of manufacturer's approach 21
4.1.1 Description of manufacturers search strategy and comment on whether the search strategy was appropriate. 21
4.1.2 Statement of the inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate 23
4.1.3 Table of identified studies. What studies were included in the submission and what were excluded 23
4.1.4 Details of any relevant studies that were not included in the submission ? 30
4.1.5 Description and critique of manufacturers approach to validity assessment 32
4.1.6 Description and critique of manufacturers outcome selection .. 33
4.1.7 Describe and critique the statistical approach used 35
4.1.8 Summary statement 35
4.2 Summary of submitted evidence 36
4.2.1 Summary of results 37
4.2.2 Critique of submitted evidence syntheses 40
4.2.3 Summary 49
5 ECONOMIC EVALUATION 50
5.1 Overview of manufacturer's economic evaluation 50
5.1.1 Populations 51
5.1.2 Perspective and time horizon 52
5.1.3 Treatment and comparator 52
5.1.4 Natural history 54

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

5.1.5 Treatment effectiveness within the submission 54
5.1.6 Health related quality of life 56
5.1.7 Resources and costs 57
5.1.8 Discounting 57
5.1.9 Sensitivity analyses 58
5.1.10 Model validation 58
5.2 Critique of approach used 58
5.2.1 Treatment effectiveness 58
5.3 Results included in manufacturer's submission 65
5.3.1 Mixed treatment comparison (MTC) 66
5.3.2 Resources and costs 72
5.3.3 Sensitivity analyses 75
5.4 Comment on validity of results presented with reference to methodology used 76
5.5 Summary of uncertainties and issues 78
6 Additional work undertaken by the ERG 79
7 Discussion 80
7.1 Summary of clinical effectiveness issues 80
7.2 Summary of cost effectiveness issues 80
7.3 Implications for research 83
8 References 97

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Abbreviations

ACR	American College of Rheumatology
BRAM	Birmingham rheumatoid arthritis model
CG	Clinical guideline
CI	Confidence intervals
CRP	C Reactive Protein
CS	Corticosteroids
DAS	Disease Activity Scores
DMARDs	Disease modifying anti-rheumatic drugs
EMEA	European Medicines Evaluation Agency
EMBA	Embase (database section)
EMYY	Embase year (database section)
EQ-5D	Euroqol five dimensions
ERG	Evidence review group
ESR	Erythrocyte Sedimentation Rate
EULAR	European League Against Rheumatism
FACIT-F	Functional Assessment of Chronic Fatigue Illness Therapy
GH	General health
HAQ	Health Assessment Questionnaire
ICER	Incremental cost effectiveness ratio
IR	Inadequate responder
iv	Intravenous
MEIP	Medline In Process
MEYY	Medline year (database section)
MRA	Myeloma receptor antibody
MTC	Mixed treatment comparison
MTX	Methotrexate
NOAR	Norfolk Arthritis Register
NSAIDs	Non-steroidal anti-inflammatory drugs
OL	Open label
OR	Odds ratio
PICO	Patients, intervention, comparator, outcomes
PSA	Probabilistic sensitivity analysis
QALY	Quality-adjusted life year
RA	Rheumatoid arthritis
RCT	Randomised controlled trial
RR	Relative risk
SD	Standard deviation
SE	Standard error
SF	Short Form (questionnaire)
sqrt	Square root
STA	Single technology appraisal
TA	Technology appraisal
TNF-α	Tumour necrosis factor alpha
UK CRN	United Kingdom Clinical Research Network
VAS	Visual analogue scale

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

1 SUMMARY

1.1 Scope of the submission

The NICE scope of the submission is to appraise the clinical and cost effectiveness of tocilizumab alone or in combination with methotrexate (MTX) within its licensed indication for the treatment of moderate to severe rheumatoid arthritis (RA). The comparators are management strategies involving DMARDs without tocilizumab, including conventional DMARDs and biologic agents including adalimumab, etanercept, infliximab and rituximab.

1.2 Summary of submitted clinical effectiveness evidence

The submission provides evidence from four RCTs of tocilizumab versus placebo, one RCT of tocilizumab versus MTX and two longer term single arm extension studies following patients up to 3 or 5 years. There were no head to head comparisons of tocilizumab versus other biologic DMARDs. Evidence from a mixed treatment comparison (MTC) gave information on the relative effectiveness of tocilizumab compared to other biologic DMARDs.

Results from the RCTs suggested that tocilizumab was more effective than placebo and more effective than MTX for ACR20, ACR50 and ACR70 and other outcome measures. Results from the MTC suggested that tocilizumab had higher relative effectiveness than rituximab and than data for TNF-α inhibitors combined.

1.3 Summary of submitted cost effectiveness evidence

No evaluations of the cost effectiveness of tocilizumab were found from literature searches. The submission included a de novo individual sampling model with a hypothetical cohort of 10,000 patients with moderate to severe RA and an inadequate response to traditional DMARDs or to one or more TNF-α inhibitors in addition. The intervention in the model was tocilizumab 8mg/kg added to a sequence of biologic and conventional DMARDs compared to the same sequence without tocilizumab. The outcomes were ACR scores from the MTC to obtain HAQ scores and then to calculate EQ-5D

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

using a quadratic equation derived from work published in a conference abstract. The costs were largely driven by the high treatment costs for biologic DMARDs. The perspective was NHS, a lifetime horizon was used and the discount rate was 3.5% per year for costs and utilities. Scenario and some probabilistic sensitivity analyses were undertaken. The model was constructed in Visual Basic for Applications within MS Excel.

1.4 Commentary on the robustness of submitted evidence

The NICE Specification for Manufacturer/Sponsor Submission states “A submission should be as succinct and informative as possible. It is expected that the main body of the submission will not usually exceed 75 pages.” The submission for this STA was 263 pages long, plus an embedded file with the MTC of 111 pages. The submission was repetitive, difficult to understand and very complicated.

1.4.1 Strengths

The submission makes a convincing case of the superior effectiveness of tocilizumab against placebo in European patients with moderate to severe RA who had an inadequate response to MTX and other conventional DMARDs in three RCTs and who had an inadequate response to TNF inhibitors or were intolerant of these drugs in one RCT. Also a single head to head trial against MTX in European patients is reported and showed better efficacy for tocilizumab. The trials were sufficiently large to obtain statistically significant results. It is unlikely that any RCT evidence was missed.

1.4.2 Weaknesses

There were no RCTs of tocilizumab against any other biologic DMARDs. Due to the lack of relevant clinical evidence, a MTC was conducted against TNF-α inhibitors, rituximab and abatacept. The MTC combined the effectiveness of TNF-α inhibitors (etanercept, adalimumab and infliximab) and included an RCT on etanercept that should have been excluded according to the submission description of the decision problem. The de novo economic model made a number of questionable assumptions, particularly in terms of the rebound effect following withdrawal of treatment, HAQ score progression on

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

long-term treatment and the relationship between HAQ and EQ-5D scores, which are likely to be highly influential. The probabilistic sensitivity analysis gave a remarkable lack of variation around cost.

1.4.3 Areas of uncertainty

The effectiveness of tocilizumab relative to other biologic DMARDs is uncertain. The economic model investigated sequences of treatment with or without tocilizumab. Only two of many possible sequences that could be used to treat RA are considered. Currently NICE allows use of a TNF inhibitor in patients who have had an inadequate response to two DMARDs including MTX. Therefore a sequence of subsequent therapies could include sulfasalazine, hydroxychloroquine and azathioprine, assuming for example that a patient had had an inadequate response to MTX alone. The sequences described in the model exclude some reasonable treatment options. The impact of doing so on cost effectiveness analyses is uncertain. Similarly some of the assumptions made in cost effectiveness analyses can be challenged and the impact of these assumptions on costs is unclear. An important assumption in the DMARD sequences considered is that a second TNF inhibitor would not be used, based on previous NICE guidance. This guidance however has been challenged recently and use of two consecutive TNF inhibitors is common in practice.[1] It is uncertain whether inclusion of more than one TNF inhibitor in treatment sequences could influence cost effectiveness of tocilizumab.

1.5 Key issues

Currently it is highly unlikely that tocilizumab would replace conventional DMARDs such as MTX in the treatment pathway. Therefore the main clinical decisions could be:

1. Whether tocilizumab could be an alternative to a TNF-α inhibitor (etanercept, adalimumab or infliximab) as the first biologic DMARD to try after 2 or more conventional DMARDs have failed to control symptoms (another TNF-α inhibitor may be tried if the first has failed due to side effects). The DMARD-IR model explored this possibility.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

2. Whether tocilizumab could replace rituximab as a biologic DMARD to try after a TNF-α inhibitor has failed to control symptoms. The TNF-IR model explored this possibility. (NB it is not mandatory that rituximab should be tried if TNF-α inhibitors have failed).
3. Whether tocilizumab could be added to the pathway after failure of a TNF-α inhibitor and rituximab . The submission, surprisingly, does not explore this possibility.

1. According to the reported MTC, tocilizumab has a higher relative effectiveness than combined TNF-α inhibitors (etanercept, adalimumab and infliximab). The MTC was reanalysed and when the three drugs were examined separately they did not have similar estimated relative effectiveness. Etanercept had a lower relative effectiveness than tocilizumab (1.65 vs 3.19) if all RCTs were included, but if one RCT was removed etanercept had a higher relative effectiveness (5.32 vs 3.19). So a key question is whether the Klareskog RCT should have been included or excluded from the MTC. As it was the only RCT to specifically mention that it included patients who were likely to benefit from MTX treatment, rather than having failed MTX treatment, it was different to the other RCTs. The submission decision problem statement was quite clear in that they were only going to investigate “adults with moderate or severe active rheumatoid arthritis (RA) who have either responded inadequately to or, or who were intolerant to, previous therapy with one or more disease modifying anti-rheumatic drugs (DMARDS)”. Therefore the Klareskog RCT should not have been included in the MTC.

CIC Since etanercept has been costed in the submission to be the same price as tocilizumab 8mg/kg then the decision as to which one to use will be based on differential effectiveness and factors such as ease of use, administration costs and adverse effects, rather than drug cost. Etanercept is delivered by subcutaneous injection by patients or carers at home whereas tocilizumab at present is given by iv infusions in a health care facility. It is unclear whether etanercept or tocilizumab have more or worse side effects.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

2. CIC TA126 guidance is that “Rituximab in combination with methotrexate is recommended as an option for the treatment of adults with severe active RA who have had an inadequate response to or intolerance of other disease modifying antirheumatic drugs (DMARDs), including treatment with at least one tumour necrosis factor α (TNF-α) inhibitor therapy. According to the MTC, tocilizumab has a higher relative effectiveness than rituximab. Rituximab was costed in the submission to be less expensive than tocilizumab (£4,980 vs £9,295). It is unclear whether the estimated additional effectiveness of tocilizumab is worth the additional cost.

2 BACKGROUND

2.1 Critique of manufacturer’s description of underlying health problem

The information about RA in the submission is brief:

P18

RA is the most common of the incurable and potentially disabling chronic systemic inflammatory autoimmune diseases. Affecting approximately 0.5-1% of the population worldwide, the onset of disease occurs in adults in their fourth and fifth decade, at a time when they are most economically active. The disease, which is 2.5-fold more prevalent in women than in men, is characterized by symmetric synovitis and erosive arthritis, often rapidly progressive with joint damage apparent soon after the onset of symptoms. This feature typically leads to a progressive decline in functional status and work disability. Patients with RA not only suffer chronic severe disability, but are also likely to die prematurely. Anemia, a common extra-articular manifestation of RA with characteristics of anemia of chronic disease, is estimated to occur in approximately 30% of patients. It reduces patients’ quality of life and is associated with excess morbidity and mortality. Improvement in hemoglobin is associated with reduction in systemic inflammation and decrease in disease activity.

The NICE scope states that “approximately 30 to 40% have moderate to severe disease despite treatment with conventional disease modifying anti-rheumatic drugs (DMARDs)”. The submission gives an estimate of the proportion of RA patients who might be eligible for Tocilizumab. In this data it is assumed that approximately 25 % of patients will fail treatment with conventional DMARDs and a similar number will fail TNF inhibitors (pages 189-191). Data from the BSR Biologics register indicates that around 33% of patients cease their primary TNF inhibitor for a variety of reasons. This

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

suggests that the number of patients eligible for tocilizumab may be underestimated in the submission.

2.2 Critique of manufacturer's overview of current service provision

The submission section on current service provision is very brief:

> From p21
>
> One area of uncertainty that is currently under review is the suitability of TNF cycling in the event that a patient has an inadequate response due to lack of efficacy. Within this submission it is assumed that the cycling of TNF in the event of lack of efficacy is not permitted consistent with NICE TA36.
>
> Secondly the current first TNF inhibitor of choice may vary across the NHS; however the impact of modifying the base case assumption (etanercept) upon the final cost effectiveness estimates is evaluated within the submission.

With reference to sequential use of TNF-α inhibitors, NICE decided, in TA130 to split the decision about TNF-α inhibitor use i.e. they approved first use as in TA36 but since the appeal on sequential TNF-α inhibitor use was upheld additional analyses of sequential use were commissioned[2]. TA130 guidance is that "An alternative TNF-α inhibitor may be considered for patients in whom treatment (with a TNF-α) is withdrawn due to an adverse event before the initial 6-month assessment of efficacy".[3] The proportions of patients with RA on a TNF inhibitor being given a second TNF-α inhibitor were described recently by Hyrich et al from the BSR Biologics Register.[1]

NICE have very recently published a guideline on the treatment of RA[4] which indicates current service provision.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

3 Critique of manufacturer's definition of decision problem

The manufacturer's statement of the decision problem is as follows:

	Final scope issued by NICE	Decision problem addressed in the submission
Population	Adults with moderate to severe rheumatoid arthritis	1. Adults with moderate to severe active rheumatoid arthritis (RA) who have either responded inadequately to, or who were intolerant to, previous therapy with one or more disease modifying anti-rheumatic drugs (DMARDs). 2. Adults with moderate to severe active rheumatoid arthritis (RA) who have either responded inadequately to, or who were intolerant to, previous therapy with one or more tumour necrosis factor (TNFα) antagonists.
Intervention	Tocilizumab alone or in combination with methotrexate	Tocilizumab in combination with methotrexate (MTX) followed by the current treatment sequence. Tocilizumab will be additive to the assumed existing standard of care / treatment strategy. **DMARD-IR Indication**: i. Tocilizumab + MTX ii. TNFα inhibitor (etanercept assumed most commonly used) iii. Rituximab iv. Leflunomide v. Gold vi. Cyclosporine vii. Palliative care **TNF-IR Indication**: i. Tocilizumab + MTX ii. Rituximab iii. Leflunomide iv. Gold v. Cyclosporine vi. Palliative care
Comparator(s)	Management strategies involving DMARDs without tocilizumab, including treatment with:	**1.DMARD-IR indication** Tocilizumab is licensed in the management of moderate to severe active RA patients who have had an inadequate response or intolerance to one or more

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

	• conventional DMARDs • biologic agents including adalimumab, etanercept, infliximab and rituximab	DMARDs. The current treatment sequence identified for this patient population according to current NICE guidance and therefore will form the assumed comparator sequence is: i. TNFα inhibitor (etanercept currently assumed to be most commonly used) ii. Rituximab iii. Leflunomide iv. Gold v. Ciclosporine vi. Palliative care **2. TNF-IR indication** Tocilizumab is licensed in the management of moderate to severe active RA patients who have had an inadequate response or intolerance to one or TNF inhibitors. The current treatment sequence identified for this patient population according to current NICE guidance and therefore will form the comparator sequence is: i. Rituximab ii. Leflunomide iii. Gold iv. Ciclosporine v. Palliative care
Outcomes	The outcome measures to be considered include: • disease activity • physical function • joint damage/radiographic progression • joint replacement • pain • mortality • fatigue • health-related quality of life • adverse effects of treatment	As well as the stated outcome measures, the inhibition of disease progression will be considered and evaluated in its own right. Specific outcome measures highlighted will be American College of Rheumatology (ACR) scores, Disease Activity Scores (DAS), EULAR scores, Health Assessment Questionaire (HAQ) score, Fatigue (FACIT-F) score, Short Form (SF-36) scores and the Sharp radiographic assessment scores.
Economic Analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of	Two reference cases reflecting the DMARD IR and TNF IR populations will be presented. The same economic model and structure will be utilised for both ICERs. In both analyses the cost-

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

	incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	effectiveness of the Tocilizumab treatments will be expressed in terms of incremental cost per quality-adjusted life year. The economic model will be an individual sampling model (ISM) similar to that used within the rituximab RA single technology appraisal. The key issues/drivers within the model that Roche anticipate forming a large part of the committee's discussions when considering previous RA appraisals relates to **long term HAQ progression.** Roche propose to utilise the actual observed HAQ data from within its phase III trials to inform this rate. After the end of the trial follow-up, an assumption will be required. This will be informed by previous NICE RA technology appraisals. Secondly Roche will re-estimate the **relationship between HAQ and utilities** through using its patient level trial data which permits the mapping of HAQ directly to the EQ-5D instrument. Previous NICE appraisals methods relied upon mapping via the HUI-3 instrument.
Subgroups to be considered	None identified	None identified
Special considerations, including issues related to equity or equality	Guidance will only be issued in accordance with the marketing authorisation. Where evidence allows, subgroup analysis may be carried out in sero-positive and sero-negative patients or any other bio-markers that may define subgroups	No comment

3.1 Population

Both the NICE scope and the market authorisation specify moderate to severe RA. However, the submission goes on to restrict it to patients who have responded inadequately to, or who were intolerant to, previous treatment with

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

conventional DMARDs or TNF-α inhibitors in line with the economic model they have constructed.

3.2 Intervention

Tocilizumab is a monoclonal antibody that is directed against interleukin-6 receptors. It is also called RoActemra, Actemra, atlizumab (before 2005) and myeloma receptor antibody (MRA). MRA is used in place of tocilizumab in some of the submission and the clinical trial reports. The multiple names can cause some confusion.

The licensed indication for tocilizumab is:

> (p3) (RoActemra) in combination with methotrexate (MTX) is indicated for the treatment of moderate to severe active rheumatoid arthritis (RA) in adult patients who have either responded inadequately to, or who were intolerant to, previous therapy with one or more disease modifying anti-rheumatic drugs (DMARDs) or tumour necrosis factor (TNF) antagonists. In these patients, RoActemra can be given as monotherapy in case of intolerance to MTX or where continued treatment with MTX is inappropriate.

EMEA marketing authorisation was granted on 16th January 2009.[5]

Tocilizumab given alone has not been investigated in this submission. Only tocilizumab given with MTX has been investigated, particularly in the economic model, and this is not made clear in the decision problem. Tocilizumab is given every 4 weeks by intravenous injection. The licensed dose is 8mg/Kg (minimum dose 480mg) and this dose does not vary if it is given alone or with MTX. For a 70Kg person, the dose would be 560mg. Tocilizumab is available in IV ampoules of 80mg, 200mg and 400mg. Tocilizumab is not recommended for use in children below 18 years of age due to insufficient data on safety and efficacy. There is no dose adjustment for people aged over 65.[6]

The decision problem does not explicitly discuss adverse effects. Tocilizumab is an immunosuppressant so suppresses innate immunity to bacterial and viral infections. Reported side effects of tocilizumab include infections such as upper respiratory tract, skin and gastro-intestinal infections. There are also increased rates of gastrointestinal problems (diarrhoea, mouth ulceration, vomiting, abdominal pain, gingival pain, oral pain and flatulence), skin disorders (rash, dermatitis and pruritus, skin ulcers), headaches, dizziness,

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

conjunctivitis and oedema. Approximately 6.5% of patients get infusion reactions, mostly hypertension, rashes and pruritis but rarely anaphylaxis. No increased rates of malignancies were found in the trial results so far but long term studies are ongoing. There is no evidence to suggest an increase or decrease in mortality rates with tocilizumab.[7]

3.3 Comparators

The NICE scope specifies the comparators to be Management strategies involving DMARDs without tocilizumab, including treatment with:

- conventional DMARDs
- biologic agents including adalimumab, etanercept, infliximab and rituximab

Conventional DMARDs currently in use include sulfasalazine, gold salts, hydroxycloroquine , azathioprine, cyclosporin, leflunomide, and MTX.[8] Corticosteroids are used commonly in combination with conventional DMARDs and this use has been enshrined by recent NICE guidance on the management of RA.[4] Penicillamine and chloroquine are very rarely used these days and gold salts are waning in use because of a high frequency of drug toxicity. The normal dose of MTX is between 7.5 and 20 mg weekly.

Relevant biologic comparators to tocilizumab are abatacept (Bristol Myers Squibb), adalimumab (Abbott), etanercept (Wyeth), infliximab (Schering Plough) and rituximab (Roche). Abatacept and anakinra are not approved by NICE for use in RA.

In the decision problem, a comparator sequence has been chosen. This is just one of numerous possible sequences of treatments that could be tried in RA. In clinical practice, treatment naïve patients in acute RA are started on steroids and MTX. Recently published NICE guidance for the management of RA recommends that two DMARDs (one of which must be MTX) and corticosteroids are given at the outset in people with moderate or severe RA. Corticosteroids may be discontinued in time. Patients who are intolerant of this regimen or who do not respond have another DMARD substituted.[4] Up to 30-40% of patients discontinue MTX due to toxicity.[9,10] Others discontinue due to lack of efficacy. MTX may then be substituted by the conventional DMARDs

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

leflunomide, gold, sulfasalazine, hydroxychloroquine, or ciclosporin depending on which DMARD has been combined with MTX and assuming that two DMARDs have been combined at the outset. Patients who have failed to respond to two DMARDs (including MTX) are then eligible for a TNF inhibitor according to current NICE guidance.

The list of available DMARDs given above indicates that the comparator sequences in the submission have omitted several possible treatment options in both the DMARD-IR group and the TNF-IR group. Of the three currently available DMARDs adalimumab and etanercept are generally preferred over infliximab because both drugs may be administered at home whereas infliximab requires day-case facilities. Patients intolerant of a first TNF inhibitor may try a second TNF inhibitor. NICE guidance for use of rituximab stipulates that patients must have tried at least two DMARDs and a TNF inhibitor before use. (NB A meta-analysis of RCTs of etanercept versus MTX showed that etanercept was marginally more effective (RR 1.24 (1.11 to 1.39)).[11]

3.4 Outcomes

The submission includes the following:

> From p9
>
> As well as the stated outcome measures, the inhibition of disease progression will be considered and evaluated in its own right.
>
> Specific outcome measures highlighted will be American College of Rheumatology (ACR) scores, Disease Activity Scores (DAS), EULAR scores, Health Assessment Questionaire (HAQ) score, Fatigue (FACIT-F) score, Short Form (SF-36) scores and the Sharp radiographic assessment scores

The stated outcome measures in the NICE scope were disease activity, physical function, joint damage/radiographic progression, joint replacement, pain, mortality, fatigue, health-related quality of life and adverse effects of treatment.

There is no subsequent mention of joint replacement in the submission. End stage joint disease is an important problem for patients and this is captured to a certain extent by functional scores such as the HAQ. Previous economic

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

models have attempted to capture the impact of joint failure and the costs of replacement surgery.[11,12]

Much of the submission focuses on ACR scores and HAQ scores. There is also some mention of Disease Activity Scores (DAS-28) but although they are mentioned in the description of the MTC they are not used in the economic model:

P133
Roche is mindful that existing NICE guidance for RA biologic therapies define response and stopping rules according to DAS. However this outcome is not publicly reported for the other comparator drugs and therefore was not considered a practical endpoint upon which to define stopping rules within the model. Previous submissions to NICE for RA biologic treatments have focused on ACR response endpoints

3.4.1 ACR outcome and its components

The American College of Rheumatology (ACR) have developed criteria that are commonly known as ACR scores (ACR20, ACR50, ACR70). The ACR20 is a change score which counts the percentage of patients who have improved by 20% in a combination of measures – 1, tender and swollen joint counts and 2, improvement in three of the following five parameters:

- Acute phase reactant
- Patient assessment
- Physician assessment
- Pain scale
- Disability/functional questionnaire

The selection of three out of five parameters in each patient will vary over time because at each visit the best three are selected. As the ACR20 uses the best three from five parameters, the HAQ score (see below) may not contribute to the ACR20 at each follow up point.

In one of the included tocilizumab RCTs (the Option RCT), 66 joints were assessed. Acute phase reactants were C Reactive Protein (CRP) and Erythrocyte Sedimentation Rate (ESR). The patients' and physician's global assessment was by 100mm visual analogue scale. The pain scale was also measured by 100mm VAS (but for some patients in the Ambition RCT

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

(another of the included tocilizumab RCTs) the scale was 96mm long, but this is unlikely to have had much impact on outcomes). The disability/functional questionnaire used was the Stanford Health Assessment Questionnaire – Disease Index (HAQ or HAQ-DI). This consists of 20 questions referring to 8 component sets: dressing/grooming, arising, eating, walking, hygiene, reach, grip, and activities. For each question, people are asked to indicate whether, over the past week, they were able to do it without any difficulty, with some difficulty, with much difficulty or were unable to. HAQ-DI scores range from 0 to 3. Scores of 0 to 1 are considered to be mild to moderate difficulty, 1 to 2 are moderate to severe disability and 2 to 3 are severe to very severe disability.

An average HAQ score reported in a population based study was 0.49 and in RA was 1.2.[13] In another study of 1109 French RA patients, the distribution was roughly normal with a median of 1.25 and mean of 1.32 (SD 0.77).[14] The HAQ in RA patients and the general population tend to increase by age.[15] However, the HAQ score rises dramatically in the general population after the age of 70 (see Appendix 5 of the submission) but the HAQ score changes for RA patients in this age group are even higher.[15]

A minimally important difference in HAQ score in RA patients has been defined as 0.22.[16] The thresholds for change in total HAQ score were classed in the Ambition RCT (WA 17824) as follows:

P105 Ambition trial report.
A clinically meaningful improvement was prospectively defined as a reduction from baseline in HAQ-DI score of at least 0.25 points; this is twice the lowest change in score values (ie, 0.125) that can be derived from the instrument. When assessed as a group, it was estimated that a group mean change from baseline score of -0.22 represents the minimal clinically important difference (MCID) (ie, the mean change from baseline that most closely correlates with other measures of patient-reported clinical improvement)

In the Radiate RCT (WA18062) HAQ score improvement was categorised as a decrease from baseline of either -0.25, -0.30 or -0.50 and HAQ score worsening was also categorised using these three different thresholds (see p75 submission and Table 11 on p76).

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

3.4.2 Disease Activity Score

The Disease Activity Score (DAS-28) is a continuous measure of disease activity in RA rather than the categorical response criteria of the ACR. It includes the proportion of tender and swollen joints (out of 28 joints), the ESR and general health or global disease activity from a 100mm VAS. These are fed into the following formula (usually using a computer package)

DAS28=0.56xsqrt(tender28) + 0.28xsqrt(swollen28) + 0.7xln(ESR) + 0.014xGH

The result is a number from 1 to10 where a higher number is worse disease. A DAS-28 score of 2.6 or less is considered to represent remission, 3.2 or less is low disease activity and 5.1 or more is high disease activity. However, in clinical practice, the DAS score may not be as useful tool for decision making in individual patient encounters.[17]

3.4.3 Other outcome measures

EQ-5D would be a suitable outcome to measure in RA and was measured in two of the trials in the submission. These results are discussed in section 4.1.6.2.

Fatigue is a common problem in RA. It may be measured by the Functional Assessment of Chronic Fatigue Illness Therapy (FACIT) scale which is a 13-item questionnaire. Although the FACIT scale was used in the submitted trials, the results were not discussed for the short term follow-up, only for the long term extension studies (see submission p87-8).

3.5 Time frame

RA is a long term condition necessitating life-long therapy for many patients. Conventional DMARDs are generally slow acting drugs (taking up to 6 months for maximum efficacy) but corticosteroids are commonly used as an adjunct to DMARDs to promote a rapid clinical improvement and perhaps for enhancing the efficacy of conventional DMARDs.[8] Biologic DMARDS such as TNF inhibitors appear to have a much faster impact which may manifest as an observable improvement within 2 weeks. This may mean lesser reliance on corticosteroids in the short term but such differences are unlikely to have any material effect on economic modelling.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

3.6 Other relevant factors

The scope states that "Where evidence allows, subgroup analysis may be carried out in sero-positive and sero-negative patients or any other bio-markers that may define subgroups". Sero-positive and sero-negative rheumatoid factor status was reported in the trial reports but did not influence response to treatment sufficiently to necessitate sub-group analyses. This accords with our clinical experience with conventional DMARDs and TNF inhibitors and indeed is not surprising as conventional DMARDs and TNF inhibitors are widely used in sero-negative arthritides such as ankylosing spondylitis and psoriatic arthritis. The submission decision problem section states that that they did not identify any subgroups to be considered. Population studies have suggested a differential response to treatment with MTX in patients who have antibodies to citrullinated antibodies.[18] Again, however, basing individual patient treatment on this factor is unlikely to be clinically useful.

4 CLINICAL EFFECTIVENESS

4.1 Critique of manufacturer's approach

4.1.1 Description of manufacturers search strategy and comment on whether the search strategy was appropriate.

Summary from the manufacturer's submission:

The following sources were searched between 20th and 21st January 2009: MEDLINE 1993 to date (MEYY), Medline In Process latest eight weeks (MEIP), EMBASE 1993 to date (EMYY), EMBASE latest eight weeks (EMBA), Cochrane Library, EULAR abstracts (European League Against Rheumatism) 2002 - 2008, ACR (American College of Rheumatology) abstracts 2002 - 2008.
Internal study reports and regulatory submission documents were accessed through company databases.

The searches were limited to humans and clinical trials. Data from clinical studies conducted in Japan by Chugai were not included as data in this patient population was not considered relevant to European patients.

Details are provided in Appendix 2, section 9.2 of the submission.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Comments:

- Search strategies were detailed in full, with the exception of Cochrane Library and the company databases and were constructed on terms capturing the intervention of interest, the condition and the study design.
- The terms in MEDLINE (MEYY) were combined in a way which made the searches appear restrictive. The alternative terms for the intervention: ‘actemra’ and ‘atlizumab’ were not used in either MEYY or MEDLINE In Process. Both of these factors increased the likelihood of missing relevant studies.
- Not all terms for the intervention were used in EMBASE ie ‘actemra’ and ‘tocilizumab’ were omitted from EMYY and ‘actemra’ and ‘atlizumab’ from EMBA. Also, in EMYY the terms were restricted to ‘Major Descriptors’ which may have made searches too restrictive. As before, both of these factors increased the likelihood of missing relevant studies.
- The full Cochrane Library search strategy was not presented but described as having used text-words only. Including index terms would have provided a more comprehensive search.
- As publicly available trials registers were not searched, additional on-going studies may have been overlooked.
- The company databases which were searched are not described.
- Not all terms for the intervention were used in searches of EULAR and ACR ie ‘atlizumab’ and ‘actemra’ were omitted from the former and ‘actemra’ from the latter.

Despite the limitations of the searches, those done by the ERG of MEDLINE, EMBASE and Cochrane Library (see Appendix 1) verified that all relevant published studies were identified.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

4.1.2 Statement of the inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate.

The stated inclusion and exclusion criteria were very unclear (see box below).

From P29
Inclusion criteria

Published papers or abstracts which evaluated the following were included:

- Tocilizumab (or atlizumab prior to 2005) was the major focus of the paper.
- Rheumatoid arthritis was a major focus of the paper.
- Patient population consisted of patients who had responded inadequately or who were intolerant to one or more DMARDs or TNF antagonists, to be consistent with the EU licence for tocilizumab, including dose
- Controlled clinical studies
- Documents relating to humans

Exclusion criteria

Published papers or abstracts which evaluated the following were excluded:

- Any papers providing a review, update or commentary on data published elsewhere were excluded
- Any papers which only mentioned tocilizumab within a discussion of treatments for rheumatoid arthritis were excluded
- Papers covering the use of tocilizumab in Castleman's disease, juvenile idiopathic arthritis, other autoimmune diseases or other off-licence indications were excluded
- Clinical studies conducted in Japanese patients were not included, as data generated in this patient population was not considered sufficiently relevant to European patients.
- Animal studies or *in vitro* research
- Case reports

4.1.3 Table of identified studies. What studies were included in the submission and what were excluded

4.1.3.1 Included trials

There were three groups of included studies,

- Five RCTs discussed in the clinical effectiveness section (Option WA17822, Lithe WA17823, Ambition WA17824, Radiate WA18062, Toward WA18063), of which 3 RCTs (Option WA17822, Lithe WA17823, Toward WA18063) were used in the meta-analysis (see Table 3)

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

- Two long term single arm extension studies of tocilizumab (WA18695, WA18696) were used to establish 5 year data (see Table 4)

- Nineteen RCTs were used in the mixed treatment comparison (MTC) of tocilizumab (3- Option WA17822, Lithe WA17823, Toward WA18063), adalimumab (4), infliximab (3), etanercept (4), abatacept (3) and rituximab (2) (see economic section).

The basic design of the five included RCTs and the two extension studies are shown in the table below.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table of included studies from submission p41

	WA17822	WA17823	WA17824	WA18062	WA18063	WA18695	WA18696
Design and Duration	DB, R, PC: 24-week	DB, R, PC; year 1 DB, year 2 OL	DB, DD, R, PC: 24-week	DB, R, PC: 24-week	DB, R, PC: 24-week	OL extension study; approxi 5 years*	OL extension study; approximately 5 years*
Patient Population	Moderate to severe active RA in MTX inadequate responders	Moderate to severe active RA in MTX inadequate responders	Active RA; MTX naïve or MTX discontinued but not due to lack of efficacy or toxic effect	Moderate to severe active RA in patients with inadequate response to anti-TNF agent(s)	Moderate to severe active RA in patients with inadequate response to DMARDs	Patients completing treatment in WA17822	Patients completing treatment in WA17824, WA18062, WA18063, WP18663
Treatment	**3 arm study**: Tocilizumab: 4 or 8 mg/kg or placebo iv every 4 weeks + MTX 10-25 mg/week	**3 arm study**: Tocilizumab: 4 or 8 mg/kg or placebo iv every 4 weeks + MTX 10-25 mg/week	**2 arm study**: Tocilizumab: 8 mg/kg iv every 4 weeks **or** MTX 7.5-20 mg/week (po) **Substudy includes 3rd arm:** Placebo (8 weeks placebo then 16 weeks TCZ 8 mg/kg)	**3 arms**: Tocilizumab: 4 or 8 mg/kg or placebo iv every 4 weeks plus MTX 10-25 mg/week	**2 arms**: Tocilizumab: 8 mg/kg or placebo iv every 4 weeks plus standard DMARD(s)	**1 arm**: Tocilizumab: 8 mg/kg iv every 4 weeks plus MTX	**1 arm**: Tocilizumab: 8 mg/kg iv every 4 weeks alone or plus MTX / other DMARD(s)
Escape therapy	Week 16: TCZ 8 mg/kg	Week 16 onwards: TCZ 4 or 8 mg/kg	Substudy only, up to Week 8: TCZ 8 mg/kg	Week 16: TCZ 8 mg/kg	Week 16: adjustment of background DMARD	-	-
Total Randomized Patients	623	1196	673	499	1220	537**	1902**
Primary Endpoint at Week 24	ACR20 response rate	ACR20 response rate	ACR20 response rate	ACR20 response rate	ACR20 response rate	Long term safety/efficacy	Long term safety/efficacy

DB = double blind, R = randomized, PC = placebo controlled, DD = double dummy, OL = open label

* Or when tocilizumab becomes commercially available in the participating country, or when the sponsor decides to discontinue the study.

** Patients were not randomized into WA18695 and WA18696, but enrolled from studies WA17822, WA18063, WA18062 and WA17824

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

4.1.3.2 Patients

The NICE scope specifies adults with moderate to severe rheumatoid arthritis. Severity in RA can be measured by several questionnaires including the Stanford Health Assessment Questionnaire – Disease Index (HAQ or HAQ-DI) and the Disease Activity Score (DAS-28). These have been discussed in a previous section. In practice, clinicians tend to use the DAS-28 primarily for regulatory purposes i.e. to request funding for high cost drugs even if they don't use it in routine practice or for routine decision making (personal observations). .

The baseline DAS-28 mean scores in the five included RCTs was given on p45 and was between 6.5 and 6.8 (where 5.1 or more is high disease activity). The baseline spread of DAS-28 scores is not given in the submission. The baseline HAQ-DI mean scores in the five included RCTs were also given on p45 and were between 1.5 and 1.7 (where scores between 1 and 2 are moderate to severe disability). The baseline spread of HAQ-DI scores is not given in the submission. From the trial report, in the Ambition RCT the baseline minimum and maximum DAS-28 scores were 3.64 to 9.14 (mean 6.7, SD 0.88-1.0) and the HAQ scores varied between 0 and 3 (mean 1.5-1.6, SD 0.63-0.65). Similar spreads were seen in the other RCTs, from looking at these outcomes in the trial reports.

The mean age of patients in the five included RCTs was between 50 and 54 years. The number of prior conventional DMARDs they received was lower than would be the case in the UK (1.5-1.9 vs 4 (see p 62 submission). However it is likely that published studies may not reflect current use of biologics and TNF inhibitors in the UK i.e. earlier use of these drugs currently would mean that fewer conventional DMARDs are tried before trying a biologic agent.

4.1.3.3 Intervention

The intervention given in the five included RCTs was tocilizumab 8mg/kg every 4 weeks. In three RCTs (Option (WA17822), Lithe (WA17823) and Radiate (WA18062)) a third arm was given tocilizumab 4mg/kg. This 4mg/kg dose is not licensed in the UK for regular treatment but can be used for the

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

management of specific adverse events and for liver enzyme abnormalities, low neutrophil counts or low platelet counts.[6] The submission focuses on the 8mg/kg arms of the RCTs.

4.1.3.4 Comparators

The comparators used in the effectiveness section were versus placebo where both arms also were treated with MTX (Option WA17822, Lithe WA17823, Radiate WA18062) or conventional DMARD (Toward WA18063) or versus MTX (Ambition WA17824). There was no direct comparison to other biologic DMARDs but as comparators were not mentioned in the inclusion criteria it could be assumed that any RCTs of tocilizumab versus another biologic DMARD would have been included.

The dose of MTX used in the Option (WA17822), Lithe (WA17823) and Radiate (WA18062) RCTs was between 10-25mg weekly. The dose of MTX used in the Ambition (WA17824) RCT was stated to be between 7.5-20mg weekly (see p41 submission). The total mean cumulative dose in the MTX arm of the Ambition (WA17824) RCT was 140.9 capsules and the median dose was 161capsules (see p1045 trial report). Since the trial lasted 24 weeks, this represents an approximate average weekly dose of 6-7 capsules per patient. If each capsule was 2.5mg this represents an average dose of 15-17.5mg. In this RCT, 87% of patients received 21 - 24 weeks of treatment. None of the RCTs mentioned intensive dosing with MTX.

4.1.3.5 Outcomes

The stated outcome measures in the NICE scope were disease activity, physical function, joint damage/radiographic progression, joint replacement, pain, mortality, fatigue, health-related quality of life and adverse effects of treatment. The Manufacturer's description of inclusion criteria on p29 of the submission is not explicit as to outcome measures. Only some of the above outcomes were reported.

The outcome used to meta-analyse the three RCTs was ACR20, ACR50 and ACR70 including changes over the 24 week follow up. Other outcomes given were mean change in DAS28 (clinical remission), radiographic changes,

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

FACIT fatigue and SF-36. (EQ-5D was listed in the Option and Lithe trial protocols also). The submission also gives elements of the ACR score including mean (and standard error) HAQ-DI (Health assessment questionnaire disability index). The economic model uses the HAQ score drop based on ACR responses (p148).

Excluded trials

There are five other RCTs of tocilizumab in RA (see Table 1). Two of these are described as supporting studies on p26-7 of the submission. The Charisma RCT (LRO301) enrolled Europeans and the results were used to justify the sample size required in the Option (WA17822) and Radiate (WA18062) RCTs (see p56-8 submission). The ACR20 result at week 16 was 63% for the 8mg/kg tocilizumab monotherapy arm compared to 74% for the tocilizumab 8mg/kg plus MTX arm and 41% in the MTX plus placebo arm. Three other RCTs were listed on p26-7 but were not mentioned further in the submission. They are all in Japanese patients and ACR20 results are in Table 1. In response to clarification, a further two studies were sent – one very small RCT and one long term extension study, also in Japanese patients.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 1. Other RCTs

Trial number (name) design	Comparison	Patients	ACR20 results
LRO301 (Charisma) RCT	Tocilizumab (several doses inc 8mg/Kg) vs tocilizumab + MTX vs MTX	European, incomplete response to MTX Total N= 359	At 16 weeks 63% tocolizumab 8mg/kg only 41% MTX 74% tocilizumab 8mg/kg + MTX
MRA009JP (Nishimoto 2004) RCT	Tocilizumab only 4mg or 8mg/Kg vs placebo	Japanese, Inadequate response to DMARD N=164	At 12 weeks 78.2% Tocilizumab 8mg/Kg, 11.3% placebo
MRA012JP (Samurai) RCT	Tocilizumab only 8mg/Kg vs DMARD	Japanese, inadequate response to DMARD N=306	At 52 weeks 78% tocilizumab, 34% DMARD
MRA213JP (Satori) RCT	Tocilizumab only 8mg/Kg vs MTX 8mg/week	Japanese, inadequate response to MTX N=127	At 24 weeks 80.3% tocilizumab, 25% MTX
Nishimoto 2003 RCT	Tocilizumab only 8mg/kg or 4mg/kg or 2mg/kg	Japanese , failed to respond to at least one DMARD or immunosuppressant or unable to continue due to adverse reactions N=16	At 24 weeks >80% (whole group)
Stream (Nishimoto 2008) Single arm extension	Tocilizumab only 8mg/kg for 12 weeks then dose varies	Japanese, previously in MRA009JP N=143 (94 at 5 years)	At 5 years ACR20 84%

It is unclear why some of the results from the Charisma RCT were not used in the systematic review as it apparently met their stated inclusion criteria. The Japanese RCTs also met their stated inclusion criteria but the explanation given as to why the Japanese RCTs were not included was because:

From p24

The phase II/III Japanese studies have been excluded due to the significant variation in RA clinical management compared to the EU and the applicability of the results to the UK population as a whole

However, there are no details given in the submission about what this significant variation in RA clinical management actually is. We assume that at least in part this is due to use of lower doses of MTX and widespread use of

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

DMARDs such as bucillamine (related to penicillamine).[19] There is no mention in the submission of the ethnic origin of patients in the five included RCTs.

4.1.4 Details of any relevant studies that were not included in the submission ?

The ERG searched Current Controlled Trials *meta*Register, UK CRN Clinical Research Portfolio and ClinicalTrials.gov. Additional ongoing studies were found as follows:

Trial	Status	No. of arms	Patient Group	Drugs	Comp-arator	Completion Date
WA 21488	recruiting	2	Mostly Europe-wide but also the US, Brazil, Israel, the Russian Federation and Thailand; ≥18 yrs with active RA, IR to prior MTX	Tocilizumab + MTX	Tocilizumab + Placebo	Jan 2012
WL 21469	recruiting	1	Germany; ≥18 yrs with active RA, IR-DMARD/anti-TNF-α	Tocilizumab + traditional DMARD	N/A	Sep 2009
WL 21753	recruiting	2	China;18-70 yr with active RA, IR-DMARD.	Tocilizumab + traditional DMARD	Placebo + DMARD	Mar 2010
ML 21136	ongoing, not recruiting	2	The US and Puerto Rico; ≥18 yrs with moderate-severe RA, IR-DMARD	Tocilizumab + traditional DMARD	Placebo + DMARD	Jun 2009
ML 22012	ongoing, not recruiting	1	Finland; ≥18 yrs with IR to current non-biologic DMARDs	Tocilizumab only, or in combination with current non-biologic DMARD	N/A	Jun 2009
ML 21530	recruiting	1	South and Central America; ≥18 yrs with moderate-severe active RA	Tocilizumab + MTX	N/A	Sep 2010
MA 21573	recruiting	1	Mostly Europe-wide, but also the US, Canada, Australia, India and Saudi Arabia; ≥18 yrs	Tocilizumab only	N/A	Mar 2010

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

			with severe active RA, IR to current non-biologic DMARDs and/or anti-TNF-α			
MRA 226 JP	recruiting	Case-only, 52-week follow up (prospective) study	Japan; ≥ 20 yrs with RA and achieved low disease activity in clinical trials of MRA (tocilizumab) and stopped treatment	N/A	N/A	not stated
MRA 225 JP	ongoing, not recruiting	1	Japan; ≥ 20 yrs with RA; non-randomized, open label, uncontrolled, safety/efficacy study	Tocilizumab	N/A	Jan 2009*
MRA 215 JP	ongoing, not recruiting	1	Location not stated; 20-75 yrs with RA; participated in Study MRA 213 JP	Tocilizumab	N/A	not stated
MRA 214 JP	ongoing, not recruiting	1	Location not given; ≥ 20 yrs patients from MRA (tocilizumab) group of previous study	Tocilizumab	N/A	not stated
MRA 010 JP	ongoing, not recruiting	1	Location not given; ≥ 20 yrs with RA; participated in Study MRA 009 JP	Tocilizumab	N/A	Dec 2009
MRA 222 JP	ongoing, not recruiting	1	Location not given; 20-75 yrs with RA; participated in Study MRA 220 JP or MRA 221 JP.		N/A	Jan 2009

An additional RCT was described in the clarifications submission, p 9

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

The trial identifier is WA19923, *A Mechanism of Action study to evaluate the effects of IL-6 receptor blockade with tocilizumab (TCZ) on lipids, arterial stiffness, and markers of atherogenic risk in patients with moderate to severe active rheumatoid arthritis (RA).*

It is a randomized controlled trial of 70 patients comparing TCZ 8 mg/kg or placebo added to MTX in DMARD inadequate responder patients. The purpose of the study is to explore the mechanistic effects of tocilizumab on a variety of artherogenic indices to investigate the hypothesis that tocilizumab has a positive effect on the pattern of cardiovascular markers seen in RA patients. There were 2 UK participating centers, Glasgow (Prof McInnes) and Newcastle (Prof Isaacs). The clinical trial report is anticipated at the end of 2009.

4.1.5 Description and critique of manufacturers approach to validity assessment

Validity assessment was by the use of the following standard questions and all five included trials were discussed together for each of the questions:

P59

- How was allocation concealed?
- What randomisation technique was used?
- Was follow up adequate?
- Were the individuals undertaking the outcomes assessment aware of allocation?
- Was the design parallel group or cross-over?
- Was the RCT conducted in the UK (or were one or more centres of the multinational RCT located in the UK)? If not, where was the RCT conducted, and is clinical practice likely to differ from UK practice?
- How do the included in the RCT participants compare with patients who are likely to receive the intervention in the UK? Consider factors known to affect outcomes in the main indication, such as demographics, epidemiology, disease severity, setting.
- For pharmaceuticals, what dosage regimens were used in the RCT? Are they within those detailed in the Summary of Product Characteristics
- Were the study groups comparable?
- Were the statistical analyses used appropriate?
- Was an intention to treat analysis undertaken?
- Were there any confounding factors that may attenuate the interpretation of the results of the RCT(s)?

The submission mentions long term follow up of five years but the two 5 year follow studies have a single arm only with patients given tocilizumab. No comparator of placebo, conventional DMARDs or biologic agents will be available so it is unclear how these 5 year studies can establish effectiveness. It is debatable whether follow up in the five included studies (24 weeks) is

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

adequate. It is not possible to tell whether any initial treatment effect wanes after time with this short length of follow up.

4.1.6 Description and critique of manufacturers outcome selection

4.1.6.1 HAQ score measurement irregularities

It is important to note that some of the HAQ questionnaires in the RCTs were not correctly completed. This is an extract from the Option trial (WA17822) report (p85):

> Mapi Research Institute, an international health outcomes organization, was approached to conduct a linguistic analysis of the affected questionnaires. Mapi Research Institute performed the linguistic validation work for most of the existing translations of the HAQ-DI and provided the linguistically valid translations used in this study. Based on the recommendations of the Mapi Research Institute, the following actions were taken:
> • For issues that Mapi Research Institute deemed not to affect content validity or acceptability, data were kept in the analysis
> • For issues relating to potentially low impact typographical or spelling/grammar errors, data were kept in the analysis as the errors were considered minor
> • For issues that Mapi Research Institute deemed as having the ability to possibly affect content validity or acceptability, data were excluded from the analysis.
> As a result, some questions and in some cases whole questionnaires were excluded from the analyses. HAQ-DI is an ACR responder component and contributes to the primary endpoint, therefore, as a conservative approach, in the event of any ambiguity, the questionnaires were removed from the analysis. This results in the assessment of an ACR responder being based on the SJC and TJC as major criteria and on only four of the possible five minor criteria (physician's global assessment of disease activity, patient's global assessment of disease activity, patient's assessment of pain, and CRP [or ESR if CRP was missing]). For such assessments, there would be a lower probability of demonstrating a response than with all five minor criteria available. A total of 105 patients from sites in Canada and Mexico had all their HAQ-DI data excluded from the analysis. This affected an equal number of patients in each treatment group (35 patients per group). A further 18 patients from sites in France had grip questions 1 and 2 excluded from the analysis (affecting 4, 6 and 8 patients in the placebo + MTX and tocilizumab 4 mg/kg + MTX and 8 mg/kg + MTX groups, respectively)

From the Lithe (WA17823) trial report

> P126
> A total of 62 patients (23 placebo + MTX, 19 tocilizumab 4 mg/kg + MTX, 20 tocilizumab 8 mg/kg + MTX) had all their HAQ-DI data excluded from the analysis. A further 16 patients (5 vs 7, 4 respectively) had individual questions excluded from the analysis

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

From the Ambition (WA17824) trial report

P106

As a result, some questions were excluded from the analyses. HAQ-DI is an ACR responder component and contributes to the primary endpoint, therefore, as a conservative approach, in the event of any ambiguity, the questionnaires were removed from the analysis. This results in the assessment of an ACR responder being based on the SJC and TJC as major criteria and on only four of the possible five minor criteria (physician's global assessment of disease activity, patient's global assessment of disease activity, patient's assessment of pain, and CRP [or ESR if CRP was missing]). For such assessments, there would be a lower probability of demonstrating a response than with all five minor criteria available. A total of eight patients from four sites in Canada had grip question 1 excluded from the analysis. All patients were in the placebo controlled substudy analysis group (3 patients in the placebo/tocilizumab group, 2 patients in the MTX group and 3 patients in the tocilizumab group).

From the Toward (WA18063) trial report

P85 Based on the recommendations of the Mapi Research Institute, one patient from a site in South Africa had all her HAQ-DI data excluded from the analysis and one patient from a site in the USA had all baseline, week 2 and week 4 HAQ-DI data excluded from the analysis.

The Radiate (WA18062) trial report mentions the Mapi Research Institute and the issues of excluded HAQ data on p5662 but does not say how many were excluded. However, many graphs include the sentence: "No imputation used for missing HAQ score", which suggests that some were missing.

4.1.6.2 Eq-5D outcomes measured

Two of the included RCTs measured EQ-5D scores (Option and Lithe) and it is unclear why these results have not been used. The Eq-5D results from the Option trial are shown in the table below and in Appendix 2. The results for the Lithe trial were not in the trial report, but "were collected for use in separate pharmacoeconomic reports." (p127 trial report). The clarification questions asked for numerical data for the figure mapping HAQ score and EQ-5D, but B14 response stated "The individual patient data….included in the analyses used to drive the HAQ-utility mapping mechanism can be found in the CSRs of the corresponding trials". A request was made for Roche to supply the Lithe trial (WA17823) EQ-5D results. Unfortunately they did not give them for tocilizumab and placebo groups separately at follow up.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 2. Eq-5D results from Option and Lithe RCTs

	Option trial (WA17822)		Lithe trial (WA17823)	
	Tocilizumab 8mg/Kg	Placebo	Tocilizumab 8mg/Kg	Placebo
Baseline mean (SD)	0.3929 (0.32360)	0.3908 (0.32943)	0.4143 (0.3983 to 0.4303) (mean +/- 1 SE)	
Follow up mean (SD)	0.6713 (0.23697)	0.5337 (0.31803)	Not supplied	Not supplied
Number of patients	205	204	401	394

4.1.6.3 Variation in outcomes between RCTs

Figure 7 on page 66 gives the ACR results for three of the RCTs Option (WA17822), Lithe (WA17823) and Toward (WA18063). It is noticeable that the Lithe RCT has worse results than the other two RCTs at 24 weeks, particularly for ACR70 results (13% vs 22% and 21%), but there is nothing to explain why this should be so.

4.1.7 Describe and critique the statistical approach used

The RCTs used standard approaches to statistical analyses. For individual RCT results, per protocol and ITT analyses were available in the trial reports. Only the ACR20, ACR50 and ACR70 results for Option (WA17822), Lithe (17823) and Toward (WA18063) RCTs were presented in the submission.

4.1.8 Summary statement

The submitted evidence is relevant to the decision problem. The search strategy for the systematic review was sufficiently complete to find all relevant published RCTs of tocilizumab in RA patients. The inclusion and exclusion criteria for the systematic review of tocilizumab studies are unclear and imprecise. Five RCTs were included but between 1- 5 other RCTs could also have been included. Critical appraisal of the five included RCTs was reasonably thorough. The background characteristics of patients in the

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

included RCTs were reported in the submission but the results of the RCTs were not.

4.2 Summary of submitted evidence

Five RCTs (Option WA17822, Lithe WA17823, Ambition WA17824, Radiate WA18062, Toward WA18063) were included in the clinical effectiveness section of which 3 RCTs (Option WA17822, Lithe WA17823, Toward WA18063) were used in the meta-analysis and the MTC. Two long term single arm extension studies of tocilizumab (WA18695, WA18696) were used to establish 5 year data for the economic model.

Baseline characteristics for the five included RCTs are shown in the table below (from p45 of the submission with additional details of corticosteroid and NSAID use from p61-2)

	WA17822		WA17823		WA18063		Pooled DMARD IR		WA17824		WA18062	
	TCZ 8mg + MTX	PLO + MTX	TCZ 8mg + MTX	PLO + MTX	TCZ 8mg + MTX	PLO + MTX	TCZ 8mg + MTX	PLO + MTX	TCZ 8mg + MTX	PLO + MTX	TCZ 8mg + MTX	PLO + MTX
Female (%)	85	78	82	83	81	68	82	82	83	81	84	79
Age, Mean, Yrs	51	51	53	51	53	54	53	52	51	50	54	53
Duration RA, Mean, Yrs	7.5	7.8	9.3	9.0	9.8	9.8	9.3	9.1	6.4	6.3	12.6	11.4
RF Positive (%)	83	71	83	82	78	75	80	77	74	75	79	75
DAS28, Mean	6.8	6.8	6.6	6.5	6.7	6.6	6.7	6.6	6.8	6.8	6.8	6.8
SJC/TJC, Mean	20/32	20/32	17/29	17/28	20/30	18/29	19/30	18/29	19/32	23/35	19/32	19/30
CRP, Mean, mg/dL	2.6	2.4	2.3	2.24	2.6	2.6	2.5	2.4	3.0	2.9	2.8	3.7
HAQ, Mean	1.6	1.7	1.5	1.5	1.5	1.5	1.5	1.5	1.6	1.5	1.7	1.7
No. Prior DMARDs, Mean	1.5	1.7	1.6	1.6	1.6	1.6	1.6	1.6	1.2	1.1	1.9	2.1
MTX Dose, Mean, mg/Wk	14.5	14.9	15.4	15.0	14.7	15	15.0	15.1	-	-	15.7	16.5
Weight (kg)			74	72	74	73	73	73	72	73	74	75
Oral CS	55		61		51		55		NR		52	
Oral NSAIDs	65		71		72		70		NR		62	

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

4.2.1 Summary of results

4.2.1.1 Results of the five included RCTs

Table 3 has details of the five included RCTs. The results presented in the table are for the 8mg/Kg tocilizumab arms only. ACR20 results were obtained from trial reports for consistence as results for Ambition (WA17824) and Radiate (WA18062) RCTs were not presented in the submission but were presented for the other three RCTs. The trial report and submission ACR20 results for Option (WA17822), Lithe (WA17823) and Toward (WA18063) were compared. In Lithe (WA17823) the placebo rate was 27% in the submission and 24.7% in the trial report. DAS-28 and HAQ results were also obtained from the trial reports.

Selected other results are given in the submission for Lithe (WA17823) and for Radiate (WA18062) RCTs (see pp 68-91), but some of this is per protocol rather than ITT results. It is unclear why these results were presented but not others.

There are also some longer term results of ACR50, DAS-28 and HAQ to 84-100 weeks (see p87-91 submission) but very few patients seem to contribute to these outcomes. Also these RCTs are only two of the four RCTs that contribute to the long term extension study WA18696 (see next section and Table 4).

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 3. Included Tocilizumab RCTs

Trial number (name)	Comparison Number (Total = 3330)	Meta analysis	Patients	ACR20 results	HAQ results (mean change from baseline	DAS-28 mean (SD)
WA17822 (Option)	Tocilizumab n=205 vs placebo n=204	Yes	Inadequate response to methotrexate	At 24 weeks 59% tocilizumab 27% placebo	At 24 weeks -0.55 tocilizumab -0.34 placebo	At 24 weeks 3.40 (1.46) tocilizumab 5.32 (1.25) placebo
WA17823 (Lithe)	Tocilizumab n=401 vs placebo n=394	Yes	Inadequate response to methotrexate	At 52 weeks 55.8% tocilizumab 24.7% placebo	At 52 weeks -0.6 tocilizumab -0.4 placebo	At 52 weeks 2.77 (1.36) tocilizumab 4.44 (1.31) placebo
WA17824 (Ambition)	Tocilizumab n=288 vs MTX n=284	No	Mostly MTX naïve	At 24 weeks 70.6% tocilizumab 52.1% MTX	At 24 weeks -0.70 tocilizumab (n=259) -0.52 MTX (n=250)	At 24 weeks (pp) 3.49 (1.51) tocilizumab 4.67 (1.50) MTX
WA18062 (Radiate)	Tocilizumab n=174 vs placebo n=160	No	Refractory to TNF-α inhibitor	At 24 weeks 50.0% tocilizumab 10.1% placebo	At 24 weeks -0.39 tocilizumab -0.05 placebo	At 24 weeks 3.57 (1.51) tocilizumab 5.60 (1.47) placebo
WA18063 (Toward)	Tocilizumab n=805 vs placebo n=415	Yes	Inadequate response to current DMARD therapy	At 24 weeks 60.8% tocilizumab 24.5% placebo	At 24 weeks -0.43 tocilizumab -0.20 placebo	At 24 weeks 3.56 (1.62) tocilizumab 5.50 (1.45) placebo

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

4.2.1.2 Results of extension studies

Two single arm tocilizumab extension studies were used to show the long term benefits of tocilizumab but neither had placebo or other control groups.

Table 4. Tocilizumab single arm extension studies

Trial number (name)	Comparison	Treatment given (to all patients)	Patients	Notes
WA18695	None (OL extension)	8mg/kg every 4 weeks + MTX	N= 537 From Option (WA17822),	Open label
WA18696	None (OL extension)	8mg/kg as monotherapy or with MTX or other DMARD	N= 1902 From Ambition (WA17824), Radiate (WA18062), Toward (WA18063), WA18663*	Open label
* WA18663 is not described in the submission, The trial number may be WP18663 (see p41 of submission) but this was not described in the submission either. A google search for Tocilizumab and WP18663 found a US drug interaction RCT by Roche of tocilizumab + MTX + simvastatin vs tocilizumab + MTX. This was a short term study (3 months) of target less than 100 patients and with a completion date of April 2008.[20]				

The total patient numbers in each of the extension studies was given on page 41 of the submission and add up to 2439, but the submission also states on page 114 that there were 2,562 patients in the extension studies. It also states on p114 that 2,439 had completed the core studies and entered the extension studies, suggesting that the remaining 123 patients were from study WA18663 (or WP18663). The number of patients completing the 24 weeks treatment in the Option (WA17822), Ambition (WA17824), Radiate (WA18062) and Toward (WA18063) RCTs was 2,633 (see p28), so not all of these entered the extension studies. It can be assumed that most of the dropout (n= 194) was due to lack of efficacy or lack of tolerance to tocilizumab.

It is useful to note that the number of patients providing HAQ scores for the extension trials, as given in the submission on p103 and p150 started with 873 patients at 24 weeks.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

4.2.1.3 Adverse events

Adverse events results are presented for all five RCTs combined and were described to be low (see table below from p116 of the submission). Adverse events for tocilizumab were compared to those for MTX. It is unclear whether the adverse event rate is higher or lower for tocilizumab than for other biologic DMARDs. For long term data, 574 patients contributed to the safety analyses at 18 months (see p114). This seems to be fewer patients than those contributing to the HAQ scores at 72-84 weeks as given on pages p103 and p150 of the submission. The submission states that the mean and median extent of tocilizumab exposure was 1.08 years (see p115). The risks therefore of longer term treatment with tocilizumab are unknown.

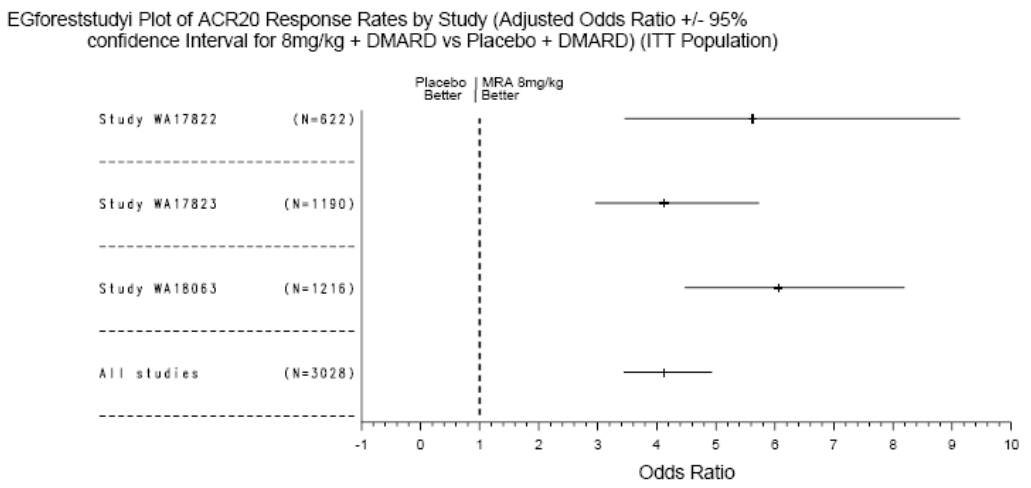
Number of patients (%)	Placebo + DMARD N=1170	MTX N=284	4 mg/kg + MTX N=774	8 mg/kg + DMARD N=1582	8 mg/kg N=288	All TCZ N=2644
Any AEs	733 (62.6%)	220 (77.5%)	547 (70.7%)	1134 (71.7%)	230 (79.9%)	1911 (72.3%)
AE rates per 100 patient years (95% CI)	377.34 (360.6,394.6)	449.70 (414.5,487.1)	472.24 (449.6,495.8)	462.37 (447.2,478.0)	491.73 (455.7,529.9)	468.44 (456.5,480.7)
Severe AEs	97 (8.3%)	19 (6.7%)	68 (8.8%)	138 (8.7%)	20 (6.9%)	226 (8.5%)
Any SAEs	62 (5.3%)	8 (2.8%)	46 (5.9%)	95 (6.0%)	11 (3.8%)	152 (5.7%)
SAE rates per 100 patient years (95% CI)	14.79 (11.6,18.5)	11.22 (6.3,18.5)	14.79 (11.0,19.5)	15.26 (12.6,18.3)	8.58 (4.4,15.0)	14.38 (12.3,16.7)
AEs leading to withdrawal	28 (2.4%)	15 (5.3%)	38 (4.9%)	74 (4.7%)	11 (3.8%)	123 (4.7%)
AEs leading to dose interruption	84 (7.2%)	63 (22.2%)	103 (13.3%)	194 (12.3%)	56 (19.4%)	353 (13.4%)
Deaths	4 (0.3%)	1 (0.4%)	-	2 (0.1%)	3 (1.0%)	5 (0.2%)

4.2.2 Critique of submitted evidence syntheses

The meta-analysis was of three RCTs only – Option (WA17822), Lithe (WA17823) and Toward (WA18063) because the participants in these RCTs had an inadequate response to MTX or DMARD. The comparator in the meta-analysis was placebo. The meta-analysis is shown below:

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

From p67 and p94



The plots for the three RCTs show adjusted odds ratios rather than using the numerical results from Figure 7 of the submission but no numbers are given for the point estimates and confidence intervals in the plot. The adjusted odds ratios are adjusted by a number of factors but these are not listed on p67. They may be listed in the table on p95 of the submission, but this is unclear. If so, adjusting odds ratios by 20 different factors seems rather excessive. The meta-analysis is described as “pre-specified” (p64 submission) and there is mention of a “pre-specified pooling protocol” on p92 but the reference was to a Roche internal study report that was not sent to us. We are not told if inverse variance, Mantel-Haenszel or Peto’s method of study weighting has been used in pooling.

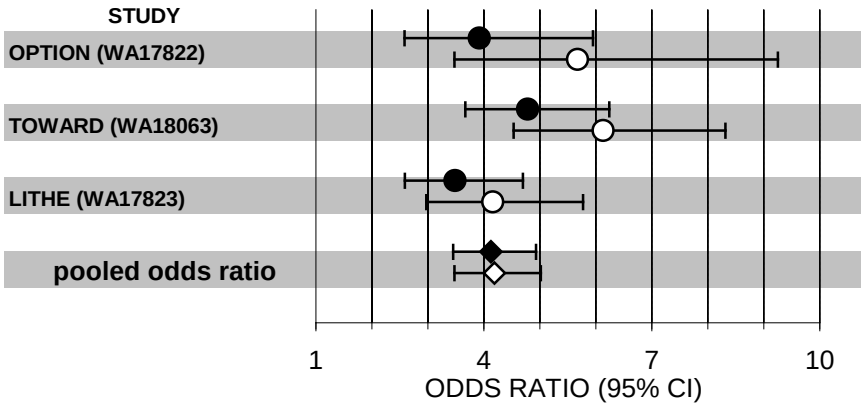
The meta-analysis result in the plot above does not appear to be based on the adjusted odds ratios from the three trials as it looks to be too far to the left. Also the scale is not logarithmic. The axis in the submission plot is labelled odds ratio but zero and minus-one appear on the axis; minus-one is an impossible value for odds ratio and makes the plot confusing. These values are familiar for log odds ratio (OR) but it is clear that the confidence intervals (CIs) correspond to an OR rather than a log OR (ie. they are not symmetrical about the point estimates). For a forest plot of OR to do this the x axis needs to be on a log scale. A plot of log OR would provide symmetrical CIs with the

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

axis linear on a log OR scale. These conventional forms of the forest plot have been ignored even though the routine was run in the SAS programme for which appropriate codes are available. Also, there are two footnotes to the plot – "odds ratios presented for studies WA17822, WA17823 and WA18063 are adjusted on site. Odds ratios presented for the pooled studies are adjusted on the study protocol". We don't understand this statement.

The results from Figure 7 give much lower odds ratios for all three trials (eg 3.9 compared to 5.6). This could be due to chance or a systematic bias. When we used the actual results from Figure 7 of the submission to estimate odds ratios and then pooled them, the result was similar to the pooled result in the Forest plot above (see Figure 1)

Figure 1. Redrawn meta-analysis ACR20

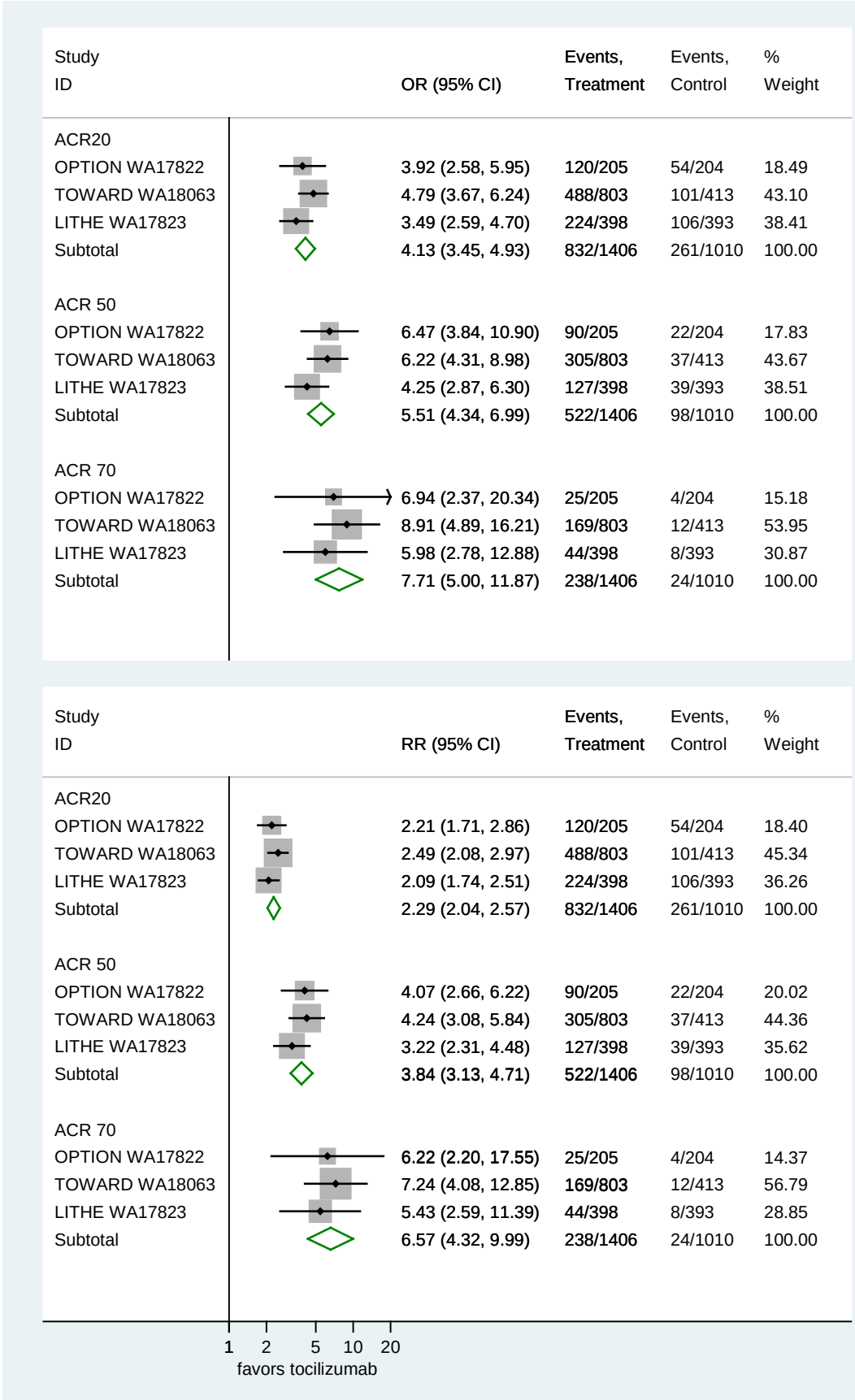


The forest plot has been redrawn to include the unadjusted odds ratios reported for the studies. The unadjusted odds ratios and their pooled estimate (Mantel-Haensel fixed effects model) are indicated by solid symbols. The hollow symbols represent the data in the submission for adjusted odds ratio and the pooled estimate provided (data read from graph).

Two more forest plots were generated using a fixed effects model (Mantel-Haenszel weighting) for all three ACR outcomes. Both odds ratio and relative risk are provided because the MS uses OR in the effectiveness section but relative risk in the MTC. Statistical heterogeneity was low (I^2 less than 20% in all cases) and lower for RR.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Figure 2. Redrawn meta-analysis of all three ACR outcomes OR and RR



Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

It is unclear if or how the meta-analysis was used subsequently in the submission. The submission used ACR 50 and ACR 70 results as well as ACR 20 scores to obtain HAQ values and thence utility estimates to feed the economic model. However no meta-analyses of these outcomes were presented.

4.2.2.1 Direct comparison of tocilizumab to conventional and biologic DMARDs

There are RCTs of direct comparisons of tocilizumab versus conventional DMARDs so these results could have been collected and presented in the submission. We present the results for these comparisons in Table 5. They show that tocilizumab is more effective than conventional DMARDs, particularly where the inclusion criteria for the RCT were that patients had an inadequate response to the DMARD.

No RCT direct comparisons to other biologic DMARDs were found.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 5. RCT comparison of tocilizumab to conventional DMARDs

	WA17824	MRA012JP	MRA213JP
Patients	Multicentre international (18 countries including Australia, Denmark, France, Italy, Spain, Norway, Portugal) Active RA who are either methotrexate naïve or who had discontinued methotrexate but not due to lack of efficacy or toxic effect	302 Japanese RA patients with inadequate response to current DMARD or immunosuppressant therapy	Japanese RA patients with inadequate response to methotrexate
Intervention	N=284, Tocilizumab only, 8 mg/kg iv every 4 weeks	Tocilizumab only, 8mg/kg iv every 4 weeks	Tocilizumab only, 8mg/kg iv every 4 weeks
Comparator	N=288, Methotrexate 7.5-20mg/week oral	Continue current therapy (methotrexate, other DMARDs or both)	Methotrexate 8mg/weekly
Outcome	24 week follow up ACR20 50, 70 response rate (results not given in submission, found in trial report pp134-9)	52 week follow up Modified Sharp erosion score, ACR20, 50 and 70	24 week follow up ACR20
Study design	Double blind double dummy	Open label, blinded evaluation	Placebo controlled
Results	ACR20 (ITT): Tocilizumab: 69.9% Methotrexate: 52.5% ACR50 (ITT): Tocilizumab: 44.1% Methotrexate: 33.5% ACR70 (ITT): Tocilizumab: 28.0% Methotrexate: 15.1% HAQ-DI=0 Tocilizumab: 20.0% Methotrexate: 5.2%	ACR20: Tocilizumab: 78% DMARD: 34% ACR50 Tocilizumab: 64% DMARD:13% ACR 70 Tocilizumab: 44% DMARD: 6% Total mean Sharp score Tocilizumab: 2.3 (1.5-3.2) DMARD: 6.1 (4.2-8.0)	ACR20: Tocilizumab: 80.3% Methotrexate: 25.0%
Comments	Also had open label extension (WA18696) but single arm, not RCT		

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

The mixed treatment comparison estimate objective specifies:

From p 13 of embedded file

Treatment for rheumatoid arthritis (RA) patients who have failed disease modifying antirheumatic drugs (DMARD-IR) including methotrexate (MTX) consists of combination therapy with an TNF_-inhibitor (etanercept, adalimumab, or infliximab) plus DMARD (including MTX). Tocilizumab, an IL-6 inhibitor, has been tested in 3 placebo-controlled trials (TOWARD, OPTION & LITHE) and was efficacious in this patient group. Since no head-to-head studies have been conducted, an indirect comparison versus TNF_-inhibitors, abatacept and rituximab has been performed.

The NICE scope specifies that one of the comparisons should be against conventional DMARDs. The comparator in the economic model is not versus conventional DMARDs but versus biologics only, even though there are RCT comparisons of tocilizumab to conventional DMARDs available with suitable ACR20, 50 and 70 results.

4.2.2.2 Past use of DMARDs

There was apparently no past use of DMARDs for some patients in three of the RCTs (Option WA17822, Lithe WA17823 and Toward WA18063) but the numbers in the clarifications document were only given combined. However, it is not clear whether this also includes no past use of MTX because the inclusion criteria for Option (WA17822), Lithe (WA17823) and Radiate (WA18062) were that they had to be taking MTX for at least 12 weeks immediately prior to baseline (see p43 submission). Only Ambition (WA17824) had an inclusion criterion that patients could be MTX naïve or not treated within 6 months prior to randomisation (see p43 submission). The Ambition Trial report states that 39.9% in the tocilizumab group and 45.4% in the MTX group were DMARD naïve (see p118 trial report). The combined result in the clarifications document was that 29.9% of 1406 in the tocilizumab arms and 23.1% of 1010 in the placebo arms had not had past use of DMARDs (see table 10, clarifications submission). Therefore more than 20% of the patients in these trials may not have met the UK licensed indication but the patients were equally distributed between the RCT intervention and control arms. (The total number of patients in each of the trials in the 8mg/Kg tocilizumab arms was 1411 (205 (WA17822), 401 (WA17823) and 805 (WA18063)) and in the placebo arms was 1013 (204 (WA17822), 394 (WA17823) and 415 (WA18063)).

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

The impact of having DMARD-naïve patients is possibly that there would be more initial benefit from treatment so the impact of tocilizumab would be attenuated. The subgroup analysis of ACR20 response at week 24 for 2416 pooled tocilizumab trial participants from the Option, Lithe and Toward RCTs on page 217-8 of the submission showed a lower OR for the tocilizumab compared to placebo arms in DMARD naïve patients.

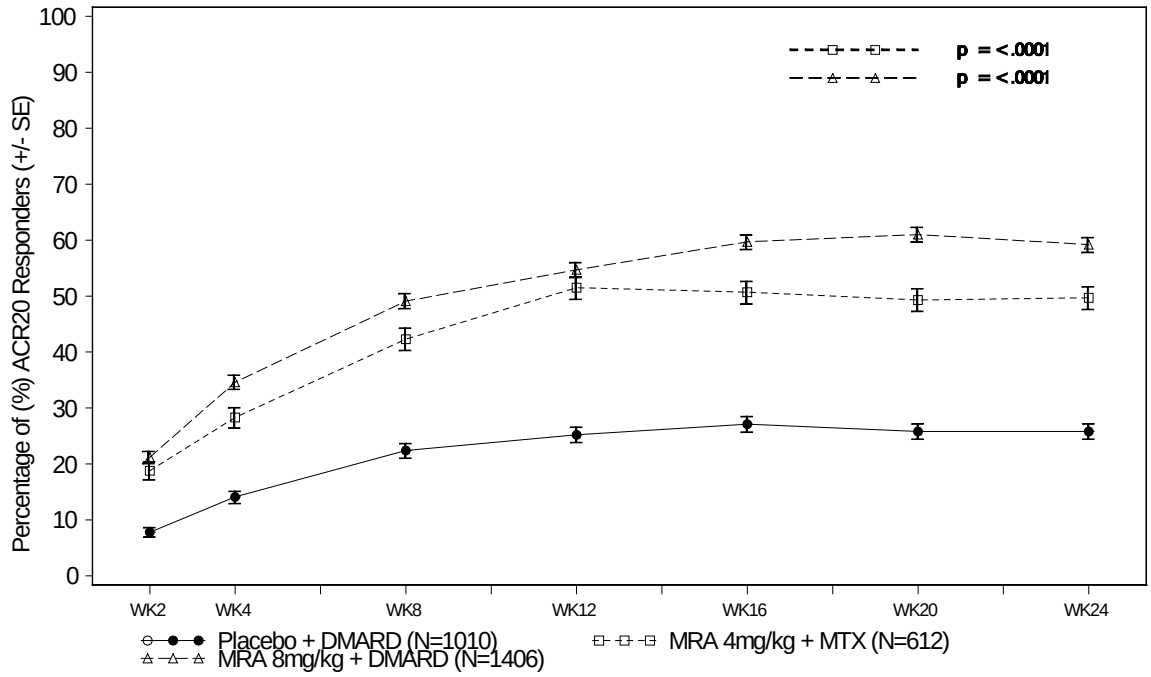
4.2.2.3 Early impact of tocilizumab

There appears to be a systematic difference in ACR20 scores at 2 weeks between intervention groups and control group in several of the trials. This is not clearly described in the submission when it presents the results of the RCTs of tocilizumab versus placebo but can be seen in the meta-analysis result presented on p97 of the submission. This early effect is consistent with the finding that CRP levels show a drop by week 2 after tocilizumab administration.[7]

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

See p 97

EGacr20pli ACR20 Response Rates by Visit - 6 Month Pooled Data (ITT Population)



Cochran-Mantel-Haenszel analysis was used to calculate p-values. All comparisons to placebo + DMARD. LOCF used for joint counts, no imputation used for missing HAQ Score, CRP, ESR and VAS assessments. CRP is used primarily for the calculation of the ACR response, if missing, ESR will be substituted. Patients who receive escape therapy, withdraw prematurely or where an ACR can not be calculated, will be set to 'Non Responder'.

Program : $PROD/cd11935h/EGacr.sas / Output : $PROD/cd11935h/reports/EGacr20pli.cgm
30JUL2007 19:06

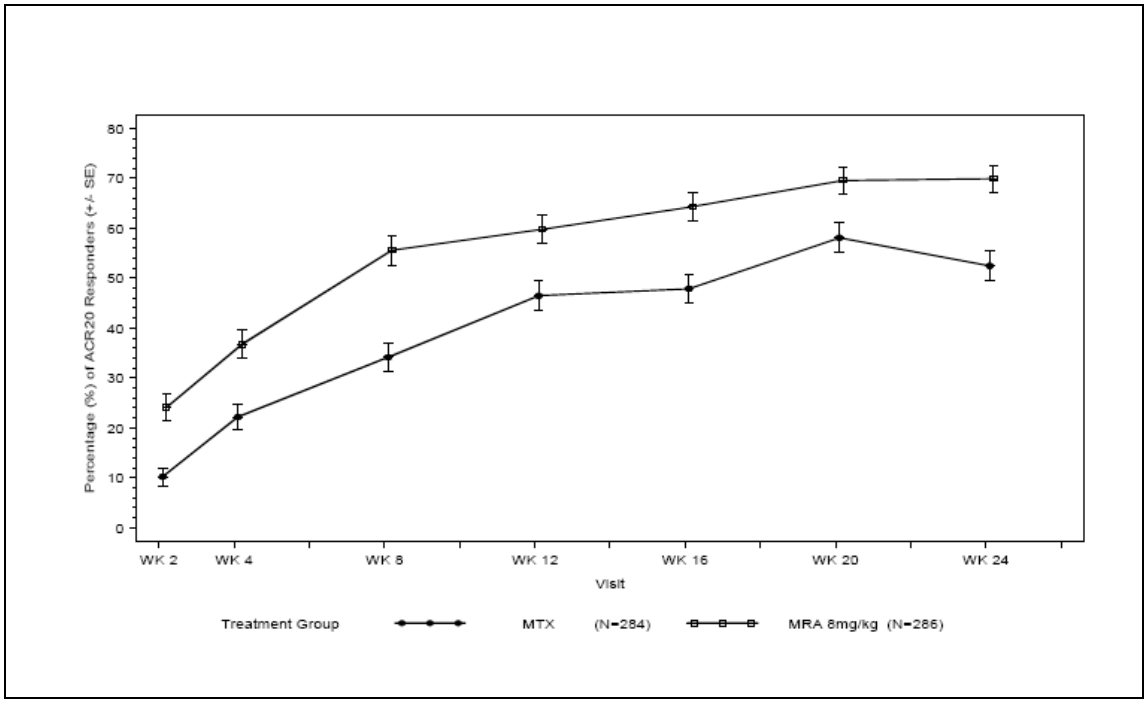
-

It is also shown in the RCT of tocilizumab versus methotrexate (Ambition, WA17824)

See p80-81

The proportion of patients with an ACR20 response at Weeks 2, 4, 8, 12, 16, 20 and 24 is shown in Figure 15 below for the ITT population. Similar plots of ACR50 and ACR70 responses are presented in Figure 16 below and Figure 17. ACR20 response rates in the tocilizumab group were consistently higher than those in the MTX group over the course of the study. Clear separation between the tocilizumab and MTX groups was observed from as early as Week 2 for ACR20 response, at which point the response rate was 24.1% in the tocilizumab group vs. 10.2% in the MTX group. The ACR20 response rates increased over time in both the tocilizumab and MTX groups before stabilizing at Week 20 to Week 24 and decreasing in the MTX group.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.



However, any comparison of tocilizumab vs MTX which incorporates this early effect would not be a fair comparison as MTX can take several weeks to have an effect. Most conventional DMARDs are slow to start working in RA, unlike biologic DMARDs.

4.2.3 Summary

In summary, the submission includes estimates of effectiveness of tocilizumab versus placebo and it is possible to estimate the effectiveness of tocilizumab versus conventional DMARDs. However, the clinical question is the effectiveness of tocilizumab versus other biologic DMARDs as it is most likely to be one of these that it could replace in the treatment pathway. There is no direct evidence of tocilizumab versus any biologic DMARDs. Therefore a mixed treatment comparison has been developed for the submission which attempts to compare tocilizumab to some other biologic DMARDs and this is described below in the economic section of this report.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

5 ECONOMIC EVALUATION

5.0.0 Description of manufacturer's search strategy and comments

Summary from the manufacturer's submission:

The review updated and extended the searches in the review by Chen et al 2006. This review was included as were ten already identified studies from the report. Two search strategies (Appendix 5 and 6) from the review by Chen et al 2006 were combined and adapted for use in: MEDLINE, EMBASE, MEDLINE In Process, HTA database (Cochrane Library 2008 Issue 4), NHS EED (Cochrane Library 2008 Issue 4) and HEED between 24th December 2008 and 6th January 2009. Only articles in English were included. Searches for Adalimumab, Etanercept and Infliximab were limited to 2005 onwards, no time limits were applied to searches for the other interventions (Abatacept, Golimumab, Certolizumab Pegol and Rituximab).

Comment:

- The searches were sound overall and were unlikely to have missed relevant studies.

5.1 Overview of manufacturer's economic evaluation

As there were no UK-based economic evaluations of tocilizumab, Roche conducted a *de novo* economic model. The table below summarises the key features of the model.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 6. Overview of economic model including the source of the information in terms of the numbered location in the manufacturer's submission (CIC)

Property	Description	Location
Type of model	Individual sampling model with a hypothetical cohort of 10,000 patients.	7.2.6.1
Code used	Visual Basic for Applications within MS Excel front end.	7.2.6.1
Time horizon	Patient lifetime.	7.2.6.5
Discount rate	3.5% a year for costs and utilities.	7.2
Perspective	NHS and Personal Social Services (effectively NHS).	7.2
Population	Two patient cohorts representing moderate to severe RA patients with an inadequate response to one or more traditional DMARDs or one or more anti-TNF-α agents respectively (see Section 7.2.2.1 of the submission).	7.2.3
Intervention	Tocilizumab + MTX added to a treatment sequence.	7.2.3
Comparator	Treatment sequence without tocilizumab + MTX.	7.2.3
Treatment effect	Initial drop in HAQ score (magnitude depending on type of ACR response), followed by long-term change in HAQ score while on treatment.	7.2.6.1
Health related quality of life	Change in EQ-5D, using a quadratic relationship fitted to tocilizumab trial data.	7.2.8.3
Costs	Main treatment cost is cost of drug; tocilizumab assumed to have the same acquisition cost as etanercept but a higher administration cost.	7.2.9.1
Sensitivity analyses	Scenario and probabilistic sensitivity analyses.	7.2.11

5.1.1 Populations

Two patient cohorts were considered, consistent with the licensed indication of tocilizumab (see Section 7.2.2.1 of the submission on p135-6):

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

1) Moderate to severe RA patients who have had an inadequate response to one or more traditional DMARDs.
2) Moderate to severe RA patients who have had an inadequate response to one or more anti-TNF-α agents.

It should be noted that no subgroup analysis of patients intolerant to MTX was conducted, even though these patients are specifically mentioned in the scope. The submission assumes that tocilizumab will be administered together with MTX so the submission has not shown any evidence of cost-effectiveness in a MTX-intolerant population.

5.1.2 Perspective and time horizon

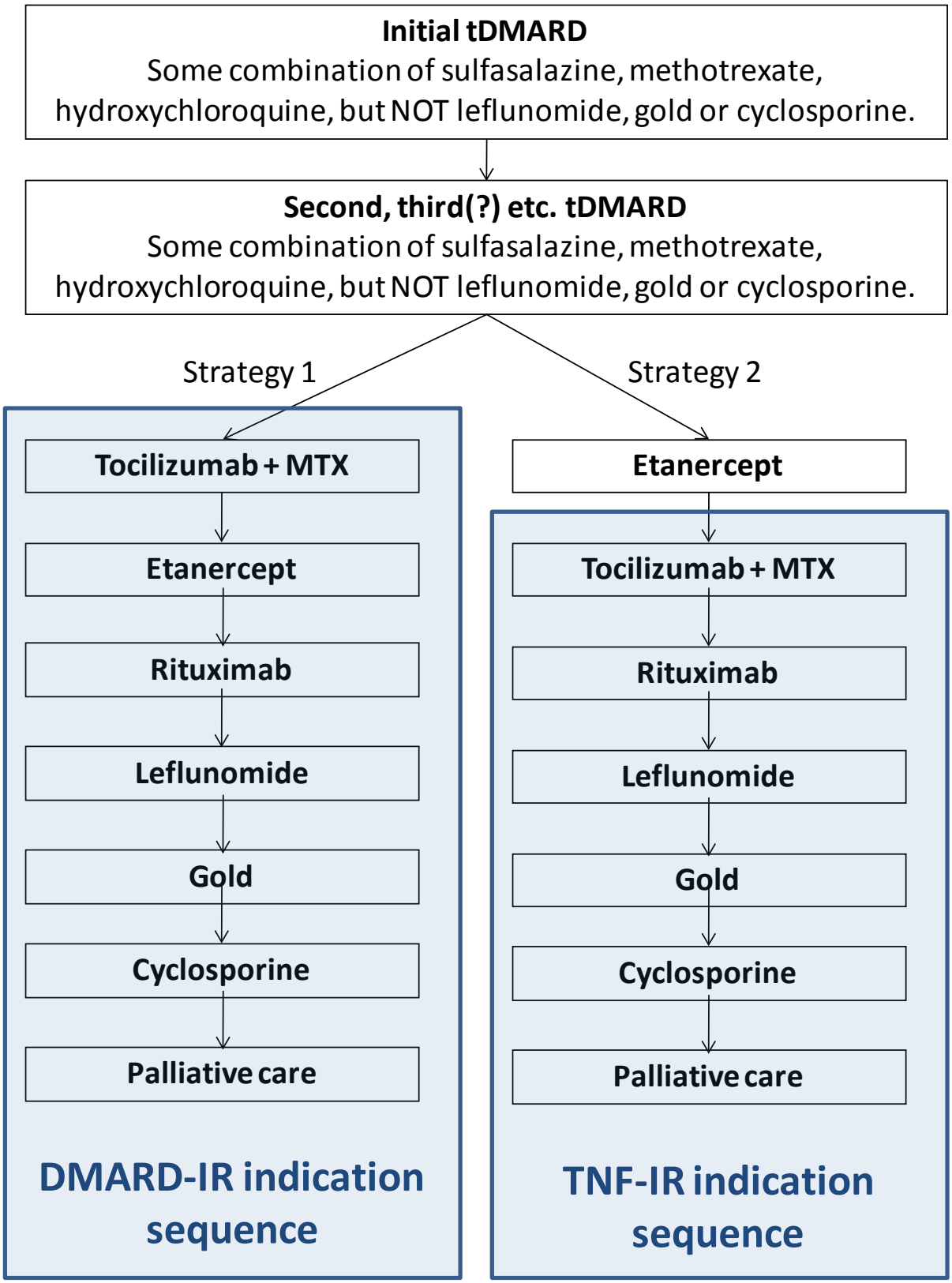
Costs were considered from an NHS and Personal Social Services perspective (in practice NHS only), consistent with the NICE reference case. A time horizon over a patient lifetime was applied, as is appropriate for a chronic disease.

5.1.3 Treatment and comparator

The model estimated the incremental cost-effectiveness analysis of adding tocilizumab + MTX to the treatment sequence given either a DMARD-IR indication or a TNF-IR indication. The actual sequence of treatments used in the tocilizumab arms of the two versions of the model is summarised in Figure 3 below. There is only one comparator sequence, which is the same as either sequence below but without the tocilizumab plus MTX option.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Figure 3. Comparison of the two tocilizumab strategies used in the economic model



Note that the two arms of the figure (Strategy 1 and Strategy 2) indicate the treatment sequence under the two indications (DMARD-IR and TNF-IR); the comparator is the same sequence with tocilizumab + MTX removed.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Only the section of the sequence shaded in blue is actually modelled, with separate models used for the DMARD-IR and TNF-IR indications – the remaining sections of the pathway are implied. Each treatment is assumed to be given to indicated patients for minimum of six months and for as long as the clinician is satisfied that patients are achieving benefit (defined as achieving an ACR 20 or greater response). If a treatment is stopped, then the patient is moved to the next treatment in the sequence. The numeric estimates of the treatment effects in the model were from a mixed treatment comparison (MTC) of patient responses in RCTs of biologic DMARDs.

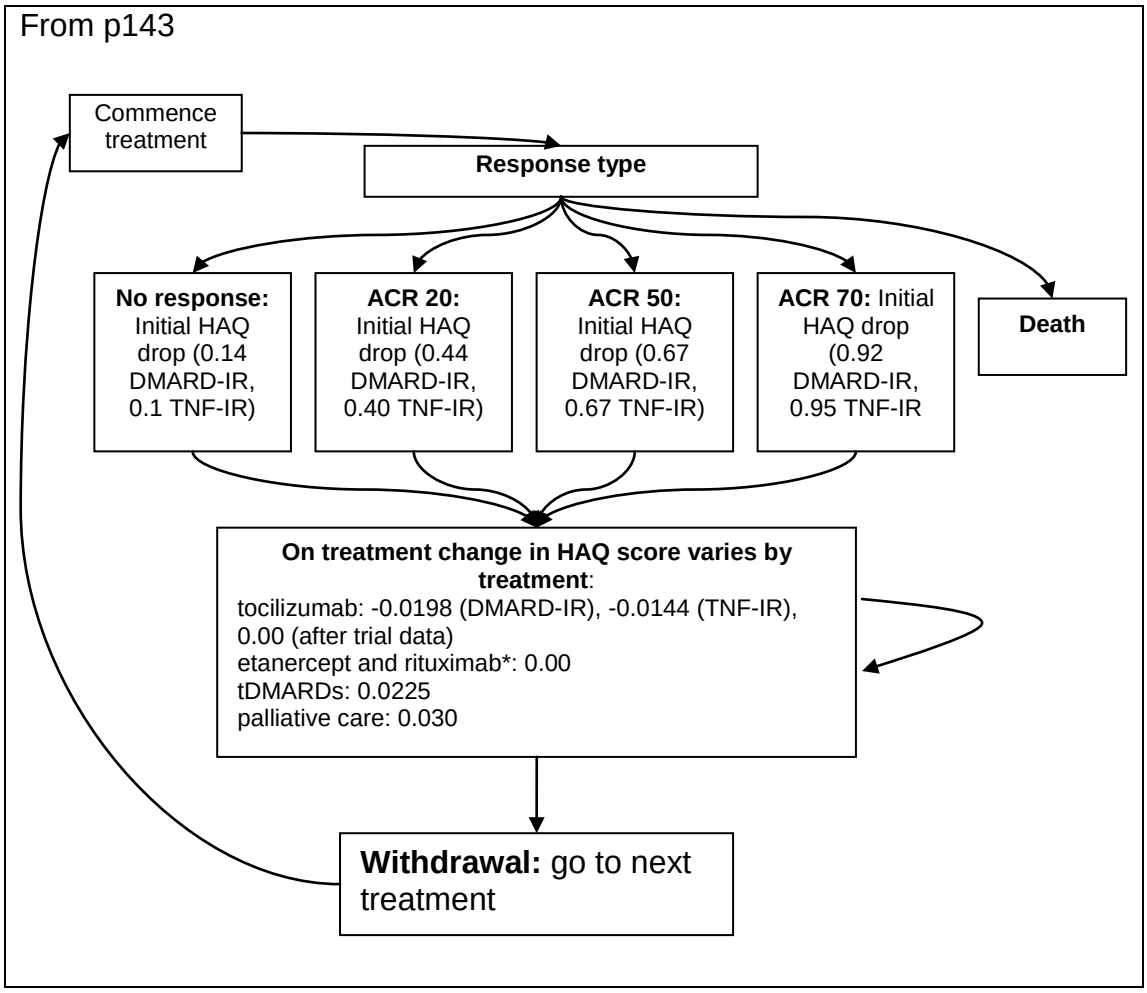
5.1.4 Natural history

The Roche model does not describe the natural history of RA. The model only describes disease progression in terms of response to treatment (measured in terms of ACR 20, ACR 50 and ACR 70 criteria, as well as HAQ and EQ-5D scores).

5.1.5 Treatment effectiveness within the submission

Treatment response was modelled in terms of achievement of ACR 20, ACR 50 or ACR70, and also in terms of changes in patients' HAQ scores. As there were no head-to-head trials comparing tocilizumab to other biologic DMARDs, the model used ACR response rates from a MTC outsourced to a consultancy (Mapi Values). The figure below indicates the way treatment response was represented.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.



The steps involved in modelling the effect of treatment are summarised in the following steps:

1) At the start of a specific treatment, a patient may have had no clinical response or achieved an ACR 20, ACR 50 or ACR 70 response. The probability of each of these responses for different treatments was based on the adjusted response rates (given in Tables 36-39 on pages 145-6) from the accompanying MTC.

2) Each type of response was associated with an initial drop (i.e. improvement) in HAQ score. The HAQ score improvement associated with each ACR response was obtained using analyses from four tocilizumab trials (OPTION (WA17822), LITHE (WA17823) and

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

TOWARD (WA18063) combined for the DMARD-IR version and RADIATE (WA18062) for the TNF-IR version) and assumed to be the same for other treatments (except presumably palliative care, although this is not mentioned). It is not specified if the meta-analysis presented on p67 and p94 was used for the DMARD-IR version. The meta-analysis was only for ACR20 so if it was used, the ACR50 and ACR70 meta-analyses are missing. If it wasn't used then it is unclear why it was presented.

3) In each six monthly cycle, a patient's HAQ score was assumed to change depending on the kind of treatment the patient received. If the patient was on tocilizumab, the HAQ score was assumed to drop, based a decreasing HAQ score observed in long-term follow-up trial data from studies WA18695 and WA18696. This is described as 5-year follow up data in table 5 of the submission on p41 but described as 3 year follow up data on p149. Beyond the limit of long term extension study data (~180 weeks), it was assumed that the HAQ score would stay constant. For etanercept and rituximab, no change in HAQ score was assumed. For traditional DMARDs and palliative care, an increase in HAQ score was assumed.

4) For every six monthly cycle, patients were assumed to have a probability of being withdrawn from treatment of 0.10 for biologic DMARDs (including tocilizumab) and 0.27 for traditional DMARDs (see Table 33 of the submission on p134).

5) When a patient was withdrawn, a "rebound" was assumed, which caused an increase in HAQ equivalent to the initial drop in HAQ score reported in step 2.

6) If a patient was withdrawn, the patient was moved to the next treatment in the cycle until he or she reached the palliative care stage.

5.1.6 Health related quality of life

The model represented patients' health states in terms of their HAQ scores. These were mapped to health-related quality of life weights measured using

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

the EQ-5D instrument. The relationship between these two measures was estimated by using linear regression on outcomes from two DMARD-IR tocilizumab trials (OPTION (WA17822) and LITHE (WA17823)).

5.1.7 Resources and costs

Tocilizumab is given every 4 weeks by intravenous injection. The licensed dose is 8mg/Kg. The minimum dose is 480mg. For a 70Kg person, the dose would be 560mg. Tocilizumab is available in IV ampoules of 80mg, 200mg and 400mg. CIC The provisional NHS price is £1.28/mg. Therefore the cost of an 80mg ampoule would be £102.4, 200mg £256 and 400mg £512. The submission states that the cost of tocilizumab would be £9,295 per annum. This would be 13 injections of £715 or 558.59mg, ie very close to 560mg for a 70Kg person. The annual cost for a 70Kg person without vial sharing would still be £9,295 if a 400mg and two 80mg vials were used but £9,984 if a 400mg and a 200mg vial were used instead. It is possible that the person drawing up the injection would opt for opening two vials rather than three in actual practice.

A list of unit costs of resources used in the validation is provided (see Table 45 of the submission on p166). Tocilizumab is priced so that its average annual acquisition cost is exactly the same as that for etanercept (£9,295), but is assumed to have higher administration and monitoring costs (£1,843 a year or 16% of total medical costs) as it is delivered by infusion rather than subcutaneous injection.

The impact of drug-related adverse events on quality of life and medical resources was assumed to be negligible (see section 7.2.7.4 of the submission on p160).

5.1.8 Discounting

Both costs and health outcomes were discounted at a rate of 3.5%, consistent with the NICE reference case.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

5.1.9 Sensitivity analyses

Two types of sensitivity analyses were conducted:

1) A scenario analysis looked at the effect on the overall ICER of changing particular assumptions in the model, such as using a different biologic DMARD instead of etanercept in the treatment sequence.

2) A probabilistic analysis looked at the effect of varying parameters governing treatment effect and costs.

Univariate sensitivity analysis was not conducted.

All the scenarios in the scenario analysis, and over 99% of samples in the probabilistic analysis, lay below a threshold of £30,000 per QALY gained.

5.1.10 Model validation

The submission states that an independent reviewer verified the input sources, programming and face validity of the model results. However, details of the validation procedures and the identity of the reviewer were not provided. It has been assumed that the reviewer was in-house and hence independent in terms of the process of model construction rather than in terms of not having any conflicts of interest.

5.2 Critique of approach used

A summary of the model and critical appraisal of its features can be found in Appendix 3.

5.2.1 Treatment effectiveness

5.2.1.1 Use of pooled parameters

The model combined effectiveness parameters for different anti-TNF-α agents. The submission justified this on the basis of the lack of differentiation in efficacy of these agents in published NICE guidelines (HTA 130 section 4.3.3)[3] and published research. However, the literature reviewed by the submission suggested that the anti-TNF-α agents do not have the same efficacy. For instance, the 6-month withdrawal rate from treatment reported in

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

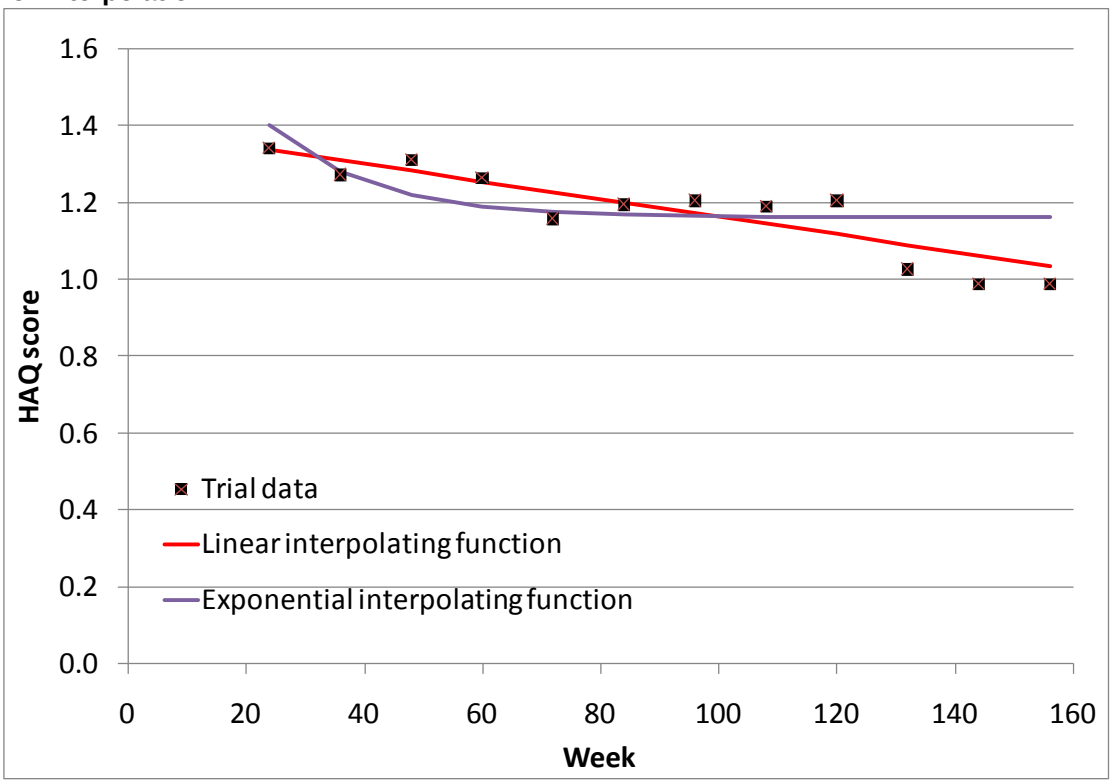
a study by Geboreck and co-workers[21] was 8% for etanercept and 12% for infliximab. In the Roche model however, the two rates were combined to produce an overall rate of 10%. Similarly, in the MTC (see MTC section below), the ACR adjusted response rates for all anti-TNF-α agents analysed (adalimumab, etanercept and infliximab) were combined, although etanercept appeared to be the most efficacious (see MTC section below). Etanercept appeared to be both the most effective anti-TNF-α treatment option and the most commonly used, so this practice artificially reduces the effectiveness of the existing treatment schedule.

5.2.1.2 Modelling of HAQ score progression on long term treatment

The data points in Figure 37 and 38 on p 150 and 152 are difficult to interpolate because of the small numbers towards the end of the follow-up period. For instance, in Figure 38, if the data points after week 120 with extremely large confidence intervals are excluded then the HAQ score appears to stabilise after about week 70. Equally, one could argue that an exponential would be a better function to use for fitting the data to on theoretical grounds (since the effect of tocilizumab on HAQ scores appears to decay over time). We were unable to obtain individual-level patient data to fit curves to, but if the mean HAQ score was assumed to have a normal sampling distribution, then fitting an exponential to the data and weighing the goodness of fit by the inverse of the standard error of the data means would give the following curve that is less dependent on the uncertain data points towards the end of the follow up period (see Figure 4).

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Figure 4. Modelling HAQ score progression on treatment using two different functions for interpolation.



Our purpose is not to argue for an exponential fit, but to suggest that a case could be made for a number of functional forms, all of which could fit data points. This is important because the HAQ progression assumptions are central to the cost-effectiveness estimates. Here the submission argued that tocilizumab, alone out of all available DMARDs, was able to improve patients' HAQ scores while they were on treatment. The results used to justify this were from their trials, but any models fitted to the data need to be constructed extremely carefully because even a slight change in predictions will have a large change in the incremental cost-effectiveness ratio.

Note that explanation in the clarifications document (A14) is very unclear as to exactly which of the graphs was correct or incorrect, and which of them were changed from the old version of the submission to the new version.

5.2.1.3 Modelling the relationship between HAQ score and EQ-5D score

We were unable to obtain the data used to obtain the best fit parameters for the curves in Figure 42 of the submission on p164. However, judging from the size of the intervals indicating standard errors in Figures 23 (p 91) and 28 (p

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

103) (showing the change in HAQ scores) as well as Figures 40 and 41 (on p 154-5, showing the change in EQ-5D scores), we would surmise that the majority of data points may lie in a narrow range. This range is likely to centre around 0.65 – 1.75 on the HAQ scale and 0.60 – 0.80 on the EQ-5D scale. Hence the extrapolation of a regression curve fitted to data mostly lying in a narrow band to the entire length of the two scales is questionable. It is worth noting that a certain amount of HAQ disability cannot be altered i.e. some residual disability may remain despite optimal control of inflammatory disease because of damaged joints from previous disease. It does appear that the entire spectrum of EQ-5D scores is used, since the scenario analysis (Table 53 on p180) indicates that removing negative utility scores increased the ICER slightly from £19,870/QALY to £20,214/QALY. Negative utility scores were used for health states that were considered to be "worse than death" – so clearly considerably worse than the relatively healthy individuals in the trials largely reporting EQ-5D scores between ~0.60 – 0.80 (see Figures 40 and 41 of the submission on pages 154-5) or the results of the Option (WA7822) RCT showing patients had an EQ-5D of 0.39 at baseline and between 0.53 and 0.67 at follow up.

Also the base case of the model does not use age as a covariate (see p164 of the submission). This is justified in the clarifications document (B13 on p 36-7) which indicates that when age is added as a covariate in the function mapping HAQ scores to EQ-5D scores, there is only a 0.017 change in utility for 20 years' difference in age. There was little change in ICERs when age was incorporated into the estimate. This lack of utility change with age is surprising as it is known that HAQ scores rise dramatically in older patients with RA[15] and may be due to multicollinearity in the data. In the clarifications document, Roche accepts the potential for multicollinearity between age and HAQ score. However, they assert that this does not reduce the predictive power or reliability of the model as a whole since the purpose of the model is not to explain the effect of age as an independent variable. While this is true, it does not address the possibility that age may be a confounding variable and hence limit the applicability of the HAQ – EQ-5D mapping equation to certain subgroups only.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

The submission used a quadratic equation to model the relationship between EQ-5D and HAQ scores. The quadratic model predicts that EQ-5D scores will be considerably lower at high HAQ scores compared to a linear model. For instance, a HAQ score of 3 corresponds to an EQ-5D score of about -0.2 with a quadratic model and 0.1 with a linear model. The quadratic model was chosen based on goodness of fit to data, and also work done by Boggs and colleagues from a database of 2070 patients.[22] This is considerably larger than the database of 240 patients that was used in a previous HTA.[11] However, this information is from a conference abstract only and to our knowledge has not been peer-reviewed. Furthermore, the manufacturers did not explore which of the two models were more appropriate from the viewpoint of parsimony. We were unable to do this ourselves since the information to which the models were fitted was not supplied.

Boggs states that "These algorithms' predictions are limited and should only be used when direct utility scores are not available".[22] EQ-5D was measured in the Option (WA17822) and Lithe (WA17823) RCTs.

With reference to peer-reviewed papers, work by Scott and colleagues found that, from a sample of 321 RA patients, HAQ showed a Gaussian distribution whereas EQ-5D was bimodal.[23] Also in treatment change studies, changes in EQ-5D and HAQ did not show a significant association (r=0.08). They state that "As HAQ and EuroQol are demonstrably not equivalent, economic evaluations of treatment cost effectiveness should not be based on EuroQol data transformed from HAQ." [23] Conversely, another study of 300 RA patients found that EQ-5D and HAQ were closely correlated at baseline (r=-0.74).[24] Kobelt investigated a cohort of 916 UK RA patients and found that as their quality of life worsened the EQ-5D became more variable (ie. the standard deviation increased).[25] It would seem sensible that this amount of heterogeneity between studies would warrant further investigation, particularly as the link between HAQ and EQ-5D is currently a key part of RA economic models.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 7. HAQ score and EQ-5D score in a UK cohort

HAQ score	UK cohort EQ-5D mean (SD)
<0.6	0.7459 (0.1402)
0.6<1.1	0.6491 (0.2053)
1.1<1.6	0.4692 (0.2678)
1.6<2.1	0.4419 (0.2873)
2.1<2.6	0.2556 (0.2908)
>2.6	0.2538 (0.3514)

5.2.1.4 Modelling of the rebound effect after withdrawal from treatment

When treatment was withdrawn due to lack of efficacy, the models assumed that patients underwent a "rebound" where their quality of life (or equivalent proxy such as HAQ score) worsened. In the BRAM model[26] (on which Roche based their own model) the rebound was assumed to return the patient to the quality of life state he or she would be in without treatment (that is, if he or she received palliative care only).

However, the Roche model made a different assumption about the rebound effect. In their submission, the rebound was assumed to be equal to the initial HAQ improvement only – that is, the patient retained any benefit from the long-term progression following treatment. This assumption disproportionately favours tocilizumab, since the Roche model assumed that patients receiving tocilizumab (but no other DMARD, whether biologic or traditional) actually improved their HAQ score over the course of treatment.

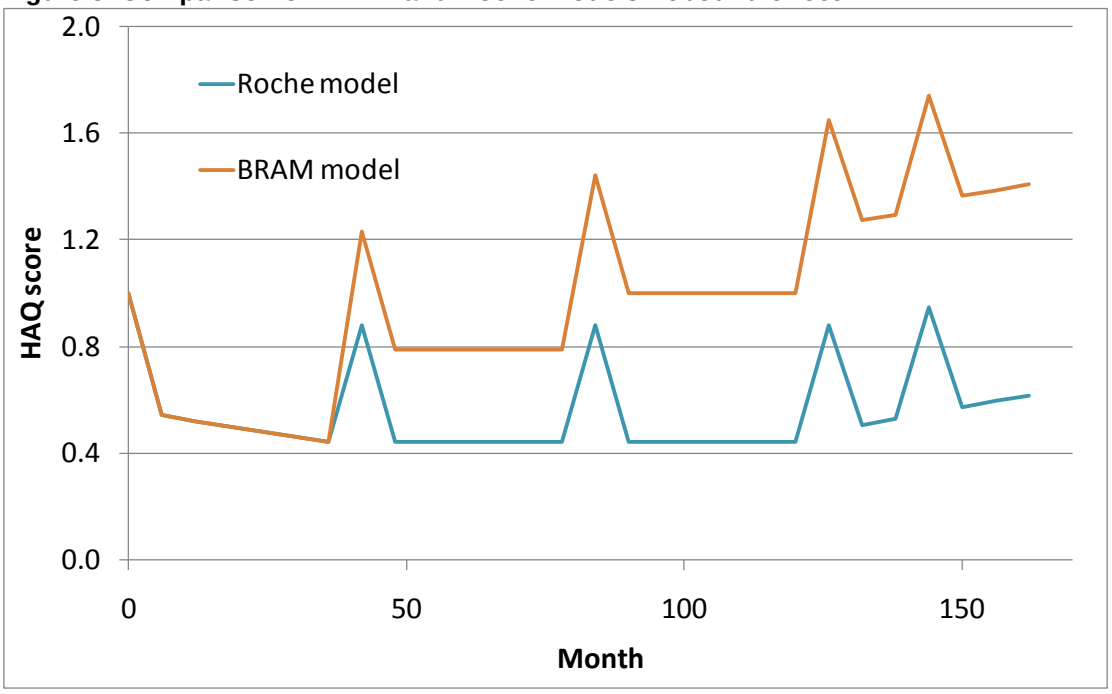
Roche's response to a query about this is twofold. Firstly, they raised the point that the action of DMARDs in delaying RA progression may itself suggest that treatment benefit does not disappear entirely at withdrawal. They also pointed to several other published models of biologic DMARDs that have used a similar assumption.

However, the comparison with other models is not entirely accurate since these models assumed that the effect of the long-term phase of biologic DMARD treatment was to stabilise HAQ scores. In contrast, the Roche model assumed that tocilizumab could actually improve HAQ scores over the course of treatment. By making both this assumption and favourable assumptions about the rebound effect, the Roche model in effect allowed tocilizumab not

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

simply to delay the course of disease progression, but to actually elicit a lasting improvement in the condition of the patient. Because this improvement lasted until the patient died, the net gain in quality of life could be dramatic. This is illustrated in Figure 5 below, which shows the difference in the HAQ score of a patient with a DMARD-IR indication being treated first with tocilizumab (for 42 months), followed by etanercept (42 months), rituximab (42 months), gold (18 months) and finally cyclosporin (18 months) using a BRAM model and the Roche model.

Figure 5. Comparison of BRAM and Roche models' rebound effect



Secondly, in the Roche model they re-ran their probabilistic sensitivity analysis adding an additional parameter in which the rebound in HAQ score was varied by a fixed proportion of the initial drop at the onset of a new treatment, with this proportion sampled from a uniform distribution between 80% and 120%. However, this additional analysis does not address the issue at all, because the rebound is still dependent on the initial drop rather than the cycle-on-cycle change in the HAQ score. Making the rebound depend on the latter would increase the size of the rebound beyond 120% in most patients in the model during their tocilizumab treatment, but make no difference in the case of patients receiving traditional DMARDs.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

5.2.1.5 Effectiveness of leflunomide, gold and ciclosporin

Table 37 on page 146 of the submission is reproduced below. It shows that patients with a DMARD-IR indication being treated with leflunomide, gold or ciclosporin are assumed to have the same ACR response rate as those on palliative case. This assumption was also made in the TNF-IR version of the model, as shown in table 39 on page 146. This assumption appears to be questionable since it seems likely that patients on these drugs would have better response rates than those on palliative care only.

Treatment	ACR 20	ACR 50	ACR 70	No Response
Tocilizumab	0.21	0.15	0.29	0.35
Etanercept	0.24	0.23	0.16	0.37
Rituximab (DMARD-IR)	0.25	0.17	0.18	0.40
Leflunomide	0.11	0.03	0.01	0.85
Gold	0.11	0.03	0.01	0.85
Ciclosporin	0.11	0.03	0.01	0.85
Palliative care	0.11	0.03	0.01	0.85
Adalimumab	0.24	0.23	0.16	0.37
Infliximab	0.24	0.23	0.16	0.37

5.3 Results included in manufacturer's submission

The Roche model suggested that adding tocilizumab to the treatment schedule for indicated patients was likely to be cost-effective, with an incremental cost-effectiveness ratio of about £20,000 per QALY gained for DMARD-IR patients and about £22,000 per QALY gained for TNF-IR patients. The more favourable cost-effectiveness profile for DMARD-IR patients was due to the fact that, in the model, tocilizumab was more cost effective than etanercept, and hence benefitted from being administered earlier (and hence discounted at a lower rate) in the treatment schedule. Sensitivity analyses suggested that all the scenarios explored and all the samples obtained during Monte Carlo sampling had ICERs below £30,000/QALY. The results are summarised in Table 8.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 8. Results of Roche economic model

	DMARD-IR indication	TNF-IR indication
Base case		
Incremental costs (£)	23,253	26,640
Incremental QALYs gained	1.17	1.210
ICER (£/QALY)	19,870	22,003
Scenario analyses		
Range of ICER values (£/QALY)	15,878 – 24,905	19,026 – 27,435
Probabilistic sensitivity analysis		
% of samples below £20,000/QALY	56.4%	22.4%
% of samples below £30,000/QALY	100%	100%

5.3.1 Mixed treatment comparison (MTC)

Roche commissioned a MTC in order to estimate ACR 20, 50 and 70 response rates to different forms of RA treatment using data from heterogeneous studies. The analysis for ACR 20 and 50 used a random effects model, but the analysis for ACR70 used a fixed effects model. This was justified on the basis that the ACR 70 treatment effects for TNFα inhibitors were relatively homogeneous. However, the Cochrane's Q statistic for the ACR 70 outcomes was borderline (p = 0.0603, see Section 6.6 on p107 of the submission), and it was not reported whether this assumption was investigated by comparing the conditions of each study included in the MTC. Use of a fixed effects model would reduce the predicted variability in treatment effect as it assumes that the only differences between studies are due to the treatment being investigated.

The MTC was conducted using a Bayesian model coded in WinBUGS. Although Roche supplied us the code used for the analysis, due to time constraints we were not able to analyse the logic of the code in order to independently reproduce the analysis. However, we have conducted a simple

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

analysis using the point estimates of treatment effect and size of each arm in the studies analysed. The point estimates from this analysis were fairly similar to the analysis presented in the submission and indeed look more favourable towards tocilizumab (largely because we did not use a pooled estimator for the response in the placebo arms).

Table 9. MTC and reanalysis of results

Treatment (v. placebo)	MTC treatment effect (base case analysis results)			Reanalysis treatment effect		
	ACR 20	ACR 50	ACR 70	ACR 20	ACR 50	ACR 70
Tocilizumab	2.06	3.60	6.75	3.19	5.36	10.83
Anti-TNF-α	1.99	3.19	3.81	2.02	2.92	4.01

Using our reanalysis, we were able to explore the data in ways that the submission did not consider. In particular, we were concerned with two aspects of the submission:

1) The submission combined the treatment effect for different anti-TNF-α agents (adalimumab, etanercept and infliximab) on the basis that NICE guidelines and published research suggest that the efficacy of these agents is similar. However, when we conducted separate analyses for each agent, we found marked differences in their treatment effects.

Table 10. Treatment effects of biologic DMARDs separately

	Relative treatment effect v. placebo			Origin
	ACR 20	ACR 50	ACR 70	
Tocilizumab	3.19	5.36	10.83	Based on 3 RCTs
Adalimumab	1.93	3.51	5.88	Based on 4 RCTs
Etanercept	1.65	2.12	3.22	Based on 4 RCTs
Infliximab	2.56	3.69	3.89	Based on 3 RCTs
This is analogous to the base case in MTC				

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

2) The reason for the low efficacy of etanercept compared to both tocilizumab and other anti-TNF-α agents was a single large trial by Klareskog and colleagues[27] with a very high response rate in the placebo arm. This was also confirmed by the clarifications document where they sent us a table of the intervention and comparator/placebo rates in a number of biologic DMARD RCTs for RA. (see table 15 on p 25 of clarifications document)

Clinical trial		% ACR20			% ACR50		
				Bio v comp p value			Bio v comp p value
Author	Protocol	Bio	Comp		Bio	Comp	
Genovese	ETAN v MTX	72	59	0.005	49	42	NS
Cohen	ANAK+MTX v PBO+MTX	42	23	0.018	24	4	0.003
Weinblatt	ADA+MTX v PBO+MTX	67	15	<0.001	55	8	<0.001
Furst	ADA+DMARD v PBO+DMARD	53	35	<0.001	29	11	<0.001
Kremer	ABAT+MTX v PBO+MTX	60	35	<0.001	37	12	<0.001
Keystone	ADA+MTX v PBO+MTX	59	24	<0.001	42	10	<0.001
St Clair	INFL+MTX v PBO+MTX	62	54	0.028	46	32	<0.001
Klareskog	ETAN+MTX v ETAN v MTX	85	75	0.009	69	43	<0.001
Edwards	RTX+MTX v MTX	73	38	0.003	43	13	0.005
	RTX+CYCLO v MTX	76	38	0.001	41	13	0.005
	RTX+CYCLO v RTX+MTX	76	73	Not shown	41	43	Not shown
WA17822	TCZ+MTX v PBO+MTX	59	27	<0.001	44	11	<0.001
WA18063	TCZ+DMARDs v PBO+DMARDS	61	25	<0.001	38	9	<0.001
WA17823	TCZ+MTX v PBO+MTX	56	27	<0.001	32	10	<0.001
WA18062	TCZ+MTX v PBO+MTX	50	10	<0.001	29	4	<0.001
WA17824	TCZ v MTX (ITT POP)	70	53	<0.001	44	34	0.0023

We have critically appraised the Klareskog RCT[27] to establish whether it was substantially different from the other RCTs in the MTC. The critical appraisal is in Appendix 4. The main differences are that it only included RA patients who might benefit from MTX and had an aggressive dosing schedule of MTX if signs and symptoms of RA reappeared. When we reanalysed the data removing the Klareskog trial, the efficacy of etanercept improved markedly, and indeed appeared to perform better than tocilizumab.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 11. Treatment effects of biologic DMARDs separately (without Klareskog RCT)

	Relative treatment effect v. placebo			Origin
	ACR 20	**ACR 50**	**ACR 70**	
Tocilizumab	3.19	5.36	10.83	Based on 3 RCTs
Adalimumab	1.93	3.51	5.88	Based on 4 RCTs
Etanercept	5.32	9.50	17.50	Based on 3 RCTs
Infliximab	2.56	3.69	3.89	Based on 3 RCTs
TNF-α	2.44	3.99	5.22	Based on 10 RCTs
This is analogous to scenario 2 in the MTC				

Although the submission discussed the removal of the Klareskog trial[27] (along with several other unrepresentative trials), the results of these separate "scenarios" are presented in a confusing and inaccurate fashion. The submission actually suggested five different scenarios, details of which are summarised in Table 12 below. The ticks indicate included RCTs in each of the scenarios and the last column explains why particular RCTs have been included or excluded. Details of these studies are in Appendix 5.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 12. Table of scenarios in MTC

	Studies							
	Van de Putte	Moreland	Combe	Furst		Klareskog	Other (13)	Rationale
				no MTX/ no DMARD	With MTX/ DMARD			
Basecase			✓		✓	✓	✓	Conventional DMARD background treatment
1	✓	✓	✓	✓	✓	✓	✓	All 18 studies included
2			✓		✓		✓	Klareskog excluded due to a high response rate in the placebo arm
3					✓	✓	✓	Combe excluded because background treatment was sulfasalazine
4					✓		✓	Both Klareskog and Combe excluded

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

There are a number of inaccuracies in the description of the scenarios:

- The text below Figure 29 of the submission on p108 indicates that because treatment arms in four trials (Van de Putte, Moreland, a subgroup of Furst and Combe)[28-32] were fundamentally different from the remaining trials, they were not included in the initial analysis. This is incorrect - the Combe trial[32] was included in the base case analysis and the results reported on page 109 and in the Appendix (Tables A4 & A5 on p251-4). What they are actually describing on p108 is Scenario 3.

- As a consequence, in the second paragraph on page 108 they report using 10 trials of anti-TNF-α agents, when in fact they used 11. They do not list the Schiff RCT[33] but this was included in all of the scenarios. The Schiff RCT also does not seem to have been counted in the abatacept part of figure 29 on p108, possibly because it is a three arm RCT of abatacept vs infliximab vs placebo.[33] The Schiff RCT is also not in the submission reference list but is in the separate MTC document reference list.

- The submission states on page 111 that three scenario analyses were performed but actually four were done. Results from Scenario 4 are not reported at all.

- There seem to be some discrepancies between the numbers in the published reports and the numbers reported in the MTC and submission Table A3 for the Furst[31] and Keystone[34] RCTs. However, they possibly are not large enough to have much impact on the results.

Most importantly, although the MTC suggested that a number of trials may not be representative and provides results for re-analyses excluding these trials, none of these results appear in the subsequent economic evaluation, whether in the base case or the subsequent sensitivity analyses.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

5.3.2 Resources and costs

5.3.2.1 RA related inpatient costs

Inpatient costs were calculated using the Norfolk Arthritis Register (NOAR) database (see Section 7.2.9.1 of the submission on p165-9). In practice, however, there is very wide practice variation, with many patients rarely requiring inpatient care. Use of the NOAR data for determining use of inpatient facilities is no longer valid for contemporary practice. This has an effect (albeit small) on the overall cost-effectiveness of tocilizumab because inpatient costs are related to patients' HAQ scores. Because tocilizumab was found to be apparently more effective at improving HAQ scores compared to other biologic DMARDs, patients receiving tocilizumab as part of their treatment cycle may have reduced inpatient costs.

5.3.2.2 Treatment costs

A number of questionable assumptions were made about treatment costs in Table 46 of the submission on p167 and elsewhere:

1) The model assumes that all patients receive tocilizumab for a minimum of 6 months (see p 133 submission). This is generous as some will stop because of side effects and others because of obvious lack of efficacy within the first 3 months
2) The administration cost each infusion of tocilizumab was assumed to be £142. This originated from the first version of the BRAM model calculations[26] using 0.5 day day-case admission cost from the 2001 version of PSSRU Unit Costs of Health and Social Care (Personal Communication Pelham Barton, University of Birmingham, April 2009). That administration cost was calculated to be £124 and the same amount was also used in the 2004 version of the BRAM model.[12] This has then been inflated from 2004 to 2008 to get to £142 which, according to the submission, has since been used in a couple of STAs including the Abatacept appraisal (TA141).[35] In current clinical practice an infusion of tocilizumab is likely to necessitate a substantially more expensive full day case admission rather than an outpatient visit as described in the submission or a half day as in the BRAM model. This

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

is because patients receiving tocilizumab need to be admitted as day cases for the infusion and monitoring for adverse effects. This also applies to infliximab and rituximab. More appropriate codes therefore are: H25 Inflammatory Spine, Joint or Connective Tissue Disorders > 69: costing £1157 and H26 Inflammatory Spine, Joint or Connective Tissue Disorders < 70: costing £642 (Personal Communication, Tejinder Malhi, University Hospital Birmingham NHS Trust, March 2009). Technically speaking these are the income the Trust gains for conducting these procedures rather than the cost, as opposed to the PSSRU which does calculate costs. However, the main point is that £142 is probably too low because it is based on a half day only and should have been inflated from 2001 rather than 2004.

3) Costs for administration of ciclosporin, gold and leflunomide appear high. The assumptions for monitoring shown in Appendix 7 of the submission do not all tally with the UK National Guidelines for the monitoring of RA drugs.[36] For example, for ciclosporin, the submission suggests 13 outpatient visits and 4 GP visits in the first 6 months whereas the Guidelines recommends fortnightly monitoring for the first 3 months then monthly monitoring, which would be 9 visits in total. For gold salts, the submission suggests weekly visits for the first 6 months whereas the guidelines recommend weekly injections until significant response. The BNF suggests that benefit should be expected after giving 300-500mg which would be 6-10 weeks thereafter. The dose of leflunomide given in table 32 of the submission on p132 is 15.2mg per day which is odd as the usual dose is 20mg per day. It is unclear if an average dose is being used.

4) The average time that a patient is on treatment with gold salts is given in Appendix 7 of the submission (p233) as 0.71 years, based on a personal communication. In practice, toxicity from gold may arise quite early and many will discontinue well before 6 months when the greatest NHS costs are incurred. Using an average of 0.71 for the whole population means that all treated patients were subject to the highest costs.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

5) ESR & CRP may not be done at each visit. The quoted costs are expensive. They have cited our previous report[12] which quoted £11.15 per test (cost from local Trust finance Department) then inflated them to 2008. However, in a later version[11] the cost for an ESR test was quoted as £3.07 (cost from a National Pathology Alliance Benchmarking Report).

6) A relatively large proportion of patients on infliximab experience dose escalation, i.e. maintenance infusions every 6-8 weeks (3mg/kg or 5mg/kg). Also some patients on adalimumab increase to weekly doses.

5.3.2.3 Adverse events

Roche did not take into account the quality of life and cost implications of treatment-associated adverse events (see section 7.2.7.4 of the submission on p160):

> From p160
>
> Adverse events observed in the trials were not included in the economic evaluation as the ones that were associated with tocilizumab treatment are assumed to generate an insignificant burden in the quality of life of the patients. In addition the treatment of the adverse events observed is unlikely to utilise a significant amount of medical resources or costs to the NHS. Adverse events have not previously been considered in NICE technology appraisals of RA for these reasons. However additive monitoring requirements for safety reasons have been included.

As described in Section 3.2 of this report, the EMEA assessment report for tocilizumab[7] indicated a number of adverse events. The report lists upper respiratory tract, skin and gastro-intestinal infections and other problems (diarrhoea, mouth ulceration, vomiting, abdominal pain, gingival pain, oral pain and flatulence), skin disorders (rash, dermatitis and pruritus, skin ulcers), headaches, dizziness, conjunctivitis and oedema as adverse events. Also approximately 6.5% of patients have infusion reactions, mostly hypertension, rashes and pruritis but rarely anaphylaxis.[7]

When queried, Roche justified this assumption based on the following arguments:

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

1) The rate of AEs in the tocilizumab arm of the clinical trials was not significantly different from the rate in other arms (where patients received other DMARDs).

2) The overall survival of patients in the tocilizumab and non-tocilizumab arms of the clinical trials was not significantly different.

However, the comparison with other DMARDs is not directly relevant because patients in the tocilizumab and non-tocilizumab arms spent a similar amount of time on other DMARDs. Instead, the AE rate should be compared to that of palliative care. An overall comparison of patient survival also does not give the complete picture since tocilizumab may be associated with AEs that do not reduce survival but which may reduce their health-related quality of life. Roche argued that it was difficult to make a direct comparison with palliative care since "what happens to patients in the palliative care stage is largely unknown". However, lack of data is not a reason for making the assumption of no effect, particularly since biologic DMARDs are well known to be associated with AEs.

5.3.3 Sensitivity analyses

5.3.3.1 Use of scenario sensitivity analyses

The Roche model presented a series of "scenario analyses" (Section 7.3.3.1 on p179-85) in which single changes were made to assumptions in the model in order to explore their effects on the ICER. They found that under none of these scenarios did the resulting ICER exceed £30,000 per QALY gained. However, the use of "scenario analyses" avoids having to consider changes to several assumptions within a probabilistic framework. The incorporation of these scenarios within the probabilistic sensitivity analysis would increase overall uncertainty and hence cause a larger proportion of scenarios to exceed the £30,000 per QALY threshold.

In response, Roche argued that each of these scenarios was a "structural change in the model and cannot be inserted as a variable in the PSA". This is a statement of doubtful validity because many of these scenarios can be explored using parameters. For instance, the two scenarios about the slope of

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

the HAQ equation after the last follow up can be converted into an overall slope parameter. Even scenarios which have no parameters can be included in the probabilistic sensitivity analysis by placing suitable probability distributions on their likelihood of occurring.

5.4 Comment on validity of results presented with reference to methodology used

In order to assess the impact of the most important shortcomings and debatable points identified in the Roche model, we have re-estimated the base-case incremental cost-effectiveness ratio (ICER) using the model supplied to us by Roche, but altering key assumptions as described below:

1) Use of unpooled parameters. We have decreased the withdrawal rate of etanercept from 10% (based on a pooled estimate for both etanercept and infliximab) to 8% (based on the estimate for etanercept alone).

2) Rebound effect following withdrawal from treatment. We have modified the rebound effect to incorporate a BRAM model-like rebound. This means that when treatment is withdrawn, a patient's HAQ score worsens to what it would be in the absence of treatment (rather than simply erasing the initial improvement).

3) Adverse events. We have incorporated a quality adjusted life year (QALY) loss of 0.05 for every cycle a patient is on any DMARD treatment (except for palliative care) as a back-of-the-envelope representation of adverse events.

Other assumptions that are likely to have a large impact on the ICER including HAQ score progression on long-term treatment and the relationship between HAQ and EQ-5D scores. We did not estimate the impact on the ICER of changing these assumptions because such changes required complex programming which we were unable to implement within the time frame available to evaluate the submission.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

The table below shows the change in the incremental direct medical costs, QALYs and ICER when each of these alterations to the model is made. The change when all three alterations are made is also presented. Outcomes are given for the model of tocilizumab administered to DMARD-IR patients; when tocilizumab is administered to TNF-IR patients, the resulting ICER is predicted to be higher.

Table 13. Effects of changed parameters on ICER

Scenario (DMARD-IR indication)	Incremental direct medical costs (£)	Incremental QALYs	ICER (£)
Roche model	23,253	1.17	19,870
Unpooled parameters	22,887	1.1349	20,166
BRAM model-like rebound	23,253	0.9588	24,252
Adverse events	23,253	0.9441	24,629
All three changes simultaneously	22,910	0.7069	32,410

While the ICER remains below £30,000 per QALY gained when the three changes are made separately, simultaneously altering the three assumptions causes the ICER to exceed £30,000 per QALY.

Although we have also highlighted issues with the MTC in our report, we found that improving the treatment effect of tocilizumab in terms of ACR response, or conversely worsening the effect of other drugs in the RA treatment sequence, did not actually increase the ICER. This was because, in the Roche model, even non-responders who commenced treatment obtained a substantial improvement to their HAQ scores (as summarised in Figure 35 of the submission, on p143). However, the MTC is important in terms of the overall picture it paints of the relative effectiveness of different treatments. The current MTC seems to suggest that patients on tocilizumab demonstrate a far superior ACR response than patients on other anti-TNF-α treatments, whereas are re-analysis of the data suggests that this may not be the case.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

5.5 Summary of uncertainties and issues

- The Roche model suggested that adding tocilizumab to the treatment schedule for indicated patients was likely to be cost-effective, with an incremental cost-effectiveness ratio of about £20,000 per QALY gained for DMARD-IR patients and about £22,000 per QALY gained for TNF-IR patients. Sensitivity analyses suggested that all the scenarios explored and all the samples obtained during Monte Carlo sampling had ICERs below £30,000/QALY. However, there are shortcomings in the way the model handled several issues relating to treatment effectiveness of biologic and conventional DMARDs, adverse events, resources and costs, as well as sensitivity analysis.

- Of the shortcomings, issues around the treatment of effectiveness (particularly in terms of the rebound effect following withdrawal of treatment, HAQ score progression on long-term treatment and the relationship between HAQ and EQ-5D scores) are likely to be highly influential.

- Making changes to three model assumptions (using an etanercept-specific withdrawal rate, using a rebound effect following withdrawal of treatment similar to that used in a previous model, and including quality of life detriment due to adverse events) caused the DMARD-IR ICER to exceed £30,000 per QALY gained.

- CIC Since etanercept has been costed in the submission to be the same price as tocilizumab 8mg/kg, the decision as to which one to use will be based on differential effectiveness and factors such as ease of use and side effects, rather than cost. The reanalysis of the MTC found that etanercept had a lower relative effectiveness than tocilizumab (1.65 vs 3.19). However if the Klareskog RCT is removed then etanercept had a higher relative effectiveness (5.32 vs 3.19). So a key question is whether the Klareskog RCT should have been included or excluded from the MTC. It is the only RCT in the MTC to specifically mention that it included patients who were likely to benefit from MTX treatment, rather than having failed MTX treatment. The NICE scope is

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

not specific on this point but the submission is quite clear in section 2 that the investigation is limited to “adults with moderate or severe active rheumatoid arthritis (RA) who have either responded inadequately to or, or who were intolerant to, previous therapy with one or more disease modifying anti-rheumatic drugs (DMARDS)”. Therefore according to their definition of the decision problem the Klareskog RCT[27] should not have been included in the MTC. Also, the relative effectiveness of the TNF-α antagonists was not similar so they should not have been combined.

- Etanercept is delivered by subcutaneous injection whereas tocilizumab is given by iv injection so etanercept is easier to administer. Both may cause adverse effects but without head to head RCT information it would be very difficult to say whether one is worse than the other. However, including tocilizumab additively to the existing treatment schedule is likely to increase patients’ overall risk of adverse events.

- CIC Adalimumab was costed in the submission to be more expensive than tocilizumab (£9,857 vs £9,295) and was found in the MTC to have less relative effectiveness. Infliximab was costed in the submission to be slightly less expensive than tocilizumab (£8,812 vs £9,295) and was found in the MTC to have less relative effectiveness. Rituximab was costed in the submission to be less expensive than tocilizumab (£4,980 vs £9,295) and was found in the MTC to have less relative effectiveness.

6 Additional work undertaken by the ERG

Additional work done by the ERG includes the following:

- Considerable sifting through trial reports to find information that should have been made explicit in the submission
- Checking of search strategies and searches to ensure no studies were missed
- Redrawing of meta-analyses

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

- Critical appraisal of a key paper in the MTC to establish whether it should have been omitted or not.
- Rerunning MTC to establish relative effectiveness of TNF-α DMARDs
- Rerunning economic model to vary some of the assumptions made in the existing model

7 Discussion

7.1 Summary of clinical effectiveness issues

The submission provides evidence from four RCTs of tocilizumab versus placebo, one RCT of tocilizumab versus MTX and two longer term single arm extension studies following patients up to 3 or 5 years. There were no head to head comparisons of tocilizumab versus other biologic DMARDs. Evidence from a mixed treatment comparison (MTC) gave information on the relative effectiveness of tocilizumab compared to other biologic DMARDs.

Results from the RCTs suggested that tocilizumab was more effect than placebo and more effective than MTX for ACR20, ACR50 and ACR70 and other outcome measures. Results from the MTC base case suggested that tocilizumab had higher relative effectiveness than rituximab and than combined TNF-α inhibitors.

The main issues include:

- not having RCTs directly comparing biologic DMARDs
- not having long-term RCT follow up results
- not using directly measured EQ-5D results where these were available
- inappropriately combining TNF-α inhibitors
- inappropriately including the Klareskog RCT[27] in the MTC

The effects of some of these issues are further explored in Table 14.

7.2 Summary of cost effectiveness issues

The submission included a de novo individual sampling model with a hypothetical cohort of 10,000 patients with moderate to severe RA and with an inadequate response to traditional DMARDs or in addition to one or more

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

TNF-α inhibitors. The intervention in the model was tocilizumab 8mg/kg added to a sequence of biologic and conventional DMARDs compared to the same sequence without tocilizumab. The outcome used were ACR scores from the MTC used to obtain to HAQ scores and then to EQ-5D using a quadratic equation derived from work in a conference abstract. The costs were largely driven by the high treatment costs for biologic DMARDs. The perspective was NHS, a lifetime horizon was used and the discount rate was 3.5% per year for costs and utilities. Scenario and some probabilistic sensitivity analyses were undertaken. The model was constructed in Visual Basic for Applications within MS Excel.

The model suggested that adding tocilizumab to the treatment sequence was likely to be cost-effective, with an ICER of about £20,000 per QALY gained for DMARD-IR patients and about £22,000 per QALY gained for TNF-IR patients. All of their sensitivity analyses gave ICERs below £30,000/QALY.

There were shortcomings in the way the model handled several issues relating to the treatment effectiveness, adverse events, resources and costs, as well as sensitivity analysis. Some of these we have explored by modifying the model to estimate ICERs of the intervention under different assumptions. However, the model was too complicated to readily change other parameters. A list of issues, changes that could be made and estimates of the impact of others are shown in Table 14. These are (mostly) in addition to the scenario analyses given on page 180 (DMARD-IR) and p183 (TNF-IR) of the submission.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Table 14. Summary of cost effectiveness issues

General issue	Details for this submission	Effect on ICER (£)
(Roche base case DMARD-IR)		19,870
Measurement of effectiveness	Use of MTX unpooled parameters of TNF-α inhibitors	20,166
	The treatment sequence for the tocilizumab arm always has one more DMARD in it	Makes tocilizumab seem more effective
	ACR results directly from RCTs rather than from MTC	Unknown
	Allowing for long term tocilizumab effects from single arm studies where placebo and tocilizumab arm RCT patients all now given tocilizumab	Unknown
	Use of EQ-5D results from Option RCT	See Appendix 3
Conversion to utility	Use of an exponential fit to HAQ progression graph	Large impact increase ICER
	Removal of negative utility scores (submission scenario 1 DMARD-IR)	20,214
Model structure	Conversion to BRAM model-like rebound	24,252
Adverse events	Incorporating comparison of adverse events tocilizumab to palliative care	24,629
Tocilizumab costs	Use of a 400mg and a 200mg vial (likely in practice) costing £9,984 rather than current £9,295 if a 400mg and two 80mg vials were used	Increase ICER
Other costs	Increased administration costs	Increase ICERs compared to etanercept, adalimumab
	Removal of high estimates of monitoring costs for gold, ciclosporin etc	Likely little effect
Combination	Using MTX unpooled parameters of TNF-α inhibitors plus BRAM model-like rebound plus Adverse events compared to palliative care	32,410

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

7.3 Implications for research

The most urgent research required is a series of RCTs directly comparing the biologic DMARDs with each other in order to determine which is the most clinically effective. Follow up should be sufficiently long to establish whether effectiveness wanes after 6 months on treatment. These RCTs should measure quality of life using Eq-5D and costs so cost effectiveness can also be established.

There needs to be a systematic review of the correlation between HAQ and EQ-5D scores to determine the best fit between the two and to explain the heterogeneity found between studies.

There also needs to be much more information available on the long term safety of tocilizumab.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Appendix 1. Searches undertaken by the ERG

Clinical effectiveness:

MEDLINE (Ovid) 1950 to January Week 2 2009
1 (tocilizumab or actemra or atlizumab).mp.
2 Arthritis, Rheumatoid/
3 rheumatoid arthritis.tw.
4 2 or 3
5 1 and 4

EMBASE (Ovid) 1988 to 2009 Week 04
1 (tocilizumab or actemra or atlizumab).mp.
2 Rheumatoid Arthritis/
4 2 or 3
5 1 and 4

Ovid MEDLINE(Ovid) In-Process & Other Non-Indexed Citations January 28, 2009
1 (tocilizumab or actemra or atlizumab).mp.

Cochrane Library 2009 Issue 1
#1actemra OR tocilizumab OR atlizumab
#2rheumatoid next arthritis
#3MeSH descriptor Arthritis, Rheumatoid, this term only
#4(#3 OR #2)
#5(#1 AND #4)

On-going studies

Sources: ClinicalTrials.gov; Current Controlled Trials *meta*Register and NIHR UK Clinical Research Network Database.
Search terms: Tocilizumab, Actemra, Altizumab

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Appendix 2. EQ-5D results from Option RCT (WA17822)

Below are the Euroqol EQ-5D results from the Option RCT trial report (p778,780). They indicate that tocilizumab 8mg/kg is more effective than placebo at all time points after baseline.

Table 15. EQ-5D results from Option RCT (WA17822) (VAS scale ITT by visit)

	TCZ (8mg/kg) + MTX		*Placebo + MTX*	
	Number	*Mean (SD)*	*Number*	*Mean (SD)*
Baseline	202	49.0 (21.17)	202	48.7 (18.95)
Week 8	195	63.8 (19.01)	198	54.1 (20.64)
Week 16	194	67.4 (19.20)	190	54.5 (22.52)
Week 24	172	71.1 (19.9)	123	59.9 (22.76)

Table 16. EQ-5D results from Option RCT (WA17822) (Single index utility score ITT by visit)

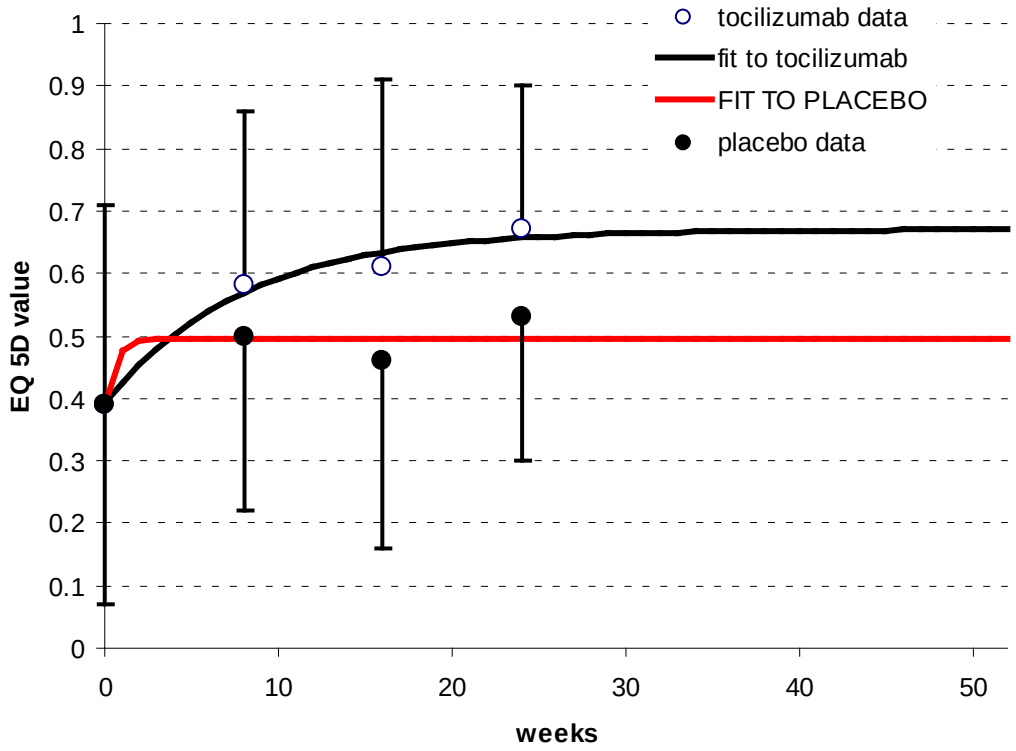
	TCZ (8mg/kg) + MTX		*Placebo + MTX*	
	Number	*Mean (SD)*	*Number*	*Mean (SD)*
Baseline	197	0.39 (0.32)	197	0.39 (0.32)
Week 8	188	0.58 (0.28)	189	0.50 (0.31)
Week 16	186	0.61 (0.30)	181	0.46 (0.34)
Week 24	168	0.67 (0.23)	122	0.53 (0.32)

In the submission, the main biologic DMARD drugs in the economic model were rituximab and etanercept, although the model cycles through a range of DMARDs with or without tocilizumab. Assuming that the main clinical decisions will be whether to use tocilizumab rather than rituximab or whether to use tocilizumab rather than etanercept, the following rough calculations may be useful.

The EQ-5D results were plotted and curves fitted and extrapolated to one year using the equation $Y = a + b + (1- (\exp-\lambda t))$. The resulting graph is shown below (see Figure 6). The error bars are standard deviation so it can be seen that will be a wide margin of error around any resulting estimate.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Figure 6. Graph of EQ-5D results from Option RCT (WA17822) (Single index utility score ITT by visit)



CIC The resulting difference in effectiveness between tocilizumab 8mg/kg and placebo was 0.1342. The difference in average yearly cost between tocilizumab 8mg/kg and rituximab according to the submission is £9,295-£4,890 = £ 4,405. Since rituximab is more effective than placebo, the difference in effectiveness will be less than 0.1342 so the ICER of tocilizumab compared to rituximab will be more than £32,824. The difference in average yearly cost between tocilizumab 8mg/kg and infliximab according to the submission is £9,295-£8,812 = £ 483. Since infliximab is more effective than placebo, the difference in effectiveness will be less than 0.1342 so the ICER of tocilizumab compared to infliximab will be more than £3,599.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Appendix 3. Quality Assessment using ScHARR-TAG economic modelling checklist

Factor	Appraisal
Title	RoActemra (tocilizumab) NICE STA submission (with clarifications document)
A statement of the problem	What is the incremental cost-effectiveness of adding tocilizumab (alone or in combination with MTX) to the treatment sequence of adults with moderate to severe RA with either a DMARD-IR or TNF-IR indication?
A discussion of the need for modelling	Modelling is required for the following reasons: • To extrapolate long-term changes in outcomes beyond the follow-up period of the clinical trials. • To obtain comparable outcomes of the effect of different RA therapies, in terms of both disease-specific (ACR response rate, HAQ scores) and generic quality of life (QALYs) measures. • To test the robustness of conclusions to changes in key parameters and assumptions. **Assessment by ERG:** The decision to use modelling was appropriate given the data constraints.
A description of the relevant factors and outcomes	Relevant factors and outcomes are the following: • Initial response to treatment (measured in terms of ACR 20, 50 and 70 response rates). • Long-term progression while on treatment (measured in terms of HAQ scores. • Withdrawal from treatment (measured in terms of six monthly rates). • Generic health-related quality of life while on treatment (measured in terms of EQ-5D scores). **Assessment by ERG:** The factors and outcomes used in the model appear to be defensible.
A description of model including: type of model; time	Individual sampling model with a hypothetical cohort of 10,000 patients. Patient lifetime time horizon. NHS and Personal Social Services (effectively NHS) perspective.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

frame; perspective; and setting	UK secondary care setting. **Assessment by ERG:** The general modelling framework chosen was appropriate to the decision problem and consistent with the NICE reference case.
A description of data sources, with description of respective strengths and weaknesses	Data sources used to model the effectiveness of different forms of treatment were the following: • Primary outcome data obtained from four clinical trials conducted by the manufacturer (ACR response rates, HAQ score progression, EQ-5D scores). These data are described in the manufacturer's submissions to NICE and to the EMEA. • Published data about the effectiveness of other RA treatments in the literature (ACR response rates). • Conclusions about treatment effectiveness from critical appraisals and other economic models in the literature (probability of treatment withdrawal, HAQ score progression). Costs of care were obtained largely from standard databases of health care costs. The strength of this approach is that it allowed the totality of evidence from several studies and appraisals to be combined in a single model. However, there were a number of shortcomings in the data which necessitated further modelling assumptions. These included the following: • Treatment effectiveness from different studies needed to be compared using an MTC. • There were no primary data about patients' HAQ scores between follow-up times and beyond the limit of the long term extension study (about 180 weeks). Similarly, there are no data about patient rebound when treatment was withdrawn. Hence model assumptions had to be made about both aspects of long-term HAQ progression. • There were no primary data about the effect of treatment (other than tocilizumab) on health-related quality of life.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

	Hence HAQ scores had to be mapped on to EQ-5D scores using data from the tocilizumab to characterise the mapping function. **Assessment by ERG:** There were a number of shortcomings in the way the data were used in the Roche model: • The MTC was based on a number of debatable assumptions about which studies to include. • The cost of administration for tocilizumab was underestimated. • Effectiveness parameters for different anti-TNF-α agents were combined, even though etanercept is the most commonly used and also appears to be the most effective anti-TNF-α treatment option. • The regression model used to interpolate HAQ scores is extremely sensitive to assumptions about the interpolating function, but this was not varied in sensitivity analysis. • The link between HAQ scores and EQ-5D, although necessary, required a number of unfounded assumptions.
Key assumptions relating to model structure and data stated	7) At the start of a specific treatment, a patient may have had no clinical response or achieved an ACR 20, ACR 50 or ACR 70 response. 8) Each type of response was associated with an initial drop (i.e. improvement) in HAQ score. 9) In each six monthly cycle, a patient's HAQ score was assumed to change (improve, worsen or undergo no change) depending on the kind of treatment the patient received. 10) For every six monthly cycle, patients were assumed to have a probability of being withdrawn from treatment. 11) When a patient was withdrawn, a "rebound" was assumed, which caused an increase in HAQ equivalent to the initial

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

	drop in HAQ score reported in step 2. 12) If a patient was withdrawn, the patient was moved to the next treatment in the cycle until he or she reached the palliative care stage. **Assessment by ERG:** The broad structural assumptions used in modelling were appropriate and similar to previous models used to evaluate the cost-effectiveness of biologic DMARDs. However, the parameterisation of the model has shortcomings which are discussed in other parts of this table.
Disease specific factors included within modelling (Items to be specified in conjunction with expert clinical input)	• Initial ACR response to treatment. • Long-term progression while on treatment in terms of HAQ scores. • Withdrawal from treatment. • Rebound in HAQ scores after treatment withdrawal. **Assessment by ERG:** The shortcomings in the way these factors were represented include the following: • Adverse events were not included. • The rebound in HAQ scores was optimistic and far smaller than that used in a previous cost-effectiveness model constructed for NICE (the BRAM model).
Validation	An independent reviewer verified the input sources, programming and face validity of the model results. **Assessment by ERG:** Details of the validation procedures and the identity of the reviewer were not provided. It has been assumed that the reviewer was in-house and hence independent in terms of the process of model construction rather than in terms of not having any conflicts of interest.
Results	The Roche model suggested that adding tocilizumab to the treatment schedule for indicated patients was likely to be cost-effective, with an incremental cost-effectiveness ratio of about £20,000 per QALY gained for DMARD-IR patients and about £22,000 per QALY gained for TNF-IR patients.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

	Assessment by ERG: Making changes to three model assumptions (using an etanercept-specific withdrawal rate, using a rebound effect following withdrawal of treatment similar to that used in a previous model, and including quality of life detriment due to adverse events) caused the DMARD-IR ICER to exceed £30,000 per QALY gained.
Sensitivity analysis results	In sensitivity analyses, the incremental cost-effectiveness ratio of adding tocilizumab to the treatment schedule for indicated patients ranged from £16,000 - £25,000 for DMARD-IR patients and £19,000 - £27,000 for TNF-IR patients. A number of "scenarios" were explored in which some model assumptions were altered; **Assessment by ERG:** The use of "scenario analyses" alongside conventional probabilistic analyses avoids having to consider changes to several assumptions within a probabilistic framework. Hence the results of the sensitivity analysis should be regarded as conservative.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

CRITICAL APPRAISAL SKILLS PROGRAMME
making sense of evidence about clinical effectiveness

Critical appraisal of the TEMPO trial. (Klareskog L, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double blind randomised controlled trial. The Lancet 2004;363:675-81)

General comments

- Three broad issues need to be considered when appraising a trial.
  - *Are the results of the trial valid?*
  - *What are the results?*
  - *Will the results help locally?*
- The 11 questions are adapted from: Guyatt GH, Sackett DL, Cook DJ, Users' guides to the medical literature. II. How to use an article about therapy or prevention. JAMA 1993; 270: 2598-2601 and 271: 59-63

These materials were developed by the CASP* team in Oxford.

* *CASP (Critical Appraisal Skills Programme)* helps health service decision-makers develop skills in appraising evidence about clinical effectiveness. It works with local programmes for evidence-based health care. Its core funding comes from Anglia and Oxford regional office.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

A/ Are the results of the trial valid?

Screening Questions

1	**Did the trial address a clearly focused issue?** *An issue can be 'focused' in terms of* *- the population studied* *- the intervention given* *- the comparator given* *- the outcomes considered*	**Yes** X	**Can't tell** O	**No** O
		Population – patients aged 18 years or older with disease duration of 6 months to 20 years with active RA with ACR functional class I–III and less than satisfactory response to at least one DMARD other than MTX. They could have been treated previously with MTX but not if they had lack of response or clinically important side effects. Excluded were patients treated with MTX in the previous 6 months, previous treatment with etanercept or other TNF-α antagonists, immunosupressives within 6 months of screening, biologic agents within 3 months of screening, any other DMARD, corticosteroid within 4 weeks of baseline visit, active infections or other comorbidities. ***Intervention*** 3 arms: 1. etanercept only (25 mg twice a week subcutaneously and oral placebo once a week), 2. MTX only (7·5 mg escalated to 20 mg oral capsules weekly within 8 weeks if patients had any painful or swollen joints, and placebo subcutaneous injections twice a week), 3. etanercept plus MTX (combination of 25 mg subcutaneous etanercept injections twice a week and MTX capsules once a week). ***Outcomes*** to week 52 – ACR response (ACR-N) area under the curve (AUC) over the first 24 weeks. The primary radiographic endpoint was change from baseline in total joint damage and was assessed with the modified Sharp score at 52 weeks.		
2	**Was the assignment of patients to treatments randomized?**	**Yes** x	**Can't tell** o	**No** o
		686 patients were randomly assigned; centralised telephone randomization was used.		
3	**Were all of the patients who entered the trial properly accounted for at its conclusion** *- was follow up complete?* *- were patients analysed in the groups to which they were randomised?*	**Yes** o	**Can't tell** o	**No** x
		A study flow chart is given. Analysis was by intention to treat. 38 patients discontinued in the combination arm, 53 in the etanercept only arm and 69 in the MTX arm.		

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Detailed Questions

4	**Were patients, health workers and study personnel 'blind' to treatment?** *- were the patients* *- were the health workers* *- were the study personnel*	**Yes** O	**Can't tell** X	**No** O It was described as a double-blind RCT with identical-appearing injectable and oral treatments. Although Sharp score was determined blind, there was no description of how or whether the other outcomes were measured by blinded outcome assessors.
5	**Were the groups similar at the start of the trial?** *In terms of other factors that might effect the outcome such as age, sex, social class*	**Yes** O	**Can't tell** O	**No** X Slightly fewer in the etanercept only arm had corticosteroid use at baseline and this group also had a higher DAS, estimated yearly progression in total Sharp score and mean CRP.
6	**Aside from the experimental intervention, were the groups treated equally?**	**Yes** X	**Can't tell** O	**No** O MTX was the comparator. Whilst etanercept was given in stable doses, a dose-escalation scheme was used for MTX. This, according to the authors, was to assure that therapeutic doses of the drug were used in the study. As the MTX treatment was blinded, it can be assumed that patients receiving placebo MTX also had this dose escalation although this is not mentioned specifically. Otherwise, all three groups appeared to be treated equally.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

B/ What are the results?

7 How large was the treatment effect? *What outcomes are measured?* **8 How precise was the estimate of the treatment effect?** *What are its confidence limits?*	1. ACR20 at week 52 combination 85% (95% CI 80–89) etanercept only 75% (69–80) MTX 76% (70–81) 2. ACR50 at week 52 combination 69% (95% CI 63–75) etanercept only 48% (42–55) MTX 43% (36–49) 3. ACR70 at week 52 Combination 43% (95% CI 36–50) Etanercept only 24% (19–30) MTX 19% (14–25) 4. ACR-N AUC at 24 weeks combination 18·3%-years [95% CI 17·1–19·6] etanercept only 14·7%-years [13·5–16·0] MTX 12·2%-years [11·0–13·4] 5. any infections combination 67% etanercept only 59% MTX 64%

C/ Will the results help locally?

9 Can the results be applied to the local population? *Do you think that the patients covered by the trial are similar enough to your population?*	**Yes** O	**Can't tell** X	**No** O
	Issues are as follows: 1. The dose schedule used in the trial was slightly unusual in that it used an aggressive treatment schedule. The authors suggest that the aggressive dose escalation scheme was used to ensure that therapeutic doses were used. 2. The enrolment of patients included the requirement that they were appropriate candidates for MTX treatment, rather than had already failed treatment, meaning that the trial had a fair comparison between etanercept and MTX.		
10 Were all clinically important outcomes considered? *If not, does this affect the decision?*	**Yes** O	**Can't tell** O	**No** X
	No measurement of EQ-5D No measurement of costs		
11 Are the benefits worth the harms and costs? *This is unlikely to be addressed by the trial. But what do you think?*	**Yes** O	**Can't tell** X	**No** O

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

Appendix 5. Details of MTC RCTs
Table 17. Details of MTC RCTs

Source	N	Patients	Interventions	Treatment effect response (%)						Background treatment	Trial duration
				ACR 20		ACR 50		ACR 70			
				Active	Placebo	Active	Placebo	Active	Placebo		
Adalimumab											
van de Putte et al (2004)	544	Patients with active RA despite DMARDs	Adalimumab 40mg every other week sc. injection vs. Adalimumab 40mg weekly sc. injection vs. Adalimumab 20mg every other week sc. injection vs. Adalimumab 20mg weekly sc. injection vs. Placebo	46%	19%	22%	8%	12%	2%	No MTX/ No DMARD	26 weeks
Furst et al, 2003 (STAR)	636	Patients with active RA despite DMARDs	Adalimumab 40mg every other week sc. injection vs. Placebo	53%	35%	29%	11%	15%	3%	With and without DMARD/MTX	24 weeks
Etanercept											
Klareskog et al, 2004 (TEMPO-I)	686	Patients with active RA despite previous DMARDs	Etanercept 25mg 2/w sc injection plus MTX vs. Etanercept 25mg 2/w sc. injection vs.MTX	82%	72%	58%	40%	35%	15%	MTX	52 weeks (26 weeks interim)
Combe et al, 2006	254	Patients with active RA despite sulfasalazine	Etanercept 25mg sc. 2/w injection + sulfasalazine 2-3 mg/day. vs Etanercept 25mg sc. 2/w injection. vs Sulfasalazine 2-3mg/day	74%	28%	52%	14%	25%	2%	No MTX/ No DMARD (sulfasalazine)	24 weeks
Moreland et al, 1999	234	Patients with active RA despite DMARDs	Etanercept 25mg 2/w sc. injection vs. Etanercept 10mg 2/w sc. injection. vs Placebo	59%	11%	40%	5%	15%	1%	No MTX/ No DMARD	26 weeks

Interventions in grey indicate dosages not included

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

8 References

1 Hyrich KL, Lunt M, Watson KD, Symmons DP, Silman AJ. Outcomes after switching from one anti-tumor necrosis factor agent to a second anti-tumor necrosis factor agent in patients with rheumatoid arthritis: Results from a large UK national cohort study. *Arthritis & Rheumatism* 2007; **56**(1):13-20.

2 Longson,C. Notification of the decision to split the final appraisal determination. URL: http://www.nice.org.uk/guidance/index.jsp?action=download&o=38691 accessed 16th March 2009

3 NICE. Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis. London: National Institute for Health and Clinical Excellence; 2007. Report No. TA130.

4 National Collaborating Centre for Chronic Conditions. Rheumatoid arthritis: national clinical guideline for management and treatment in adults. London: Royal College of Physicians; 2009.

5 Anon. European public assessment report (EPAR) Roactemra EPAR Summary for the public. London: European Medicines Agency; 2009. Report No. EMEA/H/C/955.

6 Anon. Annex 1. Summary of product characteristics (RoActemra). London: European Medicines Agency; 2009.

7 Anon. Assessment report for RoActemra . London: European Medicines Agency; 2009. Report No. EMEA/26276/2009.

8 Anon. British national formulary. 57 ed. London: BMJ Group and RPS Publishing; 2009.

9 Drosos AA. Methotrexate Intolerance in Elderly Patients with Rheumatoid Arthritis. What Are the Alternatives? *Drugs and Aging* 2003; **20**(10):723-736.

10 Grover MLHAB, Hay EM, Shadforth MF. Adverse reactions to disease modufying anti-rheumatic drugs in clinical practice. *Q Journal of Medicine* 2001; **91**:309-319.

11 Chen Y-F, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, *et al*. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. *Health Technology Assessment* 2006; **10**(42).

12 Barton P, Jobanputra P, Wilson J, Bryan S, Burls A. The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. *Health Technology Assessment* 2004; **8**(11).

13 Bruce B, Fries JF. The Health Assessment Questionnaire (HAQ). *Clinical and Experimental Rheumatology* 2009; **23**(supp 39):S14-S18.

14 Sany J, Bourgeois P, Saraux A, Durieux S, Lafuma A, Daures JP, *et al*. Characteristics of patients with rheumatoid arthritis in France: a study of 1109 patients managed by hospital based rheumatologists. *Annals of the Rheumatic Diseases* 2004; **63**:1235-1240.

15 Sokka T, Kautiainen H, Hannonen P, Pincus T. Changes in Health Assessment Questionnaire Disability Scores Over Five Years in Patients

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

with Rheumatoid Arthritis Compared With the General Population. *Arthritis & Rheumatism* 2006; **54**(10):3113-3118.
16 Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelheimer DA. Minimally important difference between patients with rheumatoid arthritis: the patient's perspective. *Journal of Rheumatology* 1993; **20**:557-560.
17 Wolfe F, Michaud K, Pincus T, Furst D, Keystone E. The disease activity score is not suitable as the sole criterion for initiation and evaluation of anti-tumor necrosis factor therapy in the clinic: discordance between assessment measures and limitations in questionnaire use for regulatory purposes. *Arthritis & Rheumatism* 2005; **52**(12):2873-2879.
18 Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. *The Lancet* 2009; **373**:659-672.
19 Koike R, Takeuchi T, Eguchi K, Miyasaka N. Update on the Japanese guidelines for the use of infliximab and etanercept in rheumatoid arthritis. *Modern Rheumatology* 2007; **17**:541-548.
20 Roche. Clinical trial research. http://qaroche-trials.centerwatch.com/patient/trials/trial110650.html accessed March 2009
21 Geborek P, Crnkic M, Petersson IF, Saxne T. Etanercept, infliximab and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in Southern Sweden. *Annals of Rheumatic Diseases* 2002; **61**:793-798.
22 Boggs R, Sengupta N, Ashraf T. Estimating health utility from a physical function assessment in rheumatoid arthritis (RA) patients treated with adalimumab. *International Society of Pharmacoeconomics and Outcomes Research* 2002; **abstract UT3**.
23 Scott DL, Khoshaba B, Choy EH, Kingsley GH. Limited correlation between the Health Assessment Questionnaire (HAQ) and EuroQol in rheumatoid arthritis: questionable validity of deriving quality adjusted life years from HAQ. *Annals of Rheumatic Diseases* 2007; **66**:1534-1537.
24 Ariza-Ariza R, Hernández-Cruz B, Carmona L, et a. Assessing Utility Values in Rheumatoid Arthritis: A Comparison Between Time Trade-Off and the EuroQol. *Arthritis & Rheumatism* 2006; **55**(5):751-756.
25 Kobelt G, Jönsson L, Lindgren P, Young A, Eberhardt K. Modeling the Progression of Rheumatoid Arthritis. *Arthritis & Rheumatism* 2002; **46**(9):2310-2319.
26 Jobanputra P, Barton P, Bryan S, Burls A. The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation. *Health Technology Assessment* 2002; **6**(21).
27 Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *Lancet* 2004; 363:675-681.
28 van der Putte LB, Atkins C, Malaise M, Sany J, Russell AS, van Riel PL, *et al*. Efficacy and safety of adalimumab as monotherapy in patients with rheumatoid arthritis for whom previous disease modifying antirheumatic drug treatment has failed. *Annals of Rheumatic Diseases* 2004; **63**(5):508-516.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.

29 van der Putte LB, Rau R, Breedveld FC, Kalden JR, Malaise MG, van Riel PL, *et al*. Efficacy and safety of the fully human anti-tumour necrosis factor alpha monoclonal antibody adalimumab (D2E7) in DMARD refractory patients with rheumatoid arthritis: a 12 week, phase II study. *Annals of Rheumatic Diseases* 2003; **62**(12):1168-1177.
30 Moreland LW, Schiff MH, Baumgartner SW, Tindall EA, Fleishmann RM, Bulpitt KJ, *et al*. Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial. *Annals of Internal Medicine* 1999; **16**(130(6)):478-486.
31 Furst D, Schiff MH, Fleishmann RM, Strand V, Birbara CA, Compagnone D, *et al*. Adalimumab, a fully human anti tumor necrosis factor-alpha monoclonal antibody, and concomitant standard antirheumatic therapy for the treatment of rheumatoid arthritis: results of STAR (Safety Trial of Adalimumab in Rheumatoid Arthritis). *Journal of Rheumatology* 2003; **30**(12):2563-2571.
32 Combe B, Codrenu C, Fiocco U, Gaubitz M, Geusens PP, Kvein TK, *et al*. Etanercept and sulfasalazine, alone and combined, in patients with active rheumatoid arthritis despite receiving sulfasalazine: a double-blind comparison. *Annals of Rheumatic Diseases* 2006; **65**(10):1357-1362.
33 Schiff M, Keiserm M, Codding C, Songcharoen S, Berman A, Nayiager S, *et al*. Efficacy and safety of abatacept or infliximab vs placebo in ATTEST: a phase III, multi-centre, randomised double blind, placebo-controlled study in patients with rheumatoid arthritis and an inadequate response to methotrexate. *Annals of Rheumatic Diseases* 2008; **67**:1096-1103.
34 Keystone EC, Kavenaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, *et al*. Radiographic, clinical and functional outcomes of treatment with adalimumab (a human anti-tumour necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomised placebo-controlled, 52 week trial. *Arthritis & Rheumatism* 2004; **50**(5):1400-1411.
35 NICE. Abatacept for the treatment of rheumatoid arthritis. London: National Institute for Health and Clinical Excellence; 2008. Report No. TA141.
36 Anon. National guidelines for the monitoring of second line drugs. London: British Society for Rheumatology; 2000.

Copyright 2009 Queen's printer and controller of HMSO. All rights reserved.