Golimumab for the treatment of rheumatoid arthritis after failure of previous disease-modifying antirheumatic drugs: A Single Technology Appraisal

Produced by School of Health and Related Research (ScHARR), The University of Sheffield

Authors Rachel Jackson, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Jonathan Tosh, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Sarah Davis, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Ruth Wong, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Matt Stevenson, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
John Stevens, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Correspondence to Rachel Jackson, Research Fellow, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Date completed 6th September 2010

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 08/118/01 STA

Declared competing interests of the authors

None

Acknowledgements

David L. Scott (Professor of Clinical Rheumatology, Department of Rheumatology, King's College London School of Medicine, Weston Education Centre, Cutcombe Road, London, SE5 9RS, UK) and Raashid Luqmani (Consultant Rheumatologist, Department of Rheumatology, Nuffield Orthopaedic Centre, Windmill Road, Oxford, OX3 7LD, UK) provided clinical advice and commented on draft materials during the project. Eva Kaltenthaler (Managing Director, ScHARR Technology Assessment Group), Sally Doss (NICE) and Zoe Garrett (NICE) commented on a draft version of the report. Abdullah Pandor (ScHARR) provided methodological advice. The authors wish to thank all of the above. We would also like to thank Andrea Shippam, Programme Administrator, ScHARR, for her help in preparing and formatting the report.

Declared competing interests of the clinical advisors

Dr Raashid Luqmani	Non-personal pecuniary interests: Pfizer provided support to travel to EULAR, advisory board to GSK (vasculitis), advisory board to Roche (vasculitis), Roche provided support to travel to ACR.
Professor David L. Scott	None declared

Rider on responsibility for report

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

This report should be referenced as follows:

Jackson R, Tosh J, Davis S, Wong R, Stevenson M, Stevens J. Golimumab for the treatment of rheumatoid arthritis after failure of previous disease-modifying antirheumatic drugs: A Single Technology Appraisal. ScHARR, The University of Sheffield, 2010.

Contributions of authors

Rachel Jackson acted as project lead and systematic reviewer on this assessment, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Jonathan Tosh acted as health economist on this assessment, critiqued the manufacturer's economic evaluation and contributed to the writing of the report. Sarah Davis critiqued the

manufacturer's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project. Ruth Wong critiqued the search methods in the submission and contributed to the writing of the report. Matt Stevenson contributed to the writing of the report and provided general guidance. John Stevens assisted in the critique of the mixed treatment comparison and indirect comparison work, and provided general guidance.

Table of contents

LIST OF TABLES AND FIGURES

FIGURES

List of Abbreviations

ACR	American College of Rheumatology
ADA	Adalimumab
AE	Adverse event(s)
AS	Ankylosing spondylitis
BSRBR	British Society for Rheumatology Biologics Registry
CI	Confidence interval
CIC	Commercial in confidence
CRD	Centre for Reviews and Dissemination
CTZ	Certolizumab (also referred to as certolizumab pegol)
DMARD	Disease-modifying antirheumatic drug
EQ-5D	EuroQol 5-dimensions
ERG	Evidence Review Group
ETN	Etanercept
GOL	Golimumab
HRQoL	Health related quality of life
IC	Indirect comparison
ICER	Incremental cost-effectiveness ratio
IFX	Infliximab
ITT	Intention to Treat
IV	Intravenous
MS	Manufacturer's submission
MTC	Mixed treatment comparison
MTX	Methotrexate
NAO	National Audit Office
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NSAID	Non-steroidal anti-inflammatory drug
QALY	Quality adjusted life year
PsA	Psoriatic arthritis
RA	Rheumatoid arthritis
RCT	Randomised controlled trial
RR	Relative risk
SAE	Serious adverse event(s)
SI	Serious infection(s)
TNF-α	Tumor necrosis factor-alpha

VAS	Visual analogue scale
Vs	Versus

1 SUMMARY

1.1 Scope of the submission

The manufacturer's submission provides evidence on the clinical and cost-effectiveness of golimumab within its licensed indication and in comparison with those interventions licensed and recommended by NICE at the time the manufacturer was invited to make its submission (April 2010). However several relevant comparators covered by two recently published NICE Technology Appraisals[1,2] have been excluded from the decision problem. The comparison against drugs for use after the failure of a TNF-α inhibitor has been restricted to rituximab when several other interventions (adalimumab, etanercept, infliximab, abatacept, tocilizumab) are now recommended by NICE for those patients for whom rituximab therapy is contraindicated or withdrawn because of an adverse event. Tocilizumab is also recommended for patients who have responded inadequately to rituximab.

The manufacturer's submission considers patients who have never received a TNF-α inhibitor (DMARD experienced population) separately from patients who have had prior therapy with a TNF-α inhibitor (TNF-α experienced population). Whilst it is reasonable to assess the clinical effectiveness and cost-effectiveness of golimumab in each of these positions within the treatment pathway, the submission does not go on to address the question of whether it is more cost-effective to use golimumab after there has been a failure to respond to DMARDs or after there has been a failure to respond to a TNF-α inhibitor. Neither does it assess whether golimumab is cost-effective compared to tocilizumab after failure to respond to rituximab. However, there are no trials for golimumab in this population on which to base this analysis.

The submission does not adequately address all of the outcomes specified within the scope. The clinical effectiveness review focuses on measures of disease activity and functional status (measured using the ACR criteria) which are then used to model the efficacy of interventions within the economic model. Mortality and adverse event data are also provided but efficacy data was lacking within the submission for pain, fatigue, progressive joint damage, extra-articular disease manifestations, and health related quality of life. The clinical advisors to the ERG noted the absence of data from the key golimumab trials on progressive joint damage, which has been available for assessments of other TNF-α inhibitors.[3]

1.2 Summary of submitted clinical effectiveness evidence

The manufacturer reports efficacy for the two populations (DMARD experienced and TNF-α inhibitor experienced). The comparators evaluated on clinical effectiveness by the manufacturer are adalimumab, certolizumab, etanercept and infliximab for the DMARD experienced population and rituximab for the TNF-α inhibitor experienced population.

The efficacy for interventions in the DMARD population was estimated using both a mixed treatment comparison and with meta-analyses. The efficacy for interventions in the TNF-α inhibitor experienced population was estimated using an indirect comparison methodology.

Efficacy in the DMARD experienced population: Using a random effects meta-analysis, the relative risk estimate for ACR20 response at 24 weeks was lower for golimumab than for the comparators although each intervention had wide 95% confidence intervals, and thus a definitive judgement on efficacy could not be made. These conclusions also applied when using a fixed effects meta-analysis, although the midpoint relative risk for golimumab may no longer be the lowest from all interventions. Evaluations of ACR50 and ACR70 response rates produced similar conclusions to that for ACR20 in that no definitive conclusion could be made.

One of the clinical advisors to the ERG highlighted potential issues that may influence such comparisons between interventions, for example changes in patient populations over time and study design issues, for example with respect to non-responders being withdrawn in the placebo arm of the certolizumab trials leading to raised treatment group response rates.

Efficacy in the TNF-α inhibitor experienced population: Estimated relative risks for ACR20, ACR50 and ACR70 were lower for golimumab vs. rituximab at 24 weeks, although considerable variation was present in these estimates, with wide 95%CI for each ACR response.

Safety in the DMARD experienced population: Meta-analyses were conducted for selected safety outcomes. Golimumab was estimated to have more serious adverse events than all comparators except certolizumab, although there was considerable uncertainty as shown by the wide confidence intervals. The estimated rate of serious infections for golimumab was similar to the rate for infliximab and etanercept, which were the lowest for the interventions, although all had wide confidence intervals. Golimumab was estimated to have the fewest injection site reactions and discontinuations due to adverse events, although the values for all interventions were subject to considerable uncertainty.

Safety in the TNF-α inhibitor experienced population: The estimated relative risks obtained for serious adverse events were similar for golimumab and rituximab, although there was considerable uncertainty. The relative risk estimate for serious infections was slightly lower for golimumab compared with rituximab, however both have wide confidence intervals. No data were available for rituximab for injection site reactions. The relative risk for discontinuation of treatment due to adverse events was lower for golimumab than rituximab, although the confidence intervals were wide and overlapped considerably.

These conclusions were also supported by a mixed treatment comparison for the DMARD experienced population and an indirect comparison using the Bucher method for the TNF-α inhibitor experienced population, which produced similar results.

1.3 Summary of submitted cost effectiveness evidence

The manufacturer submitted two decision-analytic Markov models built in Microsoft Excel. The models took a lifetime perspective, and evaluated sequential treatments until a patient's death. Many of the assumptions and model methods are similar to those used in other NICE Technology Appraisals of treatments for Rheumatoid Arthritis.

The two models evaluated golimumab in two different patient populations:

- DMARD experienced population – golimumab is compared to TNF-α inhibitors (etanercept, adalimumab, infliximab and certolizumab) and methotrexate in patients who have failed two DMARDs.
- TNF-α inhibitor experienced population – golimumab is compared to rituximab and methotrexate in patients who have failed on two DMARDs and a TNF-α inhibitor.
- All comparators are given with concomitant methotrexate. Methotrexate monotherapy is included as a comparator in each population as it represents the placebo arm in each indirect/mixed treatment analysis.

Patients progress to the next treatment if they do not achieve at least an ACR20 response at 6 months, or if they come off treatment due to either a lack of efficacy or an adverse event. In both models, patients progress to leflunomide, gold, azathioprine, ciclosporin and then palliative care.

While patients were within a health state, it was assumed that their disease activity increases over time. This was modelled by applying a HAQ progression rate. DMARDs, TNF-α

inhibitors, rituximab and palliative care all had differential HAQ progression rates. These rates are a source of uncertainty and sensitivity in the model results.

The models contain three health states: ACR20 responders, ACR50 responders and non-responders. The probability of being in a health state is derived from evidence synthesis of the clinical trial data. The DMARD experienced population model uses a mixed treatment comparison to provide relative risks for all comparators compared with golimumab. An indirect comparison is used in the TNF-α inhibitor experienced population model to compare golimumab and rituximab.

Costs relating to treatment, administration and hospital attendances were included in the economic model. Health utilities were estimated using a published regression function to convert HAQ to EQ-5D values. Costs and QALYs were discounted at a rate of 3.5%. The impact of parameter uncertainty was estimated in a probabilistic sensitivity analysis. Scenario analyses were run on key parameters.

An incremental analysis was performed within each population, however an incremental analysis was not possible between the populations, and so the optimal position of golimumab cannot be determined. The mean results for golimumab from the probabilistic sensitivity analyses undertaken by the manufacturer using their base case can be summarised as follows;

- DMARD experienced population – Golimumab has an ICER of approximately £26,000 per QALY compared to methotrexate. It is a dominant strategy compared to infliximab and certolizumab (more effective and less costly), however golimumab is extendedly dominated by adalimumab and etanercept.
- TNF-α inhibitor experienced population – Golimumab has an ICER of approximately£29,000 per QALY compared to methotrexate. It dominates rituximab, as it is £31 less costly and generates an extra 0.189 QALYs.

The DMARD experienced population results were generally robust to scenario analyses, however the TNF-α inhibitor experienced population model was extremely sensitive to the HAQ progression rates, and to the frequency of re-administering rituximab. Probabilistic sensitivity analysis (PSA) results were provided for the basecase results, but only deterministic results were presented for the scenario analyses.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

The submission can be considered to represent a balanced estimate of the effects of golimumab relative to comparator therapies for the efficacy and safety outcomes and comparators that were included. Clinical effectiveness review methods and results were reasonably clearly presented, with adequate systematic searches conducted. All relevant RCTs for golimumab and comparators appeared to have been included. The included golimumab trials were each of reasonable methodological quality and were considered by the clinical advisors to the ERG to have fair generalisability to the UK population. A good level of detail was provided for clinical effectiveness evidence in the MS and supplementary manufacturer responses.

The mixed treatment comparisons and indirect comparisons used appropriate trials to inform the networks of evidence. These analyses allow a comparison of all comparators within the cost-effectiveness analysis.

The economic model was generally of a high quality although some errors were identified and corrected by the ERG. Many of the sources of evidence are appropriate and the assumptions made are tested in a range of scenario analyses.

1.4.2 Weaknesses

The clinical efficacy data presented could be considered to have only partially addressed the decision problem, in that tocilizumab and abatacept were not incorporated into analyses and golimumab is not compared to other TNF-α inhibitors (adalimumab, infliximab, abatacept, etanercept) in the TNF-α inhibitor experienced population. Furthermore, not all relevant outcomes were fully addressed in the submission, for example radiological progression of joint damage.

Whilst additional meta-analyses, mixed treatment comparisons and indirect comparisons were conducted for ACR70 responses, this important outcome was not incorporated into cost-effectiveness analyses. Not including ACR70 responses is likely to have biased the results in favour of golimumab, as golimumab has a lower relative risk estimate than all but one comparator drug although the confidence intervals are wide and overlapping for all interventions.

The models do not allow a fully incremental comparison between use of golimumab after failure of DMARDs and use after failure of a previous TNF- α inhibitor, so it is not possible to identify the most cost-effective position for golimumab within the treatment pathway.

1.4.3 Areas of uncertainty

The confidence intervals and credible intervals surrounding key parameters in the model are wide and no definitive conclusion can be made regarding efficacy, serious adverse events, serious infections, injection site reactions or discontinuations after adverse events.

The absence of any available head-to-head trials for golimumab versus comparator therapies represents a source of uncertainty in the clinical effectiveness evidence base, although golimumab and comparators were evaluated against each other using mixed treatment comparisons and indirect comparisons in the MS.

The performance of golimumab and comparators were not assessed across the full range of outcomes pre-specified in the scope. Of particular interest would be the impact of golimumab vs comparator drugs in terms of radiological progression and the potential impact this may have on the cost-effectiveness estimates were this outcome to be incorporated in the model.
A lack of high quality evidence around the HAQ progression rates for TNF-α inhibitors, rituximab, DMARDs and palliative care produces uncertain estimates of comparative effectiveness in the economic model.

1.5 Key issues

The results of the analysis suggest that in the DMARD experienced population, golimumab has an ICER compared to methotrexate that is similar to that for other TNF-α inhibitors already recommended by NICE, however it is never the most cost-effective therapy. This conclusion is generally robust under the scenario analyses conducted, but must be considered in light of the fact that ACR70 responses were not included in the analysis. However, if these interventions are considered to be class and it is assumed that there is no difference in any clinical outcomes between the TNF-α inhibitors, the lowest cost intervention would be optimal.

The sensitivity of the results from the model to the uncertainty surrounding the HAQ progression rate estimates and the re-administration frequency of rituximab mean that a definitive estimate of the cost-effectiveness of golimumab in the TNF-α inhibitor experienced population cannot be concluded. However, the exploratory analyses conducted by the ERG

demonstrate than when making alternative assumptions regarding these two factors, which the ERG believe to be more plausible than those made in the manufacturer's basecase, golimumab is dominated by rituximab. This is the opposite conclusion to that found by the manufacturer.

As the manufacturer restricted their analysis to those interventions recommended by NICE at the time they were invited to make their submission and excluded some interventions for which there is now published guidance, it is not possible to assess the clinical and cost effectiveness of golimumab against all of the relevant treatments options. It is also not possible from the analyses presented to assess the most cost-effective position for golimumab within the whole treatment pathway.

2 BACKGROUND

2.1 Critique of manufacturer's description of underlying health problem

The estimates of incidence (4,102) and prevalence (221,490) for diagnosed RA in England and Wales provided by the manufacturer are based on published UK studies.[4,5] They are lower than the estimates provided by the National Audit Office (NAO)[6] which are based on a five year analysis of the General Practice Research Database. This was used by the NAO to estimate the incidence of newly diagnosed rheumatoid arthritis (26,000) and this was combined with mortality data to estimate the prevalence (580,000) using a Markov model.

The estimates of mortality and morbidity presented in the MS appear reasonable, although it is possible that older estimates of morbidity in RA may not reflect the impact of newer treatments on disease progression. One study has reported an association between etanercept and employment outcomes[7] whilst other studies[8,9] including a recent analysis of the British Society for Rheumatology Biologics Registry (BSRBR)[10] found that TNF-α inhibitor therapy did not prevent patients with RA from becoming work disabled.

2.2 Critique of manufacturer's overview of current service provision

The description of current service provision was accurate at the time the manufacturer was invited to make its submission (April 2010). The MS states that the treatment options for those failing to respond a first TNF-α inhibitor include switching to a different TNF-α inhibitor, rituximab, abatacept or tocilizumab, but that rituximab is the only one of these options currently recommended by NICE. However, in August 2010 NICE released guidance on the use of tocilizumab in RA[1] and guidance on the use of drugs (adalimumab, etanercept, infliximab, rituximab and abatacept) after failure of a TNF-α inhibitor.[2] This guidance increases the treatment options available to those patients who have failed on a TNF-α inhibitor but for whom rituximab therapy is contraindicated or unavailable because of an adverse event. Specifically adalimumab, etanercept, infliximab, abatacept, and tocilizumab, each in combination with methotrexate, are now recommended as treatment options. If rituximab therapy cannot be given because methotrexate is contraindicated or withdrawn because of an adverse event, adalimumab and etanercept, each as monotherapy, are now recommended as treatment options. There is also an additional treatment available to patients who have responded inadequately to rituximab, as tocilizumab (in combination with methotrexate) is now recommended for this indication.

There is little evidence provided in the MS on how often the different drugs are used in the UK at present. The MS estimates that there are currently 71,062 patients who are eligible for biologic therapy (Table 10 of the MS). However, this is at odds with the figure provided by the NAO who estimate that there were 11,900 biologic eligible patients in 2008-2009.[6] A previous NICE Technology Assessment Report[11] stated that by July 2009, 12,626 patients who started treatment with a TNF-α inhibitor were registered with the British Society for Rheumatology Biologics Registry (BSRBR) and that 23% had switched from a first to a second TNF-α inhibitor. The number registering as being treated with rituximab was cited as being 442 by August 2009.[11] Current service provision is therefore likely to include some use of TNF-α inhibitors after failure of a first TNF-α inhibitor and this may increase as a result of the recent guidance documents published by NICE.[1,2]

There is no mention in the report of the use of triple DMARD therapy in RA. The NICE Clinical Guideline recommends that a combination of DMARDs (including methotrexate and at least one other DMARD, plus short-term glucocorticoids) is offered to people with newly diagnosed active RA.[12] In some patients this may mean that triple DMARD therapy is offered prior to the use of a TNF-α inhibitor. The treatment sequences used in the model prior to biologic therapy may not represent current care in UK, although this is variable. The ERG's clinical advisors have advised the ERG that the sequence methotrexate, methotrexate-sulfasalazine, methotrexate- TNF-α inhibitor is not widely used. Instead it has been suggested that some clinicians may use methotrexate, methotrexate-sulfasalazine-hydroxychloroquine and then methotrexate-leflunomide (although the SPC for leflunomide excludes use in combination with methotrexate) prior to introducing a TNF-α inhibitor in combination with methotrexate.

In terms of the treatments used following biologic therapy our clinical advisors agreed that many patients will use DMARDs such as azathioprine, ciclosporin or gold as specified in the MS. However, they did not agree that the costs of drugs would be zero in patients undergoing what the manufacturer refers to as 'palliative care' as many patients would be receiving steroid injections and using analgesics which can be expensive (e.g transdermal patches).

3 CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM

A summary of the decision problem addressed in the MS is shown in Table 1 [information has been modified from that presented in Table 11 of the MS to reflect the ERGs view of the decision problem addressed in the MS].

Table 1: Decision problem as issued by NICE and addressed by the MS

	Final scope issued by NICE	**Decision problem addressed in the submission**
Population	Adults with RA who have had an inadequate response to DMARDs	Adults with moderate to severe, active RA who have had an inadequate response to DMARDs, including methotrexate (MTX)
Intervention	GOL in combination with methotrexate	Same as in final scope.
Comparator(s)	Management strategies involving DMARDs without golimumab, including treatment with: • conventional DMARDs (for example, sulfasalazine, leflunomide) • biological agents (including adalimumab, etanercept, infliximab, rituximab, tocilizumab*, certolizumab pegol, abatacept*). *Subject to ongoing appraisal at the time of scoping	Management strategies involving DMARDs without golimumab, including treatment with: • conventional DMARDs (for example, sulfasalazine, leflunomide) • biological agents (including adalimumab, etanercept, infliximab, rituximab, certolizumab pegol). Tocilizumab and abatacept are excluded from the meta-analysis and economic evaluation.
Outcomes	The outcome measures to be considered include: • disease activity • physical function • joint damage • pain • mortality • fatigue • radiological progression • extra-articular manifestations of disease • adverse effects of treatment • health related quality of life	The outcome measures addressed include: • disease activity • physical function • mortality • adverse effects of treatment • health related quality of life (trial outcomes not provided but HRQoL incorporated within economic analysis)

Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of 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.	Cost effectiveness of treatments expressed in terms of incremental cost per quality-adjusted life year. Time horizon considered is lifetime of the patient. Costs are considered from an NHS and Personal Social Services perspective.
Other considerations	If evidence allows, the appraisal will consider subgroups of people defined by the baseline severity of their RA. If the evidence allows, the appraisal will consider the costs of joint replacement therapy and hospital admissions. Guidance will only be issued in accordance with the marketing authorisation.	Subgroups include: Biologic experienced patients who discontinued treatment due to lack of efficacy The submission considers the cost of hospital admissions. Joint replacement data is not available from the pivotal trials. Submission in line with the current marketing authorisation.

3.1 Population

The population considered within the MS is one of adults with moderate to severe, active rheumatoid arthritis who have had an inadequate response to DMARDs including methotrexate. The rationale for this choice of population is that this is the population in which golimumab (in combination with methotrexate) is licensed. Golimumab is also licensed for the treatment of psoriatic arthritis and ankylosing spondylitis. The scope for this appraisal specified a similar but slightly broader population in that it specified patients with RA who have had an inadequate response to DMARDs but did not mention methotrexate specifically. However, NICE's Clinical Guideline 79[12] on the management of rheumatoid arthritis recommends that patients with newly diagnosed active RA are offered a combination of DMARDs, including methotrexate and at least one other DMARD. Therefore it seems reasonable to assume that the majority of patients with an inadequate response to DMARDs will have been offered methotrexate.

The MS considers patients who have never received a TNF-α inhibitor (DMARD experienced population) separately from patients who have had prior therapy with a TNF-α inhibitor (TNF-α inhibitor experienced population). This seems reasonable given that the treatment

options available to patients with rheumatoid arthritis under current NICE guidance differ depending on whether or not they have had prior therapy with a TNF-α inhibitor. Specifically rituximab is only recommended by NICE in patients who have had treatment with at least one TNF-α inhibitor.[13] It is therefore necessary to consider different comparators for these two groups of patients.

The scope also specified that if evidence allowed, it would consider subgroups of people defined by baseline severity. The ERG have not received any additional evidence from the manufacturer on the clinical and cost-effectiveness of golimumab in moderate and severe RA subgroups at the point of submission of this report.

3.2 Intervention

GOL is a human immunoglobulin G1κ (IgG1κ) monoclonal antibody produced by murine hybridoma cell line with recombinant DNA technology. Golimumab in combination with methotrexate has a marketing authorisation for the treatment of moderate to severe, active rheumatoid arthritis in adult patients when the response to disease-modifying anti-rheumatic drug (DMARD) therapy including methotrexate has been inadequate. It is delivered by monthly subcutaneous injection using a pre-filled syringe or pre-filled pen (autoinjector) and should be given concomitantly with methotrexate. Available data suggest that clinical response is usually achieved within 12 to 14 weeks of treatment (after 3-4 50mg doses) and that continued therapy should be reconsidered in patients who show no evidence of therapeutic benefit within this time period. In patients weighing more than 100 kg who did not achieve an adequate clinical response after 3 or 4 doses, increasing the dose of golimumab to 100 mg once a month may be considered. Continued therapy should be reconsidered in patients who show no evidence of therapeutic benefit after receiving 3 to 4 additional doses of 100 mg.[14] The dosing regimen considered in the model does not take into account the likelihood of dose increases from 50mg to 100mg and assumes all patients receive 50mg doses although the MS does contain evidence in Table 133 showing that 7% of patients are over 100kg and may therefore be eligible for a dose increase. It should be noted the SPC[14] for golimumab doesn't specifically mention its use in patients previously treated with one or more TNF-α inhibitor whereas use in this population is specifically mentioned in the SPC[15] for rituximab.

3.3 Comparators

Relevant comparators listed in the scope for this appraisal were conventional DMARDs (for example, methotrexate, sulfasalazine, leflunomide), and biological agents (including

adalimumab, etanercept, infliximab, rituximab, tocilizumab, certolizumab, abatacept). However, the MS excludes tocilizumab and abatacept from the decision problem stating that these treatments were subject to ongoing appraisal by NICE and there was no final guidance recommending their use at the time (28 April 2010) the manufacturer was invited to make its manufacturer submission. These treatments were included within the systematic review searches for comparator interventions, but were not included within the narrative findings, meta-analyses, mixed treatment comparisons or the economic evaluation within the MS.

NICE released draft guidance on tocilizumab for the treatment of rheumatoid arthritis in June 2010[16] and draft guidance on drugs (adalimumab, etanercept, infliximab, rituximab and abatacept) for the treatment of rheumatoid arthritis after the failure of a TNF-α inhibitor in June 2010.[17] Following this, the ERG requested that the manufacturer include these potential new indications for treatments within RA as comparators in the meta-analyses, mixed treatment comparisons and economic evaluation. In response to this the manufacturer reiterated that final guidance was not available for these appraisals and stated that it was not possible to add these comparators to the MS within the timeframe available. However, since this response was received from the manufacturer, final guidance from these two appraisals has been released as described in section 2.2.

The ERG considers that it would be beneficial to know how golimumab compares to other TNF-α inhibitors or to tocilizumab when these treatments are used after the failure of a first TNF-α inhibitor as these are now treatment options recommended by NICE when rituximab therapy is contraindicated or withdrawn due to an adverse event. It would also be beneficial to know how golimumab compares to tocilizumab when used after a failure to respond to at least one TNF-α inhibitor and rituximab, although it is recognised that there are no trials of golimumab in this population.

3.4 Outcomes

The main clinical effectiveness outcome which is the focus of the MS is the response to treatment as measured by the American College of Rheumatology (ACR) criteria. ACR is a composite measure which includes both subjective (e.g global assessment of disease activity using a visual analogue scale) and objective measures (e.g swollen/ tender joint counts and CRP) of disease activity and functional status. Physical function assessed using the Health Assessment Questionnaire Disability Index (HAQ-DI) is one of the ACR criteria. The ACR response is reported according to the percentage improvement across the various criteria (e.g ACR20 is defined as a 20% improvement in five of the seven criteria). The ACR response rates are reported within the submission for golimumab and comparator interventions and are

used as the measure of clinical effectiveness within the economic model. However, only ACR20 and ACR50 are used as outcomes within the economic model whilst ACR70 is omitted. This excludes a valuable source of effectiveness evidence from the analysis that could be used to generate a more accurate estimate of cost-effectiveness. The MS also reports Disease Activity Scores (DAS28) and HAQ-DI improvement from baseline for golimumab but not for all comparator interventions. Adverse event rates are also presented for golimumab and for all comparator interventions.

The scope also specified joint damage, pain, mortality, fatigue, radiological progression, extra-articular disease manifestations, and health related quality of life as relevant outcomes. Data on tender and swollen joints and pain (patient assessment using VAS) were extracted and included in the tables of baseline characteristics. However, this does not address the impact of treatment on these outcomes. Mortality data were provided, where available, for both golimumab and comparator interventions after request from the ERG. The manufacturer stated that fatigue was included within the search strategy but was not commonly reported and no data on fatigue are included within the submission. The manufacturer stated that radiological progression data were not available for the pivotal trials. [redacted] Our clinical advisors noted that evidence on reduced radiological progression is available for other TNF-α inhibitors.[3] The MS stated that extra-articular disease manifestations were not routinely reported in RCTs and our clinical advisors agreed that data on this outcome was unlikely to be available. [redacted]. Health related quality of life (HRQoL) data is incorporated in the economic model through the use of a mapping function to extrapolate EQ-5D utility scores from HAQ-DI scores. The scope specified that the economic analysis should include the cost of joint replacements if the evidence allowed but this was not included in the MS, citing the fact that this was not widely reported in clinical trials.

The economic model provides estimates of costs and QALYs for each intervention and incremental analysis was conducted to provide ICERs. However, the incremental analysis is conducted separately for the DMARD experienced and the TNF-α inhibitor experienced populations. This doesn't answer the question of whether it is more cost-effective to use golimumab after the failure of DMARDs or after the failure of a TNF-α inhibitor.

3.5 Time frame

The majority of the RCTs summarised in the MS had follow-up periods between 24 and 52 weeks. The MS stated that the two key golimumab RCTs publications included follow-up periods of 24 weeks (although evidence of longer follow-up was provided by the manufacturer in response to a request by the ERG) and a third smaller trial had follow-up lasting 52 weeks. The clinical effectiveness outcomes used within the economic model were ACR responses (ACR 20 and ACR50) at 24 weeks. The economic analysis uses a lifetime model with patients starting age at 50 (DMARD experienced) and 54 (TNF-α inhibitor experienced). Final model outcomes are assessed at 43 years after entering the model.

3.6 Other relevant factors

Golimumab is given by monthly injection which is less frequently than some of the TNF-α inhibitors (certolizumab, adalimumab, etanercept) currently recommended by NICE. Whilst a reduced dosing frequency may improve patient convenience, the clinical advisors to the ERG noted that the apparently longer half-life of golimumab may have potential implications for patient management. The effectiveness of golimumab in reducing symptoms associated with RA may vary over the dosing period, and it is unclear whether maximal and continual suppression is more or less beneficial than variable suppression. In addition a monthly dosing frequency may potentially delay the withdrawal of immune system suppression in patients receiving golimumab who present with infections and this may have implications for their management.

4 CLINICAL EFFECTIVENESS

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

An evidence-based checklist developed by McGowan *et al.* (2010)[18] was used to assess the quality of the electronic search strategies reported in the manufacturer's submission. Three types of searches were conducted to identify all relevant clinical effectiveness, adverse event and cost effectiveness studies for golimumab (and comparators) for the treatment of rheumatoid arthritis. Overall, the translation of the research question was deemed to be adequate with further recommendations. The clinical and cost effectiveness searches were comprised of population (rheumatoid arthritis) and intervention terms including a methodology filter (e.g. RCT). In the adverse events search strategies, the intervention and outcome terms (e.g. side effects) were applied without a methodology filter.

According to the MS, there were no RCTs on the adverse effects of the interventions in the adverse event searches. It is noteworthy that retrieval of adverse event information is difficult due to poor reporting and indexing of adverse event terms. Therefore, a sensitive search strategy and a wide range of sources searched was necessary to maximise the retrieval of all potentially relevant studies. TOXLINE is a specific and freely available database that can be searched for literature on adverse events. The Boolean operators that were applied in the adverse event searches to combine retrieved records from each concept was somewhat restrictive; even though there is evidence that four different approaches were tested (i.e. free-text terms for adverse events, specified adverse events, indexed adverse event terms and adverse event subheading with indexed intervention names). The operator AND was used to combine all approaches and this could explain why the resulting records numbers in both MEDLINE and EMBASE were low. The search could be made more sensitive by firstly, applying the OR operator within each approach.[19] Results from each approach are then combined with the drug intervention terms using the AND operator. The final result is the combination of records retrieved via each approach using the OR (rather than the AND operator). A methodology filter could then be applied at the end of the strategy to limit the results by study type.

In the cost effectiveness searches, it was not clear why records from the drug intervention concepts were combined with the tumour necrosis factor terms using the AND operator (statements 2-5 combined with 6-12). Again, this substantially limits the number of retrieved

records. By comparison to the clinical effectiveness searches, the OR operator was used instead.

In several of the search strategies, various MeSH headings were applied such as rheumatoid arthritis, the drug classes (e.g. tumour necrosis factor) and adverse event terms (e.g. drug toxicity). The spelling and syntax used in the strategies were adapted according to the different platforms used for Medline (PubMed, Ovid), Embase (Embase.com, Ovid) and Cochrane (Wiley Interscience and CRDweb) in the clinical effectiveness, adverse events and cost effectiveness searches. Not all the origin dates of the various platforms that supported the different databases were stated (i.e. PubMed and Embase.com).

In the Embase adverse events search strategy of the MS, statement 35 reads '26 or 34,' which retrieves 771,311 records and 5,104 records respectively. It was not clear how incorporating the adverse events search terms and biologics yielded ~5888 records. Reproducing the search according to the reported strategy did not yield the same number of records; perhaps the statement should read '27 or 34.'

Searches for the effectiveness studies were limited by study type by applying the RCT or cost effectiveness filters. The RCT filter is adequate. However, it appears that the cost effectiveness filter is not particularly sensitive. The filter comprised of two concepts (economic terms and quality of life terms) that were combined with the AND operator. It is suggested that the OR operator could have been used to combine these two concepts to improve the sensitivity of the filter (~200 records in Medline retrieved compared to 30 in the report).

All searches stated in the report were reproduced in their respective platforms with the exception of four strategies because of availability and accessibility. It was not surprising to find that the number of records retrieved by the ERG were slightly higher since these searches are up to date (up to week 30 of 2010). A total of 52 studies were included in the submission; a majority of these studies were indeed retrieved from the electronic searches via their reported search strategies.

On the whole, there appeared to be more emphasis placed on search precision rather than sensitivity in the adverse event and cost effectiveness strategies which could have been improved by the use of appropriate operators to combine the various concepts within the strategies. However, all appropriate RCTs have been included in the MS for golimumab and comparators to the best knowledge of the ERG and clinical advisors to the ERG.

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

The inclusion and exclusion criteria used in the selection of evidence for the systematic review were presented in the MS (page 29-30). The table in the MS was labelled as 'eligibility criteria used in search strategy' but was presented within the description of the study selection process (Section 5.2.1). It was not clear from the MS how many reviewers were involved in the study selection process. Best practice specifies that two reviewers be involved in the application of inclusion and exclusion criteria in order to limit bias in study selection. Details of the inclusion and exclusion criteria applied in the MS are presented in Table 2.

Table 2: Inclusion/exclusion criteria used in study selection (as presented by the manufacturer)

Inclusion criteria	**Details**
Populations	1. Adult patients (≥ 18 years) with active rheumatoid arthritis despite treatment with at least one conventional DMARD for ≥ 3 months; no previous use of TNF-α inhibitor or other biologic agents 2. Adult patients (≥ 18 years) with active rheumatoid arthritis despite treatment with at least one TNF-α inhibitor.
Interventions	Abatacept, adalimumab, certolizumab, etanercept, golimumab, infliximab, rituximab or tocilizumab compared with any other agent including placebo.
Study design	Double-blind, randomised controlled trials
Outcomes	Any of the following outcomes of interest: Measures of treatment efficacy: ACR responses, mean DAS or DAS28, number of patients achieving low DAS (<3.2) or DAS remission (<2.6), HAQ-DI Measures of safety and tolerability: adverse events, treatment discontinuations
Report characteristics	Articles for which the full text was available in English. No publication date restrictions were applied.
Exclusion criteria	**Details**
Populations	1. Conventional DMARD-naïve patients 2. Mixed populations of both DMARD experienced and TNF-α inhibitor experienced patients (>10% from each group) unless analysed separately.
Study design	Studies with no appropriate comparisons between biologic agents and other active comparators or placebo (eg. open label extensions and observational studies) Studies in which the drug of interest is not administered at the EMA-approved dose or details of dosing are not given.
Report characteristics	Reviews, systematic reviews and meta-analyses

Justification was provided in the MS for the use of the above inclusion and exclusion criteria. Only studies in which all patients had previously received conventional DMARDS were included. The inclusion of studies in which patients had received at least one conventional DMARD for at least 3 months was considered by the clinical advisors to the ERG to be appropriate.

Abatacept and tocilizumab were described in the inclusion criteria as being included interventions. However, both of these comparator drugs were omitted from the analyses. This

issue is discussed more extensively in relation to the appropriateness of the decision problem and treatment sequences examined in the MS.

The MS stated that only double-blind RCTs were included in the analysis. It was not explicitly stated in the MS whether both phase II and phase III clinical trials were eligible for inclusion. The ERG consider non-RCTs to be a valid and important source of evidence for the evaluation of adverse events. Controlled clinical trials may exclude patients at high risk from harms,[20] may be too short in terms of follow-up to detect long-term harms, may not have sufficiently large sample sizes to detect uncommon adverse events, or may not have reported them in a consistent manner.[21,22,23,24] The MS stated on page 115 that, as golimumab is a new drug, no non-RCTs or observational studies were available at the point of submission. However, it is possible that non-RCT evidence may be available for the comparator drugs which was not identified.

The outcomes listed above were considered by the ERG and clinical advisors to be appropriate. However, it is important to note that not all outcomes listed in the decision problem were present in the tabulated inclusion/exclusion criteria in the original MS (page 29), with the omission of joint damage, radiological progression, fatigue, mortality, pain, extra-articular manifestations of disease and health-related quality of life.

The exclusion of studies not available in English was a reasonable decision on the basis of available time. A single adalimumab trial reported by Huang *et al.* (2009) was described in the MS as being excluded as the full text was not available in English. The potential implications of this exclusion were not discussed.

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

A total of 39 articles reporting a total of 31 individual RCTs were described as being selected for inclusion in the meta-analyses for golimumab and comparator therapies. The MS subsequently stated that 28 and 5 RCTs (ie. a total of 33) were included in the meta-analyses for the DMARD-experienced and TNF-α inhibitor experienced populations respectively. It should be noted that there appeared to be inconsistency in the MS in terms of reporting, with a total of 32 RCTs being described as being included on pages 34 and 35 (as opposed to the 31 RCTs listed in the QUOROM flow diagram and the 33 RCTs referred to above). A complete list of all relevant included RCTs was presented in the MS (pages 32-33). No non-RCT evidence was included in the clinical effectiveness section.

Three RCTs evaluating the use of golimumab in patients with rheumatoid arthritis following the failure of previous DMARD therapies were included in the MS. Details of the study design and patient characteristics as collated by the ERG are presented in Table 3.

One phase III 4-arm randomised controlled trial (GO-FORWARD) reported by Keystone *et al.* (2009b)[25] was included in the analyses for the DMARD-experienced population. An additional 5-arm RCT by Kay *et al.* (2008)[26] was also included for the DMARD-experienced population. This study was described on page 110 as being a phase II dose-ranging study. A single phase III 3-arm RCT (GO-AFTER) by Smolen *et al.* (2009a)[27] was included for the TNF-α inhibitor experienced population. Treatment arms featuring the licensed dose and control group are focused on in this report.

Table 3: Characteristics of included studies

Study	Design and clinical trial identification codes[a]	Participants	Interventions	Concomitant medication	Outcomes	Follow up (weeks)
GO-FORWARD (Keystone *et al.*, 2009b)	Randomised, double-blind, placebo-controlled multinational phase III trial (n=444) Protocol number: C0524T06 Clinicaltrials.gov Identifier: NCT00264550	*Inclusion criteria:* Patients aged ≥18 years with active RA diagnosed according to revised 1987 criteria of the American College of Rheumatology (ACR) for at least 3 months before screening, having been on a stable methotrexate dose of 15 mg/week or greater but 25 mg/week or less during the 4-week period immediately preceding screening and having had tolerated 15 mg/ week or greater of methotrexate for at least 3 months before screening. *Study population characteristics:* *Golimumab plus methotrexate* 80.9% female, mean age 52.0 years (43.0, 57.0), disease duration 4.5 years (2.1, 9.7), 86.5% positive rheumatoid factor, 75.3% concomitant glucocorticoid therapy, median (inter-quartile range) DAS28 6.105 (5.366, 6.940), median	4-arm study, including: Golimumab plus methotrexate (n=89) Placebo plus methotrexate (n=133) Golimumab (50mg) or placebo administered subcutaneously every 4 weeks Methotrexate (≥ 15 mg) every week. Dose (mg/week) (mean, SD): [redacted]	NSAIDs, analgesics, oral prednisone ≤ 10 mg/day permitted if at stable dose for ≥ 2 weeks, other DMARDs discontinued prior to study	*Primary efficacy endpoints:* Two co-primary endpoints used : i) proportion of patients achieving ACR20 response at week 14 ii) improvement from baseline in HAQ-DI score at week 24 *Secondary endpoints (as reported on page 287 of MS):* ACR20, ACR50, ACR70 and ACR90 responses over time, including at week 24 ACR-N DAS28 calculated separately using both CRP and ESR HAQ-DI	24

Study	Design and clinical trial identification codes[a]	Participants	Interventions	Concomitant medication	Outcomes	Follow up (weeks)
		(inter-quartile range) HAQ-DI 1.355 (1.000, 1.875) *Placebo plus methotrexate* 82.0% female, mean age 52.0 years (42.0, 58.0), disease duration 6.5 years (3.1, 11.9), 81.2% positive rheumatoid factor, 65.4% concomitant glucocorticoid therapy, mean DAS28 6.111 (5.260, 6.574), HAQ-DI 1.250 (0.750, 1.750) Prior mean number of DMARDs not specified, mean duration not specified (methotrexate ≥ 3 months) for both golimumab 50 mg and placebo arms	[redacted]		Serum samples taken at baseline and week 24 assayed for presence of antibodies to golimumab Adverse events	
Kay *et al.*, 2008	Randomised, double-blind, placebo-controlled 5-arm multinational phase II dose-ranging study	*Inclusion criteria:* Adult patients with active RA as defined by 1987 revised criteria of the American College of Rheumatology for at least 3 months before screening. Patients must have been treated with MTX at dosage of at least 10 mg/week for ≥3 months and	Treatment arms included: Golimumab (every 4 weeks) plus methotrexate (n=35)	NSAIDs, oral prednisone ≤ 10 mg/day permitted if at stable dose for ≥ 4 weeks, other	*Primary efficacy endpoint:* ACR20 response at week 16 *Secondary endpoints (as reported on page 287 of MS):* Improvement from baseline at week 16 in DAS28 score Numeric index of the ACR	52

Study	Design and clinical trial identification codes[a]	Participants	Interventions	Concomitant medication	Outcomes	Follow up (weeks)
	(n=172)[b] Protocol number: CR005263 Clinicaltrials.gov Identifier: NCT00207714	at a stable dosage for ≥4 weeks before receiving first dose of study medication. *Study population characteristics:* *Golimumab 50 mg every 4 weeks plus methotrexate* 85.7% female, mean age 57.0 years (50.0, 64.0), disease duration 8.2 years (4.1, 14.3), positive rheumatoid factor not reported, concomitant glucocorticoid therapy not reported, median (inter-quartile range) DAS28 6.4 (5.6, 7.3), median (inter-quartile range) HAQ-DI 1.7 (1.4, 2.0) *Placebo plus methotrexate* 74.3% female, mean age 52.0 years (46.0, 66.0), disease duration 5.6 years (1.4, 10.9), positive rheumatoid factor not reported, concomitant glucocorticoid therapy not reported, median (inter-quartile range) DAS28 6.3 (5.7, 7.0), median (inter-quartile range HAQ-DI 1.3 (0.9, 1.9)	Placebo plus methotrexate (n=35) Golimumab (50 mg) or placebo administered subcutaneously every 2 weeks (golimumab every 4 weeks with placebo on alternate weeks)	DMARDs discontinued prior to study	response (ACR-N) at week 16 ACR20, ACR50, and ACR70 responses over time through week 52, including week 20. Serum samples taken at weeks 0, 48, 52 and 68 and were assayed for presence of antibodies to golimumab Adverse events	

Study	Design and clinical trial identification codes[a]	Participants	Interventions	Concomitant medication	Outcomes	Follow up (weeks)
		Prior mean number of DMARDs not specified, mean duration not specified (methotrexate ≥ 3 months) for both golimumab 50 mg and placebo arms				
GO-AFTER (Smolen *et al.*, 2009a)	Randomised, double-blind, placebo-controlled multinational phase III trial (n=461) [b] Protocol number: C0524T11 Clinicaltrials.gov Identifier: NCT00299546	*Inclusion criteria:* Patients aged 18 years or older with active rheumatoid arthritis diagnosed according to the criteria of the American College of Rheumatology (ACR) at least 3 months before screening. Eligible patients must have been treated with at least one dose of a TNF-α inhibitor (etanercept, adalimumab, or infliximab), the last dose of which must have been given at least 8 weeks (adalimumab or etanercept) or 12 weeks (infliximab) before the first dose of the study drug. Previous TNF-α inhibitor treatment may have been discontinued for any reason (effectiveness, intolerance or other) (inconvenience and accessibility issues most	Treatment arms included: Golimumab (n=153) Placebo (n=155) (Background DMARDS optional) Golimumab (50 mg) or placebo administered subcutaneously every 4 weeks Other DMARD doses were not specified.	DMARDs (methotrexate, sulfasalazine and hydroxychloroquine) permitted but not required if patients tolerated dose for ≥ 12 weeks and at a stable dose for ≥ 4 weeks prior to study;	*Primary efficacy endpoint:* Achievement of a 20% or greater improvement in ACR criteria for rheumatoid arthritis (ACR20) at week 14 *Secondary endpoints (as reported on pages 285-286 of MS):* ACR20 at week 24 ACR50 and ACR70 at weeks 14 and 24 Numeric index of the ACR response at weeks 14 and 24 DAS28 at weeks 14 and 24 HAQ-DI scores at weeks 14	24

Study	Design and clinical trial identification codes[a]	Participants	Interventions	Concomitant medication	Outcomes	Follow up (weeks)
		commonly cited as other issues). Patients who were receiving methotrexate, sulfasalazine, or hydroxychloroquine at baseline allowed to discontinue these drugs before starting the study. However, if they continued these drugs, the dose had to be maintained throughout the study. *Study population characteristics:* *Golimumab* 74% female, median (inter-quartile range) age 55.0 years (46.0, 63.0), median (inter-quartile range) disease duration 9.6 years (5.6, 17.2), 72% positive rheumatoid factor, concomitant glucocorticoid therapy not reported, median (inter-quartile range) DAS28 6.3 (5.6, 7.2), median (inter-quartile range) HAQ-DI 1.6 (1.1, 2.0) *Placebo* 85% female, median (inter-quartile range) age 54.0 years (46.0, 64.0), median (inter-quartile	[redacted]	prednisone ≤ 10 mg/day or NSAIDs permitted if at a stable dose for ≥ 2 weeks prior to study	and 24 Fatigue score at weeks 14 and 24 DAS response according to EULAR (DAS28≤5.1 and improvement from baseline >1.2 DAS28 remission (DAS28<2.6). Fatigue was measured using the Functional Assessment of Chronic Illness Therapy-Fatigue (FASCIT-F) questionnaire. Swollen joint count Tender joint count Patient assessment of pain Patient global assessment of disease activity Physician global assessment	

Study	Design and clinical trial identification codes[a]	Participants	Interventions	Concomitant medication	Outcomes	Follow up (weeks)
		range) disease duration 9.8 years (4.9, 17.6), 73% positive rheumatoid factor, concomitant glucocorticoid therapy not reported, median (inter-quartile range) DAS28 6.3 (5.5, 7.1), median (inter-quartile range) HAQ-DI 1.8 (1.3, 2.1) Prior mean number of DMARDs not specified, mean duration not specified (≥ 3 months) for both treatment arms			of disease activity C-reactive protein concentration Serum samples were taken at baseline and week 24 for assaying of presence of antibodies to golimumab. Safety was evaluated using a general question to every patients about the number, type and severity of adverse events (coded according to MedDRA).	

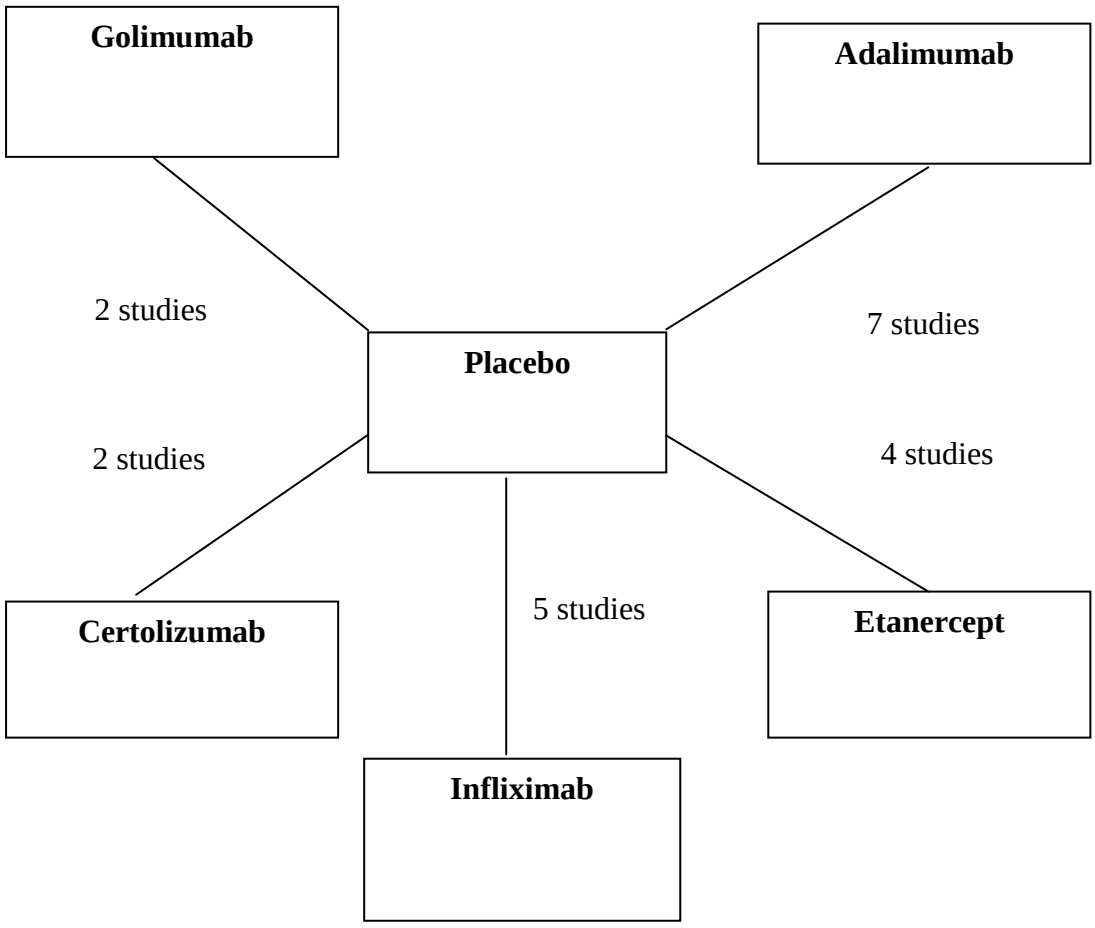
[a] *Clinical trial identification codes provided as supplementary information by manufacturer in response to request by ERG*

[b] *Intention-to-treat population provided for all treatment arms, including those not analysed in systematic review (table annotation as stated in MS)*

Limited data from the reported 52 week findings of the GO-FORWARD trial[28] were discussed in brief on page 107 of the MS.

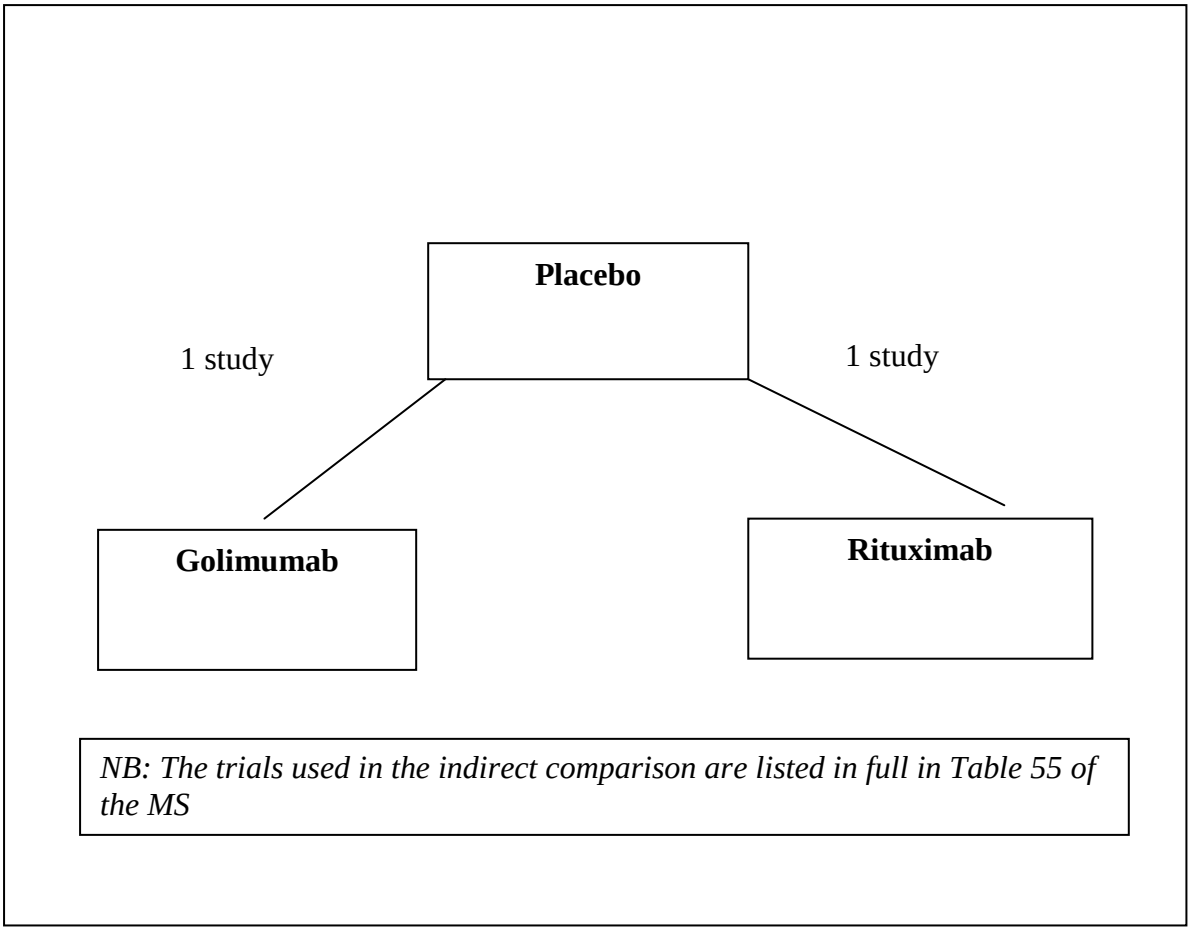
All included studies compared intervention with placebo. No head-to-head trials were available for the analysis of the efficacy of golimumab versus other comparator drugs. Therefore, a mixed treatment comparison (MTC) was undertaken for the DMARD experienced population and an indirect comparison was undertaken for the TNF-α inhibitor experienced population to estimate the effects of golimumab and comparator drugs. A table of the 20 trials used to conduct the MTC for the DMARD experienced population was presented as Table 54 on page 77 of the MS and summarised in the network diagram below. These included 2 trials for golimumab versus placebo (GO-FORWARD, Kay *et al*., 2008), 7 trials for adalimumab versus placebo, 2 trials of certolizumab versus placebo, 4 trials of etanercept versus placebo, and 5 trials of infliximab versus placebo (see Figure 1). Table 55 on page 78 of the MS outlined the 2 trials used to conduct the indirect comparison analyses of golimumab (GO-AFTER) versus rituximab (REFLEX) in the TNF-α inhibitor experienced population. The relevant network diagram is presented in Figure 2.

Figure 1: Network diagram for MTC (DMARD experienced population)



NB: The trials used in the MTC are listed in full in Table 54 of the MS

Figure 2: Network diagram for indirect comparison (TNF-α experienced inhibitor)



The MS stated (pages 64-65) that only drugs that were licensed for the specific patient population and had received NICE Technology Appraisal guidance as of the submission date were included in the meta-analyses. Therefore, rituximab, abatacept and tocilizumab studies were excluded from the DMARD experienced meta-analyses and trials for abatacept and tocilizumab were excluded from the TNF-α inhibitor experienced population analyses.

Seventy four trials were excluded from the review and were presented with reason(s) for exclusion in Table 176 (page 186) of the MS. Of these, only two were golimumab trials (Emery *et al.*, 2009; Kremer *et al.*, 2010).[29,30] It was stated in Table 176 that the GO-BEFORE study by Emery *et al.* (2009) (protocol number C0524T05, clinicaltrials.gov identifier NCT00264537) was excluded as only 50-60% of patients were DMARD-experienced. The trial reported by Kremer *et al.* (2010) (protocol number C0524T12, clinicaltrials.gov identifier NCT00361335) was excluded as non-approved dosing was used and as it was unclear what proportion of patients were TNF-α inhibitor experienced.

Inconsistency was noted in the MS in terms of the description of the certolizumab trial FAST4WARD which is included in table 16 but appears to be excluded in the remainder of the MS. A published report of the FAST4WARD study[31] reported that patients were randomised to either certolizumab 400 mg or placebo every 4 weeks and therefore no loaded dose was used. However, the SPC for certolizumab[32] stated that the recommended starting dose for adult patients with rheumatoid arthritis was 400 mg at weeks 0, 2 and 4, with a maintenance dose of 200 mg every 2 weeks. As described in Section 4.1.2., studies in which the drug of interest is not administered at the approved dose were excluded and it may be that the FAST4WARD study had been excluded by the manufacturer on this basis.

The flow diagram relating to the clinical effectiveness literature searches did not conform exactly to the PRISMA statement flow diagram (http://www.prisma-statement.org/), as the following details were missing: i) number of records identified though database searching, ii) number of additional records identified through other sources, iii) number of records after duplicates removed. However, the presented flow diagram was considered to be an adequate representation of the study selection process.

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

The ERG believes that all relevant studies were included in the MS. Repeat searches using the manufacturer's search terms were undertaken (although the ERG was not able to sift through the search results due to time constraints). However, the ERG sought advice from their clinical advisors who did not identify any additional relevant studies that should be included in the MS.

Evidence from completed studies was presented in the clinical effectiveness section. The manufacturer did not specify whether any searches for unpublished evidence were undertaken in the systematic review. The manufacturer provided details of ongoing and recruiting trials for the efficacy and safety of golimumab in combination with methotrexate in patients with rheumatoid arthritis after failure of previous antirheumatic therapy in response to a request by the ERG. Three studies were described as in the recruitment stage with estimated completion dates in 2012 (Clinicaltrials.gov identifier codes NCT01004432, NCT00973479 and NCT00975130). Study NCT00727987 was described as being ongoing with completion anticipated in 2012. The manufacturer provided additional data in abstract form for the open label extensions for 5 years of the GO-FORWARD (NCT00264550) and GO-AFTER (NCT00299546) studies (completion date 2012). Abstract data were also provided for study NCT00771251 that is ongoing with an estimated completion data of 2011. These data are described in this report. The ERG group searched ClinicalTrials.gov (http://clinicaltrials.gov) and were unable to identify any additional relevant studies findings of the clinical effectiveness or safety of golimumab in patients with rheumatoid arthritis that had reported

findings. However, it is possible that unpublished evidence relating to comparator drugs may have been missed in the identification of evidence by the manufacturer.

No non-RCT evidence was included in the systematic review. As stated earlier in this report, the ERG considers non-RCTs, including expert opinion, to be a valid and important source of evidence for the evaluation of adverse events. The manufacturer stated that no non-RCTs or observational studies of the use of golimumab in patients with rheumatoid arthritis were available at the time of submission. However, limitations in the searches conducted to identify evidence relating to adverse events (as outlined in Section 4.1.1 of this report) are such that it may be possible that useful data of the adverse events associated with golimumab and comparator drugs may have been omitted from the MS.

4.5 Description and critique of manufacturers approach to validity assessment

A formal appraisal of the validity of the golimumab RCTs was clearly presented in the MS:

- GO-FORWARD (Keystone *et al.*, 2009) (Table 177)
- Kay *et al.*, 2008 (Table 179, Table 180)
- GO-AFTER (Smolen *et al.*, 2009) (Table 178)

All the criteria listed under Section 5.4.1 (page 58) of the MS (as specified in the NICE STA Specification for manufacturer/sponsor submission of evidence)[33] were addressed in the quality assessment findings.

The ERG acknowledges that whilst the items listed in the validity assessment tool used in the MS were appropriate, several quality criteria relevant to the critical appraisal of RCTs were not taken into account. Example criteria for the assessment of the risk of bias in RCTs were described by the NHS Centre for Reviews and Dissemination (CRD).[34] The tabulated quality assessment findings for comparator RCTs (Table 180) (in which the Kay *et al.* (2008) golimumab trial was also presented) referred to the NHS Centre for Reviews and Dissemination guidance for undertaking quality assessment.[34] However, the following quality criteria recommended by the CRD were not represented in the tabulated quality assessment findings: i) Specification of eligibility criteria; ii) identification of any co-interventions with the potential to impact upon outcomes; iii) assessment of treatment compliance; iv) assessment of success of blinding; v) presence of reported point estimates and measures of variability for the primary outcome measure.

Whilst it was described in the MS (page 190) that data abstraction was undertaken by a single reviewer and checked by a second reviewer, it was not explicitly stated whether critical appraisal was conducted by a single reviewer or using consensus of multiple reviewers.

The ERG checked the quality assessment findings against the original study publications and any additional points are discussed within this Section.

The completed validity assessment tool for the three trials, as reported in the MS, is reproduced below (Tables 4,5, and 6). Responses that the ERG have highlighted and discussed in further detail are marked with an asterisk (*).

Table 4: Quality assessment of GO-FORWARD study (Keystone *et al.*, 2009) (as presented and graded by manufacturer in Table 177 of MS)

Quality criterion	Comments from MS	Grade (yes/no/not clear/N/A)
Was randomisation carried out appropriately?	Sample size considered to be adequate (with numbers of participants per group slightly exceeding numbers pre-specified in power calculation) Number of subjects randomised was stated Subjects were randomized in 3:3:2:2 ratios to 1 of 4 treatment groups: placebo plus MTX (group 1), golimumab 100 mg plus placebo (group 2), golimumab 50 mg plus methotrexate (group 3), and golimumab 100 mg plus methotrexate (group 4). Relatively even treatment balance within sites was ensured, within baseline MTX usage and within the study overall, using an adaptive stratified randomisation design.	Yes
Was the concealment of treatment allocation adequate?	Treatment allocation was made using a centralised telephone interactive voice response system	Yes
Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	Demographic characteristics of subjects at baseline were generally well balanced across treatment groups: • majority of subjects were women • majority of subjects were Caucasian (75.9%) (data not reported in original study publication and therefore could not be corroborated by ERG) • mean age was [redacted] years • mean duration of disease (8.62 years)	Yes*
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	Randomisation files containing treatment assignments for individual subjects were maintained in limited-access directories within the electronic data filing system at the central randomisation centre. Personnel having contact with study sites, including the medical monitor, remained blinded to the treatment assignment of individual subjects until the 24-week database lock. All site monitors, site personnel, and subjects remained blinded to treatment assignment until the last subject completes Week 52 evaluations and the database is locked. (Not all data reported in original study publication and therefore could not be corroborated by ERG).	Yes

Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	> 90% patients were part of follow-up assessment	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No such reference in the publication	No
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes

Table 5: Quality assessment of Kay *et al.* (2008) study (data as presented in Table 179 of MS and graded by manufacturer)

Quality criterion	Comments from MS	Grade (yes/no/not clear/N/A)
Was randomisation carried out appropriately?	Double blind RCT Power calculation Number randomised was stated	Yes
Was the concealment of treatment allocation adequate?	As this study was a double blind study presumably both subjects and investigators were blind to treatment allocation. Patients in the golimumab groups remained blinded to their dose assignment through the end of the study.	Yes*
Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	The population was somewhat heterogeneous because of the small number of patients in each treatment group, but none of the baseline characteristics of the combined Golimumab groups was significantly different from those of the placebo group ($P > 0.05$).	Yes*
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	Study was double blind so presumably both participants and outcome assessors blind to treatment allocation.	Yes
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	Through week 52, a significantly greater proportion of patients in the placebo/infliximab plus MTX group (40.0%) discontinued treatment compared with the proportion of patients in the combined golimumab plus MTX groups (21.2%) ($P <0.0217$).	Yes
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No such reference in the study publications	No
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes ITT and missing values imputed using LOCF.	Yes

Table 6: Quality assessment of GO-AFTER study (Smolen *et al.*, 2009) (data as presented in Table 179 of MS and graded by manufacturer)

Quality criterion	Comments from MS	Grade (yes/no/not clear/N/A)
Was randomisation carried out appropriately?	Adequate sample size Number randomised was stated Subjects were randomised in a 1:1.1 ratio to 1 of 3 treatment groups: placebo, golimumab 50 mg, and golimumab 100 mg. Relatively even treatment balance within sites was ensured, within baseline MTX usage and within the study overall, using an adaptive stratified randomisation design.	Yes
Was the concealment of treatment allocation adequate?	Randomised treatment allocation was done using a centralised telephone interactive voice response system	Yes
Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	Demographic characteristics of subjects at baseline were generally well balanced across treatment groups: • majority of subjects were women • most subjects were Caucasian (data not reported in original study publication and therefore could not be corroborated by ERG) • mean age was [redacted] years • mean duration of disease (12.40 years) (as reported in MS)	Yes*
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	Randomisation files containing treatment assignments for individual subjects were maintained in limited-access directories within the electronic data filing system at the central randomisation centre. Both patients and investigators were masked to treatment assignment. Personnel having contact with study sites, including the medical monitor, remained blinded to the treatment assignment of individual subjects until the 24-week database lock. Furthermore, all site monitors, site personnel, and subjects remained blinded to treatment assignment until the last subject completes Week 52 evaluations and the database is locked.	Yes
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	> 80% patients were part of follow-up assessment	No

Is there any evidence to suggest that the authors measured more outcomes than they reported?	No such reference in the publication	No
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes

Was randomisation carried out appropriately?

In the study by Keystone *et al.* (2009), randomisation was judged to have been performed appropriately. Randomisation was stratified by study site. Sample sizes were adequate, with the numbers of participants achieved per group slightly exceeding the numbers pre-specified in the power calculation.

For Kay *et al.* (2008), patients were described as being randomly allocated to 1 of 5 treatments in approximately equal proportions, with randomisation stratified by study site. Sample sizes were small. Whilst 35 patients were estimated by means of a power calculation to be required in each treatment group, 35 patients were achieved for 2 of the groups, and only 34 patients were included in the remaining 3 groups.

Smolen *et al.* (2009) states that participants were randomised in a 1:1:1 ratio to 1 of 3 treatment groups. Randomisation was stratified by study site and baseline methotrexate use. Numbers of participants per treatment group exceeded the 140 patients in each group required by the power calculation.

Was the concealment of treatment allocation adequate?

For Keystone *et al*. (2009) and Smolen *et al.* (2009), concealment of treatment allocation was adequate, being based on assignment using a centralised telephone interactive voice response system.

In the table above, the manufacturer appears to confuse the issues of concealment of treatment allocation with blinding in the description of Kay *et al*. (2008). No mention was made in the original study publication of the method of randomisation or the concealment of treatment allocation.

Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?

In Keystone *et al.* (2009), the population groups could be considered to be well-balanced for a number of factors, including gender, age, number of swollen joints and HAQ-DI. The methotrexate dose at baseline for all groups was a median of 15.0 (IQR) (15.0 to 20.0) mg/week. However, for some factors a number of imbalances were noted by the ERG, including disease duration being slightly shorter for group 3 (golimumab 50 mg/week plus methotrexate) (median 4.50 (IQR 2.1 to 9.7) years vs. median values of 5.9 to 6.7 years for other groups). However, the number of tender joints was slightly higher for group 3

(golimumab 50 mg/week plus methotrexate) (median 26.0, (IQR 16.0 to 39.0) vs. 21.0 to 23.0 for other treatment groups).

The manufacturer noted that, for Kay *et al.* (2008), the study population displayed heterogeneity, potentially as a result of the small sample size in each treatment group, but that none of the baseline characteristics of the combined golimumab groups were observed to be significantly different from those of the placebo group ($P > 0.05$). However, between the individual treatment groups, there was considerable variation in terms of gender (range of median values 67.6 to 85.7% female), age (range of median values 48.0 to 57.5 years), disease duration (range of median values 5.6 to 9.0 years), number of swollen joints (range of median values 13 to 20) and number of tender joints (range of median values 22 to 32).

The MS stated that baseline characteristics were generally well-balanced between treatment groups in Smolen *et al.* (2009). However, the ERG noted some slight imbalances between groups, including gender (74 to 85% female), and disease duration (range of median values 8.7 to 9.8 years).

The manufacturer did not attempt to incorporate any prognostic factors into a meta-regression to explicitly control for imbalances in baseline characteristics.

Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?

The MS stated that all patients and relevant study personnel were blinded to treatment group appropriately in the study by Keystone *et al.* (2009). However, not all data was presented in the original study report and therefore could not be corroborated by the ERG.

The manufacturer assumed that since the Kay *et al*. (2008) study was described as being double blind, both participants and outcome assessors were blinded to treatment allocation. The ERG also notes that in the original study publication it was stated that patients receiving golimumab every 4 weeks received placebo injections on alternate visits to maintain blinding and that patients remained blinded to their dose assignment to the end of the trial. Therefore, whilst it was not explicitly stated in the publication whether administering professionals and/or outcome assessors were blinded, it appears that patients were blinded to their treatment group.

It was noted in the MS that in the Smolen *et al*. (2009) trial, patients and investigators were blinded to treatment assignment

Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?
No drop-out imbalances were reported for Keystone *et al.* (2009).

For Kay *et al.* (2008), the manufacturer observed that a significantly greater proportion of patients in the placebo/infliximab plus methotrexate group (40.0%) discontinued treatment vs. the proportion of patients in the combined golimumab plus methotrexate groups (21.2%) (P <0.0217) (as stated in MS) through week 52.

In the Smolen *et al.* (2009) trial, all randomised subjects were followed up in analyses.

Is there any evidence to suggest that the authors measured more outcomes than they reported?
No such evidence was described for Keystone *et al*. (2009), Kay *et al.* (2008) or Smolen *et al.* (2009) by the manufacturer in their quality assessment findings.

Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?
The primary efficacy analyses were undertaken according to the intention-to-treat principle in Keystone *et al.* (2009), with all randomised patients being analysed according to their assigned treatment group. ACR20, 50 and 70 analyses were performed using intention-to-treat analysis. It was not explicitly stated in the study publication or MS how safety data were handled.

For Kay *et al.* (2008), the MS stated that ITT analysis was used. ITT analysis was used for efficacy data at week 16 in the original publication. A last observation carried forward process was used to handle patients who did not return for evaluation or for whom there was insufficient data to calculate the ACR20 response.

Intention-to-treat analysis was utilised for efficacy and safety outcomes in the trial reported by Smolen *et al.* (2009).

Additional quality criteria recommended by CRD not represented in the MS and reported by the ERG

Specification of eligibility criteria

Eligibility criteria were clearly stated in the reports by Keystone *et al.* (2009), Kay *et al.* (2008) and Smolen *et al.* (2009).

Identification of any co-interventions with the potential to impact upon outcomes

A number of co-interventions with the potential to affect outcomes were taken into consideration in the study by Keystone *et al.* (2009), whereby patients who were taking NSAIDs or corticosteroids had to have been receiving a stable dose for ≥ 2 weeks before the first dose of study agent. Patients were excluded if they had any previous use of any TNF-α inhibitor, rituximab, natalizumab or cytotoxic agents. Patients also should not have been receiving DMARDs other than methotrexate or IV, intramuscular or intra-articular corticosteroids within 4 weeks before the first dose of study agent. Initiation or escalation of various RA medications constituted treatment failure.

Co-interventions were also identified in Kay *et al*. (2008). Patients were to have been treated with methotrexate at a stable dose of at least 10 mg/week for ≥ 3 months and at a stable dose for ≥ 4 weeks before beginning study medication. The dose of oral corticosteroids was limited and, as for NSAIDs, the dosage was to have been stable for the 4 weeks preceding study entry and must have been held stable throughout the study. Initiation of corticosteroids or NSAIDs during the study was not permitted.

A number of co-interventions were taken into account in the trail by Smolen *et al.* (2009). Patients were ineligible if they had ever received rituximab, natalizumab, had taken anakinra less than 4 weeks or alefacept or eflizumab less than 3 months before commencing study medication or had ever received cytotoxic drugs. Patients receiving concomitant DMARDs were required to have tolerated the dose for at least 12 weeks and the dose must have been stable for at least 4 weeks before the first receipt of study drug. Doses had to be maintained throughout the trial. Oral corticosteroids and NSAIDs were permitted if the doses had been stable for at least 2 weeks before commencement of the study medication.

Assessment of treatment compliance

Compliance was not explicitly discussed in the original study reports by Keystone *et al*. (2009), Kay *et al.* (2008) and Smolen *et al*. (2009).

Assessment of success of blinding

It was not stated in the original publications whether the trials reported by Keystone *et al.* (2009), Kay *et al.* (2008) or Smolen *et al*. (2009) included an assessment of whether the blinding process had been successful.

Presence of reported point estimates and measures of variability for the primary outcome measure

Effect estimates and measures of variability were clearly reported for the primary outcome measures in Keystone *et al*. (2009), Kay *et al.* (2008) and Smolen *et al.* (2009).

4.6 Description and critique of manufacturers outcome selection

The manufacturer listed the outcomes from the final scope that were addressed in this submission (Table 11, pages 26-27) as follows:

- Disease activity
- Physical function
- Joint damage
- Pain
- Mortality
- Fatigue
- Radiological progression
- Adverse effects of treatment
- Health related quality of life

The outcomes listed above were considered by the ERG and clinical advisors to be relevant and appropriate.

The following outcomes did not appear to have been addressed in the original MS:

- Joint damage
- Pain
- Mortality
- Fatigue
- Radiological progression
- Extra-articular manifestations of disease
- Health-related quality of life

Disease activity and physical function were considered to be addressed in the MS. As noted in Section 3, data on tender and swollen joints and pain (patient assessment using VAS) were presented in the tables of baseline characteristics but the impact of treatment on these outcomes was not explicitly assessed (although tender and swollen joints would be indirectly covered within the ACR responses). For the outcome of mortality, the manufacturer provided supplementary data in response to a request by the ERG. The manufacturer stated that fatigue and FACIT-F were included within the search criteria but that this parameter was 'under-reported' within identified RCTs. It was unclear from this statement whether the outcome of fatigue was available or not within the identified RCTs. It was confirmed by the manufacturer that radiological progression data were not available from the 'pivotal trials.' The clinical advisors to the ERG noted a lack of joint X-ray data in the included trials that may indicate any potential benefit golimumab may have in slowing/stopping joint erosion and disease progression. They commented that other TNF-α inhibitors, such as adalimumab and etanercept have data to show this benefit.[3] It should be noted that the FDA stated that, for GO-FORWARD, the van der Heijde Modified Sharp score (measure of radiological progression) was pre-specified as a primary outcome but was not submitted to the FDA.[35,36]

[redacted]

Extra-articular manifestations of disease were specified as an outcome for inclusion in the final scope issued by NICE. This outcome was not included in the MS, with the justification that extra-articular manifestations of disease were not routinely reported in RCTs. The clinical advisors to the ERG were satisfied with this reasoning and did not consider the omission of this outcome to have any significant implications.

Furthermore, not all outcomes listed in the decision problem were present in the tabulated inclusion/exclusion criteria in the MS (page 29).

The ERG and clinical advisors have noted that ACR70 was omitted from the original MS. ACR70 was considered by the ERG and the clinical advisors to be an important disease activity outcome for inclusion in the submission. ACR70 data were reported in the study publications for the GO-FORWARD, Kay *et al.* (2008) and GO-AFTER trials. In response to a request by the ERG, supplementary meta-analysed ACR70 data were presented for

golimumab and comparators. Supplementary mixed treatment comparisons and indirect comparisons were also presented for ACR70 responses.

Additional efficacy outcomes reported in the original study publications of the GO-FORWARD, Kay *et al.* (2008) and GO-AFTER trials but not utilised or presented fully in the MS include the following:

Kay *et al.* (2008): ACR-N at week 16 (limited data presented in clarification responses from manufacturer), DAS28 at week 16 (limited data presented in clarification responses from manufacturer), CRP level though week 52

GO-FORWARD (Keystone *et al.,* 2009): Improvement from baseline in HAQ-DI at weeks 14 and 24 (some detail on page 104 of MS), ACR90 at weeks 14 and 24, ACR-N at weeks 14 and 24, DAS28 remission and sustained remission at weeks 14 and 24 (some detail presented on page 104 of MS), EULAR responders at weeks 14 and 24.

GO-AFTER (Smolen *et al.,* 2009): ACR90 at weeks 14 and 24 (some detail presented on page 113 of MS), DAS 28 remission and score at weeks 14 and 24, HAQ-DI at weeks 14 and 24 (some detail presented on page 113 of MS), and FACIT-F score at weeks 14 and 24.

4.7 Describe and critique the statistical approach used

Kay et al. (2008)

The study by Kay *et al.* (2008) was a 5-arm phase II dose-ranging study in which patients were randomly allocated to receive subcutaneous injections of placebo plus methotrexate or 50 mg or 100 mg of golimumab every 2 or 4 weeks plus methotrexate. Upon reaching week 20, placebo group subjects began open-label treatment with IV infusions of infliximab at 3 mg/kg with induction at weeks 20, 22 and 28 and maintenance therapy subsequently every 8 weeks to week 44.

The primary endpoint selected was the proportion of patients achieving an ACR20 response at week 16. The study was powered to detect a difference in this primary endpoint when the combined golimumab groups and at least one of the individual dose groups were compared vs. placebo (rather than for the comparison of individual golimumab dose groups vs. placebo). The study authors performed simulations to evaluate the power of the Chi-squared test to detect a significant effect for the combined golimumab plus methotrexate groups compared with the placebo plus methotrexate group (α=0.05, 2-sided test). According to the assumption

that 60% of golimumab-treated subjects and 25% of placebo-treated subjects reached the primary endpoint, 35 patients were required in each treatment group to achieve >90% power.

In the primary analyses, a process was used whereby the last observation was carried forward for subjects who did not return for assessment and for whom there was not sufficient data to evaluate ACR20 response. Those subjects who had begun treatment with oral corticosteroids or disease-modifying antirheumatic drugs (other than methotrexate but including biologics), increased methotrexate or oral corticosteroid dosages above levels at baseline, or discontinued the study drug due to lack of efficacy before week 16 were classed as not having reached the primary endpoint at week 16. In the primary analyses, a 2-sided Chi-squared test was employed comparing the combined golimumab plus methotrexate groups vs the placebo plus methotrexate group. If the study authors found a statistically significant difference (α=0.05) in favour of the combined golimumab plus methotrexate groups, pairwise comparisons between each individual golimumab dose group and the placebo group were conducted. It was stated that actual observations without imputation were used in determining ACR20, ACR50 and ACR70 responses through week 52. No statistical analyses were used to assess the occurrence of adverse events. The placebo and 50 mg 4 weekly golimumab groups each contained 35 patients. Of these, 29 and 31 individuals completed 16 weeks of therapy in the placebo and golimumab groups respectively. Since the numbers of subjects in each treatment arm were small, the capacity of this study to detect meaningful differences in terms of adverse events was limited.

The manufacturer clarified that all timepoints for Kay *et al.* (2008) were extracted for week 16, since this was the primary endpoint and the latest timepoint for the ACR response values reported in the publication. Therefore, the manufacturer reported that there was no requirement to take into account the crossover in the placebo arm to infliximab treatment at week 20.

Keystone et al. (2009)

Keystone *et al.* (2009) conducted the 4-arm phase III GO-FORWARD study in which patients were randomly assigned to receive placebo subcutaneous injections plus methotrexate capsules (group 1, n=133) or a range of subcutaneously injected golimumab doses, including 100 mg golimumab every 4 weeks plus placebo capsules (group 2, n=133), 50 mg golimumab every 4 weeks plus methotrexate capsules (group 3, n=89) and 100 mg golimumab every 4 weeks plus methotrexate capsules (group 4, n=89). The co-primary endpoints used were the proportion of patients reaching an ACR20 response at week 14 and the change from baseline in the HAQ-DI score at week 24. The study was powered for these two co-primary endpoints.

Under the assumption that 55% or above of patients in group 3 and 4 and 35% of patients in group 1 would reach an ACR20 response, it was estimated that a sample size of 120 subjects in group 1 and 80 subjects in groups 3 and 4 would be required to achieve >90% power (2-sided Chi-squared test, α=0.05). When assuming that 55% of subjects in group 2 and 35% of subjects in group 1 would have an ACR20 response, it was calculated that a sample size of 120 patients in both groups 1 and 2 would be necessary to achieve >85% power (2-sided Chi-squared test, α=0.05). Such a sample size would also enable >90% power for the detection of a difference in the change in baseline HAQ-DI score between treatment groups (2-sided t test on van der Waerden normal scores, α=0.05), whilst assuming an improvement in HAQ-DI score from baseline of -0.21 for group 1, -0.47 for group 3 and -0.39 for group 4.

At week 16, those patients in groups 1, 2 or 3 with below a 20% improvement from baseline in both swollen and tender joint counts underwent a double-blind adjustment in study medication, described as early escape. Group 1 early escape patients subsequently received active 50 mg golimumab every 4 weeks whilst continuing the stable dose of methotrexate. The placebo capsules taken by group 2 early escape subjects were replaced by active methotrexate capsules at their pre-screening dose in addition to continuation of receipt of 100mg golimumab every 4 weeks. Group 3 early escape patients received an increased dose of 100 mg golimumab and continuation of active methotrexate capsules. Patients in group 4 received no medication adjustments.

Statistical analyses for the two co-primary endpoints were based on a 2-sided Chi-squared test for the analysis of ACR20 data and a 2-sided analysis of variance on the van der Waerden scores for the analysis of the HAQ-DI findings. Both were assessed using a significance level of α=0.05. For the primary efficacy analyses, data from all randomised subjects were analysed based on allocated treatment group (intention-to-treat). The authors employed a hierarchical approach to the analysis of the primary endpoints. If a significant difference between combined groups 3 and 4 vs. group 1 was observed, pairwise comparisons of groups 3 and 1 and groups 4 and 1 were conducted according to the same statistical methods. Should a statistically significant difference be observed in at least one of these pairwise comparisons, groups 2 and 1 would be compared using the same method. The co-primary endpoints were also analysed in subgroups by geographical region. Statistical analyses of adverse events data were not conducted.

Subjects who did not have all components of ACR and DAS were classed as non-responders. Subjects were additionally categorised as non-responders based on any of the following treatment failure criteria: i) commenced disease-modifying antirheumatic drugs, systematic

immunosuppressive agents or biologic RA therapies; ii) increased the methotrexate dose above baseline level; iii) commenced oral corticosteroid treatment for RA, increased oral corticosteroid dosage above baseline level or received IV or intramuscular administration of corticosteroids for RA; iv) discontinued study agent injections based on lack of efficacy. For subjects who discontinued the study agent for reasons other than lack of efficacy who returned for clinical assessments, actual observed data were used, but subjects were categorised as non-responders in the event that any of the above treatment failure criteria were met. All patients were to return for continued safety and selected efficacy assessments for 4 months following the discontinuation of the study drug for any reason.

Week 16 efficacy data from those subjects in groups 1, 2 and 3 who began early escape therapy were carried forward to week 24. Since no treatment adjustment options were applied to patients in group 4, actual observed data were used for patients using analysis rules and methods as described above.

Smolen et al. (2009)

Patients in the GO-AFTER study reported by Smolen *et al.* (2009) were randomly assigned to receive subcutaneous placebo injections (n=155), 50 mg golimumab (n=153) or 100 mg golimumab (n=153) every 4 weeks, whilst continuing stable dosages of methotrexate, sulfasalazine, hydroxychloroquine, oral corticosteroids and NSAIDs. The study was placebo-controlled to 24 weeks.

The primary endpoint was defined as ACR20 response at week 16. It was calculated that a sample size of 140 subjects in each treatment group would provide >90% power at a statistical significance level of 0.05, based on the assumption that 50% of subjects received methotrexate at baseline and ACR20 response was reached in 30% of placebo group subjects (regardless of methotrexate use), 45% of 50 mg golimumab group that used methotrexate, 40% of the 50 mg golimumab group who did not use methotrexate, 55% of the 100 mg golimumab group that used methotrexate and 50% of the 100 mg golimumab group that did not use methotrexate.

The primary endpoint was assessed using a hierarchical approach. In the event that a Cochran-Mantel-Haenszel test, stratified by methotrexate use, demonstrated a significant difference between ACR20 response in the combined golimumab groups (50 mg and 100 mg) vs placebo, pairwise comparisons were conducted for 50 mg golimumab vs. placebo and for 100 mg golimumab vs. placebo. Achievement of the primary endpoint required that the

proportions of subjects reaching ACR20 response on combined golimumab and 50 mg or 100 mg golimumab should be significantly greater than for placebo group subjects.

Secondary endpoints with discrete data were also analysed using a Cochran-Mantel-Haenszel test stratified according to methotrexate use. Secondary endpoints based on continuous data were analysed using ANOVA from the van der Waerden normal scores. Subgroup analyses of DMARD use at baseline, and number of previous TNF-α inhibitors and reasons for their discontinuation were conducted for comparison between the combined golimumab group and the placebo group.

Subjects missing all components of ACR20 or DAS response criteria were classed as non-responders. Furthermore, subjects were considered to have failed to reach the primary endpoint if they had commenced treatment with a new DMARD, systemic immunosuppressive or biologic derived therapy for RA, increased methotrexate, sulfasalazine or hydroxyquinoline dose above the baseline level for RA, commenced or increased the dosage of corticosteroid therapy, or discontinued the study agent on the basis of unsatisfactory treatment effect.

Subjects were included in analyses if the study drug was discontinued for reasons other than lack of efficacy and returned for evaluation, with patients being classed as non-responders in the presence of any of the treatment failure criteria previously outlined. All subjects were to return for safety and efficacy assessments for 4 months after discontinuation of golimumab.

All efficacy data were analysed according to the intention-to-treat principle. Safety was evaluated according to the study drug received and patients who were randomised but never treated were not included. No statistical analysis was conducted for the assessment of safety outcomes. Patients who were randomised but never treated were not included.

At week 16, those subjects who had not reached 20% improvement in terms of tender and swollen joint counts became eligible for double-blinded rescue therapy, whereby treatment was changed from placebo to 50 mg golimumab or from 50 mg to 100 mg golimumab. Subjects in the 100 mg golimumab group who were eligible for rescue therapy continued to receive this dose. Change of treatment was only permitted at this timepoint. For subjects receiving rescue therapy, efficacy data at week 16 were carried forward for analysis at week 24, with the manufacturer noting that this approach should have ensured that results were not biased by increased dosages received by patients.

Clarification was requested from the manufacturer as to whether the analyses from the GO-AFTER study were presented in original form or in re-analysed form following the exclusion of patients from a single trial site in the efficacy analyses (as referred to in the European Medicines Agency document entitled: 'Simponi: procedural steps taken and scientific information after the authorisation.' The manufacturer confirmed that the GO-AFTER data within the MS were taken directly from the clinical study report and did not exclude the 16 patients as described within the EMA document. However, the manufacturer stated that the reanalysis of efficacy data did not alter the overall key efficacy parameter conclusions but slightly changed the significance of some secondary endpoints.

4.8 Summary statement

Whilst limitations were noted in terms of the design and conduct of the clinical effectiveness search strategies, the submission appears to contain all relevant RCTs to the best knowledge of the ERG and clinical advisors to the ERG. The outcomes included in the assessment were relevant, although not all were fully addressed as discussed in Sections 3 and 4. Statistical methods were well described in the MS and the original study publications. The validity assessment criteria used to assess the included studies were considered to be satisfactory and were clearly presented, but some relevant criteria and details of the conduct of the validity assessment process were missing. The submitted evidence partially reflected the decision problem defined in the MS. However, not all comparators considered by the ERG to be relevant to the decision problem were included (as discussed in Section 3) and some useful outcomes were not addressed (as discussed in Sections 3 and 4).

4.9 Summary of submitted evidence

4.9.1 Summary of results

This section presents the main clinical efficacy and safety evidence presented in the MS and supplementary information submitted by the manufacturer.

Efficacy data

ACR20, ACR50 and ACR70 responses

A tabulated summary of the ACR20, ACR50 and ACR70 efficacy findings in terms of relative risk (as reported by the manufacturer and constructed by the ERG) is presented in this section. Relative risks for ACR response in treatment group subjects versus placebo subjects were estimated by meta-analysis. Meta-analyses of ACR70 responses were provided by the manufacturer following clarification requests from the ERG.

Etanercept meta-analyses with the exclusion of the TEMPO study were provided by the manufacturer in response to the ERG's request as it was noted in the MS that the placebo arm response was higher in the TEMPO trial vs. other studies. The manufacturer has not formally assessed whether there is correlation between treatment effect and baseline risk.

The MS noted (page 64) that, whilst most studies used concomitant administration of biologic with methotrexate, no concomitant methotrexate was permitted in 4 studies and 3 studies included a monotherapy treatment arm. Meta-analyses were also conducted to determine the effect of excluding the minority of monotherapy studies and treatment arms. The relative risk of the monotherapy group versus the original group (all studies) was calculated.

DMARD experienced population

ACR20 response at 24 weeks (DMARD experienced population)

A total of 25 studies were included in the analysis of the ACR20 response in the DMARD experienced population, of which 2 studies related to the use of golimumab (GO-FORWARD (Keystone *et al.*, 2009; Kay *et al.*, 2008).

Table 7: ACR20 response at 24 weeks (DMARD experienced population)

Active treatment	Relative risk compared with placebo generated by fixed effects meta-analysis (95%CI)	Relative risk compared with placebo generated by random effects meta-analysis (95%CI)	Estimated heterogeneity within the random effects model.
Adalimumab	1.98 (1.75, 2.24)	2.22 (1.67, 2.95)	I^2=74.9%, Chi-squared P value <0.001)
Adalimumab (excluding monotherapy arms) [a]	1.86 (1.63, 2.13)	2.05 (1.46, 2.87)	I^2=79.6%, Chi-squared P value= 0.002)
Certolizumab	4.99 (3.66, 6.78)	5.04 (3.38, 7.52)	I^2=34.2%, Chi-squared P value= 0.218)
Etanercept	1.40 (1.26, 1.55)	2.43 (0.97, 6.07)	I^2=95.1%, Chi-squared P value <0.001)
Etanercept (excluding monotherapy arms) (including TEMPO study)	1.35 (1.21, 1.51)	1.93 (0.88, 4.22)	I^2=92.0%, Chi-squared P value <0.001)
Etanercept (excluding TEMPO study) [b]	3.19 (2.33, 4.37)	3.20 (2.11, 4.87)	I^2=39.9%, Chi-squared P value =0.189)
Golimumab	1.96 (1.52, 2.53)	1.92 (1.47, 2.51)	I^2=8.3%, Chi-squared P value= 0.297)
Infliximab	2.06 (1.77, 2.40)	2.05 (1.43, 2.92)	I^2=72.7%, Chi-squared P value= 0.012)
Infliximab (excluding monotherapy arms)	2.06 (1.77, 2.4)	2.06 (1.43, 2.95)	I^2=73.3%, Chi-squared P value= 0.011)

[a] *The MS stated that the relative risk of the monotherapy group versus the original group (all studies) was calculated.*

[b] *Supplementary data provided by manufacturer in response to request by ERG*

ACR50 response at 24 weeks (DMARD experienced population)

The analysis of ACR50 response among DMARD experienced patients was based on 25 studies, of which 2 were of golimumab (GO-FORWARD (Keystone *et al.*, 2009; Kay *et al.*, 2008).

Table 8: ACR50 response at 24 weeks (DMARD experienced population)

Active treatment	Relative risk compared with placebo generated by fixed effects meta-analysis (95%CI)	Relative risk compared with placebo generated by random effects meta-analysis (95%CI)	Estimated heterogeneity within the random effects model
Adalimumab	3.35 (2.67, 4.20)	3.34 (2.55, 4.38)	I^2=20.9%, Chi-squared P value =0.277)
Adalimumab (excluding monotherapy arms)	3.37 (2.64, 4.31)	3.49 (2.40, 5.08)	I^2=48.0%, Chi-squared P value= 0.124)
Certolizumab	6.06 (3.87, 9.48)	6.32 (3.15, 12.66)	I^2=43.8%, Chi-squared P value= 0.182)
Etanercept	1.81 (1.49, 2.19)	2.98 (1.06, 834) (values as stated in MS)	I^2=82.8%, Chi-squared P value =0.003)
Etanercept (excluding monotherapy arms) (including TEMPO study)	No data presented	No data presented	No data presented
Etanercept (excluding TEMPO study) [a]	5.22 (3.04, 8.98)	5.29 (2.70, 1.40)) (values as stated in MS)	I^2=22.8%, Chi-squared P value =0.274)
Golimumab	2.90 (1.84, 4.58)	2.88 (1.83, 4.53)	I^2=0%, Chi-squared P value= 0.669
Infliximab	3.00 (2.30, 3.90)	3.06 (1.79, 5.23)	I^2=60.9%, Chi-squared P value= 0.053
Infliximab (excluding monotherapy arms)	3.01 (2.31, 3.92)	3.11 (1.80, 5.39)	I^2=62.7%, Chi-squared P value= 0.045)

[a] *Supplementary data provided by manufacturer in response to request by ERG*

ACR70 response at 24 weeks (DMARD experienced population)

Table 9: ACR70 response at 24 weeks (DMARD experienced population)

Active treatment	Relative risk compared with placebo generated by fixed effects meta-analysis (95%CI)	Relative risk compared with placebo generated by random effects meta-analysis (95%CI)	Estimated heterogeneity within the random effects model
Adalimumab	5.30 (3.56, 7.90)	4.98 (3.33, 7.44)	I^2=0%, Chi-squared P value =0.577)
Adalimumab (excluding monotherapy arms)	Not presented	Not presented	Not presented
Certolizumab	8.94 (4.23, 18.90)	8.24 (3.89, 17.44)	I^2=0%, Chi-squared P value= 0.326)
Etanercept	Not presented	Not presented	Not presented
Etanercept (excluding monotherapy arms) (including TEMPO study)	Not presented	Not presented	Not presented
Etanercept (excluding TEMPO study)	11.41 (3.19, 40.83)	11.45 (3.26, 40.20)	I^2=0%, Chi-squared P value =0.992)
Golimumab	3.75 (1.81, 7.77)	3.76 (1.82, 7.78)	I^2=0%, Chi-squared P value= 0.915)
Infliximab	3.19 (2.11, 4.83)	2.97 (1.97, 4.50)	I^2=0%, Chi-squared P value= 0.427)
Infliximab (excluding monotherapy arms)	Not presented	Not presented	Not presented

The meta-analysis findings indicated that the exclusion of the etanercept TEMPO trial resulted in raised relative risk estimates for ACR20 and ACR50 in the DMARD experienced population.

Meta-analyses were also undertaken in which the small number of monotherapy studies and treatment arms were excluded. The MS stated that the relative risk of the monotherapy group versus the original group (all studies) was calculated. Such data were presented for adalimumab, etanercept and infliximab for ACR20 response at 24 weeks and adalimumab and infliximab for ACR50 response at 24 weeks in the DMARD population. For each intervention, the exclusion of the monotherapy studies and treatment arms yielded relative risk values greater than 1 relative to the original group (all studies).

Limited longer-term data from the GO-FORWARD study were presented on page 107 of the MS. It was reported that 50 mg golimumab was clinically effective over 1 year, yielding ACR 20, 50 and 70 response rates of 64.0%, 43.8% and 24.7% respectively. It was stated that 90.6% of patients who had achieved ACR20 response at week 24 maintained this response at week 52 and that 61.4% displayed DAS28 remission (≤ 2.6) at week 52, and 36.8% with sustained DAS28 remission.

TNF-α inhibitor experienced population

ACR20 response at 24 weeks (TNF-α inhibitor experienced population)

A total of 2 studies were included for the ACR20 response in the TNFα experienced population, of which 1 study related to the use of golimumab (GO-AFTER (Smolen *et al.*, 2009)). As data were only available for one study each for golimumab and rituximab, no meta-analyses were conducted; the relative risk data for each treatment are presented.

Table 10: ACR20 response at 24 weeks (TNF-α inhibitor experienced population)

Active treatment	RR (95%CI)	P value
Golimumab	2.03 (1.34, 3.07)	0.001
Rituximab	2.85 (2.08, 3.91)	<0.001

ACR50 response at 24 weeks (TNF-α inhibitor experienced population)

As above, only 2 studies were included for the ACR50 response in the TNFα experienced population, of which 1 was of golimumab (GO-AFTER (Smolen *et al.*, 2009)). (NB: Table 50 (page 75) was apparently mislabelled as ACR 20 response at 24 weeks in MS)

Table 11: ACR50 response at 24 weeks (TNF-α inhibitor experienced population)

Active treatment	RR (95%CI)	P value
Golimumab	3.55 (1.67, 7.53)	0.001
Rituximab	5.40 (2.87, 10.16)	<0.001

ACR70 response at 24 weeks (TNF-α inhibitor experienced population)

Table 12: ACR70 response at 24 weeks (TNF-α inhibitor experienced population)

Active treatment	RR (95%CI)	P value
Golimumab	3.65 (1.39, 9.58)	0.009
Rituximab	12.14 (2.96, 49.86)	0.001

Additional evidence from the GO-AFTER study was presented on page 112 of the MS. The manufacturer reported that, among patients who had discontinued one or more previous TNF-

α inhibitors on the grounds of lack of efficacy, more patients achieved an ACR20 response at week 14 in the 50 mg golimumab group vs the placebo group (35.7% vs. 17.7%, P=0.006). Similarly, a larger proportion of patients who had discontinued one or more previous TNF-α inhibitors because of intolerance reached an ACR20 response in the 50 mg golimumab arm vs. the placebo arm (34.6% vs. 16.7%, P=0.154). The lack of statistical significance is likely to be attributable to low patient numbers in this subgroup.

Summary of efficacy findings from mixed treatment comparison and indirect comparison analyses

DMARD experienced population

Twenty RCTs were used to undertake the MTC for the DMARD experienced population (Table 54, page 77). These included 2 trials for golimumab versus placebo (GO-FORWARD, Kay *et al*., 2008), 7 trials for adalimumab versus placebo, 2 trials of certolizumab versus placebo, 4 trials of etanercept versus placebo, and 5 trials of infliximab versus placebo. The aim of the MTC was described as being to evaluate the efficacy of golimumab vs. 4 comparators using a network analysis. The network of treatments included golimumab, adalimumab, certolizumab, etanercept, infliximab and placebo.

The MS stated that results were reported as:

- Mean, median relative risk of each treatment vs. placebo and 95% credible interval
- Mean, median relative risk of golimumab vs. each treatment and 95% credible interval
- Probability that each treatment is the most effective

The manufacturer confirmed that relative risk values were presented in the following tables.

Table 13: ACR20 at 24 weeks MTC (DMARD experienced population) (as presented as Table 56 of MS)

	FIXED EFFECT MODEL (DIC=389.9)		**RANDOM EFFECT MODEL (DIC=340.1)**	
	median	**95% credible interval**	**Median**	**95% credible interval**
Golimumab	1.00	-	1.00	-
Adalimumab	1.06	0.83, 1.29	0.98	0.55, 1.46
Certolizumab	0.74	0.58, 0.90	0.72	0.41, 1.06
Etanercept	1.17	0.91, 1.47	0.93	0.51, 1.43
Infliximab	1.05	0.82, 1.29	1.05	0.57, 1.65
Placebo	2.11	1.67, 2.53	2.17	1.27, 3.00

Table 14: ACR50 at 24 weeks MTC (DMARD experienced population) (as presented as Table 57 of MS)

	FIXED EFFECT MODEL (DIC=344.1)		**RANDOM EFFECT MODEL (DIC=320.9)**	
	median	**95% credible interval**	**Median**	**95% credible interval**
Golimumab	1.00	-	1.00	-
Adalimumab	0.92	0.60, 1.36	0.90	0.40, 1.76
Certolizumab	0.65	0.41, 0.99	0.63	0.27, 1.31
Etanercept	1.45	0.92, 2.19	0.98	0.40, 1.99
Infliximab	1.03	0.66, 1.54	0.99	0.42, 2.04
Placebo	3.02	2.00, 4.35	3.22	1.54, 5.74

Table 15: ACR70 at 24 weeks MTC (DMARD experienced population) (as presented as Table 14 of supplementary responses from manufacturer)

			FIXED EFFECT MODEL (DIC=171.2)			**RANDOM EFFECT MODEL (DIC=172.7)**		
			mean	**median**	**95% credibility interval**	**Mean**	**median**	**95% credibility interval**
Golimumab	vs	placebo	4.47	4.17	2.05, 8.66	4.59	4.20	1.79, 9.68
Golimumab	vs	golimumab	1.00	1.00	-	1.00	1.00	-
Golimumab	vs	adalimumab	0.82	0.76	0.34, 1.70	0.83	0.75	0.28, 1.86
Golimumab	vs	certolizumab	0.53	0.48	0.19, 1.19	0.54	0.47	0.16, 1.35
Golimumab	vs	etanercept	0.38	0.32	0.09, 1.03	0.40	0.32	0.09, 1.15
Golimumab	vs	infliximab	1.32	1.21	0.53, 2.75	1.29	1.16	0.40, 3.00

The MS (page 80) stated that both fixed and random effects models were run for each outcome, with the most appropriate model being selected based on the DIC value, with the smaller DIC indicating better performance. Thus, the MTC analyses were interpreted using the results generated by the random effects model (for ACR20 and ACR50) and the fixed effects model (for ACR70) as appropriate (based on the model having the lower DIC value). This approach could be criticised in that the DIC values for the ACR70 response are similar and a conservative approach would be to use the random effects model as it allows for heterogeneity.

For ACR20, ACR50 and ACR70 response rates, there were no statistically significant differences when golimumab was compared with adalimumab, certolizumab, etanercept or infliximab and using the ERG-preferred random effects model, and there was considerable uncertainty. For each ACR response, golimumab was superior to placebo, with a statistically significant difference demonstrated.

Sensitivity analyses were performed for ACR20 and ACR50 responses in which the TEMPO etanercept trial was excluded due to a greater response within the placebo arm compared with other studies.

Additional sensitivity analyses in which studies did not have concomitant methotrexate in both arms were excluded did not alter the broad conclusions.

TNF-α inhibitor experienced population

Two trials were used in the indirect comparison analyses of golimumab (GO-AFTER) versus rituximab (REFLEX) in the TNF-α inhibitor experienced population (Table 55, page 78). In these analyses (based on the methods developed by Bucher *et al*.), the efficacy of golimumab and rituximab were indirectly compared with placebo as the common comparator. Relative risks and 95% confidence intervals were calculated for each study.

Table 16: ACR20 & ACR50 at 24 weeks Indirect Comparison (TNF-α inhibitor experienced population) (as presented as Table 58 of MS and Table 15 of supplementary responses from the manufacturer)

	Mean indirect estimate	
Outcome	**Golimumab vs Rituximab**	**95% confidence interval**
ACR20 at 6 months	0.71	0.42, 1.20
ACR50 at 6 months	0.66	0.25, 1.76
ACR70 at 6 months	0.30	0.05, 1.66

At no outcome measure was there a statistically significant difference between golimumab and rituximab, although the mean values always favoured rituximab.

Mortality

Mortality data were provided by the manufacturer in response to a request for further information by the ERG. Data are presented in the Appendices. No differences in mortality between golimumab and comparator drugs were apparent in the tabulated data. Further details of deaths occurring in the golimumab trials are discussed further in this Section.

[redacted]

Table 17: NCT00771251 efficacy results through week 24

Assessment	Placebo + MTX	Golimumab 50 mg + MTX	Golimumab 100 mg + MTX
Patients treated (n)	[redacted]	[redacted]	[redacted]
ACR20	[redacted]	[redacted]	[redacted]
ACR50	[redacted]	[redacted]	[redacted]
ACR70	[redacted]	[redacted]	[redacted]
Mean change from baseline in vdH-S [a] (TSS) (SD, range)	[redacted]	[redacted]	[redacted]
Median change from baseline in vdH-S (TSS) (IQ range)	[redacted]	[redacted]	[redacted]
Mean change from baseline in Joint Space Narrowing Score (SD, range)	[redacted]	[redacted]	[redacted]
Median change from baseline in Joint Space Narrowing Score (IQ range)	[redacted]	[redacted]	[redacted]
Mean change from baseline in Bone Erosion Score (SD, range)	[redacted]	[redacted]	[redacted]
Median change from	[redacted]	[redacted]	[redacted]

baseline in Bone Erosion Score (IQ range)	[redacted]	[redacted]	[redacted]

[redacted]

Critique of included efficacy data

The clinical efficacy data reported in the MS addresses only part of the appropriate decision problem, in that data and analyses for abatacept and tocilizumab were not included in the assessment. Furthermore, as discussed previously, not all outcomes specified in the scope and decision problem were addressed adequately.

The phase III GO-FORWARD study reported by Keystone *et al*. (2009) was powered for the two co-primary endpoints: the proportion of patients reaching an ACR20 response at week 14 and the change from baseline in the HAQ-DI score at week 24. It was calculated that a sample size of 120 subjects in group 1 and 80 subjects in groups 3 and 4 would be required to achieve >90% power and that that a sample size of 120 patients in both groups 1 and 2 would be necessary to achieve >85% power. These sample sizes were achieved. Patients were excluded from the GO-FORWARD trial if they had known hypersensitivity to human immunoglobulin proteins or components of golimumab, any previous use of any TNF-α inhibitor, rituximab, natalizumab or cytotoxic agents, had inflammatory disease other than RA, were pregnant, nursing or planning a pregnancy. Patients should not have received anakinra, DMARDs other then methotrexate, or IV, intramuscular, intra-articular corticosteroids within 4 weeks of the

first study dose or efalizumab within 3 months of the first study dose. Keystone *et al*. noted that the patients included in this study had rheumatoid arthritis of shorter disease duration, fewer tender and swollen joints, a greater level of physical function, and lower CRP levels than subjects of previous trials of biologics in patients with rheumatoid arthritis despite methotrexate treatment. However, the study authors suggested that the study characteristics may be representative of the reduced disease activity among patients who receive TNF-α inhibitors in clinical practice, with the shorter disease duration proposed to be reflective of clinical practice with the receipt of TNF-α inhibitors at an earlier point in the disease course. Whilst the clinical advisors to the ERG considered the study population to be generally representative of the UK population, they noted that the proportion of patients receiving concomitant glucocorticoid therapy (65.4 to 75.3%) to be higher than the UK average, which the clinical advisors to the ERG noted had been recorded as being 49% in the British Society for Rheumatology Biologics Register.[38]

Additional sources of longer-term data are available for the GO-FORWARD study. Keystone *et al.* (2010) [28] presented the 52 week findings from the open label extension of this trial and reported that the responses of patients receiving golimumab through week 24 were maintained at 52 weeks. The 104 week data from the GO-FORWARD trial were presented at EULAR 2010 (EULAR conference abstract supplied by manufacturer) and appeared to demonstrate that ACR responses were sustained at this later timepoint.

The study by Kay *et al*. (2008) was a 5-arm phase II dose-ranging study of small sample size (placebo plus methotrexate n=35; 50 mg golimumab every 4 weeks plus methotrexate n=35). A number of exclusion criteria were used during recruitment of the Kay *et al.* (2008) trial that may impact on external validity. These exclusions covered patients who had received DMARDs (eg. D-penicillamine, hydroxychloroquine, chloroquine, gold, anakinra, azathiporine, sulfasalazine and agents other than methotrexate) within 4 weeks of the first dose of study agent, had prior treatment failure due to lack of efficacy or toxicity with more than 3 of the DMARDs listed in the criterion above, had previously had treatment with infliximab or any other agent targeted at reducing TNF-α (eg. etanercept, adalimumab), used cytotoxic drugs or alkylating agents, had previously had treatment with a Prosorba column, taken leflunomide within 4 weeks of first study dose and within 3 months of study dose without drug elimination, had received treatment with an anti-CD4 antibody or been treated with any study drug within the previous 3 months or within 5 half-lives. The Kay trial was powered to detect a difference in the primary endpoint (proportion of patients achieving an ACR20 response at week 16). The study was powered to detect a difference in this primary endpoint when the combined golimumab groups and at least one of the individual dose groups

were compared vs. placebo (rather than for the comparison of individual golimumab dose groups vs. placebo). It was calculated that 35 patients were required in each treatment group to achieve >90% power. 35 patients were achieved for the placebo and 50 mg golimumab every 4 weeks treatment arms. Some heterogeneity was observed in terms of baseline characteristics, attributed by the study authors to the small sample size of the study, although these were not adjusted for. This variation was in terms of gender (range of median values 67.6 to 85.7% female), age (range of median values 48.0 to 57.5 years), disease duration (range of median values 5.6 to 9.0 years), number of swollen joints (range of median values 13 to 20) and number of tender joints (range of median values 22 to 32). The clinical advisors to the ERG considered the study population to be appropriate in terms of gender composition, age, disease duration and was considered representative of the UK population. In relation to the use of prior DMARD treatment among the study population (methotrexate, other DMARDs not specified, mean number not specified, mean duration not specified (methotrexate ≥ 3 months)), the clinical advisors to the ERG noted that it was important to be consider whether patients in this study might be viewed as failing minimal therapy rather than standard therapy (in that subjects had previously tolerated methotrexate at a dose of at least 10 mg/week for at least 3 months prior to the first dose of study drug). Therefore the study population may not be truly representative of the target population.

The manufacturer clarified that all timepoints from Kay *et al.* (2008) were extracted at week 16, as this was the primary endpoint and the latest timepoint for ACR response values reported in the original publication. Therefore, the manufacturer would seem to indicate that the data for Kay *et al.* relate to efficacy at week 16 rather than week 24. However, the ERG have noted inconsistencies between the data presented for ACR20 and ACR50 responses in terms of differing values presented in the original study publication (week 16) and in the efficacy meta-analyses in the MS (Tables 18 and 19, pages 60-61). It is therefore unclear how the efficacy raw data from Kay *et al.* (2008) have been derived and handled in the meta-analyses.

In the GO-AFTER study, potential participants were excluded if they had ever received natalizumab or rituximab, had taken anakinra within 4 weeks, or abatacept or efalizumab less than 3 months before the first dose of study agent, or had ever received cytotoxic drugs. Concomitant DMARD treatment with methotrexate, sulfasalazine, and hydroxychloroquine (alone or in combination) was permitted but not required. All patients had been previously treated with at least one dose of a TNF-α inhibitor (etanercept, adalimumab, or infliximab) at least 8 weeks (adalimumab or etanercept) or 12 weeks (infliximab) before the first dose of study agents. Treatment with previous TNF-α inhibitor may have been stopped for any

reason, cited by investigators as 'lack of effectiveness,' 'intolerance' or 'other'. This provision was considered to be appropriate by the clinical advisors to the ERG, who indicated that it was not unreasonable to cease TNF-α treatment after only one dose (for example on the basis of adverse events). The GO-AFTER trial was powered to satisfy the primary endpoint of ACR20 response at week 16. The numbers of participants obtained per treatment group exceeded the 140 patients in each group required by the power calculation. Whilst the clinical advisors to the ERG considered the study population to be reflective of the UK population, it was noted that the steroid use in this study may potentially also be higher than in the UK population. Therefore the study population may not be truly representative of the target population.

The 100 week findings from the GO-AFTER study (Smolen *et al.,* 2010, EULAR conference abstract supplied by manufacturer) were described as supporting the maintenance of the efficacy of golimumab in improvement of signs and symptoms of rheumatoid arthritis and physical function through to week 100.

The original publication of the GO-AFTER study (Smolen *et al.,* 2009) included additional data not presented in the MS. The publication presented the proportion of patients who achieved ACR20 at week 14 according to receipt of DMARD use at baseline.

Table 18: ACR20 response by DMARD use at baseline

DMARD use at baseline	Patients in placebo group who achieved ACR20 at week 14, n/N (%)	Patients in combined golimumab groups who achieved ACR20 at week 14, n/N (%)
Yes	19/107 (18%)	86/215 (40%)
No	9/48 (19%)	26/89 (29%)

Therefore, a larger difference was observed between the combined golimumab subjects and placebo group subjects in terms of ACR20 response among patients who had received DMARD at baseline.

Further evidence not presented in the MS was obtained from the original publication of the GO-AFTER study (Smolen *et al.,* 2009). 115 (25%) and 43 (9%) had received two or three TNF-α inhibitors respectively before study enrolment. Over 95% of patients were described as having been treated for 4 weeks or more with at least one TNF-α inhibitor. ACR20 response was reported according to the number of previous TNF-α inhibitors received. This

evidence indicated that golimumab demonstrated efficacy in terms of ACR20 response at week 14 following receipt of one or two previous TNF-α inhibitors, but that in patients who had failed treatment on 3 previous TNF-α inhibitors, the proportions of patients achieving an ACR20 response were similarly low among the combined golimumab groups and placebo group.

Table 19: ACR20 response by use of previous TNF-α inhibitors

Number of previous TNF-α inhibitors	Patients in placebo group who achieved ACR20 at week 14, n/N (%)	Patients in combined golimumab groups who achieved ACR20 at week 14, n/N (%)
1	18/90 (20%)	82/213 (38%)
2	7/44 (16%)	27/71 (38%)
3	3/21 (14%)	3/22 (14%)

Whilst the outcomes of health-related quality of life and fatigue were not adequately addressed in this assessment, the SPC[14] notes that in the GO-FORWARD study 'clinically significant and statistically meaningful improvements' were observed in health-related quality of life as evaluated using the physical component score of the SF-36 in golimumab-treated patients vs placebo. Furthermore, in GO-FORWARD and GO-AFTER, the SPC[14] reported that 'statistically significant improvements' were found in fatigue as assessed by functional assessment of chronic illness therapy-fatigue scale (FACIT-F).

One of the clinical advisors to the ERG noted the complexities in comparing data across the interventions in this assessment. Response rates may be influenced by changes in patient populations over time. Furthermore, the clinical advisor considered the certolizumab trials to be non-comparable with respect to trials of other TNF-α inhibitors in that patients were withdrawn at 12 weeks if there was no response, resulting in slower responses among placebo group subjects not being detected. The clinical advisor stated that this contributed to a higher ratio of ACR responses on active treatment versus placebo. Therefore, the analyses presented in the MS may not reflect this. The clinical advisor thus noted that comparing biologic therapies indirectly was problematic due to underlying differences in patient groups and trial methods.

Safety and tolerability

Adverse events occurring in the golimumab trials were described by the manufacturer in detail in the MS and in response to queries from the ERG. Section 2.7 (page 23) of the MS states that no significant adverse reactions of the treatments under assessment are known. The ERG and clinical advisors to the ERG do not concur with this statement, on the basis that a range of adverse events are known to be associated with this class of treatments.

The ERG requested further clarification from the manufacturer on the definition of serious adverse events, whereby supplementary details were provided. The manufacturer confirmed

that the majority of RCTs did not specifically define 'serious adverse events' as such but provided absolute and proportional figures for ' ≥ 1 SAE'. [redacted]

The ERG also requested clarification on the definition used for serious infections. As previously, the manufacturer stated that the majority of RCTs did not specifically define 'serious infections' but provided absolute and proportional figures for ' ≥ 1 Serious Infection'.

The classification groupings of the reported serious adverse events for GO-FORWARD, GO-AFTER and serious infections for GO-FORWARD, GO-AFTER and Kay *et al.* (2008) are listed in the Appendices.

Adverse events data presented in the original MS and supplementary information provided in response to queries from the ERG are considered for each trial in turn.

DMARD experienced population

GO-FORWARD study (Keystone et al., 2009)

In the GO-FOWARD study, the incidence and type of adverse events were reported at weeks 16 and week 24. These tables can be viewed in the Appendices.

[redacted]

Table 20: GO-FORWARD serious adverse events data (up to week 24 per treatment arm) (DMARD experienced)

GO-FORWARD treatment arm	Placebo + methotrexate	Placebo + methotrexate to golimumab 50mg + methotrexate	Golimumab 100mg + Placebo	Golimumab 100mg + Placebo to Golimumab 100mg + methotrexate	Golimumab 50mg + methotrexate	Golimumab 50mg + methotrexate to Golimumab 100mg + methotrexate	GOL 100mg + methotrexate
Number of patients with ≥1 SAE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Patient specific SAE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Longer-term data for serious adverse events (SAE) reported by Keystone *et al.* (2010) from the GO-FORWARD study through week 52 for DMARD experienced patients were provided (data as reported by the manufacturer):

Table 21: GO-FORWARD serious adverse events data through week 52

Treatment arms	**Number of serious adverse events**
Placebo + methotrexate to golimumab 50mg plus methotrexate: early escape (weeks 16-52) Placebo + methotrexate to golimumab 50mg plus methotrexate: crossover (weeks 24-52)	5 3
Golimumab 100mg + placebo	16
Early escape (weeks 16-52): golimumab 100mg + placebo to golimumab 100mg + methotrexate	7
Golimumab 50mg + methotrexate	9
Early escape (weeks 16-52): golimumab 50mg + methotrexate to golimumab 100mg + methotrexate	3
Golimumab 100mg + methotrexate	16

The manufacturer provided additional detail in terms of serious infections. [redacted]

In response to a request from the ERG for further detail on adverse events, the manufacturer also provided supplementary information on longer-term data on adverse events in the form of 104 week safety data for the open label extension of GO-FORWARD in DMARD experienced rheumatoid arthritis patients (EULAR conference abstract supplied by manufacturer):

Table 22: GO-FORWARD serious adverse events data through week 104

Treatment group	Serious adverse events (per 100 patient-years (95% CI))	Serious infections (per 100 patient-years (95% CI))
Group 1: n=133; Placebo + methotrexate	15 (6.28, 28.68)	2 (0.05, 10.14)
Group 2: n=133; Golimumab 100mg + Placebo	27 (20.37, 35.97)	6 (3.33, 11.24)
Group 3: n=89; Golimumab 50mg + methotrexate	16 (12.16, 21.47)	4 (1.76, 6.30)
Group 4: n=89; Golimumab 100mg + methotrexate	25 (19.35, 32.23)	6 (3.66, 10.39)

Active tuberculosis occurred in 2 patients, of which 1 patient was in Group 3 (golimumab 50 mg plus methotrexate) and 1 patient in Group 4 (golimumab 100 mg plus methotrexate).

Fifteen subjects discontinued the study agent through week 16 due to an adverse event. Details are provided below (as presented by the manufacturer):

Table 23: GO-FORWARD discontinuations of study agent due to adverse events at week 16

Treatment group	Number of discontinuations of study drug due to adverse event	Details of adverse events where provided
Group 1: Placebo + methotrexate	[redacted]	[redacted]
Group 2: Golimumab 100mg + Placebo	[redacted]	[redacted]
Combined Golimumab + methotrexate group (Group 3: Golimumab 50mg + methotrexate Group 4: Golimumab 100mg + methotrexate)	[redacted]	[redacted]

The manufacturer stated that, through week 24, [redacted]

[redacted]

Additional details on malignancies by treatment group were requested by the ERG and provided by the manufacturer. A total of 15 malignancies occurred through to week 104, with 1 in the placebo plus methotrexate group, 3 in the 100 mg golimumab plus placebo group, six in the 50 mg golimumab plus methotrexate group and 5 in the 100 mg golimumab plus methotrexate group, demonstrating an elevation among the golimumab plus methotrexate treatment groups (EULAR conference abstract supplied by manufacturer).

[redacted] Through week 104, 4 deaths occurred, with 1 case of sepsis, 1 case of fulminant hepatic failure, and 1 case of complicated respiratory distress in Group 2 (golimumab 100 mg plus placebo) and 1 case of circulatory insufficiency in Group 4 (golimumab 100 mg plus placebo).

[redacted]

Kay et al. (2008)

The ERG requested that details of adverse events reported in the Kay trial be provided. Supplementary information was forwarded.

The data for adverse events (with greater than 10% frequency) through week 20 of the Kay *et al.* (2008) trial are provided in the Appendices.

The manufacturer listed the following incidents as reported serious adverse events in Kay *et al.* (2008): worsening of rheumatoid arthritis activity, congestive heart failure, cardiac tamponade, lung cancer, squamous cell carcinoma, peripheral arterial occlusive disease, pyelonephritis and pneumonia. The manufacturer noted that the serious adverse events provided below (as reported through week 20) did not equal the total patients reported with ≥1 SAE as due to gaps in Kay *et al.* (2008) (no further explanation provided):

Table 24: Kay *et al.* (2008) serious adverse events data through week 20

Treatment group	**Number and type of serious adverse events (where provided)**
Placebo arm	2
Golimumab 50mg + methotrexate every 4 weeks arm (n=37)	4 (congestive heart failure, basal cell carcinoma and pneumonia)
Golimumab 50mg + methotrexate every 2 or 4 weeks arm (n=32)	3 (lung cancer, squamous cell carcinoma, pneumonia)
Golimumab 100mg + methotrexate every 4 weeks arm (n=33)	2 (basal cell carcinoma)
Golimumab 100mg + methotrexate every 2 or 4 weeks arm (n=35)	3 (cardiac tamponade and pneumonia)

No cases of tuberculosis were reported. [redacted]

Four cases of malignancy were observed in 4 patients treated with golimumab plus methotrexate.

In the Kay *et al*. trial, no deaths were reported during the 52-week study period.

TNF-α inhibitor experienced population

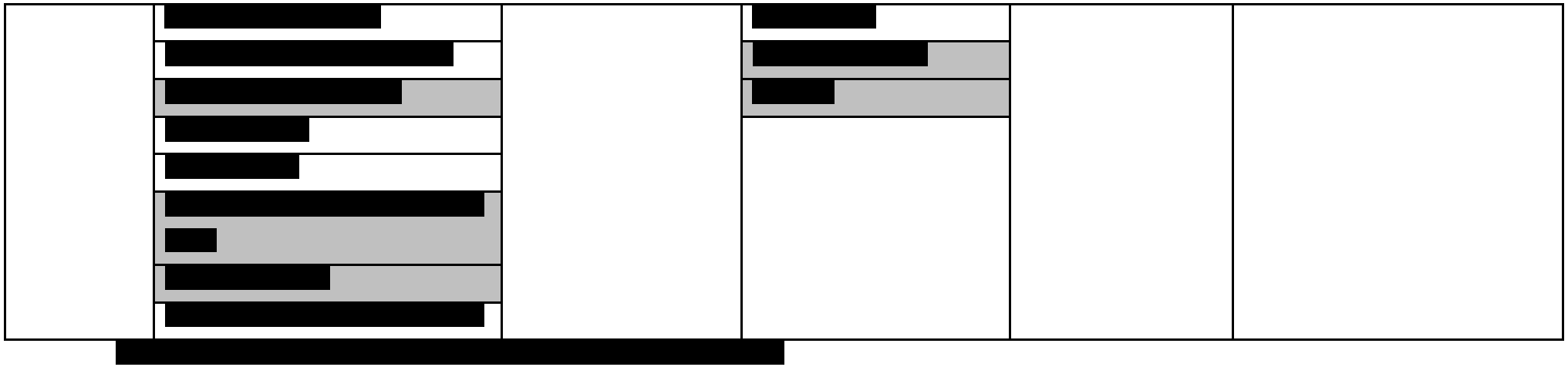
GO-AFTER study (Smolen et al., 2009)

The MS (page 113) stated that in the GO-AFTER study safety was evaluated by the incidence and type of adverse events by treatment group and that a subject with an adverse event was counted as belonging to a treatment group according to the study drug the subject was in receipt of at the time of the onset of the adverse event. The adverse events data though week 24 of the GO-AFTER trials are provided in the Appendices.

In response to a request from the ERG for further detail on adverse events, the manufacturer provided the following supplementary information. [redacted]

Table 25: GO-AFTER serious adverse events data (up to week 24 per treatment arm) (TNF-α inhibitor experienced)

GO-AFTER treatment arm	Placebo + methotrexate	Placebo + methotrexate to Golimumab 50mg + methotrexate	Golimumab 50mg + methotrexate	Golimumab 50mg + methotrexate to Golimumab 100mg + methotrexate	Golimumab 100mg + methotrexate
Number of patients with ≥1 SAE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Patient specific SAE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]



[redacted]

Further detail was also provided by the manufacturer on week 100 safety data from the open label extension of the GO-AFTER study (Smolen *et al.,* 2010, EULAR conference abstract supplied by manufacturer).

Table 26: GO-AFTER safety data through week 100

Treatment arm	**Serious adverse events at week 100**	**Serious infections at week 100**	**Injection site reactions at week 100**
Group 1: n=150; Placebo	Data not presented	Data not presented	Data not presented
Group 2: n=147; Golimumab 50mg + methotrexate	16.1%	5.0%	0.8%
Group 3: n=148; Golimumab 100mg + methotrexate	16.6%	5.7%	1.3%

No cases of tuberculosis were reported.

[redacted]

[redacted]

One death was reported during the study, whereby one patient within the placebo plus methotrexate group developed pancreatic cancer at week 23 followed by death.

[redacted]

Analyses of safety and tolerability data

The manufacturer conducted meta-analyses and mixed treatment comparisons in the DMARD and TNF-α inhibitor experienced patient populations for the following selected safety features: i) serious adverse events, ii) serious infections, iii) injection site reactions, and iv) discontinuation of treatment due to adverse events. The safety features chosen for these analyses were considered by the ERG to be appropriate. Since the timepoints at which data for each of these safety outcomes were measured were not reported in the original MS, clarification was requested from the manufacturer. The manufacturer clarified that the latest timepoint for all safety parameters had been extracted from each trial. For the majority of trials, the latest timepoint was described as being 24 weeks. The trials for which earlier or later timepoints were presented are tabulated below (as reported by the manufacturer). It was

stated that data for these timepoints across all available safety parameters had been extracted and included within the economic evaluation. It should be noted that a timepoint of 16 weeks was used for the Kay *et al.* (2008) trial, with 24 weeks applied for the GO-FORWARD and GO-AFTER studies.

Table 27: Safety timepoints other than 24 weeks by trial (as presented by manufacturer)

Intervention	Study	Safety time point
ADA	Chen 2009	12 weeks
ADA	DE019	12 months
CTZ	RAPID 1	12 months
ETN	TEMPO	12 months
GOL	Kay 2008	16 weeks
IFX	ATTEST	12 months
IFX	ATTRACT	12 months
IFX	Abe 2006	14 weeks
IFX	START	12 months

DMARD experienced population

Serious adverse events (DMARD experienced population)

Table 28: Serious adverse events in DMARD population

Active treatment	Relative risk compared with placebo generated by fixed effects meta-analysis (95%CI)	Relative risk compared with placebo generated by random effects meta-analysis (95%CI)	Estimated heterogeneity in random effects model
Adalimumab	1.04 (0.73, 1.48)	1.04 (0.70, 1.55)	I^2=12.1%, Chi-squared P value =0.337
Certolizumab (data taken from Table 66 erroneously headed as adalimumab)	2.13 (1.23, 3.67)	2.13 (1.23, 3.67)	I^2=0%, Chi-squared P value= 0.889)
Etanercept	0.87 (0.56, 1.35)	0.85 (0.55, 1.32)	I^2=0%, Chi-squared P value =0.414)
Golimumab	1.32 (0.54, 3.20)	1.31 (0.54, 3.22)	I^2=0%, Chi-squared P value= 0.630)
Infliximab	0.87 (0.63, 1.21)	0.86 (0.63, 1.20)	I^2=0%, Chi-squared P value= 0.586)

DMARD experienced population

Serious infections (DMARD experienced population)

Table 29: Serious infections in DMARD population

Active treatment	Relative risks placebo generated by fixed effects meta-analysis (95%CI)	Relative risk vs placebo generated by random effects meta-analysis (95%CI)	Estimated heterogeneity
Adalimumab	2.57 (1.14, 5.80)	2.98 (0.44, 19.92)	I^2=70.9%, Chi-squared P value =0.032)
Certolizumab	8.60 (0.50, 147.84) (data as reported in MS)	8.60 (0.50, 147.84) (data as reported in MS)	Not applicable
Etanercept	1.00 (0.48, 2.11)	1.00 (0.48, 2.11)	Not applicable
Golimumab	1.11 (0.18, 6.65)	1.10 (0.18, 6.65)	I^2=0%, Chi-squared P value= 0.863)
Infliximab	0.89 (0.47, 1.68)	0.87 (0.45, 1.65)	I^2=0%, Chi-squared P value= 0.462)

No meta-analyses were performed for serious infections associated with certolizumab or etanercept due to the inclusion of a single study for each intervention.

For certolizumab, the relative risk for serious infections with certolizumab vs. placebo was 8.60 (95%CI 0.50, 147.84) (P=0.138).

The relative risk for serious infections associated with etanercept vs. placebo based on the TEMPO study was 1.00 (95%CI 0.48, 2.11) (P=0.991). (The MS refers to certolizumab under the etanercept analysis, apparently erroneously).

DMARD experienced population

Injection site reactions (DMARD experienced population)

Table 30: Injection site reactions in DMARD population

Active treatment	Relative risk compared with placebo generated by fixed effects meta-analysis (95%CI)	Relative risk compared with placebo generated by random effects meta-analysis (95%CI)	Estimated heterogeneity in random effects model
Adalimumab	1.76 (1.40, 2.23)	2.53 (1.25, 5.14)	I^2=75.1%, Chi-squared P value =0.001)
Certolizumab	6.60 (0.87, 50.24)	5.97 (0.77, 46.26)	I^2=0%, Chi-squared P value= 0.624)
Etanercept	6.21 (3.75, 10.26)	5.33 (3.30, 8.61)	I^2=0.3%, Chi-squared P value =0.390)
Golimumab	0.95 (0.39, 2.32)	0.96 (0.39, 2.35)	I^2=0%, Chi-squared P value= 0.681)
Infliximab	No data reported	No data reported	No data reported

DMARD experienced population

Discontinuation due to adverse events (DMARD experienced population)

Table 31: Discontinuation due to adverse events in DMARD population

Active treatment	Relative risk compared with placebo generated by fixed effects meta-analysis (95%CI)	Relative risk compared with placebo generated by random effects meta-analysis (95%CI)	Estimated heterogeneity in random effects model
Adalimumab	1.73 (1.15, 2.62)	1.72 (1.12, 2.64)	I^2=0%, Chi-squared P value =0.432)
Certolizumab	2.86 (1.11, 7.33)	2.86 (1.11, 7.33)	I^2=0%, Chi-squared P value= 0.991)
Etanercept	0.83 (0.57, 1.21)	0.81 (0.56, 1.18)	I^2=0.3%, Chi-squared P value =0.570)
Golimumab	0.63 (0.19, 2.04)	0.63 (0.19, 2.05)	I^2=0%, Chi-squared P value= 0.928)
Infliximab	1.78 (1.04, 3.06)	1.66 (0.96, 2.89)	I^2=0%, Chi-squared P value= 0.530)

TNF-α inhibitor experienced population

As data were only available for one study each for golimumab and rituximab, no meta-analyses were undertaken and relative risk data for each treatment are presented in the tables below.

Serious adverse events (TNF-α inhibitor experienced population)

Table 32: Serious adverse events in TNF-α inhibitor experienced population

Active treatment	RR (95%CI)	P value
Golimumab	0.75 (0.35, 1.58)	0.445
Rituximab	0.74 (0.42, 1.30)	0.290

Serious infections (TNF-α inhibitor experienced population)

Table 33: Serious infections in TNF-α inhibitor experienced population

Active treatment	RR (95%CI)	P value
Golimumab	1.02 (0.30, 3.45)	0.975
Rituximab	1.57 (0.41, 6.00)	0.509

Injection site reactions (TNF-α inhibitor experienced population)

Table 34: Injection site reactions in TNF-α inhibitor experienced population

Active treatment	RR (95%CI)	P value
Golimumab	1.53 (0.56, 4.19)	0.409)
Rituximab	No data available	No data available

It was stated that no data were available for injection-site reactions for rituximab. Since this agent is administered intravenously, infusion reactions may be a more appropriate issue, although such reactions are not discussed in the MS.

Discontinuation due to adverse events (TNF-α inhibitor experienced population)

(NB: Table mislabelled in MS as injection site reaction meta-analyses)

Table 35: Discontinuation due to adverse events in TNF-α inhibitor experienced population

Active treatment	RR (95%CI) compared with placebo	P value
Golimumab	0.45 (0.14, 1.44)	0.180
Rituximab	2.69 (0.58, 12.55)	0.207
Tocilizumab	1.14 (0.46, 2.82)	0.772

Summary of safety and tolerability findings from mixed treatment comparison and indirect comparison analyses

DMARD experienced population

The MTC and indirect comparison findings were clearly tabulated within Section 5.8 of the MS. The definitions of serious adverse events and serious infections used by the manufacturer are described earlier in this section.

Table 36: Serious adverse events MTC findings (DMARD experienced population)

			FIXED EFFECT MODEL (DIC=263.2)			RANDOM EFFECT MODEL (DIC=263.1)		
			mean	median	95% credible interval	mean	median	95% credible interval
Golimumab	vs	placebo	1.46	1.32	0.55, 3.12	1.49	1.33	0.51, 3.39
Golimumab	vs	golimumab	1.00	1.00	-	1.00	1.00	-
Golimumab	vs	adalimumab	1.44	1.28	0.5, 3.23	1.43	1.25	0.44, 3.48
Golimumab	vs	certolizumab	0.72	0.63	0.23, 1.7	0.74	0.63	0.2, 1.92
Golimumab	vs	etanercept	1.73	1.53	0.57, 4.06	1.73	1.46	0.46, 4.52
Golimumab	vs	infliximab	1.52	1.36	0.54, 3.44	1.61	1.39	0.49, 3.96

Based on a random effects model, golimumab may have a greater number of serious adverse events than adalimumab, etanercept and infliximab and a fewer number than certolizumab, although these differences were not statistically significant.

Table 37: Serious infections MTC findings (DMARD experienced population)

			FIXED EFFECT MODEL (DIC=167.3)			RANDOM EFFECT MODEL (DIC=165.6)		
			Mean	median	95% credibility interval	mean	Median	95% credibility interval
Golimumab	vs	placebo	1.90	1.11	0.17, 8.49	2.18	1.13	0.13, 10.46
Golimumab	vs	golimumab	1.00	1.00	-	1.00	1.00	-
Golimumab	vs	adalimumab	0.78	0.43	0.05, 3.7	0.92	0.40	0.03, 4.8
Golimumab	vs	certolizumab	0.10	0.02	0, 0.7	0.16	0.02	0, 0.93
Golimumab	vs	etanercept	2.00	1.09	0.14, 9.51	3.79	1.10	0.07, 17.77
Golimumab	vs	infliximab	1.85	1.03	0.14, 8.56	2.27	0.99	0.09, 11.71
Golimumab	vs	abatacept	1.58	0.89	0.12, 7.25	2.16	0.94	0.09, 11.34
Golimumab	vs	rituximab	1.24	0.42	0.02, 7.16	2.36	0.42	0.01, 11.68
Golimumab	vs	tocilizumab	1.18	0.65	0.08, 5.5	1.4	0.6	0.05, 7.25

Using the results from the random effects model for serious infections, golimumab had significantly fewer serious infections than certolizumab. There was considerable uncertainty in the comparison between golimumab and the remaining comparators.

Table 38: Injection site reactions MTC findings (DMARD experienced population)

			FIXED EFFECT MODEL (DIC=173.4)			RANDOM EFFECT MODEL (DIC=157.9)		
			Mean	median	95% credible interval	mean	median	95% credible interval
Golimumab	vs	placebo	1.08	0.95	0.36, 2.53	1.31	0.96	0.2, 4.52
Golimumab	vs	golimumab	1.00	1.00	-	1.00	1.00	-
Golimumab	vs	adalimumab	0.55	0.48	0.18, 1.34	0.42	0.29	0.04, 1.57
Golimumab	vs	certolizumab	0.07	0.04	0.01, 0.36	0.11	0.03	0, 0.53
Golimumab	vs	etanercept	0.15	0.13	0.04, 0.38	0.17	0.11	0.02, 0.67
Golimumab	vs	infliximab	-	-	-	-	-	-

Using the results from the random effects model for serious infections, golimumab had significantly fewer injection site reactions than certolizumab or etanercept. There was considerable uncertainty in the comparison between golimumab and the remaining comparators.

No analyses were presented for infliximab. Similarly, no data had been presented for infliximab in the injection site reaction meta-analyses. The reason for this was not explicitly stated in the MS, but data may have been omitted on the basis that infliximab is administered via IV infusion. However, infusion site reactions may have been a relevant issue but were not discussed within the MS.

Table 39: Discontinuation due to adverse events MTC findings (DMARD experienced population)

			FIXED EFFECT MODEL (DIC=273.3)			RANDOM EFFECT MODEL (DIC=274.5)		
			mean	**median**	**95% credible interval**	**mean**	**Median**	**95% credible interval**
Golimumab	vs	placebo	0.70	0.59	0.15, 1.92	0.71	0.59	0.14, 2.02
Golimumab	vs	golimumab	1.00	1.00	-	1.00	1.00	-
Golimumab	vs	adalimumab	0.40	0.33	0.08, 1.14	0.41	0.33	0.07, 1.26
Golimumab	vs	certolizumab	0.27	0.20	0.04, 0.88	0.27	0.20	0.04, 0.95
Golimumab	vs	etanercept	0.88	0.72	0.17, 2.54	0.86	0.68	0.14, 2.67
Golimumab	vs	infliximab	0.37	0.30	0.07, 1.08	0.37	0.29	0.06, 1.17

Using the results from the random effects model, golimumab had significantly fewer discontinuations due to adverse events than certolizumab. There was uncertainty in the comparison between golimumab and the remaining comparators, although it is noted that the mean RR was lower for golimumab than for any other intervention.

TNF-α inhibitor experienced population

In the following table the manufacturer did not provide the results from a random effects model due to their only being one study of each intervention.

Table 40: Serious adverse events MTC findings (TNF-α inhibitor experienced population)

			FIXED EFFECT MODEL		
			Mean	**Median**	**95% credible interval**
golimumab	vs	placebo	0.79	0.74	0.34, 1.53
golimumab	vs	golimumab	1.00	1.00	-
golimumab	vs	rituximab	1.12	1.00	0.38, 2.55

Table 41: Serious infections MTC findings (TNF-α inhibitor experienced population)

			FIXED EFFECT MODEL		
			Mean	**Median**	**95% credible interval**
golimumab	vs	Placebo	1.25	1.02	0.28, 3.62
golimumab	vs	Golimumab	1.00	1.00	-
golimumab	vs	Rituximab	0.94	0.61	0.09, 3.75

Patients on golimumab are estimated to have a greater risk of serious adverse events and a lesser risk of serious infections compared with rituximab, but there was considerable uncertainty regarding these values.

Data were available for golimumab only for injection site reactions.

Table 42: Discontinuation due to adverse events MTC findings (TNF-α inhibitor experienced population)

			FIXED EFFECT MODEL		
			Mean	**Median**	**95% credible interval**
golimumab	vs	placebo	0.50	0.43	0.11, 1.33
golimumab	vs	golimumab	1.00	1.00	-
golimumab	vs	rituximab	0.23	0.15	0.02, 0.91

Golimumab was found to have fewer discontinuations due to adverse events than rituximab, with a statistically significant difference.

Critique of safety data reported

There are a number of issues that may impact upon the interpretation of the safety data reported in the MS.

The following exclusion criteria were used during recruitment of the GO-FORWARD (Keystone *et al.,* 2009) trial: patients who had a history of past or active granulomatous infection (including tuberculosis), had experienced a nontuberculous mycobacterial infection or opportunistic infection, had a serious infection, been hospitalised for an infection or treated with IV antibiotics for infection within 2 months of first dose of study drug, had a history of ongoing, chronic or recurrent infectious disease (in line with the cautionary note within the

SPC[14] relating to use of golimumab in patients with chronic or recurrent infection), were infected with HIV, hepatitis B or hepatitis C, had a history of demyelinating disease (eg. multiple sclerosis), were pregnant, had other inflammatory diseases, had received or would receive live virus or bacterial vaccination within 3 months of first study dose, during trial, or 6 months following last study dose, had a history of infected joint prosthesis or had received antibiotics for such a suspected infection, had known hypersensitivity to any other components of golimumab, had current signs or symptoms of severe, progressive or uncontrolled renal, hepatic, haematological, gastrointestinal, endocrine, pulmonary, cardiac, neurological, psychiatric or cerebral disease, had a history of congestive heart failure, lymphoproliferative disease, any known malignancy or malignancy within the previous 5 years, had a transplanted organ, or had a substance abuse problem within the previous 3 years. The GO-FORWARD study was not powered for the assessment of adverse events.

It should be noted that a number of exclusions were applied during the recruitment process of Kay *et al.* (2008) relevant to the assessment of safety and tolerability such as excluding patients that were pregnant, had experienced a serious infection during the 2 months prior to first study treatment, had a history of ongoing or recurrent infectious disease (including chronic chest infection, chronic renal infection, sinusitis, recurrent UTI, skin wound or ulcer), had experienced an opportunistic infection within 6 months prior to screening, had evidence of prior or active tuberculosis, had a chest radiograph showing evidence of malignancy, infection or abnormalities suggestive of tuberculosis, had other inflammatory diseases (eg. ankylosing spondylitis) or a history of known demyelinating disease (eg. multiple sclerosis), were infected with HIV, hepatitis B or C, had a history of clinically significant adverse reactions to murine or chimeric proteins, had received or expected to receive live virus or bacterial vaccinations within 3 months of first study dose, during the study or up to 3 months after the study dose, had a history of infected joint prosthesis or had received antibiotics for a suspected infection of a joint prosthesis, had current signs or symptoms of severe, progressive or uncontrolled renal, hepatic, haematological, gastrointestinal, endocrine, pulmonary, cardiac, neurological, psychiatric or cerebral disease, had a history of congestive heart failure, lymphoproliferative disease, any known malignancy or malignancy within the previous 5 years, had a transplanted organ, or had a substance abuse problem within the previous 3 years. It should also be noted that the Key *et al.* trial was not powered to detect adverse events and therefore that the capacity of this study to detect adverse events was considerably limited by the small sample size.

The following exclusion criteria with the potential to effect the adverse event findings were used during participant selection in the GO-AFTER trial (Smolen *et al.*, 2009): patients who

had experienced a serious adverse reaction to a previous TNF-α inhibitor, those who had a serious infection less than two months before screening or an opportunistic infection less than 6 months before screening or a history of chronic infection, patients with other inflammatory disease or demyelinating disease, those who had a history of latent or active granulomatous infection (except tuberculosis that had been treated prophylactically in the past 3 years), patients with congestive heart failure, or severe, progressive uncontrolled renal, hepatic, haematological, gastrointestinal, endocrine, pulmonary, cardiac, neurological, psychiatric, or cerebral disease or had a transplanted organ or malignancy in the previous 5 years. GO-AFTER was not powered for the assessment of adverse events. It should be noted that patients who had discontinued TNF-α inhibitors due to serious adverse events were excluded, therefore the findings from this trial could not be considered generalisable to such patients.

Such exclusions in the recruitment of the golimumab trials should be taken into consideration with regards to the generalisability of these findings to clinical practice.

A report by the Centre for Drug Evaluation and Research on the FDA website described a comparative safety analysis (golimumab with and without methotrexate) and indicated that, whilst overall the monotherapy and combination therapies are similar in terms of safety, there was a suggestion that the monotherapy group may have a more favourable safety profile.[36]

Adverse drug reactions from safety data from phase IIb and phase III clinical trials from 2578 golimumab-treated RA, PsA and AS patients vs. 751 control group patients as presented in the SPC[14] are summarised in the Appendices.

According to the SPC,[14] no overall differences in adverse events, serious adverse events and serious infections occurred in patients aged 65 years or older (n=155) in phase II studies of golimumab for rheumatoid arthritis, psoriatic arthritis and ankylosing spondylitis vs younger patients, but caution should be used in the treatment of the elderly, particularly in relation to the development of infections.

The FDA website highlighted the risk of serious infections (leading to hospitalisation or death including tuberculosis, bacterial sepsis, invasive fungal and other opportunistic infections), hepatitis B reactivation, malignancies (particularly the increased risk of lymphomas and other malignancies in children and adolescents treated with TNF-α inhibitors, heart failure, cytopenias, demyelinating disease and new-onset psoriasis. Furthermore, the FDA also described an increased risk of serious infection when golimumab is used with another TNF-α

inhibitor, abatacept or anakinra (and therefore concomitant use of these therapies was not recommended).

The European Medical Agency[39] reviewed the evidence for golimumab in RA, PsA and AS and reported that in the 5 phase III studies through week 16, patients who had been receiving corticosteroids at baseline were more likely to experience a serious infection (1.4% placebo, 2.1% 50 mg golimumab, 2.5% 100 mg golimumab) vs. subjects who were not (1.1% placebo, 0.3% 50 mg golimumab, 1.1% 100 mg golimumab). Seven cases of tuberculosis were also reported in the 5 phase II studies, reflecting the importance of the close monitoring of patients. The Agency considered the adverse event profile of golimumab to be similar to that of other TNF-α inhibitors.

Data was identified relating to the development of malignancies within the included golimumab trials. The SPC[14] notes that 'the potential role of TNF-blocking therapy in the development of malignancies is not known...a possible risk for the development of lymphomas, leukaemia or other malignancies in patients treated with a TNF-antagonist cannot be excluded.' The SPC[14] also noted that malignancies (of which approximately half were lymphomas, with others including rare cancers typically associated with immunosuppression) had been reported in children, adolescents and young adults (up to 22 years of age) in the post-marketing setting. It was also reported in the SPC[14] that in a clinical trial assessing the use of golimumab in patients with severe, persistent asthma, more malignancies occurred in golimumab-treated patients vs control group patients, but the significance of this occurrence was unknown.

A small number of deaths were reported in the submitted evidence for the 3 golimumab RA trials. The European Medical Agency[39] described the evidence for golimumab in RA, PsA and AS and stated that a total of 13 deaths occurred during the golimumab studies according to available data, of which 12 subjects were treated with golimumab.

[redacted]. However, the FDA noted that transaminase elevations have been associated with TNF-α inhibitors and should be taken into consideration during treatment. The original study publication of Kay *et al.* (2008) also reported a higher proportion of golimumab-treated patients with abnormal alanine aminotransferase levels (above 75 IU/litre and increased by ≥ 100%):

Table 43: Numbers of patients with abnormal alanine aminotransferase levels through week 52

	Placebo/ Infliximab	50 mg golimumab every 4 weeks	50 mg golimumab every 2/4 weeks	100 mg golimumab every 4 weeks	100 mg golimumab every 2/4 weeks
Number of patients with abnormal alanine aminotransferase levels	1 (2.9%)	3 (8.1%)	4 (12.5%)	1 (3.0%)	3 (8.6%)

One clinical advisor to the ERG also noted that the treatment history of individual patients may also complicate the assessment of efficacy and safety, with multiple drug use over time varying by patient and resulting in various effects.

No non-RCT evidence was included in the systematic review of clinical evidence. As stated earlier in this report, the ERG consider non-RCTs, and expert opinion, to be a valid and important source of evidence for the evaluation of adverse events. It is stated on page 227 under the description of inclusion and exclusion criteria for adverse events that evidence from non-randomised trials would be considered only when the information was not available in RCTs. The manufacturer stated that no non-RCTs or observational studies of the use of golimumab in patients with rheumatoid arthritis were available at the time of submission. However, it is unclear whether any non-randomised trials were identified for the comparator drugs. It should be noted that only limited safety outcomes were listed under the inclusion criteria on page 227 that do not appear to reflect the range of adverse events outcomes included in the assessment.

In response to a request for further detail on the adverse events searches and results, the manufacturer provided supplementary information. The Appendices include the adverse event trials which were identified from the adverse events search results. However, it was unclear from the MS and supplementary information provided by the manufacturer what the study characteristics of these included trials were and how the evidence was used in the assessment. Further clarification was requested from the manufacturer, who stated that adverse events studies described as 'included' (see Appendices) were reviewed for relevant safety data (serious adverse events, serious infections, injection site reactions and discontinuations due to

adverse events). The manufacturer also noted in their further clarification that no trials which primarily assessed the safety outcomes of the interventions were identified that had not already been identified from the clinical effectiveness systematic review. No further information or clarification was presented.

The manufacturer used for a quality assessment checklist was adapted from Downs & Black (1998) for the trials described as being included in the submission (presented in Appendices).

It was stated in the MS that no trials were identified for which the primary aim was the assessment of the safety and tolerability of golimumab or comparator interventions.

For the discontinuation of treatment due to adverse events in the TNF-α inhibitor experienced population, data were presented for tocilizumab (page 100 of MS). It is unclear why data were reported for tocilizumab in this case but for no other outcome or analyses. Similarly, it was unclear why results were presented for tocilizumab, abatacept and rituximab in the MTC findings for serious infections (page 92 of MS).

It is crucial to undertake further monitoring and reporting of long-term safety outcomes for the use of golimumab in rheumatoid arthritis, particularly in terms of serious adverse events (including malignancies), serious infections and deaths.

Since RCTs have a limited ability to assess drug safety and tolerability, there is a requirement of such evidence to be supplemented by other sources of data, including post-marketing surveillance studies, to facilitate the longer-term follow-up of larger numbers of patients and the collection of data relating to the target population treated in clinical practice. One clinical advisor to the ERG noted the value of registers such as the British Society for Rheumatology Biologics Register (BSRBR) in the collection of longer-term safety data for the evaluation of such biologic therapies as golimumab.

4.10 Critique of submitted evidence syntheses

Meta-analyses were conducted by the manufacturer for the efficacy outcomes of ACR20, ACR50 and ACR70 responses at 24 weeks for both DMARD and TNF-α inhibitor experienced patient populations (ACR70 response meta-analyses were provided following a request for additional data by the ERG). Meta-analyses were also performed for selected safety and tolerability outcomes, including i) serious adverse events, ii) serious infections, iii) injection site reactions, and iv) discontinuation of treatment due to adverse events. The

efficacy and safety outcomes chosen for these analyses were considered by the ERG to be appropriate. Meta-analysis methods and findings were reasonably clearly presented. Both fixed and random effects models were used to generate relative risks for efficacy and safety outcomes.

Only biologic therapies licensed for the particular patient population and which had received NICE Technology Appraisal Guidance as of the date of submission were included by the manufacturer within the meta-analyses. Therefore, rituximab, abatacept and tocilizumab were excluded as comparators from the DMARD experienced meta-analyses as the NICE final decision was still pending. Abatacept and tocilizumab were also excluded as comparators from the TNF-α inhibitor experienced meta-analyses on the same basis. As stated in Section 3, the ERG considered abatacept and tocilizumab to be valid comparators for inclusion in this assessment.

The manufacturer presented an MTC for the DMARD experienced population and an indirect comparison for the TNF-α inhibitor experienced population. The results are presented in Section 5.6 of the MS, and are used in the cost-effectiveness analyses to provide comparison between the relevant set of comparators.

DMARD experienced population MTC

The MTC allows the indirect comparison of golimumab, methotrexate, adalimumab, infliximab, etanercept and certolizumab. A literature search was performed to identify relevant trials to provide direct evidence in the network (Section 4.2.1). The MTC estimates two clinical outcomes, ACR20 and ACR50, both at week 24. Supplementary ACR70 MTC analyses were included in response to a request by the ERG.

20 trials inform the evidence network, with all trials including a comparison between a comparator and placebo, except for the ATTEST study (as described in the MS) which compares MTX vs infliximab vs abatacept. This comparison is not correctly indicated in Table 54 (p.77) of the MS. All five comparator TNF-α inhibitors have more than one trial compared to placebo.

The manufacturer developed both a fixed-effect and random-effect Bayesian model, with the variance in the random-effect model informed by a vague prior. The precision is assigned a gamma distribution, and all other parameters are given vague non-informative normal priors. No other details or justification regarding the selection of the prior distributions is given in the submission.

The manufacturer's basis for selecting either a fixed or random-effect model for use in the economic analysis is made solely on the reported Deviance Information Criterion (DIC). In general, a lower DIC indicates a better goodness-of-fit, as well as favouring models with a smaller number of parameters. However, DIC should be regarded as a tool to assist model building in conjunction with expert opinion. In addition, small differences in DIC should not be taken to mean that the model with the smaller DIC provides the most appropriate results for making inferences. The ERG would have preferred the use of a random effects model to be used in analyses to allow for heterogeneity.

The manufacturer chose the random-effect model over the fixed-effect model when selecting which data to use in the economic model. Median RR estimates were similar between the two model specifications, with an expected increase in the uncertainty in the random-effect model. The credible intervals for golimumab versus comparator TNF-α inhibitors crossed the line of no effect for both ACR20 and ACR50 when using the random-effect model.

TNF-α inhibitor experienced population indirect comparison

The manufacturer undertook an indirect comparison of golimumab and rituximab using placebo as the common comparator. Two RCT trials inform the indirect comparison (GO-AFTER and REFLEX), and the Bucher *et al.* method was used to calculate the relative risk and associated variance of golimumab versus rituximab. A mixed treatment analysis was not included within the MS. The ERG requested clarification on why this was not done and requested this analysis to be provided. The rationale given by the manufacturer was that it was not necessary given that only two comparisons against placebo were available and that the model may have struggled to estimate all parameters. The ERG believe this to be a reasonable approach. A fixed effect MTC was provided which had similar results but with slightly smaller confidence intervals.

Sensitivity Analyses

Sensitivity analyses were undertaken in which the TEMPO etanercept trial (Klareskog *et al.*, 2004) was excluded. The manufacturer justified this on the basis that the placebo arm response was high in the TEMPO trial vs. other studies. The ERG considered the undertaking of these sensitivity analyses to be appropriate. The manufacturer did not formally assess whether the treatment effect is dependent on the baseline response. To explore the influence of this study, the manufacturer excluded the TEMPO trial from the mixed treatment analyses. To provide further detail on the impact of the exclusion of the TEMPO etanercept study the ERG requested that analyses be provided showing the impact of excluding the TEMPO study

from the etanercept versus placebo meta-analyses. These meta-analyses are presented in Section 4.9.

As described by the manufacturer, most studies reporting data for the DMARD experienced population administered the biologic DMARD in conjunction with methotrexate. It was noted that for four studies no concomitant administration of methotrexate was permitted and that three studies included a monotherapy treatment arm. In order to assess the impact of the minority of monotherapy studies and monotherapy treatment arms on the relative risk estimate, the manufacturer performed separate meta-analyses and mixed treatment comparisons excluding any monotherapy trial groups from the analysis. These analyses were considered by the ERG to be reasonable.

It should be noted that a comment was made within the assessment report for the use of tocilizumab in rheumatoid arthritis (TA 198)[1] that the etanercept trial by Combe *et al.* included a combination of etanercept and sulfasalazine and this trial was therefore excluded from the analyses in the tocilizumab assessment on the basis that the treatment arms between trials were different from the remaining trials, which evaluated combination therapy as treatment with a biologic agent and methotrexate. This sensitivity analysis was not conducted in the current assessment for golimumab.

[redacted]

4.11 Summary

ACR response findings (DMARD population)

The manufacturer reports efficacy for the two populations (DMARD experienced and TNF-α inhibitor experienced). The comparators evaluated on clinical effectiveness by the manufacturer are adalimumab, certolizumab, etanercept and infliximab for the DMARD experienced population and rituximab for the TNF-α inhibitor experienced population.

The efficacy for interventions in the DMARD population was estimated using both a mixed treatment comparison and with meta-analyses. The efficacy for interventions in the TNF-α inhibitor experienced population was estimated using an indirect comparison methodology.

Efficacy in the DMARD experienced population: Using a random effects meta-analysis, the relative risk estimate for ACR20 response at 24 weeks was lower for golimumab than for the comparators although each intervention had wide 95% confidence intervals, and thus a definitive judgement on efficacy could not be made. These conclusions also applied when using a fixed effects meta-analysis, although the midpoint relative risk for golimumab may no longer be the lowest from all interventions. Evaluations of ACR50 and ACR70 response rates produced similar conclusions to that for ACR20 in that no definitive conclusion could be made.

One of the clinical advisors to the ERG highlighted potential issues that may influence such comparisons between interventions, for example changes in patient populations over time and study design issues, for example with respect to non-responders being withdrawn in the placebo arm of the certolizumab trials leading to raised treatment group response rates.

Efficacy in the TNF-α inhibitor experienced population: Estimated relative risks for ACR20, ACR50 and ACR70 were lower for golimumab vs. rituximab at 24 weeks, although considerable variation was present in these estimates, with wide 95%CI for each ACR response.

Safety and tolerability findings (DMARD population)

[redacted]

[redacted]

Meta-analyses were conducted for selected safety outcomes. Golimumab was estimated to have more serious adverse events than all comparators except certolizumab, although there was considerable uncertainty as shown by the wide confidence intervals. The estimated rate of serious infections for golimumab was similar to the rates for infliximab and etanercept, which were the lowest for the interventions, although all had wide confidence intervals. Golimumab was estimated to have the fewest injection site reactions and discontinuations due to adverse events, although the values for all interventions were subject to considerable uncertainty.

Safety and tolerability findings (TNF-α inhibitor population)

[redacted]

The estimated relative risks obtained for serious adverse events were similar for golimumab and rituximab, although there was considerable uncertainty. The relative risk estimate for serious infections was slightly lower for golimumab compared with rituximab, however both values have wide confidence intervals. No data were available for rituximab for injection site reactions. The relative risk for discontinuation of treatment due to adverse events was lower for golimumab than rituximab, although the confidence intervals were wide and overlapped considerably.

These conclusions were also supported by a mixed treatment comparison for the DMARD experienced population and an indirect comparison using the Bucher method for the TNF-α inhibitor experienced population, which produced similar results.

The submission can be considered to represent a balanced estimate of the clinical effects of golimumab relative to comparator therapies for the efficacy and safety outcomes and comparators that were included. However, the clinical efficacy data presented only partially addressed the decision problem, in that tocilizumab and abatacept were not incorporated into analyses. Furthermore, not all relevant outcomes were addressed adequately in the full submission, for example radiological progression.

5 ECONOMIC EVALUATION

This section appraises the economic evidence submitted by the manufacturer in their report, and in response to points of clarification from the ERG. The critical appraisal is based on a review of their submission report, as well as an examination of their economic models. The quality of the economic evaluation is assessed, and key assumptions and possible limitations are highlighted. The manufacturer provided extra information and sensitivity analyses in response to the ERG's clarification letter. These details and results are presented within the relevant section. Additional analysis undertaken by the ERG to investigate remaining uncertainties is also presented in Section 6.

5.1 Overview of manufacturer's economic evaluation

A summary of the manufacturer's economic evaluation and signposts to relevant sections of their MS are reported in Table 44.

The manufacturer evaluated golimumab in two different patient populations:

1. DMARD experienced population – golimumab is compared to TNF-α inhibitors (etanercept, adalimumab, infliximab and certolizumab) and methotrexate in patients who have failed two DMARDs.
2. TNF-α inhibitor experienced population – golimumab is compared to rituximab and methotrexate in patients who have failed on two DMARDs and a TNF-α inhibitor.

   All comparators are given with concomitant methotrexate. Methotrexate monotherapy is included as a comparator in each population as it represents the placebo arm in each indirect/mixed treatment analysis.

The economic evaluation included comparators approved by NICE in previous appraisals. For the DMARD experienced population these were adalimumab, infliximab and etanercept (TA130) and certolizumab (TA186). In the TNF-α inhibitor experienced population this was rituximab (TA126). Technologies being appraised by NICE at the time of the manufacturer's submission were not included as comparators.

Each analysis was reported in the MS, and a Microsoft Excel Markov model for each analysis was submitted.

Table 44: Manufacturer economic evaluation overview

Model element	Assumption	Evidence/Justification	Signpost to MS
Models provided	Two cost-utility analyses were provided, with one evaluating golimumab (GOL) plus methotrexate (MTX) in a DMARD experienced population, and the other evaluating GOL + MTX in a TNF-α inhibitor experienced population. A Markov framework was used to develop the economic models.	Two RCTs, one for each population group, provided clinical effectiveness estimates for golimumab versus placebo. Mixed treatment comparison and indirect treatment comparison methods were used to evaluate golimumab against relevant comparator treatments.	Section 6.2.1, p121
Health states	The model incorporated three health states, no response (sub ACR20), ACR20 and ACR 50.	The basis for the manufacturer's justification of this approach is that a published economic evaluation of rituximab[40] uses the same health states.	Section 6.2.4, p123
Comparators	**DMARD experienced population:** GOL, Adalimumab, Infliximab, Etanercept, MTX, Certolizumab Pegol. **TNF-α inhibitor experienced population:** GOL, Rituximab, MTX All therapies are in combination with MTX	The DMARD experienced population comparators were selected due to existing NICE Guidance recommending adalimumab, etanercept, infliximab (TA130) and certolizumab pegol (TA186) which identifies current UK practice for third line therapy as one of these four TNF-α inhibitors. MTX was included as a comparator due to direct comparative trial data. The TNF-α inhibitor experienced population comparator, rituximab, was chosen due to existing NICE Guidance recommending its use after failure on one previous TNF-α inhibitor (TA126). Again, MTX was included as a comparator due to direct comparative trial data being available.	Section 6.2.5, p123
Health Related	The model estimates a HAQ score for each health state,	The manufacturer does not report Health Related Quality of Life data from the	Section 6.4.3,

Quality of Life	and a published regression analysis comparing HAQ to EQ-5D is used to estimate utility values.	published trials, and justifies the use of a HAQ to EQ-5D regression analysis due to a strong correlation being found[41] and due to being a widely used method in previously published economic models.	p139
Adverse Events	Adverse events are modelled indirectly, as a trigger for patients to withdraw from treatment (trial data at 24 weeks, or long-term drop-out rate from observation studies for subsequent cycles). Costs and utility outcomes are assumed to be equivalent and therefore not modelled.	The manufacturer assumes a TNF-α inhibitors are a class, with equivalent costs and utilities associated with adverse events.	Section 6.4.8, p142
Mortality	The model assumes an equal risk of death, irrespective of what treatment patients are receiving	UK general population lifetables[42] were used to estimate a mortality risk for each model cycle. The proportion of males and females recruited in the golimumab trials were used to estimate a weighted average mortality risk by gender.	Section 6.3.3
Resource Use and Costs	Resource utilisation and costs are presented for each treatment regimen modelled. The costs include acquisition and administration costs, as well as staff time and inpatient hospital visits. (Tables 145 and 145 in the MS).	Resource use was estimated in consultation with two expert clinicians in the UK. The Birmingham Rheumatoid Arthritis Model (BRAM), the ACR and British Society for Rheumatology (BSR) were reviewed to check consistency	Section 6.5.1, p147
Discount Rate	Costs and QALYs were discounted at 3.5% per annum.	Assumption as per the NICE Guide to the Methods of Technology Appraisal[43]	Section 6.2.6 p126
Time Horizon	A lifetime time horizon (45 years). A half-cycle correction is incorporated	Patients starting age of 50 years (DMARD experienced) and 54 years (TNF-α inhibitor experienced).	Section 6.2.6, p126
Sensitivity Analysis	Probabilistic sensitivity analysis and scenario analysis were conducted.	Parameters were based on published literature or assumptions, and along with the distributional form were justified.	Section 6.6, p 154

5.1.1 Natural history

The model estimates a patients disease level based on their Health Assessment Questionnaire Disability Index (HAQ) score. Baseline HAQ is derived from the baseline characteristics of the GO-FORWARD and GO-AFTER trials. The model reassigns a HAQ score every cycle (6 months).

When patients are assigned either the no-response, ACR20 or ACR50 health state, their HAQ is re-assigned to reflect the level of response achieved. The HAQ scores for the three health states are estimated from the respective trial, and all non-methotrexate arms of the models are assigned the golimumab change in HAQ.

If a patient has achieved either an ACR20 or ACR50 response, they are assumed to remain in that health state until they withdraw from treatment due to a loss of efficacy or an adverse event. Patients cannot, for example, move from an ACR20 to an ACR50 response in a subsequent cycle. While the patient is in a health state until withdrawal, a constant risk of their HAQ worsening is applied. The manufacturers estimate that patients receiving DMARDs see their HAQ worsen at a rate of 0.045 per year. They assume patients receiving palliative care have a HAQ rate double that of DMARD therapy, at 0.09 per year. Patients receiving TNF-α inhibitor therapy have no worsening of their HAQ while on treatment. They justify their selection of these values based on the NICE TNF-α inhibitor appraisal.[44] Whilst the ERG note that these could be seen as supported values by the appraisal committee, it is essential to highlight the uncertainty around these estimates. The manufacturer's rate of 0.09 per year for patients receiving palliative care is inconsistent with the rate of 0.06 per year assumed in the TA130 appraisal.[3]

For rituximab, the manufacturer assumes a HAQ progression rate equal to those of DMARDs, and not to TNF-α inhibitors. This is not noted in their MS, and the ERGs clinical advisors suggest this underestimates the benefit of rituximab. After requesting clarification with the manufacturer, they confirmed that they have assumed rituximab to have a HAQ progression rate equal to conventional DMARDs. Neither the tocilizumab STA, nor the sequential TNF-α inhibitor MTA provide evidence that the committee accept differential rates of HAQ progression between biologics. The sensitivity of the results to this assumption is explored in Section 6.

In the first cycle of the model, the rate of discontinuation of treatment due to an adverse event is directly estimated from the trial data. For the longer-term extrapolation, Weibull models are fitted to two long term trials to estimate the time spent on each treatment. Kristensen *et al.*

(2006)[45] is used to provide estimates of long term withdrawal for infliximab and etanercept. Edwards *et al* [46] is used to provide estimates of long-term withdrawal for methotrexate. The models were fitted to data extracted from the published Kaplan-Meier curves. The parameters are given in Table 137 of the MS. The adjusted R squared values are 0.93, 0.97 and 0.98 for the infliximab, etanercept and methotrexate regressions respectively. The correlation between the shape and scale parameters is incorporated in the model using the variance covariance matrix and a multinormal distribution. The Kristensen *et al.* (2006)[45] study is a 5 year observational study, and so extrapolation beyond 5 years is likely to result in additional uncertainty in the cost-effectiveness estimates. As there was no evidence on long-term withdrawal for golimumab, adalimumab, certolizumab and rituximab, these were assumed to have equal withdrawal rates to infliximab. The ERG's clinical advisors suggested that there was no theoretical reason why golimumab would have a withdrawal rate different to other TNF-α inhibitors. They did comment that the 20 year mean time spent on 1st line methotrexate based on the Edwards *et al* paper[46] did seem to be very long. Leflunomide is estimated from the paper by Geborek *et al*[47] and other subsequent DMARD therapies are also estimated from the Edwards *et al.* paper, although in this case Weibull models are not fitted and instead an exponential survival curve is assumed.

5.1.2 Treatment effectiveness within the submission

A treatment's effectiveness was modelled as the probability of an ACR20 or ACR50 response at 24 weeks, as well as the length of time spent on a treatment after response. The manufacturer assumes that a patient can withdraw from a treatment due to a loss of efficacy or an adverse event, and when this occurs the patients HAQ rebounds by the initial gain achieved on treatment.

The model samples absolute values of ACR20 and ACR50 responders for golimumab and placebo (methotrexate) from the GO-FORWARD and GO-AFTER trials, for the DMARD experienced and TNF-α inhibitor experienced populations respectively. For the DMARD experienced population, it does not use the MTC analysis which incorporates the Kay *et al.*[26] study and therefore this evidence has been omitted from the model, with no justification given. We conducted an exploratory analysis in section 6 to see whether using the MTC analysis in the economic evaluation to estimate the ACR20 and ACR50 outcomes for placebo relative to golimumab significantly alters the cost-effectiveness estimates.

The model uses the estimated risk ratios from the MTC results for the other comparators. The ACR20 and ACR50 rates for these comparators are sampled independently, and so any

correlation between these outcomes is not captured in the analysis. This methodology seems inappropriate, as essentially ACR is a distribution of responses and ACR20 and ACR50 are therefore correlated. In cases where this independent sampling results in there being more people achieving ACR50 than ACR20, the model forces the number achieving ACR20 to zero. A better approach would have been to have estimated the probability of achieving ACR20 and the probability of achieving ACR50 in those patients achieving ACR20.

The non-response health state is estimated as a remainder (pNR=1-pACR50-pACR20).

The manufacturer has included ACR20 and ACR50 response states, but has omitted ACR70 from the model. As reported in Section 4.2.1, the majority of trials report this outcome as well as ACR20 and ACR50. ACR70 is often seen as a remission state, with patients reporting an improvement in 70% of their affected joints. Omitting this outcome means that a valuable source of clinical effectiveness data has been ignored. When asked to justify the exclusion of this outcome from the model the manufacturer stated that there was no statistically significant difference between the golimumab and comparator interventions in terms of the ACR70 response rate and that incorporating this outcome would only add an additional element of uncertainty around the model outputs. However there is not a statistically significant difference between golimumab and other biologics in terms of ACR50 or ACR20, and these are included in the analysis making this rationale inconsistent with the approach taken. The MTC shows that the mean and median RR estimates for golimumab were inferior compared to adalimumab, certolizumab and etanercept in the DMARD experienced population, and compared to rituximab in the TNF-α inhibitor experienced population. Therefore omitting ACR70 response rates from the analysis has ignored relevant data and is likely to have biased the analysis in favour of golimumab.

The manufacturer does not use the Convergence Diagnostic and Output Analysis (CODA) samples from the WinBugs MTC analysis in the economic evaluation. This would maintain correlation between the efficacy parameters within the PSA. The manufacturer justifies this approach based on their belief that it was improbable that a large correlation would exist between efficacy parameters. Instead the manufacturer uses the median and 95% credible intervals from the MTC and samples from these independently for each comparison within the model.

5.1.3 Health related quality of life

[redacted]. The manufacturer does not undertake a systematic review of HRQoL data in this population.

To derive utility values for the cost-utility analysis, the manufacturer uses a HAQ to utility regression analysis to transform HAQ scores into EQ-5D values. They justify this approach due to the fact that a strong correlation between HAQ and HRQoL has been shown,[41] and that this method has been widely used in other RA cost-utility analyses. The manufacturer's search identified five studies that estimate a relationship between HAQ and HRQoL, and identified Hurst (1997) as the most relevant due to estimating HAQ to EQ-5D, and due to it reporting the full results required for this analysis. This function was used in the BRAM[44] model for the appraisal of adalimumab, infliximab and etanercept (TA130).

The Hurst analysis provides a univariate linear regression with the following form:

$$EQ5D = 0.862 - 0.327 * HAQ$$

With HAQ on a scale of 0-3, the EQ-5D estimates are bounded at 0.862 and at -0.119. The standard errors around the intercept and gradient from this regression (0.034 and 0.0201 respectively) were used in the model to capture uncertainty related to the regression parameters. However, these were each individually sampled from a normal distribution rather than using a multivariate normal to capture correlation between the intercept and gradient.

5.1.4 Resources and costs

Pre-treatment, treatment and monitoring costs were calculated and applied on a first-cycle (6 month) and subsequent cycle basis.

DRUG COSTS

Drug costs were calculated by doses administered, either in the first cycle (6 month) or in subsequent cycle. The drug costs applied in the model are summarised in Table 45. Methotrexate was administered as at 7.5mg weekly dose, which seems low compared to the GO-FORWARD and GO-AFTER trials, which had approximately 15mg doses alongside golimumab. Because methotrexate is a both a comparator and a concomitant therapy to TNF-α inhibitors, it is unlikely to significantly change the incremental cost or the resulting ICERs. After clarification with the manufacturer, they ran a sensitivity analysis with methotrexate

administered 25mg weekly. In the DMARD experienced population this made little difference to the ICERs. In the TNF-α inhibitor experienced population the results changed slightly, and saw rituximab move from a dominated to an extendedly dominated strategy when compared to golimumab. This highlights the sensitivity of the results in this population.

Table 45: Acquisition and administration costs associated with the technology in the economic model [copied from Table 146 of MS]

	Cost per dose	No. doses per first 6 months	No. doses post 6 months	Treatment cost first 6 months	Treatment cost pot 6 months	Cost per administration first 6 months	Total cost first 6 months	Total cost post 6 months
Golimumab	£774.58	6	6	£4,647.48	£4,647.48	£34.00	£4,681.48	£4,647.48
Adalimumab	£357.50	13	13	£4,647.50	£4,647.50	£34.00	£4,681.50	£4,647.50
Infliximab[A]	£419.62	13.35	8.6775	£5,601.93	£3,641.25	£55.00	£6,336.18	£4,118.52
Etanercept	£89.38	52	52	£4,647.76	£4,647.76	£34.00	£4,681.76	£4,647.76
Rituximab[B]	£873.15	6	4	£5,238.90	£3,492.60	£76.00	£5,694.90	£3,796.60
Certolizumab[C]	£357.50	6	13	£2,145.00	£4,647.50	£34.00	£2,179.00	£4,647.50
Leflunomide	£1.70	194.5	182.5	£331.43	£310.98	£0.00	£331.43	£310.98
Gold	£11.23	26	26	£291.98	£291.98	£0.00	£291.98	£291.98
Azathioprine	£0.17	547.5	547.5	£93.08	£93.08	£0.00	£93.08	£93.08
ciclosporin	£2.12	365	365	£773.80	£773.80	£0.00	£773.80	£773.80
Methotrexate	£0.12	78	78	£9.36	£9.36	£0.00	£9.36	£9.36

(A) Cost per dose based on 73kg patient, 194.91mg IFX (2.67 vials with wastage). No doses per first 6 months based on 2.67 vials (average full vials – BSRBR). Cost per administration based on SPC (1 hr infusions if initial 3 well received); (B) No doses based on 6 month dosing frequency. Cost per administration based on SPC (1st infusion ~3hrs, subsequent infusions ~2hrs); (C) No doses per first 6 months adjusted for PAS.

The manufacturer assumes rituximab is re-administered every 6 months. They support this assumption with two European surveys (Appendix 8.15). These surveys contain a substantial number of UK physicians. However, the results of the survey are unlikely to fully reflect UK clinical practice. The surveys have small size, and report estimates of 76% (Synovate, n = 50) and 63% (Therapy Knowledge, n=49) of physicians surveyed administering rituximab at least every 6 months. The Synovate data reports that in 12% of cases rituximab was re-dosed every 2 weeks, which does not seem realistic. There may have been an error of interpretation about the difference between the two infusions in a single course rather than the dosing frequency between courses. If the survey was misinterpreted then the results are likely to be inaccurate. Also, the survey results do not allow us to estimate an average re-treatment time. The SPC[15] for rituximab recommends re-dosing every 6-12 months and clearly states that re-dosing should be no more frequent than every 16 weeks. The TA126 and TA195 NICE guidance specified that re-treatment should occur no more frequently than every 6 months[2,13]. It is fair to assume some clinicians will re-administer every 6 months, and some more infrequently.

Therefore it is inappropriate to assume re-dosing occurs every 6 months and, after consultation with clinical experts, the more likely average re-dosing schedule is every 9 months. This re-dosing schedule is explored in Section 6 by the ERG.

All other dosages are estimated in line with their licensed indication and prices match the BNF59 (March 2010). Infliximab dosage is calculated by weight 3mg/kg, and a weighted average from the British Society for Rheumatology Biologics Registry (BSRBR) is used to calculate a mean number of vials used of 2.67. Vial wastage is assumed.

Because the model estimates an average drug cost for the first cycle, and for subsequent cycles, this allows the decision model to fit into a relatively simple Markov framework. However this methodology means that the cost estimate per cycle is an approximation, rather than a precise calculation.

ADMINISTRATION COSTS

Administrations costs were applied to the estimated doses required. Originally the models incorporated 2006 Reference Costs and 2008 Unit Costs. However, more recent costs were available. The manufacturer incorporated 2008 Reference Costs and 2009 Unit Costs in their clarification response document and these results are reported and used as the starting point for the ERG's analyses in section 6. The results are generally consistent, with only minor changes to the costs of treatment and no substantial change in the incremental cost.

Golimumab, along with adalimumab, etanercept and certolizumab are administered subcutaneously during a one hour specialist nurse consultation (PSSRU 2009 – £36).

Infliximab is administered as a two hour infusion per dose (1 hour infusion subsequently), and rituximab as a three hour infusion (2 hour infusion subsequently).

The ERG requested the manufacturer to undertake sensitivity analyses with different values for the administration of infliximab and rituximab. These assumptions have been a source of uncertainty in previous appraisals. The results of this analysis are shown in Table 46.

Table 46: Manufacturer's sensitivity analysis with revised administration costs

Parameter	Current analysis		Previous appraisal - Review of TA126 and TA141 (FAD guidance)	
	Basecase	Source	Previous value	Appraisal
Infliximab administration	£55	Specialist nurse (£34) + Per hour admin cost (£21)	£284.73	NHS reference costs
Rituximab administration	£76	Specialist nurse (£34) + 2 x Per hour admin cost (2 x £21)	£284.73	

The manufacturer used a previous value of £284.73, which is referred to in the FAD document (Section 4.2.21). The ERG was unable to find more information about the actual source of this estimate.

The BRAM independent assessment group model used a value of £141.83 according to Section 4.2.25 of the FAD. The manufacturer ran sensitivity analyses with £284.73 applied for both infliximab and rituximab, and unsurprisingly they are not cost-effective treatments compared to golimumab.

ADVERSE EVENTS

Adverse events and unscheduled hospital stays are not modelled directly in the model. Instead, to account for the significant number of patients requiring inpatient stays, a multivariate regression function[48] is used. This regression function estimates hospital days, as a function of a patient's age, baseline utility, disease duration, number of previous DMARDs and whether a patient is on a TNF-α inhibitor.

5.1.5 Discounting

Discounting was applied to costs and QALYS, at 3.5% per annum as required in the NICE Guide to the Methods of Technology Appraisal.[43]

5.1.6 Sensitivity analyses

The manufacturer undertakes scenario analyses to estimate the uncertainty around structural assumptions. In particular, the manufacturer investigates the impact of changing the 'rebound' assumption for when patients withdraw from a treatment. In the base case analysis, a patients' HAQ rebounds by an amount equal to the original gain. This is equivalent to a patients' HAQ

returning to their baseline score. In the scenario analysis, the rebound is equivalent to the HAQ score equal to the natural history of primary non-responders. This implies that a patient loses all the benefit accrued by TNF-α inhibitor therapy.

The manufacturer also performs a range of one-way sensitivity analyses, as well as probabilistic sensitivity analysis (PSA). The PSA was run for 2,000 simulations; however no justification for this number of runs was given. After clarification with the manufacturer, it was shown that the mean estimates were stable at 2,000 runs.

PSA results are presented, in the form of CEACs within the MS, although mean costs and QALYs from the PSA were provided in response to a request from the ERG. Only deterministic results were presented in the MS for the scenario analyses undertaken.

The distributions used for the PSA are described in MS section 6.3.6. Beta distributions are applied to the ACR response probabilities and adverse event discontinuation rates, log-normal distributions to the risk ratios and costs. Normal distributions are applied to the HAQ parameters (although truncated on 0-3 scale), utility algorithm and standardised mortality rate. Gamma distributions are assigned for the cost of chest X-rays and rheumatology visits. It is not clear why log-normal distributions have been chosen for some costs and gamma distributions for others.

5.1.7 Model validation

The MS states that the model was validated by an experienced programmer. This was to check the accuracy of inputs, undertake top down testing, testing key sensitivities and checking the submission document. [redacted] The ERG group have not fully validated both submitted models, and have instead assumed that apart from the treatment comparisons and patient populations chosen, that there is no difference between the two models. The manufacturer confirmed this in response to the ERGs clarification document.

5.2 Critique of approach used

The use of a Markov framework to model the probability and time spent in health states means the model is rather complex. The selected treatments have been hard-coded into the

model, which means that full validation of the programming is very time consuming. A number of minor errors have been found in the computation of the Markov states in the model. One error is that costs are accrued for patients in the dead state for infliximab but not for other treatments, suggesting that a programming error has been made in the infliximab arm. This has been corrected in Section 6. In addition, the method used to calculate HAQ decrements in the certolizumab Markov sheet differs from that used in the Markov sheets for the other TNF-α inhibitors. This has been corrected in Section 6.

Ideally, the ERG would incorporate the ACR70 response rate data to more accurately evaluate the cost-effectiveness of golimumab. This would use all available clinical evidence in evaluating golimumab. As discussed, omitting ACR70 is likely to have biased the results in favour of golimumab. The model framework developed by the manufacturer means that incorporating this additional health state would require almost complete rebuilding of the model and so was not possible for either the ERG or manufacturer to undertake.

A more appropriate methodology would be to use individual patient sampling techniques. This would allow easier structural changes and more varied sensitivity analysis to be performed. The methods are transparent and can be modelled in Microsoft Excel, and have been widely used in RA economic evaluations.[49]

As discussed in Section 3.3, the analysis does not allow the optimal position of golimumab within the treatment sequence to be evaluated. A conclusion could be that golimumab is cost-effective in both the DMARD and TNF-α inhibitor experienced populations. However, the comparison between golimumab in these two positions cannot be undertaken.

In the DMARD experienced population, failure on TNF-α inhibitor treatment sees patients progress to leflunomide. However, this completely contradicts the comparison made in the TNF-α inhibitor experienced population, where rituximab is the established treatment. The ERG asked the manufacturer in their clarification letter to alter the DMARD experienced model so that all patients progress to rituximab, and then onto leflunomide and subsequent therapies. The results of this sensitivity analysis are provided below. The results only alter slightly, as adding an extra line of treatment for all comparators is unlikely to substantially alter the incremental costs or QALYs between therapies.

Table 47: Manufacturer's sensitivity analysis – Rituximab after TNF-α inhibitor in DMARD experienced population model

	Comparator 1	**Comparator 2**	**Comparator 3**	**Comparator 4**	**Comparator 5**	**Comparator 6**
1st Line	Golimumab	adalimumab	Infliximab	etanercept	Certolizumab	Methotrexate
2nd Line	Rituximab	rituximab	Rituximab	rituximab	rituximab	rituximab
3rd Line	leflunomide	leflunomide	Leflunomide	leflunomide	leflunomide	leflunomide
4th Line	Gold	Gold	Gold	Gold	Gold	Gold
5th Line	azathioprine	azathioprine	Azathioprine	azathioprine	azathioprine	azathioprine
6th Line	ciclosporin	ciclosporin	Ciclosporin	ciclosporin	ciclosporin	ciclosporin
7th Line	Palliative Care	Palliative Care	Palliative Care	Palliative Care	Palliative Care	Palliative Care

Table 48: DMARD experienced population – Rituximab included in treatment pathway after TNF-α inhibitor failure

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	40,855	4.451	-	-
Infliximab	74,660	5.545	Dominated by golimumab	30,900
Certolizumab	71,542	5.690	Extendedly dominated by etanercept	24,768
Adalimumab	77,817	5.720	Dominated by golimumab	29,127
Golimumab	72,379	5.726	Extendedly dominated by etanercept	24,725
Etanercept	74,208	6.133	19,829	19,829

5.3 Results included in manufacturer's submission

This section presents the main results from the manufacturer's economic evaluation.

DMARD experienced population

The deterministic result for the DMARD experienced population is presented in Table 150 of the MS. They report an incremental analysis, however the method of presentation is slightly confusing and so the results have been presented below in a revised table.

Table 49: DMARD experienced population – manufacturer's basecase results

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	35,869	4.569	-	-
Infliximab	69,899	5.651	Dominated by golimumab	31,451
Certolizumab	73,571	5.768	Dominated by golimumab	31,444
Adalimumab	66,875	5.792	Extendedly dominated by etanercept	25,352
Golimumab	67,747	5.827	Extendedly dominated by etanercept	25,340
Etanercept	74,208	6.133	24,513	24,513

[§] Indicates cost-effectiveness when all other biologics contraindicated

Table 49 shows that infliximab and certolizumab are both dominated by golimumab, because golimumab is more effective and less costly. The remaining strategies all have very similar ICERs when compared to methotrexate, at around £25,000. The incremental analysis shows that adalimumab and golimumab are both extendedly dominated by etanercept. Etanercept generates the most QALYs of any strategy, but at a lower cost per QALY ratio. The full incremental analysis shows that etanercept is the optimal strategy, with an ICER of £21,000 compared to golimumab. The analysis also shows that golimumab is a cost-effective strategy when compared to infliximab and certolizumab, which are already recommended by NICE.

Figure 3: The results from the manufacturer's deterministic basecase in the DMARD experienced population

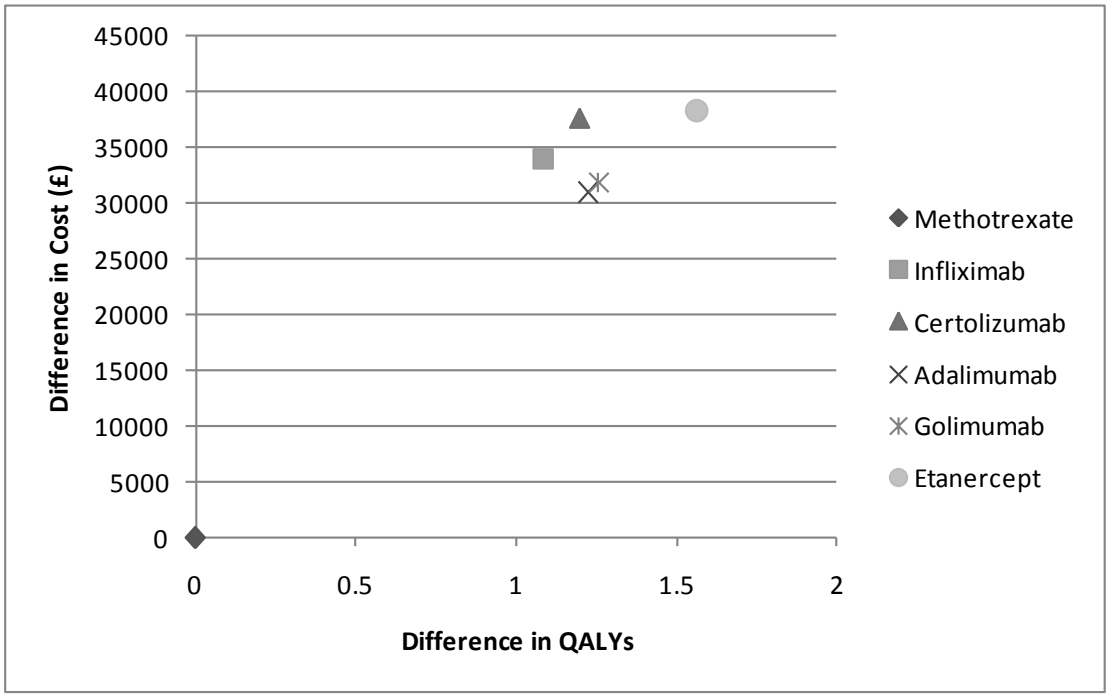


Figure 3 shows the incremental costs and QALYs compared to methotrexate. It shows that adalimumab and golimumab dominate infliximab and certolizumab, and it also shows how etanercept extendedly dominates adalimumab and golimumab (due to the lower slope from the origin).

In the manufacturer's deterministic basecase analysis, golimumab dominates some treatments already recommended by NICE, however it is not the optimal strategy.

Basecase PSA results for the DMARD experienced population

The manufacturer presents the results of the PSA, in the form of a CEAC within the MS, although mean costs and QALYs from the PSA were provided in response to a request from the ERG.

For the DMARD experienced population, the cost-effectiveness results based on the mean costs and QALYs from the PSA are consistent with the deterministic analysis in that golimumab dominates some treatments already recommended by NICE, however, it is not the most cost-effective strategy as it is extendedly dominated by etanercept. Golimumab has a mean ICER of £25,800 relative to methotrexate. At a willingness to pay of £20,000 per QALY, golimumab is the most cost-effective intervention in 5% of PSA samples whilst for

comparison methotrexate is most cost-effective in 56% of samples followed by etanercept in 17% of samples. At £30,000 per QALY golimumab is the most cost-effective intervention in 8% of samples. For comparison, etanercept is most cost-effective in 32% of samples followed by methotrexate which is most cost-effective in 24% of samples.

TNF-α inhibitor experienced population

The results for the manufacturer's deterministic base case analysis of golimumab in a TNF-α inhibitor experienced population are provided in Table 151 of the MS. Again, the ERG have provided a revised table below to make the incremental analysis more clear.

Table 50 shows that rituximab is dominated by golimumab, as golimumab is less costly and more effective. Golimumab compared to methotrexate has an ICER of £28,000.

Table 50: TNF-α inhibitor experienced population – manufacturer's basecase results

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	33,673	3.129	-	-
Rituximab	50,206	3.523	Dominated by golimumab	41,961
Golimumab	50,175	3.712	28,305	28,305

[§] Indicates cost-effectiveness when all other biologics contraindicated

Figure 4: The results from the manufacturer's deterministic basecase in the TNF-α inhibitor experienced population

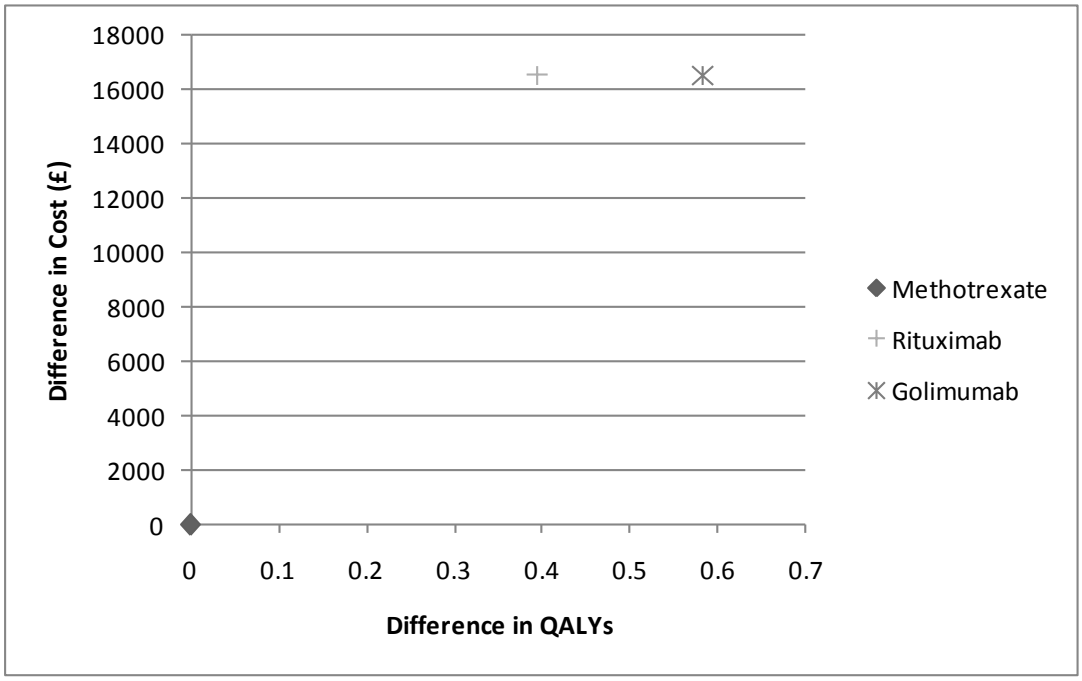


Figure 4 shows how golimumab dominates rituximab due to the provision of 0.19 QALYs at a slightly lower cost (£31).

Basecase PSA results for the TNF-α inhibitor experienced population

The manufacturer presents the results of the PSA, in the form of a CEAC within the MS, although mean costs and QALYs from the PSA were provided in response to a request from the ERG.

In the TNF-α inhibitor experienced population, rituximab is extendedly dominated by golimumab based on the mean costs and QALYs from the PSA and golimumab has an ICER of £29,100 compared to methotrexate. At a willingness to pay threshold of £20,000 per QALY golimumab is most cost-effective in 5% of PSA samples. For comparison methotrexate is most cost-effective in 90%. At a willingness to pay threshold of £30,000 per QALY, golimumab and methotrexate have a similar probability of being most cost-effective (46% and 44% respectively).

SENSITIVITY ANALYSIS

The ERG repeated the scenario analysis performed by the manufacturer, as indicated in Table 152 of the MS. The table is replicated below, and the ERG confirmed the estimated ICERs for golimumab vs methotrexate, which are based on the deterministic model.

Table 51: DMARD experienced population – scenario analyses

Variable	Basecase	Sensitivity analysis	ICER vs Methotrexate
Base case analysis			25,346
Time Horizon	43 years (lifetime)	5 years	95,809
		10 years	56,221
Discount rate	3.5% for costs and QALYs	0% costs 0% QALYs	19,247
		0% costs 3.5% QALYs	31,191
		3.5% costs 0% QALYs	15,640
Hospitalisation costs	Included	Excluded	32,382
Age	50 years	45 years	23,272
		60 years	32,681
Efficacy of golimumab (ACR20)	0.213	-20%	26,041
		+20%	24,786
Efficacy of golimumab (ACR50)	0.382	-20%	27,505
		+20%	23,900
Efficacy of golimumab (ACR20 and ACR50)	0.213 / 0.382	-20%	28,692
		+20%	23,532
SMR	1.65	2.5th (1.34)	24.382
		97.5th (1.98)	26,317
HAQ progression	0 TNF-α inhibitor's, 0.0225 DMARDs, 0.0450	0 for all	132,906
		0.0225 for all	115,795
		0 TNF-α inhibitors, 0.0225 non-TNF-α inhibitors	33,219
		0.015 TNF-α inhibitors, 0.0225	39,055

	Palliative Care	DMARDs, 0.0450 Palliative Care	
Baseline HAQ score	1.41	-50%	25,323
		+50%	25,366
Golimumab annual acquisition cost	Equivalent to adalimumab	-20%	18,797
		+20%	31,895
Natural history HAQ progression*	0.0719	0.1018	£39,491
Long term withdrawal	Equal to infliximab	Equal to etanercept	24,965

Table 52: TNF-α inhibitor experienced population – scenario analyses

Variable	Basecase	Sensitivity analysis	ICER vs Methotrexate
Base case analysis			28,286
Time Horizon	43 years (lifetime)	5 years	115,012
		10 years	61,537
Discount rate	3.5% for costs and QALYs	0% costs 0% QALYs	21,040
		0% costs 3.5% QALYs	32,367
		3.5% costs 0% QALYs	18,387
Hospitalisation costs	Included	Excluded	41,254
Age	54 years	49 years	25,132
		64 years	39,466
Efficacy of golimumab (ACR20)	0.157	-20%	30,677
		+20%	26,631
Efficacy of golimumab (ACR50)	0.183	-20%	30,677
		+20%	26,631

Efficacy of golimumab (ACR20 and ACR50)	0.157 / 0183	-20%	31,908
		+20%	26,329
SMR	1.65	2.5th (1.34)	26,838
		97.5th (1.98)	29,740
HAQ progression	0 TNF-α inhibitors, 0.0225 DMARDs, 0.0450 Palliative Care	0 for all	146,172
		0.0225 for all	126,515
		0 TNF-α inhibitors, 0.0225 non-TNF-α inhibitors	36,067
		0.015 TNF-α inhibitors, 0.0225 DMARDs, 0.0450 Palliative Care	44,245
Baseline HAQ score	1.41	-50%	28,267
		+50%	28,302
Golimumab annual acquisition cost	Equivalent to adalimumab	-20%	19,966
		+20%	36,598
Natural history HAQ progression	Equal to gain	Equal to natural history	£42,237
Long term withdrawal	Equal to infliximab	Equal to etanercept	27,928

The manufacturer undertakes a scenario analysis around the HAQ rebound assumption when a patient withdraws from treatment. In the base case analysis, the patients' HAQ rebound is equal to the gain achieved. This allows for structural benefit accrued whilst on a treatment to be incorporated. The manufacturer undertakes a sensitivity analysis with the HAQ rebound causing the HAQ to return to the population natural history level. This is a conservative assumption that suggests that there are no long-term benefits of active interventions. In the DMARD experienced population, the incremental analysis (see Table 53 and Figure 5) shows that infliximab and etanercept are dominated strategies, and golimumab and adalimumab are extendedly dominated by certolizumab. The QALY gain of certolizumab in this scenario analysis is better than that for etanercept whereas in the basecase analysis the opposite is true. This difference is counterintuitive as all TNF-α inhibitors should be similarly affected by this change in assumption. This discrepancy lead to the ERG identifying differences between the Markov sheets used in the model for the individual TNF-α inhibitors. The modelling of HAQ

decrements for certolizumab (columns "ER:EX" of sheet "Comp6") differs from the method used for comparator drugs which appears to be an error as no difference is expected from the methods reported in the MS. This error is corrected in section 6.

Table 53: Rebound back to natural history HAQ (DMARD experienced population)

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	35,869	4.489	-	-
Infliximab	69,899	5.174	Dominated by golimumab	49,678
Adalimumab	66,875	5.276	Extendedly dominated by certolizumab	39,397
Golimumab	67,747	5.297	Extendedly dominated by certolizumab	39,452
Etanercept	74,208	5.554	Dominated by certolizumab	35,999
Certolizumab	73,571	5.675	31,789	31,789

[§] Indicates cost-effectiveness when all other biologics contraindicated

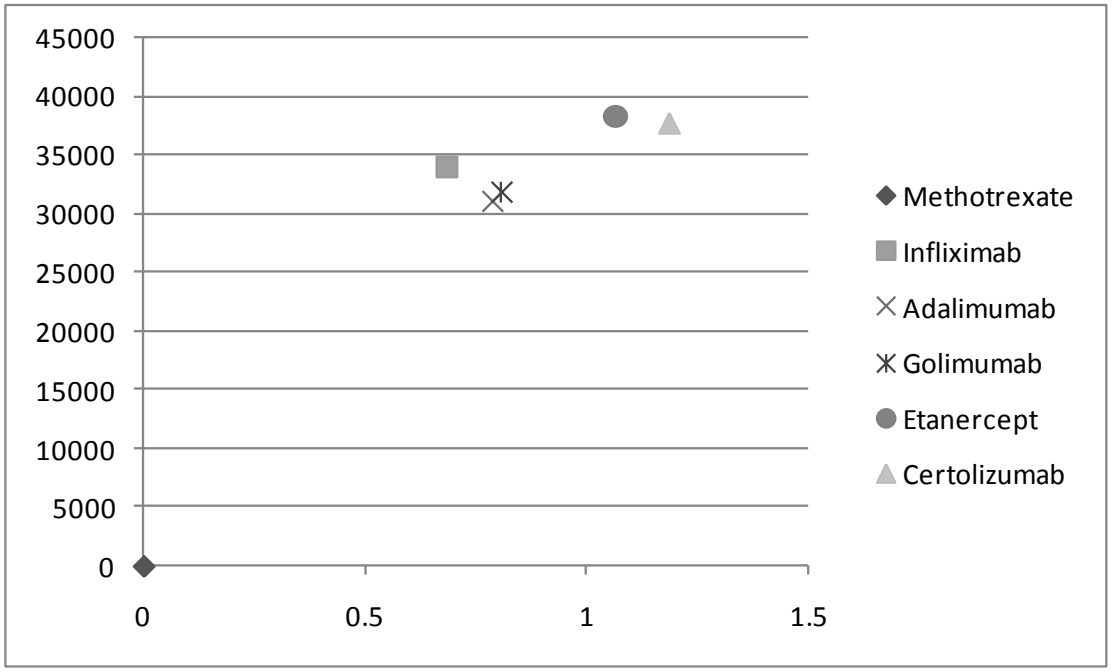
The ERG were unable to replicate the HAQ rebound sensitivity analysis results provided by the MS. The MS results can be found in Table 155. The results estimated when the ERG ran the sensitivity analysis are provided in Table 54. The conclusion in the TNF-α inhibitor experienced population are not significantly altered by this structural sensitivity analysis in that golimumab is still more cost-effective than rituximab (see Table 54). This is the case for both the ERG and the MS results, although the ERG results suggest that golimumab fully, rather than extendedly dominates rituximab.

Table 54: Rebound back to natural history HAQ (TNF-α experienced population)

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	33,673	3.095	-	-
Rituximab	50,206	3.393	Dominated by golimumab	-
Golimumab	52,175	3.486	42,237	42,237

[§] Indicates cost-effectiveness when all other biologics contraindicated

Figure 5: The results from the manufacturer's deterministic basecase in the DMARD experienced population



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

The model results (total costs and QALYs, time in states, HAQ scores and incremental costs and QALYs) appear plausible given the parameter inputs. The manufacturer does not provide a full comparison and face validation, although Table 149 in the MS represents a partially filled comparison to the clinical evidence. The ERG have validated the model by altering parameters and checking results. The validation found some programming errors in the model which are explored in Section 6. These did not change the conclusion that golimumab has an ICER compared to methotrexate that is comparable to other TNF-α inhibitors, but that golimumab is not the most cost-effective TNF-α inhibitor.

5.5 Summary of uncertainties and issues

The uncertainty in the parameter estimates has been appropriately incorporated in the PSA. The PSA provides similar mean estimates of costs and QALYs compared to the deterministic results. The CEACs provided in the MS, which indicate the likelihood that an intervention is the most cost-effective within the probabilistic analyses are reproduced in Figures 6 and 7 for reference. In the DMARD experienced population the probability of being most cost-effective was less than 33% over a threshold range of £20,000 to £30,000 per QALY for all biologics.

In the TNF-α inhibitor experienced population, rituximab has a probability of less than 10% of being the most cost-effective over the threshold range of £20,000 to £30,000 per QALY. Methotrexate has a high probability of being most cost-effective at £20,000 per QALY and golimumab has a high probability of being most cost-effective at a threshold of £30,000 per QALY.

Figure 6: Manufacturer's basecase analysis CEAC (DMARD experienced)

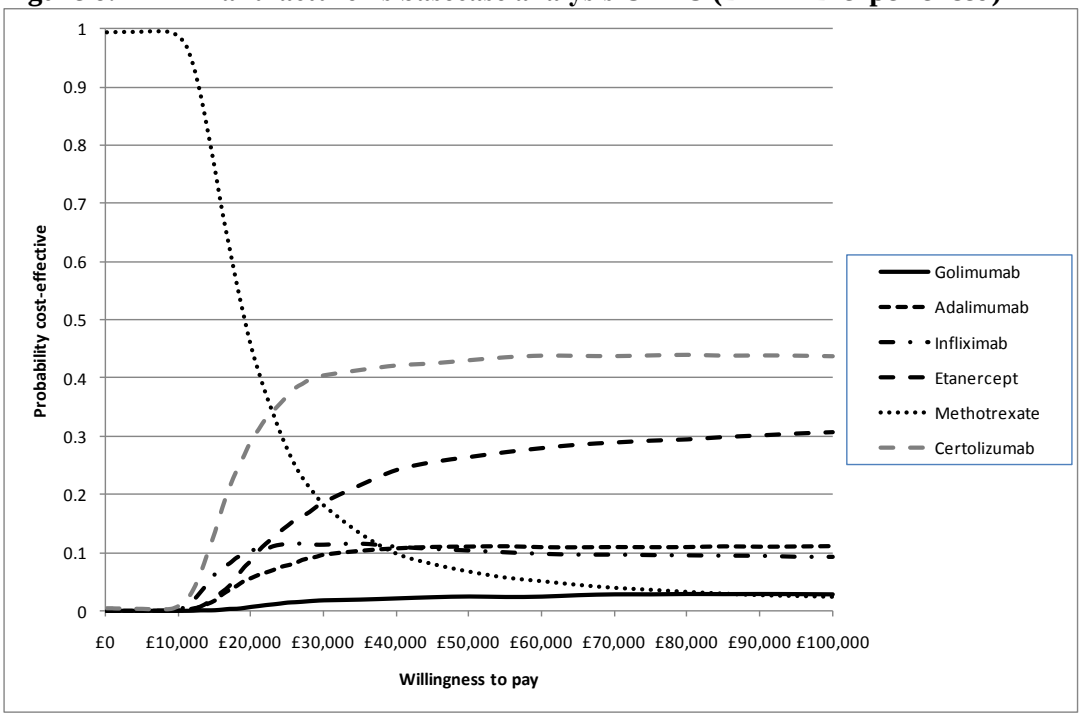
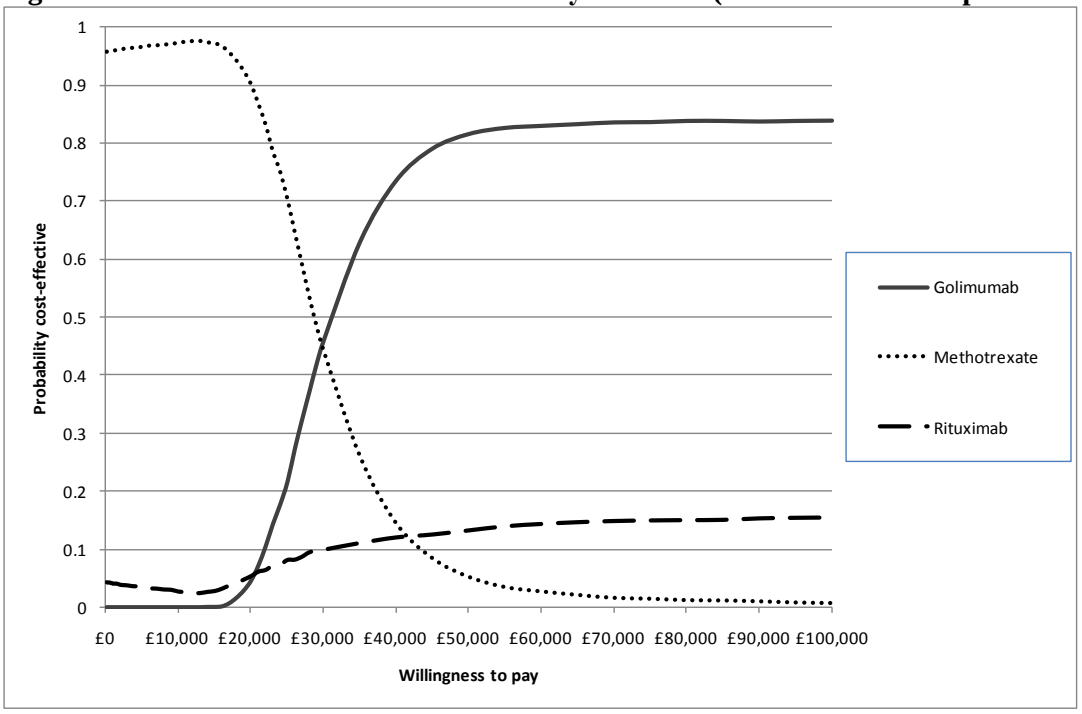


Figure 7: Manufacturer's basecase analysis CEAC (TNF-α inhibitor experienced)



The manufacturer's scenario analyses highlight key uncertainties around the HAQ progression rate, and the rebound after withdrawal assumption. The effect of these uncertainties are explored in additional work undertaken by the ERG.

6 ADDITIONAL WORK UNDERTAKEN BY THE ERG

For a number of the issues raised by the ERG, the manufacturer has undertaken one-way sensitivity analysis and provided the results in their response to the ERG clarification letter. Where appropriate these results have been reported in the previous section. However the ERG was not convinced that the manufacturer's basecase analysis or any of their one-way analyses represented the most plausible set of assumptions for this analysis. Therefore, the ERG has undertaken a step-wise analysis for a list of alternative assumptions, to derive a final set of results that represents the ERGs 'base case' analysis. The results are presented in a step-wise fashion to ensure the results change as expected.

The ERG made the following amendments to the DMARD experienced model:

1. Uses the manufacturer's model with updated unit and reference costs
2. Infliximab costs corrected and HAQ decrements in certolizumab arm corrected
3. Uses the mixed treatment comparison to estimate the relative risk of placebo compared to golimumab

The ERG made the following amendments to the TNF-α inhibitor experienced model:

4. Uses the manufacturers model with updated unit and reference costs
5. Rituximab HAQ progression equal to TNF-α (Progression rate = 0).
6. Rituximab re-administered every 9 months

Details of each change are listed below. In the follow section all results and ICERs are reported. The final section provides a conclusion on the ERG basecase results.

1. Uses the manufacturers model with updated unit and reference costs

The original model used 2006 Reference Case costs, and 2008 Unit Costs. However more recent costs are available, and the ERG asked the manufacturer to incorporate these. The manufacturer incorporated 2008 Reference Costs and 2009 Unit Costs in their clarification response document, and these estimates are used as the starting position for the step-wise revisions to the model. These revised unit costs are carried forward into analysis 2.

2. Infliximab costs and HAQ decrements in certolizumab arm corrected

The Markov model sheets contain and error for infliximab in the DMARD experienced population. This error incorrectly attributes a cost when a patient dies. The modelling of HAQ decrements for certolizumab differs from the method used for comparator drugs which appears to be an error as no difference is expected from the methods reported in the MS. As each of these errors only affect a single comparator, both these errors are corrected here as

their individual impact can be seen in the results table. These changes are carried forward into analysis 3.

3. Uses the mixed treatment comparison to estimate the relative risk of placebo compared to golimumab

The basecase model used the event rates from the GO-FOWARD trial to estimate the probability of ACR response and the probability of withdrawal due to adverse event at 6 months in the golimumab and placebo arms. However it used the mixed treatment comparison to estimate the rate of these events in the other comparator arms. This approach excluded the evidence from the Kay *et al.* [26] study from the economic model. In this analysis we use the mixed treatment comparison, which incorporates the Kay *et al.* [26] study, to estimate the probability of these outcomes for placebo which is used to populate the methotrexate arm of the economic model.

4. Uses the manufacturers model with updated unit and reference costs

As described above (1) for the DMARD experienced population. This was also done for the TNF-α inhibitor experienced population. These changes were carried forward to analysis 5.

5. Rituximab – zero HAQ progression rate

The basecase model assumes that rituximab has a HAQ progression rate equal to DMARDs (0.045) as opposed to TNF-α inhibitors (zero). The ERG believes that this assumption underestimates the benefit of rituximab, and so have altered the model to assume that rituximab has no HAQ progression rate (equal to that of TNF-α inhibitors). This change is carried forward into analysis 6.

a) Rituximab – 0.03 HAQ progression rate

Evidence suggests that patients on active treatment progress at a rate of 0.03 per year. This rate is incorporated for rituximab only in this analysis and is not carried forward to analysis 6.

b) Rituximab and TNF-α inhibitors – 0.03 HAQ progression rate

The 0.03 HAQ progression rate is applied to both rituximab and all TNF-α inhibitors. This rate is incorporated for rituximab only in this analysis and is not carried forward to analysis 6.

6. Rituximab re-treated every 9 months

The assumption that rituximab was re-administered every 6 months is discussed previously. The ERG believes that 9 months on average is more reflective of current clinical practice. This assumption was incorporated in the model, with every patient receiving 4 x 50ml doses in the first 6 months, and then an average of 2.6 doses every subsequent cycle.

Analyses 3 and 6 represent the scenarios considered by the ERG to contain the most plausible assumptions for the DMARD and TNF-α inhibitor experienced populations respectively. The probabilistic analyses for scenarios 3and 6 are presented after the deterministic scenario analyses in Tables 63 and 64.

RESULTS

The step-wise changes to the models show logical changes in the results. Analyses 3 and 6 represent the scenarios considered by the ERG to contain the most plausible assumptions for the DMARD and TNF-α inhibitor experienced populations respectively.

Updating the unit and reference costs, had little impact on the incremental costs between the treatments, and so the resulting ICER's did not change substantially. Correcting the infliximab costs reduces the total cost of infliximab treatment, and it is no longer dominated by adalimumab. Correcting the HAQ progression in the certolizumab arm meant that it became the most cost-effective intervention instead of etanercept. Using the meta-analysis rather than the GO-FORWARD study alone to inform the golimumab versus methotrexate comparison did not substantially alter the results. The probabilistic analysis for the ERGs preferred scenario agrees with the deterministic analysis in that golimumab is still extendedly dominated. The ICER for golimumab compared to methotrexate is £25,000 per QALY.

As expected, the model is extremely sensitive to the assumptions made regarding rituximab and TNF-α inhibitor HAQ progression. Finally, if rituximab is assumed to be re-administered every 9 months, then it dominates golimumab. The probabilistic analysis for this scenario also estimates that rituximab dominates golimumab and rituximab is the most cost-effective strategy in over 85% of PSA samples when considering an ICER threshold between £20,000 and £30,000 per QALY.

The results of these extra analyses highlight key sensitivities in the model, most notably around the re-treatment rate of rituximab, and the HAQ progression rate of rituximab relative to that of golimumab.

DMARD experienced population

1. Uses the manufacturers model with updated unit and reference costs

Table 55: DMARD experienced population – updated costs

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	39,589	4.569	-	-
Infliximab	75,904	5.649	Dominated by adalimumab and golimumab	33,628
Certolizumab	76,868	5.768	Dominated by adalimumab and golimumab	31,086
Adalimumab	70,376	5.790	Extendedly dominated by etanercept	25,211
Golimumab	71,229	5.825	Extendedly dominated by etanercept	25,193
Etanercept	77,548	6.131	24,302	24,301

[§] Indicates cost-effectiveness when all other biologics contraindicated

2. Infliximab costs and HAQ decrement in certolizumab arm corrected

Table 56: Infliximab costs and HAQ decrement in certolizumab arm corrected

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	39,589	4.569	-	-
Infliximab	66,144	5.649	Extendedly dominated by certolizumab	24,589
Adalimumab	70,376	5.790	Extendedly dominated by certolizumab	25,211
Golimumab	71,229	5.825	Extendedly dominated by certolizumab	25,193
Etanercept	77,548	6.131	Dominated by certolizumab	24,301
Certolizumab	76,868	6.341	21,040	21,040

[§] Indicates cost-effectiveness when all other biologics contraindicated

3. Uses the mixed treatment comparison to estimate the relative risk of placebo compared to golimumab

Table 57: DMARD experienced population – MTC to estimate relative risk of placebo compared to golimumab

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	39,611	4.550	-	-
Infliximab	66,144	5.649	Extendedly dominated by certolizumab	24,137
Adalimumab	70,376	5.790	Extendedly dominated by certolizumab	24,800
Golimumab	71,229	5.825	Extendedly dominated by certolizumab	24,794
Etanercept	77,548	6.131	Dominated by certolizumab	23,990
Certolizumab	76,868	6.341	20,800	20,800

[§] Indicates cost-effectiveness when all other biologics contraindicated

TNF-α experienced population

4. Uses the manufacturers model with updated unit and reference costs

Table 58: TNF-α inhibitor experienced population – updated costs

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	37,134	3.129	-	-
Rituximab	53,530	3.522	Dominated by golimumab	41,622
Golimumab	53,519	3.711	28,115	28,115

[§] Indicates cost-effectiveness when all other biologics contraindicated

5. Rituximab HAQ progression zero

Table 59: TNF-α inhibitor experienced population – Rituximab HAQ progression rate equal to zero

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	37,134	3.129	-	-
Golimumab	53,519	3.711	Extendedly dominated by rituximab	28,115
Rituximab	53,530	3.898	21,306	21,306

[§] Indicates cost-effectiveness when all other biologics contraindicated

5a. Rituximab HAQ progression 0.03

Table 60: TNF-α inhibitor experienced population - Rituximab HAQ progression rate equal to 0.03

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	37,134	3.129	-	-
Rituximab	53,530	3.647	Dominated by golimumab	31,621
Golimumab	53,519	3.711	28,115	28,115

[§] Indicates cost-effectiveness when all other biologics contraindicated

5b. Rituximab and Golimumab HAQ progression 0.03

Table 61: TNF-α inhibitor experienced population – Rituximab and Golimumab HAQ progression rate equal to 0.03

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	37,134	3.129	-	-
Golimumab	53,519	3.501	Extendedly dominated by rituximab	44,997
Rituximab	53,530	3.647	31,621	31,621

[§] Indicates cost-effectiveness when all other biologics contraindicated

6. Rituximab re-administered every 9 months

Table 62: TNF-α inhibitor experienced population – Rituximab re-administered every 9 months

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	37,134	3.129	-	-
Golimumab	53,519	3.711	Dominated by rituximab	28,115
Rituximab	44,897	3.898	10,088	10,088

[§] Indicates cost-effectiveness when all other biologics contraindicated

Probabilistic results for the ERG's preferred scenarios

Table 63: DMARD experienced population – ERG basecase probabilistic results

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	39,701	4.622	-	-
Infliximab	67,528	5.792	Extendedly dominated by certolizumab	23,774
Golimumab	71,530	5.889	Extendedly dominated by certolizumab	25,123
Adalimumab	72,824	5.968	Extendedly dominated by certolizumab	24,604
Etanercept	80,096	6.307	Dominated by certolizumab	23,966
Certolizumab	79,185	6.518	20,828	20,828

[§] Indicates cost-effectiveness when all other biologics contraindicated

Table 64: TNF-α inhibitor experienced population - ERG basecase probabilistic results

Technology	Total Costs (£)	Total QALYs	Incremental Analysis – comparison made to next least effective non-dominated strategy ICER (£/QALY)	ICER (£/QALY) versus Methotrexate[§]
Methotrexate	37,162	3.262	-	-
Golimumab	53,565	3.831	Golimumab is dominated by rituximab	28,804
Rituximab	43,697	3.892	10,359	10,359

[§] Indicates cost-effectiveness when all other biologics contraindicated

Figure 8: CEAC for DMARD experienced population – ERG basecase

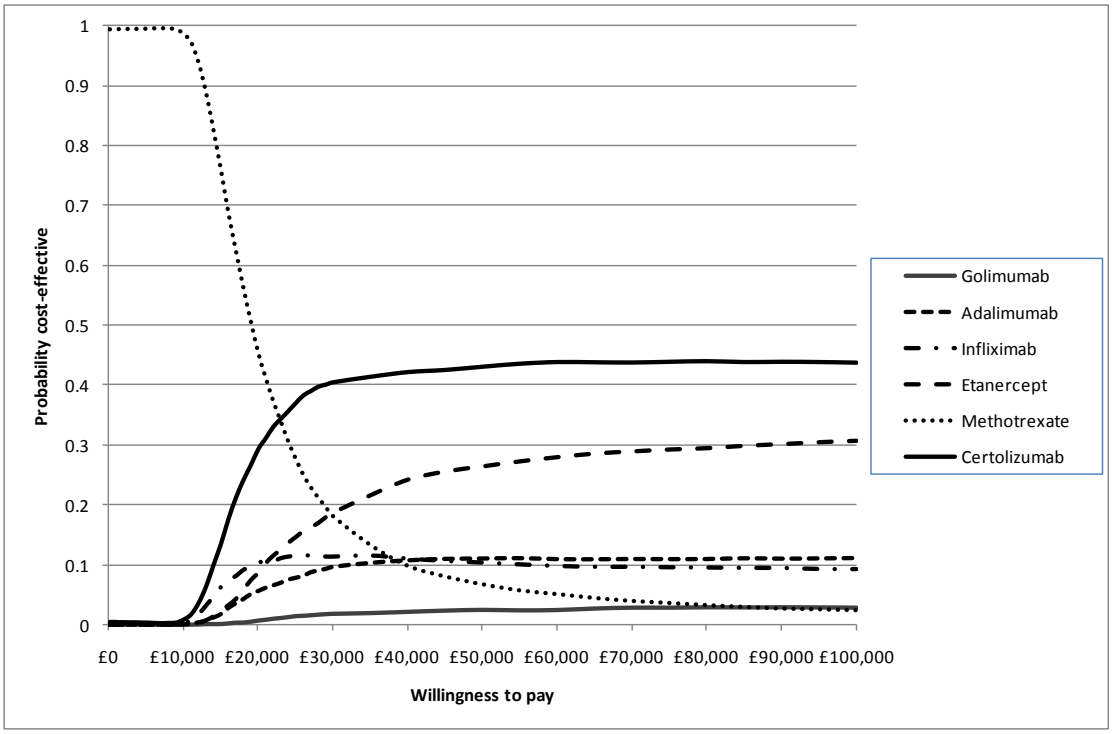
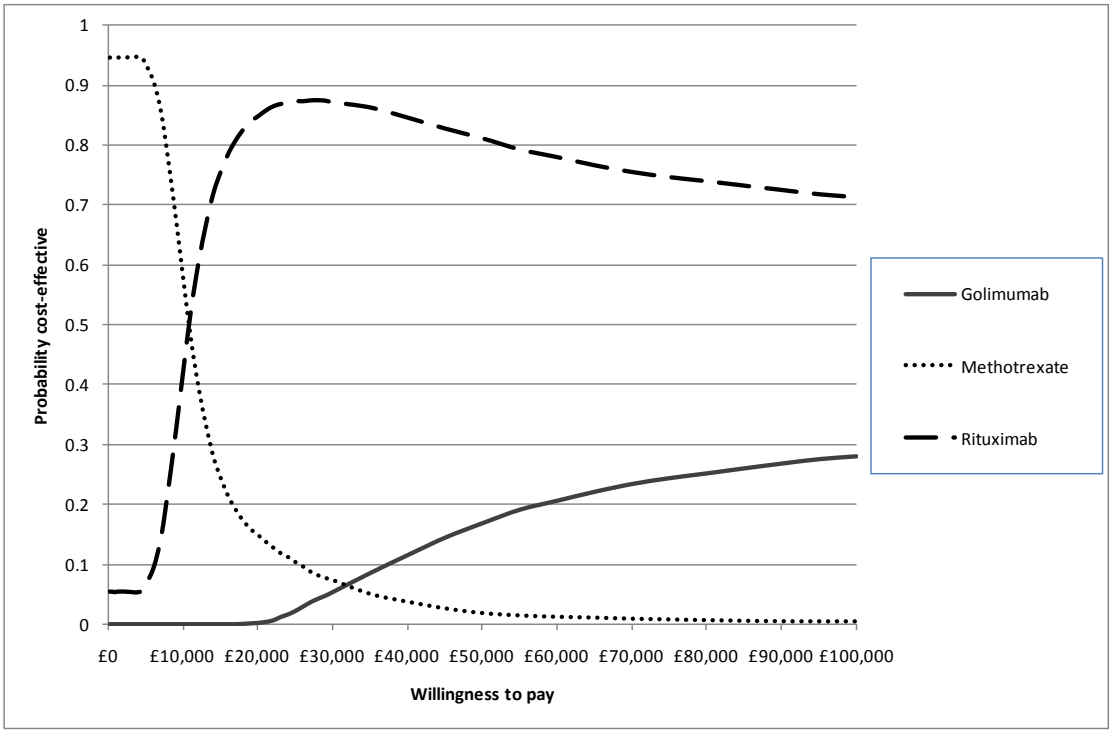


Figure 9: CEAC for TNF-α inhibitor experienced population - ERG basecase



7 DISCUSSION

7.1 Summary of clinical effectiveness issues

The submission can be considered to represent a balanced estimate of the effects of golimumab relative to comparator therapies for the efficacy and safety outcomes and comparators that were included. Clinical effectiveness review methods and results were reasonably clearly presented, with adequate systematic searches conducted. All relevant RCTs for golimumab and comparators appeared to have been included. The included golimumab trials were each of reasonable methodological quality and were considered by the clinical advisors to the ERG to have fair generalisability to the UK population. A good level of detail was provided for clinical effectiveness evidence in the MS and supplementary manufacturer responses.

The mixed treatment comparisons and indirect comparisons used appropriate trials to inform the networks of evidence. These analyses allow a comparison of all comparators within the cost-effectiveness analysis.

The clinical efficacy data presented could be considered to have only partially addressed the decision problem, in that tocilizumab and abatacept were not incorporated into analyses and golimumab is not compared to other TNF-α inhibitors (adalimumab, infliximab, abatacept, etanercept) in the TNF-α inhibitor experienced population. Furthermore, not all relevant outcomes were addressed, for example radiological progression of joint damage.

The confidence intervals and credible intervals surrounding key parameters in the model are wide and no definitive conclusion can be made regarding efficacy, serious adverse events, serious infections, injection site reactions or discontinuations due to adverse events.

The absence of any available head-to-head trials for golimumab versus comparator therapies represents a source of uncertainty in the clinical effectiveness evidence base, although golimumab and comparators were evaluated against each other using mixed treatment comparisons and indirect comparisons in the MS.

The performance of golimumab and comparators were not assessed across the full range of outcomes pre-specified in the scope. Of particular interest would be the impact of golimumab vs comparator drugs in terms of radiological progression and the potential impact this may have on the cost-effectiveness estimates were this outcome to be incorporated in the model.

The potential implications of the monthly dosing frequency should be taken into consideration, in terms of both patient benefit and harm. The reduced dosing frequency relative to comparator therapies may improve patient convenience. The SPC[14] states that golimumab should be discontinued in the event of a patient developing a serious infection or sepsis. However, it is not clear whether the longer dosing frequency of golimumab could potentially delay the withdrawal of immune system suppression in patients receiving golimumab who present with infections and may have implications for their management.

7.2 Summary of cost effectiveness issues

The results of the manufacturer's analysis suggest that in the DMARD experienced population, golimumab has an ICER versus methotrexate which is comparable to that of other TNF-α inhibitors. In the manufacturer's deterministic basecase analysis golimumab dominates some treatments already recommended by NICE. However, it is not the most cost-effective strategy. This is also true when using the mean results from the probabilistic sensitivity analysis. The results are generally robust under the scenario analyses conducted, but must be considered in light of the fact that ACR70 responses were not included in the analysis. The manufacturer's failure to incorporate ACR70 is likely to have biased the results in favour of golimumab, as golimumab has a lower midpoint RR estimate than all but one comparator drug, although the confidence intervals are wide and overlap those of the comparator drugs.

The DMARD experienced population model reflects the clinical evidence, which shows significant uncertainty in the relative effectiveness of golimumab compared to other TNF-α inhibitors. The price and dosing frequency mean that the total cost is likely to be broadly similar to comparator TNF-α inhibitors. However, if these interventions are considered to be a class and it is assumed that there is no difference in any clinical outcomes between the TNF-α inhibitors, the lowest cost intervention would be optimal.

The ERG ran a series of step-wise sensitivity analysis to derive their best estimate of the cost-effectiveness of golimumab in both populations. In the DMARD experienced population the results did not alter significantly following these changes. Correcting an error found in the infliximab arm meant that infliximab was comparable to the other TNF-α inhibitors. Correcting an error in the certolizumab arm meant that certolizumab had the greatest QALY gains rather than etanercept. In the ERG's preferred scenario, the ICER for golimumab

compared to methotrexate is still comparable to that for other TNF-α inhibitors but it is still not the most cost-effective strategy.

In the TNF-α inhibitor experienced population, the manufacturer reports that golimumab is both more effective and less costly than rituximab. However the uncertainty surrounding the HAQ progression rate estimates and the re-administration frequency of rituximab means that there is considerable uncertainty in the cost-effectiveness of golimumab in the TNF-α inhibitor experienced population. The ERG's analysis in the TNF-α inhibitor experienced population found that if the HAQ progression rate of rituximab was assumed to be equal to that of golimumab and if rituximab was assumed to be re-administered every 9 months instead of every 6, then rituximab was more effective and less costly than golimumab.

This sensitivity of the results in the TNF-α inhibitor experienced population reflects the uncertainty and variability of rituximab re-administration in the UK. The SPC[15] recommends re-administering every 6-12 months, and so the manufacturer's 6 month assumption seems inappropriate and an estimate using 9 months is probably more reflective of current practice. The sensitivity of the results also highlights the uncertainty of disease progression while on treatments. Further research is required on HAQ progression rates for patients receiving treatment.

7.3 Implications for research

No direct evidence exists for comparison of the biologic therapies under consideration in this assessment. Large-scale head to head trials of golimumab versus comparator therapies conducted in DMARD experienced and TNF-α experienced patients with moderate and severe rheumatoid arthritis in which follow-up is maintained over at least 52 weeks may be of significant value in assessing the most effective and safe therapy for use in these patient populations. However, the value of any further research should be assessed using expected value of sample information analysis to determine whether the benefits of the additional information that would be gained would be expected to outweigh the costs of gathering further data[50]. There is also a requirement for longer-term safety data relating to the use of golimumab in rheumatoid arthritis. Uncertainty still remains surrounding the long-term effectiveness of treatments for RA. It is still unclear how patients' disease progresses while on treatment and how their disease level changes when removed from treatment.

APPENDICES: FURTHER INFORMATION ON ADVERSE EVENTS

Appendix 1: Mortality data at 6 and 12 months for identified RCTs (as presented by manufacturer)

Study name	Intervention, dosing	Safety population	Mortality	
			6 months	12 months
(Kim *et al.*, 2007)	Placebo subcutaneously every other week + methotrexate	63	-	-
	Adalimumab 40 mg subcutaneously. every other week + methotrexate	65	-	-
(van de Putte *et al.*, 2004)	Placebo subcutaneously weekly	110	-	-
	Adalimumab 40 mg subcutaneously. every other week (with placebo injected on the alternate week)	113	-	-
ARMADA	Placebo subcutaneously. every other week + methotrexate	62	-	-
	Adalimumab 40 mg subcutaneously. every other week + methotrexate	67	-	-
CHANGE	Placebo subcutaneously. every other week	87	-	-
	Adalimumab 40 mg subcutaneously every other week	91	-	-
DE019	Placebo subcutaneously every other week + methotrexate	12	0 (12 weeks)	-
	Adalimumab subcutaneously 40 mg every other week + methotrexate	35	0 (12 weeks)	-
STAR	Placebo subcutaneously every other week + background DMARDs	318	0	-
	Adalimumab subcutaneously 40 mg every other week + background DMARDs	318	1	-
(Chen *et al.*, 2009)	Placebo subcutaneously every week + methotrexate	200	-	0
	Adalimumab subcutaneously 40 mg every other week (placebo on non-treatment weeks) + methotrexate	207	-	1
RAPID 1	Placebo subcutaneously weeks 0, 2, 4, every 2 weeks thereafter + methotrexate (oral)	199	-	1
	Certolizumab subcutaneously 400 mg weeks 0, 2, 4, 200 mg every 2 weeks thereafter + methotrexate (oral)	392	-	2
RAPID 2	Placebo subcutaneously weeks 0, 2, 4, every 2 weeks thereafter + methotrexate (oral)	125	0	
	Certolizumab subcutaneously 400 mg weeks 0, 2, 4, 200 mg every 2 weeks thereafter + methotrexate (oral)	248	1	
TEMPO	Placebo subcutaneously twice weekly + methotrexate (oral)	228	-	1

	Etanercept subcutaneously 25 mg twice weekly + placebo (oral)	223	-	1
	Etanercept subcutaneously 25 mg twice weekly + methotrexate (oral)	231	-	1
(Combe *et al.*, 2006)	Placebo subcutaneously twice weekly + sulfasalazine (oral)	50	0	-
	Etanercept subcutaneously 25 mg twice weekly + placebo (oral)	103	0	-
	Etanercept subcutaneously 25 mg twice weekly + sulfasalazine (oral)	101	0	-
(Moreland *et al.*, 1999)	Placebo twice weekly for 26 weeks	80	-	-
	Etanercept subcutaneously 25 mg twice weekly for 26 weeks	78	-	-
(Weinblatt *et al.*, 1999)	Placebo subcutaneously twice weekly for 24 weeks + methotrexate (oral or subcutaneously)	30	0	-
	Etanercept subcutaneously 25 mg twice weekly for 24 weeks + methotrexate (oral or subcutaneously)	59	0	-
GO-FORWARD	Placebo subcutaneously + M methotrexate (oral)	134	0	-
	Golimumab subcutaneously 50 mg every 4 weeks + methotrexate (oral)	212	0	-
(Kay *et al.*, 2008)	Placebo (subcutaneously every 2 weeks); open-label infliximab 3 mg/kg at week 20, 22, 28, every 8 weeks thereafter + methotrexate (oral ≥ 10 mg/week)	34	0	0
	Golimumab 50 mg/4 weeks; placebo every other 2 weeks + methotrexate (oral ≥ 10 mg/week)	37	0	0
ATTEST	Placebo intravenously. (all infusion days) + methotrexate	110	0	-
	Infliximab intravenously. 3 mg/kg day 1, 14, 43, 85, every 56 days + methotrexate	165	1	2
ATTRACT	Placebo intravenously. week 0, 2, 6, every 4 weeks after + methotrexate (oral)	86	-	3
	Infliximab intravenously. 3 mg/kg week 0, 2, 6, every 4 weeks after + methotrexate (oral)	86	-	-
	Infliximab intravenously. 3 mg/kg week 0, 2, 6, every 8 weeks after + methotrexate (oral)	88	-	-
START	Placebo intravenously. week 0, 2, 6, 14 + methotrexate (oral); other DMARDs as necessary	361	-	1
	Infliximab intravenously. 3 mg/kg week 0, 2, 6, 14 + methotrexate (oral); other DMARDs as necessary	360	-	1
(Abe *et al.*, 2006)	Placebo intravenously. week 0, 2, 6 + methotrexate (oral)	47	0	-
	Infliximab intravenously. 3 mg/kg week 0, 2, 6 + methotrexate (oral)	49	0	-

Appendix 2: Classification of serious adverse events

GO-FORWARD and GO-AFTER both included the following reported serious adverse events (with classification groupings):

Infections and infestations

- •Sepsis
- •Urinary tract infection
- •Arthritis bacterial
- •Cellulitis
- •Lower / upper respiratory tract infection
- •Arthritis infective
- •Subcutaneous abscess
- •Bronchitis
- •Pneumonia
- •Gastroenteritis
- •Herpes zoster
- •Pelvic inflammatory disease

Gastrointestinal disorders

- •Gastric ulcer
- •Colitis
- •Diarrhoea
- •Nausea
- •Vomiting

Musculoskeletal and connective tissue disorders

- •Arthritis
- •Arthralgia
- •Bursitis
- •Rheumatoid arthritis
- •Acquired claw toe
- •Toe deformity

Neoplasms benign, malignant and unspecified (including cysts and polyps)

- •Breast cancer
- •Bowen's disease
- •Squamous cell carcinoma
- •Lymphoma

Respiratory, thoracic and mediastinal disorders

- •Pulmonary embolism

Cardiac disorders

- •Myocardial infarction
- •Angina pectoris
- •Coronary artery disease

Endocrine disorders

- •Goitre

Injury, poisoning and procedural complications

- •Femur fracture

- •Dislocation of joint prosthesis
- •Laceration

Vascular disorders

- •Deep vein thrombosis
- •Aortic thrombosis

Nervous system disorders

- •Cerebrovascular accident
- •Paraesthesia

Renal and urinary disorders

- •Renal disorder

Blood and lymphatic system disorders

- •Anaemia

Hepatobiliary disorders

- •Hepatotoxicity

Metabolism and nutrition disorders

- •Diabetes mellitus inadequate control
- •Diabetic ketoacidosis

Appendix 3: Reported serious infections

GO-FORWARD, GO-AFTER and Kay *et al.* (2008) all included the following reported serious infections:

- Bacterial arthritis
- Bronchitis
- Cellulitis
- Colitis
- Diarrhoea
- Fever
- Gastroenteritis
- Infected cystic lymphangioma
- Infective arthritis
- Lower / upper respiratory tract infection
- Lung disorder
- Pelvic inflammatory disease
- Pneumonia
- Sepsis
- Sinusitis
- Skin laceration
- Subcutaneous abscess
- Urinary tract infection
- Urosepsis

Appendix 4: GO-FORWARD adverse events data to week 16 (before early escape) (as presented in original MS, page 106)

Assessment item	Placebo (Group 1)	Golimumab 50mg (Group 3)
	n=133	n=89
Average duration of follow-up (weeks)	15.9	16.1
Average exposure (no of administrations)	3.9	3.9
Adverse events		
Urinary tract infections	1 (0.8%)	0 (0.0%)
Cellulitis	0 (0.0%)	1 (1.1%)
Subcutaneous abscess	0 (0.0%)	1 (1.1%)
Bursitis	1 (0.8%)	0 (0.0%)
Rheumatoid arthritis	0 (0.0%)	1 (1.1%)
Myocardial infarction	1 (0.8%)	0 (0.0%)
Goitre	7 (5%)	5 (3%)
Hypertension	2 (1%)	5 (3%)
Infections	32 (24.1%)	25 (28.1%)
Serious adverse events	3 (2.3%)	5 (5.6%)
Serious infections	1 (0.8%)	2 (2.2%)
Injection-site disorders	3 (2.3%)	4 (4.5%)
Malignancies	0 (0.0%)	0 (0.0%)
Data are number of patients (%) unless stated		

Appendix 5: GO-FORWARD adverse events data to week 24 (as presented in original MS, pages 106 to 107)

Assessment item	Placebo (Group 1)	Golimumab 50mg (Group 3)
	n=133	n=89
Average duration of follow-up (weeks)	21.1	22.6
Average exposure (no of administrations)	5.1	5.5
Adverse events	90 (67.7%)	65 (73.0%)
Upper respiratory tract infection	9 (6.8%)	11 (12.4%)
Cough	7 (5.3%)	6 (6.7%)
Headache	5 (3.8%)	5 (5.6%)
Nasopharyngitis	6 (4.5%)	4 (4.5%)
Rash	4 (3.0%)	5 (5.6%)
Bronchitis	3 (2.3%)	3 (3.4%)
Abdominal pain upper	4 (3.0%)	3 (3.4%)
Diarrhoea	4 (3.0%)	4 (4.5%)
Infections	37 (27.8%)	28 (31.5%)
Serious adverse events	5 (3.8%)	6 (6.7%)
Serious infections	1 (0.8%)	2 (2.2%)
Injection-site reactions	0 (0.0%)	3 (3.4%)
Data are number of patients (%) unless stated		

Appendix 6: Kay *et al*. (2008) adverse events (with greater than 10% frequency) reported through week 20 (as presented by the manufacturer)

	Placebo	Golimumab 50mg every 4 weeks	Golimumab 50mg every 2 weeks	Golimumab 100mg every 4 weeks	Golimumab 100mg every 2 weeks
Patients treated	34	37	32	33	35
Average duration of follow-up (weeks)	18.2	17.8	20.0	18.5	20.2
Average exposure (number of administrations)	8.7	8.5	9.7	8.5	9.6
Patients with ≥1 adverse event	29 (85.3%)	34 (91.9%)	24 (75.0%)	29 (87.9%)	31 (88.6%)
Adverse events with frequency of ≥10%					
Nausea	1 (2.9%)	2 (5.4%)	7 (21.9%)	6 (18.2%)	8 (22.9%)
Headache	7 (20.6%)	6 (16.2%)	5 (15.6%)	7 (21.2%)	3 (8.6%)
Injection site erythema	4 (11.8%)	5 (13.5%)	2 (6.3%)	3 (9.1%)	10 (28.6%)
Rheumatoid arthritis	7 (20.6%)	6 (16.2%)	3 (9.4%)	4 (12.1%)	3 (8.6%)

Appendix 7: GO-AFTER adverse events data to week 24

	Placebo	Golimumab 50mg
	n=155	n=152*
Number of injections, mean (SD)	4.4 (1.3)	5.2 (1.2)
Patients reporting adverse events	112 (72%)	101 (66%)
Common adverse events		
Upper respiratory tract infection	10 (6%)	11 (7%)
Nasopharyngitis	11 (7%)	12 (8%)
Rheumatoid arthritis	16 (10%)	9 (6%)
Cough	5 (3%)	11 (7%)
Diarrhoea	7 (5%)	5 (3%)
Arthralgia	8 (5%)	6 (4%)
Sinusitis	7 (5%)	5 (3%)
Hypertension	2 (1%)	5 (3%)
Infections	51 (33%)	53 (35%)
Serious adverse events	15 (10%)	11 (7%)
Serious infections	5 (3%)	5 (3%)
Injection-site reactions	6 (4%)	9 (6%)
Malignancies	1 (1%)	1 (1%)
Data are number of patients (%) unless stated		
* patient randomised but excluded prior to treatment not included		

Appendix 8: Inclusion and exclusion of adverse event trials (as presented by the manufacturer)

Author	Year	Title	Inclusion/Exclusion
Alldred	2001	Etanercept in rheumatoid arthritis	Excluded - review
Baeten	2003	Systematic safety follow up in a cohort of 107 patients with spondyloarthropathy treated with infliximab: a new perspective on the role of host defence in the pathogenesis of the disease?	Excluded - different indication
Bongartz	2009	Etanercept therapy in rheumatoid arthritis and the risk of malignancies: a systematic review and individual patient data meta-analysis of randomised controlled trials	Excluded - review
Brown	2002	Tumor necrosis factor antagonist therapy and lymphoma development: twenty-six cases reported to the Food and Drug Administration	Included
Burmester	2009	Adalimumab safety and mortality rates from global clinical trials of six immune-mediated inflammatory diseases	Included
Criscione	2002	Tumor necrosis factor-alpha antagonists for the treatment of rheumatic diseases	Excluded - review
Fleischmann	2005	Long term safety of etanercept in elderly subjects with rheumatic diseases	Included
Flendrie	2003	Survival during treatment with tumour necrosis factor blocking agents in rheumatoid arthritis	Included
Genovese	2009	Safety of biological therapies following rituximab treatment in rheumatoid arthritis patients	Excluded - different population
Hyrich	2004	Anti-tumour necrosis factor alpha therapy in rheumatoid arthritis: an update on safety	Excluded - review
Kaur	2007	Pneumocystis jiroveci (carinii) pneumonia after infliximab therapy: a review of 84 cases	Excluded - review
Keane	2001	Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent	Excluded - review
Klapman	2003	A lupus-like syndrome associated with infliximab therapy	Excluded - different indication
Klareskog	2006	A long-term, open-label trial of the safety and efficacy of etanercept (Enbrel) in patients with rheumatoid arthritis not treated with other disease-modifying antirheumatic drugs	Included
Koike	2009	Japan College of Rheumatology 2009 guidelines for the use of tocilizumab, a humanized anti-interleukin-6 receptor monoclonal antibody, in rheumatoid arthritis	Excluded - not intervention of interest

Lee	2007	Cutaneous side-effects in patients with rheumatic diseases during application of tumour necrosis factor-alpha antagonists	Excluded - different indication
Listing	2005	Infections in patients with rheumatoid arthritis treated with biologic agents	Included
Mayordomo	2002	Pulmonary miliary tuberculosis in a patient with anti-TNF-alpha treatment	Exclude - review
Mikuls	2003	Lessons learned in the use of tumor necrosis factor-alpha inhibitors in the treatment of rheumatoid arthritis	Excluded - review
Mohan	2004	Tuberculosis following the use of etanercept, a tumor necrosis factor inhibitor	Exclude - review
Nam	2010	Current evidence for the management of rheumatoid arthritis with biological disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of RA	Excluded - review
Neven	2005	Adverse events in patients with rheumatoid arthritis treated with infliximab in daily clinical practice	Excluded - letter; insufficient data
Saba	2008	Adalimumab-induced acute myelogenic leukemia	Exclude - review of single case study; n=1
Salliot	2007	Infections during tumour necrosis factor-alpha blocker therapy for rheumatic diseases in daily practice: a systematic retrospective study of 709 patients	Included
Scheinfeld	2004	A comprehensive review and evaluation of the side effects of the tumor necrosis factor alpha blockers etanercept, infliximab and adalimumab	Excluded – review
Voulgari	2005	Infliximab therapy in established rheumatoid arthritis: an observational study	Included

Appendix 9: Adverse event quality assessment checklist (as presented by the manufacturer)

Clinical Study	**Brown 2002**	**Burmester 2009**	**Fleischmann 2005**	**Flendrie 2003**	**Klareskog 2006**	**Listing 2005**	**Salliot 2007**	**Voulgari 2005**
Objective stated?	Y	Y	Y	Y	Y	Y	Y	Y
Main outcomes described?	Y	Y	Y	Y	Y	Y	Y	Y
Patient characteristics described ?	Y	Y	Y	Y	Y	Y	Y	Y
Interventions of interest described ?	Y	Y	Y	Y	Y	Y	Y	Y
Distributions of principal confounders in each group of subjects to be compared described?	NC	Y	Y	NC	Y	Y	N	N
Main findings described?	Y	Y	Y	Y	Y	Y	Y	Y
Were the participating subjects representative of the entire population from which they were recruited?	NC	Y	Y	Y	Y	Y	Y	Y
Subjects blinded?	N	Y	Y	NA	Y	N	NA	N
Investigators blinded?	N	Y	Y	NA	Y	N	NA	N
Time periods between intervention / outcome same for cases and controls?	Y	Y	Y	NC	Y	N	N	Y
Appropriate statistical tests conducted?	Y	Y	Y	Y	Y	Y	Y	Y
Compliance with the interventions reliable?	Y	NC	Y	NC	NC	NC	Y	Y
Reliability and validity of main outcome measures?	N	Y	Y	N	N	N	N	N
Lost to follow-up accounted for?	N	Y	Y	N	Y	Y	N	Y
Sufficient power?	NC	NC	Y	NC	Y	Y	N	NC

Appendix 10: Safety data from Phase IIb and Phase III clinical trials of golimumab-treated RA, PsA and AS patients (as presented in the SPC)[14]

Infections and infestations	
Very common:	Upper respiratory tract infection (nasopharyngitis, pharyngitis, laryngitis and rhinitis)
Common:	Bacterial infections (such as cellulitis), viral infections (such as influenza and herpes), bronchitis, sinusitis, superficial fungal infections,
Uncommon:	Septic shock, sepsis, tuberculosis, lower respiratory tract infection (such as pneumonia), opportunistic infections (such as invasive fungal infections [histoplasmosis, coccidioidomycosis, pneumocytosis], bacterial, atypical mycobacterial infection and protozoal), pyelonephritis, abscess, arthritis bacterial, bursitis infective
Rare:	Hepatitis B reactivation
Neoplasms, benign, malignant and unspecified	
Uncommon:	Neoplasms (such as skin cancer, squamous cell carcinoma and melanocytic naevus)
Rare:	Lymphoma
Not known:	leukaemia*
Blood and lymphatic system disorders	
Common:	Anemia
Uncommon:	Leukopenia, thrombocytopenia,
Rare:	Pancytopenia
Not known:	Aplastic anemia*
Immune system disorders	
Common:	Allergic reactions (bronchospasm, hypersensitivity, urticaria), autoantibody positive
Endocrine disorders	
Uncommon:	Thyroid disorder (such as hypothyroidism, hyperthyroidism and goiter)
Metabolism and nutrition disorders	
Uncommon:	Blood glucose increased, lipids increased
Psychiatric disorders	
Common:	Depression, insomnia
Nervous system disorders	
Common:	Dizziness, paresthesia, headache
Uncommon:	Demyelinating disorders, balance disorders, dysguesia
Eye disorders	
Uncommon:	Visual disorders (such as blurred vision and decreased vision acuity), conjunctivitis, eye allergy (such as pruritis and irritation)
Cardiac disorders	
Uncommon:	Congestive heart failure (new onset or worsening), arrhythmia, ischemic coronary artery disorders
Vascular disorders	
Common:	Hypertension

Uncommon:	Thrombosis (such as deep venous and aortic), Raynaud's phenomenon, flushing
Respiratory, thoracic and mediastinal disorders	
Uncommon:	Asthma and related symptoms (such as wheezing and bronchial hyperactivity)
Rare:	Interstitial lung disease
Gastrointestinal disorders	
Common:	Constipation, dyspepsia, gastrointestinal and abdominal pain
Uncommon:	Gastrointestinal inflammatory disorders (such as gastritis and colitis), gastrooesophageal reflux disease, stomatitis
Hepatobiliary disorders	
Common:	Alanine aminotransferase increased, aspartate aminotransferase increased
Uncommon:	Cholelithiasis, hepatic disorders
Skin and subcutaneous tissue disorders	
Common:	Alopecia, dermatitis, pruritus, rash
Uncommon:	Psoriasis (new onset or worsening of pre-existing psoriasis, palmar/plantar and pustular), urticaria
Musculoskeletal and connective tissue disorders	
Rare:	Lupus-like syndrome
Renal and urinary disorders	
Uncommon:	Bladder disorders
Rare:	Renal disorders
Reproductive system and breast disorders	
Uncommon:	Breast disorders, menstrual disorders
General disorders and administration site conditions	
Common:	Pyrexia, asthenia, injection site reaction (such as injection site erythema, urticaria, induration, pain, bruising, pruritus, irritation and paraesthesia), impaired healing, chest discomfort
Injury, poisoning and procedural complications	
Uncommon:	Bone fractures
*: Observed with other TNF-blocking agents, but not observed in clinical studies with golimumab.	

8 REFERENCES

1. National Institute for Health and Clinical Excellence Tocilizumab for the treatment of rheumatoid arthritis. NICE technology appraisal guidance 198. 2010.
2. National Institute for Health and Clinical Excellence Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor. NICE technology appraisal guidance 195. Part review of NICE technology appraisal guidance 36. Review of NICE technology appraisal guidance 126 and 141. 2010.
3. Chen, Y. F., Jobanputra, P., Barton, P., Jowett, S., Bryan, S., Clark, W., Fry-Smith, A., and Burls, A. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. *Health Technol.Assess.* 2006; **10** iii-xiii, 1.
4. Symmons, D., Turner, G., Webb, R., Asten, P., Barrett, E., Lunt, M., Scott, D., and Silman, A. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. *Rheumatology* 2002; **41** 793.
5. Rodriguez, L. A. G., Tolosa, L. B., Ruigomez, A., Johansson, S., and Wallander, M. A. Rheumatoid arthritis in UK primary care: incidence and prior morbidity. *Scandinavian Journal of Rheumatology* 2009; **38** 173-177.
6. National Audit Office Services for People With Rheumatoid Arthritis. 2009.
7. Yelin, E., Trupin, L., Katz, P., Lubeck, D., Rush, S., and Wanke, L. Association Between Etanercept Use and Employment Outcomes Among Patients With Rheumatoid Arthritis. *Arthritis and Rheumatism* 2003; **48** 3046-3054.
8. Wolfe, F., Allaire, S., and Michaud, K. The prevalence and incidence of work disability in rheumatoid arthritis, and the effect of anti-tumor necrosis factor on work disability. *The Journal of rheumatology* 2007; **34** 2211.
9. Allaire, S., Wolfe, F., Niu, J., Zhang, Y., Zhang, B., and LaValley, M. Evaluation of the effect of anti-tumor necrosis factor agent use on rheumatoid arthritis work disability: the jury is still out. *Arthritis Rheum* 15-8-2008; **59** 1082-1089.
10. Verstappen, Suzanne M. M., Watson, Kath D., Lunt, Mark, McGrother, Katie, Symmons, Deborah P. M., Hyrich, Kimme L., and on behalf of the BSR Biologics Register Working status in patients with rheumatoid arthritis, ankylosing spondylitis and psoriatic arthritis: results from the British Society for Rheumatology Biologics Register. *Rheumatology* 1-8-2010; **49** 1570-1577.
11. Malottki, K, Barton, P, Tsourapas, A, Liu, Z, Uthman, A, Routh, K, Connock, M, Jobanputra, P, Moore, D, Fry-Smith, A, and Chen, Y-F. Adalimumab, etanercept,

infliximab, rituximaband abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor: a systematic reviewand economic evaluation. 2009.

12. National Institute for Clinical Excellence Rheumatoid arthritis: The management of rheumatoid arthritis in adults, NICE clinical guideline 79. 2009.
13. National Institute for Health and Clinical Excellence Rituximab for the treatment of rheumatoid arthritis. NICE technology appraisal guidance 126. 2007.
14. MSD Simponi 50mg solution for injection. Summary of Product Characteristic. 2010.
15. Roche Mabthera 100mg and 500mg concentrate for solution for infusion. Summary of Product Characteristic. 2010.
16. National Institute for Clinical Excellence Tocilizumab for the treatment of rheumatoid arthritis: Final Appraisal Determination. 2010.
17. National Institute for Health and Clinical Excellence Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor (part review of NICE technology appraisal guidance 36, review of NICE technology appraisal guidance 126 and 141): Final Appraisal Determination . 2010.
18. McGowan, J., Sampson, M., and Lefebvre, C. An evidence based checklist for the peer review of electronic search strategies. *Evidence Based Library and Information Practice* 2010; **5** 149-154.
19. Golder, S. Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Information and Libraries Journal* 2006; **23** 3-12.
20. Rothwell, P. M. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* 1-1-2005; **365** 82-93.
21. Ray, W. A. Population-based studies of adverse drug effects. *N.Engl.J.Med.* 23-10-2003; **349** 1592-1594.
22. Kaufman, D. W. and Shapiro, S. Epidemiological assessment of drug-induced disease. *Lancet* 14-10-2000; **356** 1339-1343.
23. Vandenbroucke, J. P. Benefits and harms of drug treatments. *BMJ* 3-7-2004; **329** 2-3.
24. Dieppe, P., Bartlett, C., Davey, P., Doyal, L., and Ebrahim, S. Balancing benefits and harms: the example of non-steroidal anti-inflammatory drugs. *BMJ* 3-7-2004; **329** 31-34.
25. Keystone, E. C., Genovese, M. C., Klareskog, L., Hsia, E. C., Hall, S. T., Miranda, P. C., Pazdur, J., Bae, S. C., Palmer, W., Zrubek, J., Wiekowski, M., Visvanathan, S., Wu, Z., and Rahman, M. U. Golimumab, a human antibody to tumour necrosis factor {alpha} given by monthly subcutaneous injections, in active rheumatoid arthritis despite methotrexate therapy: the GO-FORWARD Study. *Ann.Rheum.Dis.* 2009; **68** 789-796.

26. Kay, J., Matteson, E. L., Dasgupta, B., Nash, P., Durez, P., Hall, S., Hsia, E. C., Han, J., Wagner, C., Xu, Z., Visvanathan, S., and Rahman, M. U. Golimumab in patients with active rheumatoid arthritis despite treatment with methotrexate: a randomized, double-blind, placebo-controlled, dose-ranging study. *Arthritis Rheum.* 2008; **58** 964-975.
27. Smolen, J. S., Kay, J., Doyle, M. K., Landewe, R., Matteson, E. L., Wollenhaupt, J., Gaylis, N., Murphy, F. T., Neal, J. S., Zhou, Y., Visvanathan, S., Hsia, E. C., and Rahman, M. U. Golimumab in patients with active rheumatoid arthritis after treatment with tumour necrosis factor alpha inhibitors (GO-AFTER study): a multicentre, randomised, double-blind, placebo-controlled, phase III trial. *Lancet* 18-7-2009; **374** 210-221.
28. Keystone, E., Genovese, M. C., Klareskog, L., Hsia, E. C., Hall, S., Miranda, P. C., Pazdur, J., Bae, S. C., Palmer, W., Xu, S., and Rahman, M. U. Golimumab in patients with active rheumatoid arthritis despite methotrexate therapy: 52-week results of the GO-FORWARD study. *Ann.Rheum.Dis.* 2010; **69** 1129-1135.
29. Emery, P., Fleischmann, R. M., Moreland, L. W., Hsia, E. C., Strusberg, I., Durez, P., Nash, P., Amante, E. J., Churchill, M., Park, W., Pons-Estel, B. A., Doyle, M. K., Visvanathan, S., Xu, W., and Rahman, M. U. Golimumab, a human anti-tumor necrosis factor alpha monoclonal antibody, injected subcutaneously every four weeks in methotrexate-naive patients with active rheumatoid arthritis: twenty-four-week results of a phase III, multicenter, randomized, double-blind, placebo-controlled study of golimumab before methotrexate as first-line therapy for early-onset rheumatoid arthritis. *Arthritis Rheum.* 2009; **60** 2272-2283.
30. Kremer, J., Ritchlin, C., Mendelsohn, A., Baker, D., Kim, L., Xu, Z., Han, J., and Taylor, P. Golimumab, a new human anti-tumor necrosis factor alpha antibody, administered intravenously in patients with active rheumatoid arthritis: Forty-eight-week efficacy and safety results of a phase III randomized, double-blind, placebo-controlled study. *Arthritis Rheum.* 2010; **62** 917-928.
31. Fleischmann, R., Vencovsky, J., van-Vollenhoven, R. F., Borenstein, D., Box, J., Coteur, G., Goel, N., Brezinschek, H. P., Innes, A., and Strand, V. Efficacy and safety of certolizumab pegol monotherapy every 4 weeks in patients with rheumatoid arthritis failing previous disease-modifying antirheumatic therapy: the FAST4WARD study. *Annals of the Rheumatic Diseases* 2009; **68** 805-811.
32. UBC Cimzia 200mg solution for injection. Summary of Product Characteristic. 2009.
33. National Institute for Health and Clinical Excellence (NICE) Single Technology Appraisal: Specification for manufacturer/sponsor submission of evidence. 2009.

34. NHS Centre for Reviews and Dissemination Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. 2001.
35. Food and Drug Administration Statistical review and evaluation: clinical studies. Golimumab in combination with methotrexate...adult patients with moderate to sevrely active rheumatoid arthritis. 2009.
36. Center for Drug Evaluation and Research Medical Review(s). Golimumab. 2009;
37. Tanaka, Y. *et al.* Golimumab, a human anti-TNF alpha monoclonal antibody administered subcutanously every four weeks in patients with actice rheumatoid arthritis despite methotrexate therapy: 24 week results of clinical and radiographic assessments. 2010.
38. Hyrich, K. L., Watson, K. D., Isenberg, D. A., and Symmons, D. P. M. The British society for rheumatology biologics register: 6 years on. *Rheumatology* 2008; **47** 1441-1443.
39. European Medicines Agency Assessment report for Simponi. 2009.
40. Kielhorn, A., Porter, D., Diamantopoulos, A., and Lewis, G. Uk cost-utility analysis of rituximab in patients with rheumatoid arthritis that failed to respond adequately to a biologic disease-modifying antirheumatic drug. *Current Medical Research and Opinion* 2008; **24** 2639-2650.
41. Hurst, N. P., Kind, P., Ruta, D., Hunter, M., and Stubbings, A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Rheumatology* 1997; **36** 551.
42. Office of National Statistics Interim Life Tables. United Kingdom. 2006.
43. National Institute for Clinical Excellence (NICE) Guide to the methods of technology appraisal. 2008.
44. Chen, Y.-F., Jobanputra, P., Barton, P., Jowett, S., Bryan, S., Clark, W., Fry-Smith, A., and Burls, A. 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** iii-138.
45. Kristensen, L. E., Saxne, T., and Geborek, P. The LUNDEX, a new index of drug efficacy in clinical practice: Results of a five-year observational study of treatment with infliximab and etanercept among rheumatoid arthritis patients in Southern Sweden. *Arthritis and Rheumatism* 2006; **54** 600-606.
46. Edwards, C. J., Arden, N. K., Fisher, D., Saperia, J. C., Reading, I., Van Staa, T. P., and Cooper, C. The changing use of disease-modifying anti-rheumatic drugs in individuals with rheumatoid arthritis from the United Kingdom General Practice Research Database. *Rheumatology* 2005; **44** 1394.

47. Geborek, P., Crnkic, M., Petersson, I. F., and Saxne, T. Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southern Sweden. *Ann Rheum Dis* 2002; **61** 793-798.
48. Brennan, A., Bansback, N., Nixon, R., Madan, J., Harrison, M., Watson, K., and Symmons, D. Modelling the cost effectiveness of TNF-alpha antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. *Rheumatology (Oxford)* 2007; **46** 1345-1354.
49. Brennan, A., Bansback, N., Reynolds, A., and Conway, P. Modelling the cost-effectiveness of etanercept in adults with rheumatoid arthritis in the UK. *Rheumatology* 2004; **43** 62-72.
50. Ades, AE, Lu, G, and Claxton, K Expected values of sample information calculation in medical decision making. *Med Decis Making* 2004; **24** 207-227.