Abatacept for the treatment of rheumatoid arthritis after the failure of conventional disease-modifying anti-rheumatic drugs

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Myfanwy Lloyd Jones, 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 Anthea Sutton, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Correspondence to	Myfanwy Lloyd Jones, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Date completed	02/02/2011

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 10/24/01STA

Declared competing interests of the authors

None.

Acknowledgements

Professor Gerry Wilson, Professor of Rheumatology, Academic Unit of Rheumatology, Department of Infection and Immunity, The University of Sheffield Medical School, Dr Mark Quinn, Clinical Director, Rheumatology, York Teaching Hospitals NHS Foundation Trust, and Dr Vivek Rajagopal, Acute Physician and Rheumatologist, West Suffolk Hospital, Bury St Edmunds, provided clinical advice and commented on draft materials during the project. Abdullah Pandor (Research Fellow, 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 Mark Quinn	None declared
Dr Vivek Rajagopal	None declared
Professor Gerry Wilson	Non-personal pecuniary interests: Involved in multiple pharma-supported clinical drug trials (no personal funding involved) Invited to attend the American College of Rheumatology AGM (Nov 2011) as a guest of Roche Member of the CHM Expert Advisory Group for Immunology & Rheumatology

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:

Lloyd Jones M, Stevenson M, Stevens, J, Sutton A. Abatacept for the treatment of rheumatoid arthritis after the failure of conventional disease-modifying anti-rheumatic drugs: A Single Technology Appraisal. ScHARR, The University of Sheffield, 2011.

Contributions of authors

Myfanwy Lloyd Jones 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. Matt Stevenson acted as health economist on this assessment, critiqued the manufacturer's economic evaluation and contributed to the writing of the report. John Stevens critiqued the statistical analyses included in the manufacturer's submission, and contributed to the writing of the report. Anthea Sutton commented on the searches included in the manufacturer's submission.

Table of contents

List of abbreviations

ACR	American College of Rheumatology
AE	Adverse event
AIM	Abatacept in Inadequate responders to Methotrexate
Anti-CCP	Anti-cyclic citrullinated peptide
ARRIVE	Abatacept Researched in RA patients with Inadequate anti-TNF response to Validate Effectiveness
ASSURE	Abatacept Study of Safety in Use with other Rheumatoid arthritis thErapies
ATTAIN	Abatacept Trial in Treatment of Antitumor necrosis factor (TNF) IN adequate responders
ATTEST	Abatacept or infliximab versus placebo, a Trial for Tolerability, Efficacy and Safety in Treating RA
BHPR	British Health Professionals in Rheumatology
BSR	British Society for Rheumatology
CFB	Change From Baseline
CODA	Convergence diagnostic and output
CRP	C-reactive protein
CSR	Clinical study report
DAS	Disease Activity Score
DAS28	Disease Activity Score 28 joint count
DMARD	Disease-Modifying Anti-Rheumatic Drug
EC	European Community
ESR	Erythrocyte sedimentation rate
ERG	Evidence Review Group
GMS	Genant-modified Sharp
HAQ	Health Assessment Questionnaire
HAQ-DI	Health Assessment Questionnaire Disability Index
HRQoL	Health-related quality of life
ICER	Incremental Cost-Effectiveness Ratio
IV	Intravenous
MHAQ	Modified HAQ
MRI	Magnetic resonance imaging
MS	Manufacturer's submission
MTX	Methotrexate
NS	Not significant
NSAID	Non-steroidal anti-inflammatory drug
OMERACT	Outcome Measures in Rheumatology Clinical Trials
QALY	Quality Adjusted Life Years
RA	Rheumatoid arthritis
RAMRIS	Rheumatoid arthritis MRI score
RCT	Randomised controlled trial
RF	Rheumatoid factor
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SF-36	Medical Outcome Study Short Form 36
SJC	Swollen joint count
SPI	Sleep problem index
TB	Tuberculosis
TJC	Tender joint count
TNF-α	Tumour necrosis factor alpha
UK	United Kingdom
VAS	Visual Analogue Scale

1 SUMMARY

1.1 Scope of the manufacturer's submission

The manufacturer's submission to NICE sought to provide evidence relating to the clinical and cost-effectiveness of abatacept within its licensed indication and in comparison with those interventions licensed and recommended by NICE for the treatment of moderate to severe active rheumatoid arthritis in adults who had responded inadequately to previous therapy with one or more conventional disease-modifying anti-rheumatic drugs (DMARDs), including methotrexate. As stipulated by the NICE final scope, the submission thus differed from the licensed indication in that it excluded the use of abatacept after the failure of a TNF-α inhibitor; this use was dealt with in NICE's Technology Appraisal Guidance 195.[1] The manufacturer's statement of the decision problem further limited the scope of the submission from the population defined in the NICE final scope to patients for whom self-administered subcutaneously-injected biological agents would be inappropriate.

Although the NICE final scope identified the relevant comparators as conventional DMARDs (eg sulfasalazine, leflunomide) and biological agents (adalimumab, etanercept, infliximab, certolizumab pegol, and golimumab), the manufacturer's submission limited the comparators to conventional DMARDs on the one hand and infliximab on the other, claiming as justification that infliximab was the only biological agent other than abatacept which was administered by intravenous infusion, and that, in clinical practice, the use of abatacept would be limited to those patients with rheumatoid arthritis who were unsuited to subcutaneous pharmacotherapy. However, many patients who were identified by the submission as unsuited to subcutaneous pharmacotherapy would in fact be able to receive subcutaneous therapy administered by nursing personnel in the home.

The submission did not adequately address all the outcomes specified within the NICE final scope. No efficacy data were available relating to the extra-articular manifestations of disease, and data relating to some outcomes known to be important to patients (pain, fatigue, sleep quality, and health related quality of life) were poorly presented.

The manufacturer did not present an analysis of abatacept compared with a sequence of biologic treatments nor was there an analysis of a sequence involving both abatacept and infliximab compared with conventional DMARDs in the population of patients who could not have a subcutaneous injection. It is unclear whether this limitation was stipulated in the scope, which could be perceived as ambiguous.

1.2 Summary of clinical effectiveness evidence submitted by the manufacturer

The manufacturer's submission included a systematic review of the evidence for the clinical effectiveness of abatacept compared with either infliximab or placebo in adults with RA who had had an inadequate response to one or more conventional DMARDs, including methotrexate. The systematic review identified four relevant double-blind randomised controlled trials (RCTs). Three of these, the Kremer Phase 2b, AIM, and ATTEST studies, were of one year's duration; the duration of study IM101-119 was only four months. All four studies were placebo-controlled. The Kremer Phase 2b study used two doses of abatacept(2 or 10 mg/kg), and the manufacturer's submission presented data relating to the lower dose even though they were not pertinent because that dose is not licensed for use in adults with RA. The ATTEST study randomised participants to infliximab as well as abatacept and placebo; however, it was not powered to detect statistical differences between abatacept and infliximab. The submission included some meta-analyses. Study IM101-119 was appropriately excluded from these; it differed from the other three studies in terms of its duration, its outcome measures, and the fact that its population had less severe RA.

Efficacy

The clinical effectiveness evidence submitted by the manufacturer indicated that, relative to placebo, abatacept, at a dose of, or approximating to, 10 mg/kg, reduced disease activity, as measured by the DAS28 and ACR responses, at 6 and 12 months. Abatacept was associated with a relative risk of achieving low disease activity (DAS28 $\leq$3.2) at 12 months of 3.89 (95% CI 1.13, 13.40; p=0.03), and a relative risk of achieving remission (DAS28 <2.6) of 4.78 (95% CI 2.06, 11.09; p=0.003).

Relative to placebo, abatacept also appeared to be associated with improved physical function, as measured using the HAD-DI or MHAQ, at 6 months and 1 year, and with less joint damage at one year; however, the clinical significance of these results was not clear. It also appeared to be associated with improvements at both 6 months and 1 year in pain, morning stiffness, sleep quality, fatigue, and health-related quality of life as assessed by the physical and mental summary components of the SF-36.

As noted above, the ATTEST study was not powered to detect statistical differences between abatacept and infliximab, and a statistically significant difference was identified in only two outcomes: ACR20 response and the physical summary component of the SF-36, both at 1 year. Both results favoured abatacept.

The manufacturer's submission included a mixed treatment comparison (MTC) designed to evaluate the efficacy of abatacept plus methotrexate versus five biological DMARDs using a network analysis which included abatacept, adalimumab, certolizumab, etanercept, golimumab, infliximab, and

placebo, despite the fact that the manufacturer's definition of the decision problem claimed that infliximab was the only one of those biological agent which it was relevant to compare with abatacept since the others were administered subcutaneously. The only direct comparison in the network was abatacept versus infliximab; the other comparisons were indirect comparisons via placebo. Efficacy findings from the MTC relating to the DAS28, arguably the most clinically important outcome measure, were not presented; instead, the analysis focused primarily on the change from baseline (CFB) in the HAQ at 24/26 weeks. This suggested that abatacept plus methotrexate was expected to be more efficacious than placebo plus methotrexate, and was expected to display efficacy comparable to that of most other biologic DMARDs; the absolute CFB for biological agents in combination with methotrexate was expected to range from -0.46 for infliximab to 0.65 for certolizumab. The MTC also suggested that all biological agents considered in the submission would result in comparable proportions of ACR20/50/70 responders, although certolizumab pegol would be expected to have a slightly higher ACR20 response rate than other biological DMARDs.

Safety

The RCT evidence suggested that abatacept at a dose of, or approximating to, 10 mg/kg was not associated with a higher rate of serious adverse events than placebo at either 6 months or 1 year, and its adverse event profile appeared to be favourable compared with infliximab. However, not all of the studies reported data relating to acute infusional AEs and peri-infusional AEs.

Longer-term data incorporated into the integrated safety analyses of abatacept indicated that the incidence of serious AEs did not increase over time, and no new safety events were identified. Thus, abatacept appeared to be generally well tolerated in both the short and the longer term. However, as this conclusion was based on an analysis in which the mean exposure to abatacept was only 3.56 months, it cannot be regarded as definitive. Moreover, the submission indicated an 80% discontinuation rate from the two-year LTE of the ATTEST study, and no explanation was provided for this.

1.3 Summary of cost effectiveness evidence submitted by the manufacturer

The manufacturer estimated a cost per quality adjusted life year (QALY) gained of £29,646 when abatacept was compared with conventional DMARDs, and a cost per QALY gained of £25,355 when abatacept was compared with infliximab. Univariate sensitivity analyses comparing abatacept with conventional DMARDs showed that the model was relatively robust to most plausible changes. However, in the comparison with infliximab it was seen that the cost per QALY gained increased markedly when it was assumed that vial wastage of infliximab did not occur (£57,843) or when no dose escalation was assumed (£37,025).

Abatacept was estimated by the manufacturer to be less effective and more costly than adalimumab.

1.4 ERG commentary on the robustness of evidence submitted by the manufacturer

1.4.1 Strengths

The manufacturer undertook a systematic review of the evidence for the clinical and cost-effectiveness of abatacept in combination with methotrexate for the treatment of moderate to severe active rheumatoid arthritis in adults who had responded inadequately to previous therapy with one or more conventional DMARDs, including methotrexate. The studies which were identified and included in the review of clinical effectiveness measured outcomes which were appropriate and clinically relevant, including the majority of the outcomes listed in the final scope. Moreover, they were generally of reasonable methodological quality, although they incorporated some risk of bias because of the differential discontinuation rates in patients randomised to placebo and active treatment, and the methods used to deal with incomplete data and nonadherence to study therapy.

The conceptual model used by the model appeared robust and allowed both variability in individual response and uncertainty in the value of key model parameters to be assessed. The model contained the functionality to assess the impact of changing parameters and structural uncertainties on the incremental cost-effectiveness ratio.

1.4.2 Weaknesses and areas of uncertainty

The evidence base for the assessment of clinical effectiveness may not be complete: because of the failure of the manufacturer's Medline search strategy to identify at least one relevant publication, the failure to search some relevant databases, and in particular the restriction of the searches to publications in the English language, some relevant studies may not have been identified. Although it is probably unlikely that any major European or North American trials were overlooked for this reason, the ERG identified an apparently relevant Korean study which was not mentioned in the manufacturer's submission. Moreover, the submission excluded without adequate justification a Japanese long-term extension (LTE) study to which no reference was provided. The selection of non-RCT studies for inclusion in the review of clinical effectiveness was poorly reported. No observational studies or publications of post-marketing surveillance data were identified which were independent of the manufacturer's clinical trials programme, but it is possible that none as yet exist.

The presentation of results from the included studies displayed a number of omissions and inconsistencies, failing to refer to some publications relating to the included studies and to present all the relevant data which were available in the public domain, and at times suggesting that those data did not exist. For example, in relation to pain, one of the outcome measures specifically listed in the final scope, the submission stated that information was not available in a suitable format for

presentation[2] despite the fact that data relating to the pain component of the SF-36 were presented graphically in Figures B 18-22, and published data relating to pain were available for the Kremer Phase 2b[3] and AIM[4] studies. Again, the submission stated that DAS28 scores did not form a reported outcome in the Kremer Phase 2b study, but then reported relevant data from that study, and stated that the Kremer Phase 2b study provided little insight into the HRQoL of the enrolled patients, although detailed HRQoL results from this study had been published.[5] Moreover, where data presented in the manufacturer's submission differed from published data, explanations were generally not provided.

Relative risks, whether relating to individual studies or resulting from the meta-analysis of data from those studies, were not presented for all relevant comparisons.

The populations of the included studies had a shorter duration of RA, and had previously taken fewer conventional DMARDs, than is current standard UK clinical practice before the initiation of biological therapy. Therefore, although the submitted evidence largely reflects the decision problem defined in the final scope, the difference between the two populations is such that less benefit may be gained abatacept in UK clinical practice than in the study populations.

The construction of the mathematical model based on the conceptual model was flawed and the uncertainty associated with the cost-effectiveness of abatacept compared with infliximab or conventional DMARDs was underestimated. The model supplied by the manufacturer had numerous errors, the most important of which was the failure of the probabilistic sensitivity analyses to incorporate key parameters. The complete list of identified errors within the model is detailed more fully in the main document.

There was a considerable possibility that the manufacturer's base case scenario was favourable to abatacept compared with infliximab regarding the assumptions that dose escalation would occur for infliximab but not for abatacept and also that there would be no vial sharing for infliximab.

1.5 Summary of additional work undertaken by the ERG

The ERG amended some of the identified errors, and undertook five analyses (objective, optimistic, favourable, pessimistic, and hybrid) in order to provide the appraisal committee with additional information to evaluate the cost-effectiveness of abatacept in rheumatoid arthritis after the failure of conventional disease-modifying anti-rheumatic drugs.

Based on previous evaluations of treatments for RA (where in 63% of cases infliximab was assumed to be vial-shared), the ERG believe that the hybrid analysis would be most pertinent. The hybrid analyses estimate a cost per QALY gained of £34,569 when abatacept is compared with placebo and a

cost per QALY gained of £57,896 when abatacept is compared with infliximab, although these values would increase if abatacept was associated with dose escalation or if disutilities for more frequent hospital attendance were included.

The optimistic scenario reduces the cost per QALY gained to £31,328 when abatacept is compared with placebo and to £27,157 when abatacept is compared with infliximab, although these values are associated with 0% vial sharing for infliximab, dose escalation for infliximab and not for abatacept, and a lesser rate of serious adverse events for abatacept than for infliximab.

The pessimistic scenario increases the cost per QALY gained to £36,613 when abatacept is compared with placebo and to £85,209 when abatacept is compared with infliximab, although these values are associated with 100% vial sharing for infliximab, no dose escalation for either infliximab or abatacept, and equal rates of serious adverse events for infliximab and abatacept.

2 BACKGROUND

This report provides a review of the evidence submitted by Bristol-Myers Squibb Pharmaceuticals Ltd in support of intravenous abatacept therapy for the treatment of rheumatoid arthritis (RA) after the failure of conventional disease-modifying anti-rheumatic drugs (DMARDs). It considers both the original submission received on 23rd November 2010 and subsequent addenda supplied by Bristol-Myers Squibb Pharmaceuticals Ltd on 7th January 2011. The addenda were produced in response to the Evidence Review Group's (ERG's) queries relating to the clinical- and cost- effectiveness data. The report also summarises additional work undertaken by the ERG in correcting errors identified in the manufacturer's model and reporting ICERs.

2.1 Critique of manufacturer's description of underlying health problem

The manufacturer's description of RA is appropriate and relevant to the decision problem under consideration. The submission states that, without successful treatment, RA is associated with high morbidity: pain, fatigue, and loss of motion in joints make it more difficult for a patient with RA to remain in employment or live normally. Because of problems with activities of daily living (eg dressing, bathing, and walking), patients usually need help from family, friends or carers. The onset of RA often interferes with social roles, and may by associated with feelings of helplessness, depression, anxiety, loss of self-esteem, and other psychological difficulties, affecting sleep patterns and causing fatigue. Consequently, people with RA report a decreased health-related quality of life. In addition, family members and carers are subject to the burden of caring for a chronically ill person. The manufacturer's submission does not attempt to quantify any of these problems, although they may be implicitly incorporated in the function relating HAQ and utility. However, it states that most patients experience moderate disability within two years of diagnosis, and 30% are severely disabled after 10 years.[6] The ERG's clinical advisors consider that this statement may be true of untreated disease, but that it probably does not accurately reflect the current position relating to patients who come to medical attention.

The manufacturer's submission states that RA is also associated with increased mortality, citing Felson[7] as evidence that patients with severe RA die 3-18 years earlier than those without RA, and that death rates are highest in patients with early loss of physical function and comorbidities such as cardiovascular disease. More recently, Sokka *et al.*, have stated that mortality rates in patients with RA are approximately 1.5 times higher than in the general population. They state that the most common attributed cause of death is cardiovascular disease, and deaths from this cause occur at a younger age than in the general population. In addition, mortality rates for infection, and for pulmonary, renal, and gastrointestinal disease are higher in patients with RA than in the general population.[8]

However, there is some uncertainty relating to the manufacturer's estimate of the number of patients likely to be eligible for treatment with abatacept. Section 2.2 of the manufacturer's submission estimates that 346,357 adults in the UK have RA, whereas Table C 1 suggests that this estimate relates only to England and Wales;[6] clarification has subsequently been received that the figure related specifically to England and Wales.[2] The manufacturer's submission states that Symmons *et al*.,[9] estimated the UK prevalence of people meeting the 1987 Americal College of Rheumatology (ACR) criteria for RA to be 0.86%, and the estimate included in the submission is based on the application of this prevalence figure to the population age and sex profile for 2009. In fact, Symmons *et al*., estimated the prevalence at 0.81%, and therefore the calculation underlying the manufacturer's estimate is slightly inaccurate, but broadly consistent with the recent NICE national clinical guideline for the management and treatment of RA in adults which, on the basis of prevalence and incidence rates calculated by Symmons *et al*.,[9,10] estimated the prevalence of RA in the UK to be 400,000, and the annual incidence to be 12,000.[11] However, using more recent data, the National Audit Office has estimated that, in England alone, approximately 582,000 people over the age of 16 have RA, and that the annual incidence is approximately 26,000 new cases[12,13] – substantially higher figures than those presented by the manufacturer. These figures may, however, be overestimates, as the ERG's clinical advisors note that they are derived from figures estimated by the contributing Trusts, and the actual figure may lie between the NICE and NAO estimates.

Current NICE guidance[14] recommends that patients with RA are treated with DMARDs, drugs which attempt to inhibit the immune process underlying RA and to prevent long-term damage. A significant proportion of patients are unable to tolerate, or only experience partial benefit from, conventional DMARDs such as methotrexate, leflunomide, and sulfasalazine.[15] Such patients may benefit from newer DMARDs - biological agents which, by mimicking substances found in the immune system during an inflammatory reaction, can specifically target parts of the immune system to reduce inflammation and thus reduce the symptoms of RA. Most biological agents (eg etanercept, adalimumab, golimumab, infliximab, and certolizumab pegol) target tumour necrosis factor alpha (TNF-α), a protein produced by the body during the inflammatory response, whereas abatacept inhibits T-cell activation.[15]

Current NICE guidance[14] recommends the use of TNF-α inhibitors only in adults who have severe active RA, as measured by a Disease Activity Score 28 joint count (DAS28) of 5.1, confirmed on at least two occasions one month apart, and who have undergone trials of two DMARDs, including methotrexate (unless contraindicated). The guidance recommends the use, in such patients, of adalimumab, etanercept, or infliximab in combination with methotrexate; if a patient is intolerant of methotrexate, or methotrexate treatment is considered to be inappropriate, adalimumab or etanercept may be given as monotherapy. Rituximab, a monoclonal antibody which targets the protein CD20,

may be used in combination with methotrexate in adults with severe active RA who have had an inadequate response to, or intolerance of, other DMARDs including at least one TNF-α inhibitor.[14] The manufacturer's submission estimates that 10% of patients with RA have a DAS28 of >5.1. This estimate was said[2] to be obtained from the costing template for NICE technology appraisal guidance TA195.[1] If it is applied to the submission's estimate of the number of adults with RA, then approximately 34,656 patients in the UK would be eligible for a biological agent; its application to the NAO figures would suggest a figure of 58,200 people over the age of 16 in England alone.

However, the manufacturer's submission notes that the more recent BSR/BHPR guidelines recommend that the eligibility criteria for biological therapies be lowered to include patients with moderate RA (ie a DAS28 of >3.2 with at least three tender joints and at least three swollen joints) who have undergone trials of two DMARDs, including methotrexate (unless contraindicated).[16] The manufacturer's submission claims that patients with a DAS28 $\geq$3.2 are estimated to form 30% of the total population with RA, and that therefore, according to the BSR/BHPR guidelines, 103,907 patients in the UK would be eligible for biological therapy. The estimate that 30% of patients with RA have a DAS28 $\geq$3.2 rests on personal communications from RA specialists;[2] if it is correct, its application to the NAO estimate would suggest that approximately 174,600 people in England alone would be eligible for a biological agent on the basis of their DAS28 score. The manufacturer's submission claims that, currently, only 10% of the estimated eligible population receive an IV biological agent; again, this estimate rests on personal communications from RA specialists.[2] Two factors should be borne in mind when interpreting this claim:

- the manufacturer's estimates appear to be based solely on the DAS28 score, and do not take into account the eligibility criterion relating to the previous failure of two DMARDs
- no data are presented relating to the proportion of the estimated eligible population who receive a biological agent which is administered subcutaneously.

2.2 Critique of manufacturer's overview of current service provision

The manufacturer's overview of current service provision is broadly appropriate. The major goal of treatment for RA is to achieve disease remission or, failing that, to minimise disease activity. The submission correctly notes that current NICE guidance recommends the use of DMARDs to attempt to prevent long-term damage by inhibiting the immune process underlying RA. Other pharmacological treatments (painkillers, and anti-inflammatory drugs including steroids) may be used in parallel with DMARDs to alleviate the symptoms of RA.[14] Because a significant proportion of patients are unable to tolerate, or only experience partial benefit from, conventional DMARDs such as methotrexate, leflunomide, and sulfasalazine,[15] current NICE guidelines recommend the use of biological agents specifically in adults who have had an inadequate response to, or intolerance of, more than one DMARD. They further specify that the TNF-α inhibitors adalimumab, etanercept, and

infliximab should be used only in patients with a DAS28 of ≥5.1; no DAS28 threshold is set for the initiation of rituximab therapy, but it is stated that it should be used in patients with severe active RA (generally understood to equate to a DAS28 of ≥5.1) who have had an inadequate response to, or intolerance of, other DMARDs, including at least one TNF-α inhibitor.[14] However, as the manufacturer's submission notes, the more recent BSR/BHPR guidelines recommend treatment for patients with a DAS28 of >3.2 (ie patients with moderate RA) who have undergone trials of two DMARDs, including methotrexate.[16]

The manufacturer's submission then seeks to identify abatacept's potential niche within current service provision. Biological agents are divided into those which are administered by intravenous infusion (abatacept, infliximab, and rituximab), and those (eg etanercept, adalimumab, golimumab, and certolizumab pegol) which are administered by subcutaneous injection. Because abatacept is administered intravenously, the submission focuses on those patients for whom subcutaneous self-injection is not appropriate, whether because their manual dexterity is compromised through disease, age, or infirmity; because they find it difficult to adhere to or comply with self-injected medication; or because the clinical team feel that the patient's personal circumstances or need for closer medical supervision render self-injection inappropriate.

The manufacturer's assessment of abatacept's potential role within current service provision adheres to the current NICE guidelines in stating that biological agents should be used in patients with severe active RA only after trials of two traditional DMARDs, including methotrexate. However, it does not reflect current clinical practice in stating that only two intravenous infusion agents, infliximab and abatacept, are available: it fails to mention rituximab, probably the most commonly prescribed intravenous therapy for RA, presumably on the grounds that the current NICE guidelines state that it should be used after an inadequate response to or intolerance of other DMARDs including at least one TNF-α inhibitor, whereas the population in the scope for the current appraisal is specified as patients who have had an inadequate response to one or more conventional DMARDs including methotrexate. The ERG's clinical advisors indicate that patients for whom subcutaneous self-injection of biological agents is inappropriate are currently offered two options:

- intravenous infliximab (a TNF-α inhibitor) as first-line treatment, with rituximab as second-line treatment
- subcutaneous agents injected by the service provider.

The manufacturer's submission[6] (pp 37-38) states that, in 31% of patients, infliximab treatment is associated with a loss of response which requires dose escalation within the first year of treatment. The ERG's clinical advisors indicate that, in the past, dose escalation or increased frequency of dosing would be used under such circumstances. However, current practice for patients with RA who do not

fall within the niche market outlined in the manufacturer's submission would generally be to change to another therapeutic agent if the standard dose of infliximab was not effective.

As noted above, the manufacturer's submission states that only 10% of the estimated eligible population currently receive an IV biological agent. The ERG assumes that the eligible population referred to here is the total population which is eligible for a biological agent, not the subgroup of that population which is eligible for a biological agent and also unsuited to subcutaneous administration. The manufacturer's submission does not offer any evidence in support of the figure of 10%. The ERG's clinical advisors recognise that there is a subgroup of patients with RA who, because they are unable to inject subcutaneous drugs, are candidates for treatment with biological agents which are delivered by intravenous infusion (infliximab, rituximab or abatacept); however, from clinical experience, they would not expect this proportion to be as high as 10%.

3 CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM

A summary of the decision problem addressed by the manufacturer's submission is shown in Table 1.

Table 1: Decision problem as issued by NICE and addressed by the manufacturer's submission

	Final scope issued by NICE	Decision problem addressed in the submission	Rationale if different from the scope
Population	Adults with RA who have had an inadequate response to one or more conventional DMARDs including methotrexate (MTX)	Said in submission to be as in final scope. However, the submission further restricts its primary focus to adults with RA who have had an inadequate response to one or more conventional DMARDs including MTX and who are unsuited to subcutaneously-injected biological agents, whilst providing ICERs against etanercept, a subcutaneously-injected biological agent.	Not specified but assumed to reflect the fact that most patients would opt for therapies which can be self-injected rather than those which require hospital visits for IV infusion
Intervention	Abatacept in combination with MTX	As in final scope	Not applicable
Comparator(s)	Management strategies involving DMARDs without abatacept, including treatment with: • biological agents (adalimumab, etanercept, infliximab, certolizumab pegol, golimumab) • conventional DMARDs (eg sulfasalazine, leflunomide)	Management strategies involving DMARDs without abatacept, limited to: • infliximab • conventional DMARDs	The submission limited biological agents to infliximab on the grounds that this is the only biological agent listed in the final scope which is administered intravenously.
Outcomes	The outcome measures to be considered include: • disease activity • physical function • joint damage • pain • mortality • fatigue • extra-articular manifestations of disease • adverse effects of treatment • health-related quality of life	Said in submission to be as in final scope. However, the decision problem was limited to the following primary and secondary outcome measures: • disease activity • physical function • joint damage • mortality (reported in the safety section) • fatigue • adverse effects of treatment (reported in the safety section) • health-related quality of life	The manufacturer's clarification letter[2] provides the following information relating to outcome measures stipulated in the scope but not included in the submission: • pain: information not available in a suitable format for presentation • extra-articular manifestations of disease: not reported in the clinical trials

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.	Said in submission to be as in final scope.	Not applicable
Other considerations	Guidance will only be issued in accordance with the marketing authorisation. If the evidence allows, the appraisal will consider the costs of joint replacement therapy and hospital admissions. If the evidence allows, the appraisal will consider subgroups based on: • severity of disease activity (moderate to severe disease, and severe disease) • auto antibody status including rheumatoid factor (RF) and anti-CCP This appraisal will consider the use of abatacept only after the failure of conventional DMARDs alone. It will not include a review of the guidance in technology appraisal 195 relating to the use of abatacept after the failure of a TNF inhibitor.[1]	No subgroup analyses were conducted.	Not clear

3.1 Population

Abatacept in combination with methotrexate is licensed for the treatment of moderate to severe active rheumatoid arthritis in adults who have responded inadequately to other previous therapy with one or more DMARDs, including methotrexate or a TNF-α inhibitor.[17] The scope for this appraisal defines the patient population as all adults with RA who have had an inadequate response to one or more conventional DMARDs, including methotrexate. It therefore differs from the licensed indication in that it excludes the use of abatacept after the failure of a TNF-α inhibitor; this use is dealt with in NICE's Technology Appraisal Guidance 195.[1]

The manufacturer's statement of the decision problem further limits the population defined in the final scope to patients for whom self-administration of subcutaneously-injected biological agents is inappropriate. The clinical evidence submitted by the manufacturer matches the final scope in that it is limited to studies in patients with RA who have had an inadequate response to one or more conventional DMARDs, including methotrexate; it is not further restricted to patients for whom self-administration of subcutaneously-injected biological agents is inappropriate. However, the study populations differ somewhat from the patient population which the ERG's clinical advisors would consider eligible for abatacept therapy, in that they had a shorter duration of RA and had previously received a mean of fewer than two DMARDs, whereas current standard UK clinical practice would be to try at least two, and in many cases more, conventional DMARDs, including methotrexate, before initiating biological therapy. Consequently, the ERG's clinical advisors suggest that the results achieved with abatacept in the clinical trials are likely to be more favourable to the intervention than those which might be expected in UK clinical practice.

3.2 Intervention

Abatacept is a selective modulator of the T lymphocyte activation pathway. It acts by binding to molecules on the surface of antigen presenting cells, thus preventing full activation of the T lymphocytes (immune system cells which are involved in causing the inflammation in rheumatoid arthritis) and interrupting the inflammatory process, helping to reduce the inflammation and other symptoms of the disease. Abatacept is produced using recombinant DNA technology.[17] It is marketed in the UK by Bristol-Myers Squibb under the trade name Orencia at a list price of £242.17 per 250-mg vial.[18]

In May 2007, abatacept in combination with methotrexate received marketing authorisation within the EC for the treatment of moderate to severe active RA in adult patients who had an insufficient response to, or intolerance of, other DMARDs, including at least one TNF-α inhibitor. In May 2010, this authorisation was extended to give marketing authorisation within the EC to abatacept in combination with methotrexate for the treatment of moderate to severe active RA in adult patients who responded inadequately to previous therapy with one or more DMARDs, including methotrexate or a TNF-α inhibitor[19] – in other words, its licensed application was extended to include patients who had responded inadequately to methotrexate but had not necessarily responded inadequately to a TNF-α inhibitor.

Abatacept takes the form of a powder which is made up into a solution and then administered as an intravenous infusion which lasts 30 minutes. Dosing is by body weight, and approximates to 10mg/kg (see Table 2). Abatacept is given every two weeks for the first three doses, and then every four weeks.

If there is no response within six months, continuation of therapy should be reconsidered.[17] Thus, 14 infusions are required in the first year of treatment, and 13 in subsequent years.

Table 2: Abatacept dosage[17]

Patient body weight	Abatacept dose	Number of vials
<60kg	500mg	2
60-100kg	750mg	3
>100kg	1000 mg	4

Abatacept is administered with methotrexate, which is taken in tablet form. The manufacturer's submission[6] states that methotrexate is usually prescribed with a folic acid supplement. The submission does not describe the location of care, staff usage, administration costs, monitoring and tests associated with the use of abatacept, but makes reference to a 2002 publication by Kobelt *et al.*,[20] modelling the progression of rheumatoid arthritis.

Prior to receipt of abatacept, patients should be screened for latent tuberculosis (TB) and viral hepatitis, and treatment should not be initiated until active infections are controlled.[18]

3.3 Comparators

The NICE final scope stated that abatacept should be compared with:

- conventional DMARDs (eg sulfasalazine, leflunomide)
- biological agents (adalimumab, etanercept, infliximab, certolizumab pegol, golimumab).

The ERG's clinical advisors note that rituximab, another biological therapy for RA which is administered by IV infusion, is also commonly used in the relevant population; however, it was not included as a comparator within the scope of this assessment. The manufacturer provided clarification that its exclusion from the scope, and therefore from the submission, was because rituximab had recently been assessed within TA 195.[2] It seems more likely that its exclusion from the scope reflects the fact that its EC marketing authorisation is restricted to adults with severe active RA who have had an inadequate response or intolerance to other DMARDs, including one or more TNF-α inhibitor therapies,[21] whereas abatacept is not restricted to patients who have failed a TNF-α inhibitor.

The manufacturer's submission limits the comparators to conventional DMARDs on the one hand, and infliximab on the other, claiming as justification that infliximab is the only biological agent other than abatacept which is administered by IV infusion. Biological agents administered by subcutaneous injection (adalimumab, certolizumab pegol, etanercept, and golimumab) are excluded on the basis that, in clinical practice, the use of abatacept would be limited to those RA patients who are unsuited to subcutaneous pharmacotherapy, ie:

- patients whose compromised manual dexterity, whether resulting from RA, age, or infirmity, makes self-administration very difficult, or who find it difficult to adhere to/comply with therapy;

- patients who find it difficult to store medication for subcutaneous therapy in the home
- patients with needle phobia
- patients whom the Rheumatology Unit feel require closer supervision - for instance, medically complex cases.

It should be noted that intravenous therapy is not necessarily required for patients who find self-administration or medication storage difficult, for whatever reason: such patients may receive subcutaneous therapy administered by nursing personnel in the home or Rheumatology Unit, although it is recognised that this will negate some of the advantages associated with self-administered therapy. In relation to patients with needle phobia, the use of abatacept would reduce, but not obviate, the need for invasive treatment, replacing subcutaneous injections administered weekly or twice weekly (etanercept[18]), fortnightly (adalimumab, certolizumab pegol[18]) or monthly (golimumab[22]) with an infusion every four weeks together with blood tests every three months to monitor the effects of the associated methotrexate therapy. A greater reduction in invasive treatment, with greater patient convenience, would be obtained by the use of infliximab, which is administered by intravenous infusion every eight weeks,[18] rather than abatacept every four weeks.

3.4 Outcomes

As noted in Table 1, the outcomes reported in the manufacturer's submission are largely the same as those described in the final scope. They are listed and discussed below.

Disease activity

The included studies report disease activity using one of two composite outcome measures, the Disease Activity Score (DAS) and the ACR response criteria.

The **DAS** combines the following measures:

- the number of tender and swollen joints
- the patient's global assessment of disease activity
- a laboratory parameter - the erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP).[23]

The DAS is scored on a scale of 0 to 10; a higher score indicates higher disease activity.[15] It has been validated for both full (DAS44) and limited (DAS28) joint counts; the DAS28 excludes the foot joints. A DAS score <3.2 is held to indicate low disease activity, a score of 3.2-5.1 moderate disease activity, and a score >5.1 severe disease activity.[24] A score below 2.6 indicates remission (absence of disease activity).[25] The manufacturer's submission states that a reduction from baseline of 1.2 or more in the DAS28 indicates a clinically meaningful response.[6]

In a UK setting, the DAS28 is arguably the most important measure of disease activity because it is the measure used in clinical practice to determine the initiation of, and assess response to, therapy. Recent research has shown that, when the DAS28(CRP) (ie the DAS28 calculated using CRP as the laboratory parameter) is used, fewer patients, particularly women, have a high DAS28 score than when the DAS28(ESR) (ie the DAS28 calculated using ESR as the laboratory parameter) is used.[23] The ATTEST study used the DAS28(ESR),[26] while IM101-119,[27] the Kremer Phase 2b study,[3] and probably also the AIM study[4] used the DAS28(CRP).

The **ACR response criteria** combine the following measures:

- tender joint count
- swollen joint count
- the patient's assessment of pain measured using a Visual Analogue Scale (VAS) or Likert scale
- the patient's global assessment of disease activity (using a VAS or Likert scale)
- the physician's global assessment of disease activity (using a VAS or Likert scale)
- the patient's assessment of functional ability measured using the Health Assessment Questionnaire (HAQ) or a similar tool
- laboratory parameters (eg the ESR or the C-reactive protein (CRP) level).[15]

An ACR20 response indicates a decrease of at least 20% in the number of both tender joints and swollen joints, and a 20% improvement in at least three of the other criteria; an ACR50 response indicates a 50% improvement, and an ACR70 response a 70% improvement, using the same criteria.[28]

As may be seen, both the DAS and the ACR response criteria are largely composed of subjective measures. Moreover, the objective measures which they contain (CRP and ESR) are non-specific, and may reflect disease other than active RA.[16] Thus, to avoid detection bias when using these measures, it is essential that both participants and outcome assessors are blinded to treatment allocation.[15]

Physical function

The included studies report physical function using measures derived from the **HAQ (Health Assessment Questionnaire).** The HAQ Disability Index (HAQ-DI) is a subjective measure of patient-assessed functional ability: it comprises 20 questions about difficulties experienced with 8 categories of activities of daily living measured on a scale of 0 (no disability) to 3 (completely disabled), together with four questions about the assistance used to perform those activities. Unless aids or devices are required, the highest component score in each activity category determines the category score. The eight category scores are then averaged to produce the overall HAQ-DI score, which lies between 0 and 3: scores of 0 to 1 represent mild to moderate disability, 1 to 2 moderate to severe disability, and 2 to 3 severe to very severe disability. It has been suggested that a difference of at least 0.22 in the HAQ-DI score is required to demonstrate a clinically significant improvement.[29]

The 'Modified HAQ' (MHAQ) contains only 8 of the 20 questions included in the HAQ-DI, one relating to each activity category; it has been shown to underestimate disability relative to the HAQ-DI.[30]

Joint damage

Joint damage is reported using either the **Genant-modified Sharp (GMS) score** or the **RAMRIS system**. The GMS score is an objective measure of the progression of structural damage using radiographical measurement of bone erosion and joint space narrowing.[31] The 8-point erosion scale, scored in 0.5 point increments from 0 (normal) to 3.5 (severe), and the 9-point joint-space narrowing score, scored in 0.5 point increments from 0 (normal) to 4 (ankylosed), are combined to produce the total GMS score.[32]

The RAMRIS system uses magnetic resonance imaging (MRI) to visualise the bone and soft tissues. Because MRI can assess inflammation (synovitis) and bone marrow oedema (osteitis) as well as bone erosion, it can detect earlier changes which are predictive of later structural damage[33] and is therefore more sensitive than conventional radiography in the earlier stages of RA. The OMERACT RA MRI scoring system (RAMRIS) focuses on the wrist and metacarpopharyngeal joints.[34]

Fatigue

Fatigue has been identified by patients with RA as an area of health which is particularly important to them.[35] The AIM study, the only one of the included studies to report patient-reported fatigue, used a 100mm VAS which ranges from 0 (no tiredness) to 100 (extreme tiredness). The minimum clinically relevant difference is said to be 10 units.

Pain

The Kremer Phase 2b and AIM studies also measure the patient's global assessment of pain using a 100mm VAS.

Extra-articular manifestations of disease

Extra-articular manifestations of disease were specified as an outcome for inclusion in the final scope issued by NICE. However, this outcome was not included in the manufacturer's submission because it was not reported in the clinical trials.[2]

Mortality and adverse effects of treatment

Mortality and adverse effects, including those which are considered to be related to the study medication, are reported as safety outcomes.

Health-related quality of life

The included studies measure health-related quality of life (HRQoL) using the **SF-36** (Medical Outcome Study Short Form 36), a questionnaire designed specifically to measure self-reported health-related quality of life. It contains 36 questions which measure functional status, wellbeing, and overall health in eight dimensions (physical functioning, role limitations due to physical health, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional health, and mental health); these eight dimensions may be are aggregated to produce physical (PCS) and mental (MCS) component summary measures.[5] Results are presented on a scale of 0 to 100, with higher scores indicating better health.[24] The Kremer Phase 2b and AIM studies specified that they used norm-based methods which standardised scores to a mean of 50 and a standard deviation of 10 in the general US population.[36,5] The manufacturer's submission states that the minimum clinically relevant difference is an improvement of ≥ 3 units in the SF-36;[6] it is not clear whether this relates specifically to the physical and mental component summary measures or to any aspect of the SF-36.

Other outcomes

Other outcomes reported in the manufacturer's submission:

- Morning stiffness: the manufacturer's submission states that this is usually self-assessed, either using a scale of 0-10 (where 0 = no morning stiffness at all and 10 = extreme, severe morning stiffness), or as mild (0-2), moderate (3-6) or severe (7-10)
- Sleep quality has been identified as a key concern for patients with RA.[37] The manufacturer's submission states that sleep quality can be measured by a number of different instruments which look across different domains such as adequacy, maintenance, and initiation of sleep, and daytime functioning, as well as by instruments such as the insomnia severity index and sleep diaries.[6] Cole *et al*., consider the Medical Outcomes Study (MOS) Sleep Scale[38] to be the most appropriate tool available for the general assessment of sleep disorders in patients with chronic pain.[39] This tool contains 12 questions, the responses to which can be used, in different combinations, to produce a number of scores including two sleep problems indices (Sleep Problems Index I and Sleep Problems Index II). It was used in the AIM study[40]
- Patient compliance, measured by the number of missed infusions.

It should be noted that, with the exception of joint damage, and possibly patient compliance, the outcomes are all either wholly or primarily subjective. Consequently, the blinding of patients, care providers, and outcome assessors to treatment allocation is crucial.

3.5 Other relevant factors

The manufacturer's submission includes a substantial section on equity and equalities issues. This notes that abatacept is most suitable for those patients who require or reasonably request intravenous infusion, namely:

- Patients who cannot or will not self-administer subcutaneously, including:
  - Those who are mentally ill
  - Those who have genuinely provable clinically diagnosed needle phobias
  - Those who cannot reasonably be expected to store medications at home, including those in dysfunctional families or who live with young children
- Patients who require regular monitoring, including those with co-morbidities such as advanced heart disease, malignancies, or active infection
- Patients who would particularly benefit from regular attendance at a site which has staff available during the administration of treatment, including:
  - The aged and infirm
  - Those with special needs
  - Those with disabilities including clinical depression which prevent or discourage them from active pursuit of their best options for treatment, and require help and care in monitoring and treating their condition.

The submission states that small but significant numbers of patients fall into these categories. No further attempt at quantification is made in Section A, but in Table C1 it is estimated that, in accordance with NICE guidance that biological agents should be used only in patients with severe RA,[14] a total of 173 patients would be eligible for abatacept; use of the BSR/BHPR guidance that biological agents should be used in patients with severe or moderate RA[16] would increase this figure to 520.[6]

In addition to factors related to age and disability, the manufacturer's submission notes that race is relevant to abatacept principally in relation to the raised prevalence of TB among ethnic subgroups because abatacept should not be administered in the presence of active infection with TB, which rules out a number of current treatments for RA.

In terms of equality legislation, the manufacturer's submission claims that the use of abatacept would ameliorate inequity to patients disadvantaged by age or disability who, it is claimed, might otherwise fall outside or through the current net of treatments. This argument is open to question: as noted in section 2.2, the ERG's clinical advisors indicate that patients eligible for treatment with biological agents, but for whom subcutaneous self-injection of biological agents would be inappropriate, would

currently receive treatment either with infliximab or rituximab delivered intravenously, or with subcutaneous biological agents administered by healthcare personnel.

4 CLINICAL EFFECTIVENESS

4.1 Critique of the methods used by the manufacturer to systematically review clinical effectiveness evidence

4.1.1 Objective of systematic review, and description and critique of manufacturer's search strategy

The manufacturer's submission includes a systematic review of evidence for the clinical effectiveness of abatacept compared with either infliximab or placebo in adults with RA who have had an inadequate response to one or more conventional DMARDs including methotrexate.

Searches to identify all relevant randomised controlled trials (RCTs) were conducted in January 2010 and updated in October 2010. The search strategy utilised terms to identify the condition (RA), the intervention (abatacept), and the type of evidence (RCTs); it was said to be closely modelled on the search strategy used by Chen *et al*.,[41] and, whilst inelegant in places, appears to be largely appropriate. However, it included a term limiting findings to studies published in the English language, and no justification for this restriction is provided in the manufacturer's submission.

Three electronic bibliographic databases were searched: Medline, Embase, and the Cochrane Library. Because research registers such as the National Research Register, Clinical Trials Register, and metaRegister of Controlled Trials were not searched, key data may have been missed, particularly in relation to unpublished data and ongoing trials. Other key databases which were overlooked include the Science Citation Index (Web of Science), BIOSIS and the Conference Proceedings Index; however, two conference websites, those of the European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR), were searched for the years 2008 to 2010, and poster publications and unpublished manuscripts were provided by Bristol-Myers Squibb Pharmaceuticals Ltd. Supplementary searches such as citation searching of key papers and scanning of bibliographies of retrieved items were not reported. Nevertheless, the searches were adequate to retrieve important citations relating to all eligible studies of which the ERG and its clinical advisors are aware and, given the nature of the product and the stage of development, it is considered unlikely that the additional sources would have yielded any significant additional data.

The flow diagram relating to the clinical effectiveness literature searches (manufacturer's submission Figure B 1) is described as a QUOROM flow diagram. In fact, it more closely resembles the PRISMA statement flow diagram (http://www.prisma-statement.org/statement.htm) which replaced the QUOROM diagram in 2009.[42] The flow diagram appears to be an adequate representation of the study selection process.

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

Details of the inclusion and exclusion criteria used in the selection of evidence for the systematic review of clinical effectiveness, as specified in the manufacturer's submission (page 66), are summarised in Table 3. In the manufacturer's submission, this table was labelled 'eligibility criteria used in search strategy': however, it was presented within the description of the study selection process and clearly relates to the inclusion and exclusion criteria used at that stage.

Table 3: Eligibility criteria used in study selection (clinical effectiveness), as presented by the manufacturer[6]

Inclusion criteria	**Details**
Population	Adult patients with moderate to severe RA who inadequately responded to MTX
Intervention	Abatacept in the proposed indication
Comparators	Another biological DMARD, a conventional DMARD, or placebo (including 'do nothing' option).
Outcomes	Outcomes reported at interim time points, if necessary to enable comparisons across trials over equal time periods, and studies that include the following endpoints: • *Efficacy parameters:* Change From Baseline (CFB) in HAQ score at 24/28 and 48/54 weeks, ACR20, ACR50, ACR70 response rates at 24/28 weeks and 48/54 weeks • *Safety parameters:* Withdrawals due to adverse events at 24/28 weeks.
Study design	Human studies; published RCTs at any phase beyond Phase I that involve de novo use of the biologic therapies of interest. Open label extensions with parallel design or comparing different doses or schedules of the drug were also considered. RCTs may be blinded or unblinded.
Language restrictions	English
Exclusion criteria	**Details**
Population	Disease other than RA; patients with early RA; paediatric patients
Interventions	Other biologic therapies; conventional DMARDs
Outcomes	Laboratory measures aimed at investigating disease or treatment mechanisms; no reported relevant clinical outcome
Study design	Non-randomised and uncontrolled trials (unless an extension of an included RCT); conversion/crossover or switch studies; pharmacokinetic studies; observational studies; reviews; update or commentaries on data published elsewhere; case reports; letters to the editor; animal or *in vitro* studies
Language restrictions	Non-English

With the exception of the exclusion of non-English language studies, which is not justified, the specified inclusion and exclusion criteria are appropriate and reflect the information given in the decision problem.

4.1.3 Studies included in the clinical effectiveness review, with a table of identified studies.

In sections 5.2.1-2, the manufacturer's submission states that the electronic searches identified ten relevant publications which reported a total of three clinical trials and two long-term extension studies (LTEs). These publications included two clinical study reports (CSRs) for which, unfortunately, no references are provided. However, it is clear from Table B 3 (section 5.2.4) that the CSRs relate to the AIM and ATTEST studies. They appear to be unpublished, and it is not clear how they were identified by the electronic searches. Clarification has been obtained from the manufacturer that they were used to supplement the publicly-available and peer-reviewed full-text publications of the AIM and ATTEST studies when the latter did not report the necessary data.[2]

A further six conference abstracts, three poster presentations, and one unpublished manuscript were said to be identified by hand-searching conference websites; again, references to these citations are not given at this point. It is clear from the flow diagram (Figure B 1), but not from the text, that the six additional citations did not increase the number of studies included in the quantitative synthesis, but that they identified an additional RCT (study IM101-119[27]) which was not identified by the electronic searches because it had been published only as a conference abstract; data from this study were presented in the report but not included in the quantitative synthesis.

The electronic searches identified two additional RCTs of abatacept whose populations did not meet the review's inclusion criteria:

- A placebo-controlled study in patients who were either methotrexate-naïve or had prior exposure of $\leq$10 mg/week for $\leq$3 weeks[43]
- A placebo-controlled study in patients who had been receiving $\geq$1 biological and/or nonbiological DMARD approved for RA for at least 3 months.[44]

The manufacturer's submission excluded an LTE of phase I and phase II trials of abatacept in patients with active RA who were either methotrexate-inadequate or methotrexate-intolerant, reported only in a conference abstract for which the submission provided no reference, on the basis that the population was Japanese and therefore not of interest to the UK. No further justification of this exclusion was provided. However, the ERG's clinical advisors suggest that it may be valid because of the genetic differences between the Japanese and UK populations, and the differences which have been observed in pharmacological response and toxicity (for instance, methotrexate hepatotoxicity and leflunomide pulmonary toxicity in Japanese populations).

Details of the four RCTs included in the review of clinical effectiveness may be found in Table 4. Three (the Kremer Phase 2b, AIM, and IM101-119 studies) compared abatacept with placebo; the Kremer Phase 2b study also compared two abatacept dosing regimens. The fourth RCT, the ATTEST study, compared abatacept with both placebo and infliximab. The population, duration and outcome

measures of the three studies included in the quantitative synthesis, the Kremer Phase 2b, AIM, and ATTEST studies, differ from those found in study IM101-119.

Table 4: Characteristics of included RCTs

Study	Design and clinical trial identification codes	Participants	Intervention/comparators	Concomitant therapy	Outcomes	Follow up (weeks)
Kremer Phase 2b[28,3]	Randomised, double-blind, placebo-controlled, multicentre study Protocol number: IM101100 Clinicaltrials.gov Identifier: NCT00162266	*Inclusion criteria:* • Specified as age 18-65 years;[28] however age range goes up to 80[3] • Met ACR criteria for diagnosis of RA (functional classes I, II, or III) • ≥10 swollen joints • ≥12 tender joints • CRP level ≥1 mg/dL • Treated with MTX (10-30 mg/week) for ≥6 months, with a stable dose for 28 days prior to enrolment *Exclusion criteria:* Pregnant or nursing women	Abatacept + MTX Placebo + MTX Abatacept 2 mg/kg or 10 mg/kg Administration was by 30 minute IV infusion on days 1, 15, 30, and monthly thereafter for a total of 6 months. All participants received MTX at a level deemed appropriate by the physician (10-30 mg/week). For the first 180 days of the trial, no dosage adjustments were allowed except in cases of hepatotoxicity. Between days 180-360 the dosage could be changed provided it was <30mg/week	Leflunomide and infliximab discontinued ≥60 days before enrolment; other DMARDs other than MTX discontinued ≥28 days before enrolment. Between days 180-360, another DMARD (hydroxychloroquine, sulfasalazine, gold, or azathioprine) could be added. Corticosteroids reduced to the equivalent of ≤10 mg/day prednisone and stabilised for ≥28 days prior to day 1; between days 180-360, dose adjustment equivalent to ≤10mg/day prednisone could be made. NSAIDs permitted.	*Primary outcome measure:* ACR20 response at 6 months *Secondary outcome measures:* i) ACR50 and ACR70 responses at 6 months and 1 year ii) Improvements in individual components of the ACR core data set at 6 months and 1 year iii) pain and global assessment of disease activity (patient's and physician's) evaluated by VAS at 6 months and 1 year iv) low disease activity (DAS28<3.2) and remission (DAS28 <2.6) at 6 months and 1 year v) physical function (MHAQ score 0; improvement in MHAQ score of ≥0.22 units from baseline) at 1 year vi) adverse events at 6 months and 1 year vii) acute hypersensitivity reactions both during and following treatment administration	52
AIM[4,36,45]	Randomised, double-blind, placebo-controlled, multicentre study Protocol number: IM101102 Clinicaltrials.gov Identifier: NCT00048568	*Inclusion criteria:* • Age ≥18 years • Met ACR criteria for diagnosis of RA • RA for at least 1 year • RA persistent and active despite MTX treatment • ≥10 swollen joints • ≥12 tender joints • CRP levels ≥10.0 mg/L while receiving MTX	Abatacept + MTX Placebo + MTX Abatacept dose was by weight: patients weighing <60kg, 60-100kg, or >100kg received 500mg, 750mg or 1000mg respectively. Administration was by 30 minute IV infusion on days 1, 15, 29, and every 28 days thereafter	DMARDs other than MTX discontinued ≥28 days before randomisation. Corticosteroid use permitted at doses ≤10mg prednisone/ day, stabilised for 25 days before randomisation. Stable doses of NSAIDs also permitted.	*Primary outcome measures:* i) ACR20 response at 6 months ii) HAQ-DI improvement of ≥0.3 at 1 year iii) change from baseline in GMS joint erosion score at 1 year *Secondary outcome measures:* i) ACR50 and ACR70 responses at 6 months ii) All ACR responses, major clinical response, and protocol-	52

Study	Design and clinical trial identification codes	Participants	Intervention/comparators	Concomitant therapy	Outcomes	Follow up (weeks)
		• MTX ≥15 mg/week for ≥3 months, stable dose 28 days before enrolment *Exclusion criteria:* Patients with a positive tuberculin skin test, unless they had completed treatment for latent TB before enrolment	up to and including day 337. MTX dose was ≥15 mg/week (although 10mg/week was acceptable if the patient had a history of toxicity).	Between 6-12 months, dose modification was permitted for MTX and oral corticosteroids (equivalent of ≤10 mg/day prednisone); addition of 1 other DMARD (hydroxychloroquine, sulfasalazine, gold, or azathioprine) was also permitted.	defined extended major clinical response, at 1 year iii) Change in disease activity (DAS28) iv) improvements in physical function over 1 year using HAQ-DI v) changes in HRQoL (SF36)	
ATTEST[26]	Randomised, double-blind, double-dummy, placebo- and active-controlled, multicentre study Protocol number: IM101043 Clinicaltrials.gov Identifier: NCT00095147	*Inclusion criteria:* • Age ≥18 years • Met ACR criteria for diagnosis of RA • RA for at least 1 year • inadequate response to MTX treatment as demonstrated by ongoing disease activity (≥10 swollen joints, ≥12 tender joints, and CRP levels ≥1 mg/dl) • MTX ≥15 mg/week for ≥3 months prior to randomisation (stable for ≥28 days) • No DMARDs other than MTX received for ≥28 days • No prior experience of abatacept or anti-TNF therapy *Exclusion criteria:* Not reported	Abatacept + MTX Infliximab 3 mg/kg+ MTX Placebo + MTX Abatacept dose was by weight (patients weighing <60kg, 60-100kg, and >100kg received 500mg, 750mg or 1000mg respectively). Administration was by 30 minute IV infusion on days 1, 15, 29, and every 28 days thereafter up to and including day 337, with normal saline on day 43. Placebo was infused simultaneously over 2 hours to ensure blinding. Infliximab administration was by 2 hour IV infusion on days 1, 15, 43, 85 and every 56 days thereafter up to and including day 337, with normal saline on the remaining visit days.	Concomitant medications permitted between days 1-197 were oral corticosteroids (equivalent of ≤10 mg/day prednisone and stabilised for ≥25 out of 28 days prior to randomisation) and/or stable NSAIDs including acetyl salicylic acid, and analgesics not containing aspirin or NSAIDs. Between days 198-365, dose modification was permitted for MTX (≤25 mg/week) and oral corticosteroids (equivalent of ≤10 mg/day prednisone); hydroxychloroquine, sulfasalazine, gold, or azathioprine were also permitted.	*Primary outcome measure:* Reduction in disease activity, measured by DAS28 (ESR), for abatacept vs placebo at 6 months *Secondary outcome measures:* i) Reduction in DAS28 (ESR) with infliximab vs placebo at 6 months ii) Reduction in DAS28 (ESR) with abatacept vs infliximab at 6 months and 1 year iii) DAS28 (ESR) EULAR responses at 6 months and 1 year iv) low disease activity (DAS28 ≤3.2) and remission (DAS28 ≤2.6) at 6 months and 1 year v) ACR20, 50 and 70 responses at 6 months and 1 year vi) ≥0.3 improvement from baseline in HAQ-DI at 6 months and 1 year vii) Changes in the physical and mental component summary scores and 8 subscales of the SF-36 at 6 months and 1 year	52*

Study	Design and clinical trial identification codes	Participants	Intervention/comparators	Concomitant therapy	Outcomes	Follow up (weeks)
			Placebo was infused simultaneously over 30 minutes to ensure blinding. Treatment with placebo was limited to days 1-197; on day 198, placebo-treated patients were reallocated to abatacept, but blinding was maintained; patients initially randomised to abatacept or infliximab continued their treatment		*Tertiary outcome measure:* Comparative safety of abatacept and infliximab at 1 year	
IM101-119[27,6]	Randomised, double-blind, placebo-controlled, multicentre, multinational study Protocol number: IM101119 Clinicaltrials.gov Identifier: NCT00420199	*Inclusion criteria:* • Active RA despite MTX therapy, defined as either DAS28 ≥3.2, or ≥6 tender and ≥6 swollen joints and CRP above upper limit of normal • Clinically detectable synovitis of ≥1 wrist at screening and baseline • ≥1 erosion shown by X-ray or positive for anti-CP or RF	Abatacept ~10 mg/kg + MTX Placebo + MTX Administration on days 1, 15, 29, and every 28 days thereafter up to and including day 113		*Primary outcome measure:* Changes in wrist synovitis score as measured by the OMERACT-RAMRIS method, at 4 months *Secondary outcome measures:* • Changes in bone lesions (ie bone oedema, bone erosions) in hands/wrists • Changes in biochemical markers in bone, cartilage, and synovial fluid metabolism • Safety • Tolerability *Exploratory efficacy analyses:* • DAS28 (LDAS ≤3.2; remission <2.6)	16

* with reallocation of placebo group to abatacept on day 198 (with blinding maintained)

The manufacturer's submission also identified as relevant four non-RCT studies: the long-term extensions of the three major included RCTs (the Kremer Phase 2b, AIM, and ATTEST studies) in which all patients received abatacept and methotrexate regardless of original treatment allocation, and an integrated analysis of safety data from the three major studies included in this review, together with the ATTAIN[46] and ARRIVE[47] studies in patients with an inadequate response to TNF inhibitors, the ASSURE study in patients taking one or more traditional nonbiologic and/or biologic DMARD,[44] the Phase 2 abatacept plus etanercept study,[48] and Buch *et al.*,'s Phase 2 Mode of Action study.[49] These studies were said to be presented in two publications, four conference abstracts, three poster presentations, and one unpublished manuscript. This is inconsistent with Table B 4, which lists three publications, five conference abstracts, three poster presentations, and two unpublished reports. In turn, Table B 4 is inconsistent with the reference list in listing Schiff *et al.*, 2008 as a study reference for the LTE of the ATTEST study; in fact it is Schiff *et al.*,'s 2008 publication presents the randomised double-blind phase of the ATTEST study,[26] and the relevant reference appears to be the 2009 conference abstract in which Schiff *et al.*, presented safety data from the ATTEST LTE,[50], while a second 2009 abstract by the same authors which presented efficacy data from that LTE[51] appears to have been ignored. Details of these studies are presented in Table 5.

It should be noted that the integrated analysis includes data from patients who did not meet the criteria for inclusion in the systematic review of clinical effectiveness. However, the manufacturer felt that it was appropriate to include this analysis in the submission as evidence of the long-term safety and tolerability of abatacept. Because it includes data from the patients who participated in the included clinical trials and their LTEs, its inclusion in the manufacturer's submission together with those trials and their LTEs clearly raises the possibility of double-counting.

No observational studies or publications of post-marketing surveillance data were identified which were independent of the clinical trials programme.

Table 5: Characteristics of included non-RCT studies

Study	Relevant publications	Objective	Population	Comparison
LTE of the Kremer Phase 2b trial	Westhovens *et al*., 2009a[52]	To evaluate the efficacy and safety of abatacept + MTX over 5 years in patients with RA	Patients who completed the 1 year double-blind period of the Kremer Phase 2b study were eligible to enter the open-label LTE period in which all received a fixed dose of abatacept (~10 mg/kg) + MTX	Efficacy data presented only for patients originally randomised to abatacept and who received at least 1 infusion of abatacept during the open-label LTE. Safety assessed for all patients who received at least 1 infusion of abatacept during the open-label LTE, including those originally randomised to placebo.
	Westhovens *et al*., 2009b[53]	To evaluate the efficacy and safety of abatacept + MTX over 8 years in patients with RA		Efficacy data presented only for patients randomised to abatacept with available data at the visit of interest (as-observed). Safety assessed for all patients who received ≥1 dose of abatacept
LTE of the AIM trial	Genant *et al*., 2008[54] Kremer *et al*., 2008[55]	To evaluate the efficacy and safety of abatacept + MTX over 2 years in patients with RA	Patients who completed the 1 year AIM trial of abatacept ~10 mg/kg + MTX vs placebo + MTX were eligible to enter the open-label LTE period in which all received a fixed dose of abatacept (~10 mg/kg) + MTX	Results available for all patients, ie including treatment switch from placebo (as-observed analysis). Safety assessed for all patients who received ≥1 dose of abatacept.
	Kremer *et al*., unpublished[56]	To evaluate the efficacy and safety of abatacept + MTX over 4 years in patients with RA		
	Genant *et al*., 2009[57] Kremer *et al*., 2009[58]	To evaluate the efficacy and safety of abatacept + MTX over 5 years in patients with RA		
	Kremer *et al*., 2010[59]	To evaluate HRQoL over 5 years in patients with RA		Results summarised over time by original randomisation group using point estimates for patients who received ≥1 dose of abatacept in the LTE (as-observed data)
LTE of the ATTEST trial	Schiff *et al*., 2009[50] Schiff *et al*., 2009[51] Abatacept CSR	To evaluate the efficacy, safety and long-term tolerability of abatacept in patients who had completed the initial 12-month double-blind treatment period	Patients who completed the 1 year double-blind period (abatacept 10 mg/kg + MTX or infliximab 3 mg/kg + MTX vs placebo + MTX for 6 months, then abatacept 10 mg/kg for the 2nd 6 months) were eligible to enter the open-label LTE period in which all patients were allocated to abatacept 10 mg/kg	Results for all patients who received ≥1 dose of abatacept during the open-label period.
Integrated analysis of abatacept trials	Becker *et al*., 2010[60]	To evaluate the safety of abatacept over short- and long-term periods	Patients with RA receiving abatacept in the cumulative abatacept clinical development programme (double-blind plus open-label periods)	Safety assessed for patients who received ≥1 dose of abatacept
	Hochberg *et al*., 2010[61]			
	Smitten *et al*., 2010[62]			

The manufacturer's submission identified the following completed and ongoing studies from which additional evidence is likely to be available in the next 12 months:

- the AIM study: the submission stated that a further publication was in progress and would be submitted to a peer-reviewed journal before the end of 2010
- study IM101-119: the submission anticipated that the primary outcome and 4-month results would be available towards the end of 2010, with publications planned for the second and third quarters of 2011, and that the one-year results would be available in the first quarter of 2011, with publication following. The ERG team notes that the study completed in May 2010 (www.clinicaltrials.gov) and a document entitled "Final Clinical Study Report for Study IM101119" is readily available on the internet,[63] although it bears the caveat that it is a confidential communication and that no unpublished information contained therein may be published or disclosed without Bristol-Myers Squibb Company's prior written approval. However, despite claiming to be a final report, it only contains the 4-month results, not the one-year results.
- study IM101-179: the manufacturer's submission notes that this multicentre, open-label study is still recruiting, and that its estimated completion date is October 2011; its aim is to assess early response to abatacept with background methotrexate using power doppler ultrasonography in patients with active RA and inadequate response to methotrexate. The clinicaltrials.gov website indicates that this is not a randomised study (http://clinicaltrials.gov/ct2/show/NCT00767325?term=IM101-179&rank=1).

The manufacturer's submission also states that 4.5-year data (4-year LTE) from the ATTAIN study are due to be submitted for publication in a peer-reviewed journal before the end of 2010. This is relevant only in that the ATTAIN study, whose population was limited specifically to patients with active RA and an inadequate response to TNF inhibitors,[46] contributed data to the integrated safety analysis.

4.1.4 Details of relevant studies which were not discussed in the submission

Because of time constraints, it was not possible to conduct independent searches of the electronic databases to verify whether the manufacturer's submission included all published studies which compared abatacept with a relevant comparator in the relevant population. However, the manufacturer's Medline search strategy was rerun: it failed to identify one of the major publications from the AIM study,[36] and may therefore also have missed other relevant publications. Moreover, it is not clear whether any relevant studies were excluded by the restriction of the searches to English language papers.

The ERG team also searched ClinicalTrials.gov (http://clinicaltrials.gov), and identified one additional relevant study which was said to have available results. This was a small phase 3 study designed to demonstrate the clinical efficacy at six months of abatacept (at a dose of 10mg/kg) plus methotrexate relative to placebo plus methotrexate in Korean patients with active RA and an inadequate clinical response to methotrexate (study ref ID NCT00409838, http://clinicaltrials.gov/ct2/show/NCT00409838?term=rheumatoid+arthritis+AND+abatacept&rank=1). As noted above, the manufacturer's submission excluded another study on the basis that its population was Japanese. It is therefore possible that the Korean study was excluded for a similar reason, but this is not stated.

Two relevant articles relating to studies identified by the manufacturer's systematic review were not referenced in the manufacturer's submission. Emery *et al*.,'s study of health-related quality of life data from the Kremer Phase 2b study[5] was not identified by the manufacturer's Medline search strategy; it came to the ERG's attention because it was identified by the recent Cochrane review of abatacept for rheumatoid arthritis.[15] The manufacturer's Medline search strategy did identify the paper by Wells *et al*.,[40] which analysed the impact of abatacept on sleep quality in the AIM and ATTAIN studies, and which presented the data from the two studies separately, but for reasons which are not clear the manufacturer did not consider this paper relevant to the submission.

The manufacturer's systematic review of clinical effectiveness was limited to studies which compared abatacept with either infliximab or placebo. It is implied, but not stated, that no studies were identified which compared abatacept directly with adalimumab, certolizumab pegol, etanercept, or golimumab.

4.1.5 Description and critique of the manufacturer's approach to validity assessment

A formal appraisal of the validity of the three main RCTs was clearly presented in the manufacturer's submission. This addressed all the criteria specified in the NICE STA Specification for manufacturer/sponsor submission of evidence. However, the summary of quality assessment results presented in the submission's Table B10 differed from the full assessment presented in Appendix 3 Table 1 in indicating that the concealment of treatment allocation was adequate in the three studies, whereas both the text and Appendix 3 stated that, in all three, it was not clear whether allocation was adequately concealed. Subsequent clarification from the manufacturer stated that this was a typographical error, and that the text should read 'concealment of treatment allocation was adequate'.[2] This is consistent with the Cochrane reviewers' statement that the use of a central randomisation procedure in the Kremer Phase 2b and AIM studies indicated adequate allocation concealment;[15] although the published article relating to the ATTEST study[26] did not mention allocation concealment, and the Cochrane reviewers were unable to obtain additional information from the manufacturer, they thought it likely that concealment would have been adequate since it appeared to have been adequate

in other trials of abatacept.[15] The corrected data are presented in Table 6. It should be noted that the manufacturer's submission states, in Appendix 3, that there were unexpected imbalances in drop-outs between groups in the ATTEST study, and that it was not clear whether, in the Kremer Phase 2b study, the authors measured more outcomes than they reported (for details, see Table 6).

The manufacturer's submission did not include a formal appraisal of the validity of study IM101-119, presumably on the basis that it had only been published in abstract form. However, it should be noted that, as the manufacturer both sponsored the study and had access to the CSR, they should have had access to sufficient information to undertake such an appraisal.

Table 6: Manufacturer's quality assessment of abatacept RCTs (from manufacturer's submission Appendix 3, Table 1, corrected following clarification[2])

Trial	Kremer Phase 2b (n=339)	Grade (yes/no/not clear/NA)	AIM (n=656)	Grade (yes/no/not clear/NA)	ATTEST (n=431)	Grade (yes/no/not clear/NA)
Was randomisation carried out appropriately?	Central randomisation system. Patients were randomly assigned with use of a permuted-block size of 6. Randomly assigned in a ratio of 1:1:1	Yes	At enrolment each subject was assigned a unique sequential subject number by the Central Randomisation System (CRS). Each subject who was qualified for treatment was assigned a unique randomisation number via the CRS in the order in which subjects qualified for treatment, not in the order of study enrolment. Patients were randomly assigned in a 2:1 ratio.	Yes	At enrolment, each subject was assigned a unique (at the site level) sequential subject number via the CRS. Subject numbers were not reused. Each subject who was qualified for treatment was assigned a unique randomisation number via the CRS in the order in which subjects qualified for treatment, not in the order of study enrolment. Randomised by centre in a 3:3:2 ratio	Yes
Was the concealment of treatment allocation adequate?	Not reported	Yes	Randomisation schedules were generated and kept sealed by the sponsor's Randomisation Group until study unblinding.	Yes	Randomisation schedules were generated and kept sealed by the sponsor's Randomisation Group until study unblinding.	yes
Were the groups similar at the outset of the study in terms of prognostic factors?	Demographic and baseline clinical characteristics were similar for both treatment	Yes	Demographic and baseline clinical characteristics were similar for both treatment groups	Yes	Demographic and baseline clinical characteristics were similar for the 3 treatment groups	Yes
Were the care providers, participants and outcome assessors blind to treatment allocation?	Double-blind	Yes	Double-blind: the subjects and clinical assessor(s) were blinded to treatment assignment. The pharmacist (or qualified drug preparation person) was unblinded to study medication and dose level, and prepared the appropriate dose of active drug or placebo.	Yes	Double-blind: the subjects and clinical assessor(s) were blinded to treatment assignment. Although dosing regimens for abatacept and infliximab were different, subjects received normal saline at some dosing visits to maintain the integrity of the blinding. The pharmacist (or qualified drug preparation person) was unblinded to study medication and dose level, and prepared the appropriate dose of active drug or placebo.	Yes

Trial	Kremer Phase 2b (n=339)	Grade (yes/no/not clear/NA)	AIM (n=656)	Grade (yes/no/not clear/NA)	ATTEST (n=431)	Grade (yes/no/not clear/NA)
Were there any unexpected imbalances in drop-outs between groups?	Subject could discontinue study for multiple reasons. • A greater proportion of subjects in the abatacept arm group (86% 10mg and 78% 2mg arm) completed 169 days of treatment compared with the placebo group (66%). • Lack of efficacy was the most common reason for discontinuation in the placebo (24%) and abatacept groups (10% 10mg and 12% 2mg arm). No unexpected imbalances in drop-outs were reported by authors and/or noted by reviewer.	No	Subject could discontinue study for multiple reasons. • A greater proportion of subjects in the abatacept group (93%) completed 169 days of treatment compared with the placebo group (79%). • Lack of efficacy (15%) was the most common reason for discontinuation in the placebo group. • AEs (3%) and lack of efficacy (3%) were the most common reasons for discontinuation in the abatacept group. No unexpected imbalances in drop-outs were reported by authors and/or noted by reviewer.	No	Subject could discontinue study for multiple reasons. • The overall rate of discontinuation during the first 6 months, as well as in the study period up to one year, was highest for the infliximab group. • The main reason for discontinuation in the infliximab group was due to AEs (4.8% for infliximab, 0.9% for placebo, and 1.3% for abatacept at 6 months).	Yes
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No study protocol reviewed	Unclear	No	No	No	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?	• All statistical analyses carried out on ITT population—defined as all patients who received at least 1 treatment infusion. • Missing data for all patients who discontinued from the study due to worsening RA disease (lack of efficacy) were imputed as non-responders from the time of discontinuation, for ACR 20 response.	Yes	• All efficacy and safety analyses carried out on a modified ITT population—defined as all patients randomly assigned who received at least 1 dose of study medication. • Missing data for patients who discontinued were imputed as non-responders subsequent to treatment discontinuation for ACR20 response and HAQ-DI. • Additional sensitivity analyses were performed. • Missing annual radiographic data were imputed with linear extrapolation for discontinued patients on the basis of the baseline value and the on treatment	Yes	• All efficacy and safety analyses carried out on a modified ITT population—defined as all patients randomly assigned who received at least 1 dose of study medication. • Patients who discontinued the study prematurely were considered as non-responders subsequent to the time of discontinuation for ACR 20, 50 and 70 responses, good EULAR responses and clinically meaningful HAQ-DI responses. • Missing data for all continuous measurements (mean changes in DAS28, SF-36 and the HAQ-DI score). LDAS and DAS28-defined remission were imputed as the last observations prior to the discontinuation.	Yes

Trial	Kremer Phase 2b (n=339)	Grade (yes/no/not clear/NA)	AIM (n=656)	Grade (yes/no/not clear/NA)	ATTEST (n=431)	Grade (yes/no/not clear/NA)
	• However, patients who discontinued for other reasons had their last observations carried forward.		assessment at the time of discontinuation.			

The ERG checked the findings of the manufacturer's quality assessment against the original study publications, and the assessment of quality included in the recent Cochrane review of abatacept for rheumatoid arthritis.[15] The individual studies are discussed in turn below, following a summary of recruitment to, and discontinuation from, those studies. In both the Kremer Phase 2b and AIM studies, the discontinuation rates are noticeably higher in the placebo groups than in the abatacept groups; this is largely due to discontinuations due to lack of efficacy. By contrast, in the ATTEST study, fewer discontinuations were reported in the placebo group than in the active treatment groups; surprisingly, there were more discontinuations due to lack of efficacy in the infliximab group, and to a lesser extent the abatacept group, than in the placebo group (for details, see Table 7).

Table 7: Recruitment to, and discontinuation from, the included studies

Study	**Kremer Phase 2b[3,64]**			**AIM[4]**		**ATTEST[26]**			**IM101-119[6]**	
Number screened	544			1250		748			63	
Number rejected	205			594		317			13	
Reasons for rejection	NR									
No longer met study criteria	-			519		271			10	
Withdrew consent	-			33		34			1	
Lost to follow-up	-			3		2			0	
Administrative reasons	-			1		1			1	
Adverse events	-			1		1			0	
Poor/non-compliance	-			0		1			0	
Other	-			0		7			1	
Allocation	Placebo	Ab 2mg/kg	Ab 10mg/kg	Placebo	Ab 10mg/kg	Placebo	Ab 10mg/kg	Inflix	Placebo	Ab 10mg/kg
Number randomised	119	105	115	656		NR	NR	NR	23	27
Number who received the intervention	NR	NR	NR	219	433	110	156	165	23	27
Number completing the study	71 (59.7%)	74 (70.5%)	90 (78.3%)	162 (74.0%)	385 (88.9%)	104 (94.5%)	139 (89.1%)	141 (85.5%)	23 (100%)	26 (96.3%)
Primary reasons for study discontinuation										
Lack of efficacy	30	17	13	40	13	2	4	12	0	0
Adverse event	11	9	5	4	18	1	4	6	0	0
Withdrew consent	6	2	5	5	10	1	4	2	0	0
Death	0	1	0	1	1	1	1	1	0	0
Lost to follow-up	0	2	1	1	1	0	2	2	0	0
No longer met study criteria	0	0	0	3	2	0	0	0	0	1
Pregnancy	0	0	0	0	2	0	0	0	0	0
Poor or non-adherence	0	0	0	3	1	0	0	0	0	0
Other	1	0	1	0	0	1	2	1	0	0

The Kremer Phase 2b study was considered by the Cochrane reviewers to be at high risk of bias because the drop-out rate at 12 months exceeded 20%, and the resulting incomplete data were not felt to be addressed adequately for either efficacy or safety outcomes. The method used was imputation of missing data using the last observation carried forward: patients who discontinued the study because of worsening disease were considered to have had no response, while for those who discontinued the study for other reasons the values for the last efficacy observation were carried forward.[28] This use of two separate criteria for imputing data was considered potentially inappropriate: the Cochrane reviewers noted that, for example, if a participant did not tell investigators that the reason for no longer attending follow-up visits was worsening disease, the last observation would be carried forward, whereas in fact the patient should have been considered to have had no response. In addition, the method did not allow for the possibility that some patients might have multiple reasons for withdrawal, and might or might not share all of these with study staff.[15] It is clearly important that missing data are handled appropriately, particularly in studies like the Kremer Phase 2b study where the probability of drop-out differs between the intervention and placebo group, leading to the possibility of biased comparisons of effect.

The AIM study was also considered by the Cochrane reviewers to be at high risk of bias because of the exclusion of nonadherent patients from the efficacy analyses. Kremer *et al*., state that nine abatacept-treated patients and five placebo recipients from one site were excluded from all efficacy analyses before unblinding due to nonadherence, but were included in all safety analyses;[4] these patients seem to be additional to the four patients shown in the study flow diagram as not completing the study because of poor or nonadherence. The Cochrane reviewers therefore describe the study as having performed a 'modified intention-to-treat analysis' on subjects who received at least one infusion of study medication.[15] However, the proportion of subjects excluded for nonadherence was less than 1% of those randomised, suggesting that any consequent bias might not be substantial. Some patients were excluded from efficacy analyses because of protocol violations but included in the safety analyses, and therefore the Cochrane reviewers deemed the study to have a higher risk of bias for efficacy outcomes than for safety outcomes.[15] The imbalance between treatment groups in the proportion of patients who received additional DMARDs between 6 and 12 months (15 (3.7%) patients in the abatacept group and 25 (14.4%) in the placebo group, $p<0.001$),[4] is likely to have reduced rather than exaggerated the treatment effect.

The Cochrane reviewers judged the ATTEST study to have a low risk of bias at 6 months because the completion rate exceeded 80%.[15] However, there were some imbalances between treatment groups at baseline in that fewer patients in the placebo group tested positive for

rheumatoid factor, while the duration of methotrexate therapy was lower in abatacept group. In addition, between days 198-365, 12.8% of patients in the original abatacept arm and 17.6% of those in the original infliximab arm received either an additional DMARD or an increased dose of methotrexate/corticosteroids; this is likely to disadvantage abatacept relative to infliximab.

It is difficult to comment on study IM101-119 because of the lack of quotable information in the public domain. Thus, while it is said to be a randomised, double-blind study, no information is available regarding the method of randomisation or allocation concealment. However, such information as is available shows an imbalance between treatment groups: a higher proportion of patients in the placebo than in the abatacept group tested positive for RF and anti-CCP2 (for details, see Table 9).

In all four studies, because of the largely subjective nature of the outcome measures, the blinding of patients, clinical staff, and outcome assessors to treatment allocation is crucial. All four studies were said to be double-blind, but none undertook an assessment of the success of the blinding.

All the included studies were funded by Bristol-Myers Squibb. It should be noted that there is evidence to suggest that industry-sponsored trials overestimate treatment effects.[65]

4.1.6 Description and critique of the statistical approach used within each relevant trial

The statistical analyses used in the RCTs included in the clinical effectiveness section are summarised in Table 8. They are generally acceptable. However, it should be noted that, in the Kremer Phase 2b study, the objective is based on response rates at 6 months, whereas the main analysis with adjustment for multiple comparisons is based on response rates at 12 month. In the ATTEST study, the alternative hypothesis does not specify the size of the treatment effect which the study was designed to detect. Furthermore, the statement that "the χ^2 test was performed to evaluate the differences (and 95% CIs) between the abatacept or infliximab groups and placebo" should specify the variables. The absence of information relating to the power calculation used in study IM101-119 is unfortunate, particularly as the results suggest that it may have been underpowered in relation to all outcomes, including its primary outcome (see section 4.2.1).

Table 8: Statistical analyses used in included RCTS (data from study publications supplemented where necessary by the manufacturer's submission)

Trial	Alternative hypotheses, objectives	Statistical analysis	Sample size, power calculation	Data management, patient withdrawals
Kremer Phase 2b[28,3]	Alternative hypothesis: 25% difference in ACR20 responses between the 2 abatacept groups and the placebo group Objective: Study was designed to evaluate the % of patients with 20% improvement in ACR at 6 months	• Efficacy analyses were carried in all patients who received at least 1 dose of study medication • Differences in ACR20, ACR50, and ACR70 response rates on day 360 were analysed by comparing each abatacept treatment group with the placebo group using a Dunnett-adjusted chi-square test. ACR response rates at other time points were compared between each abatacept treatment group and the placebo group using a chi-square test unadjusted for multiple comparisons • Differences in percentage CFB to LOCF for all ACR core components were analysed using analysis of covariance with the baseline value as the covariate and without adjustment for multiple comparisons • Fisher's exact tests were used to compare incidence of AEs between the abatacept groups and placebo group • For all other endpoints, discrete variables were analysed using chi-square tests, and all continuous variables were analysed using t-tests unadjusted for multiple comparisons • All statistical tests were conducted using a 5% significance level (2-tailed)	• A sample size of 107 patients per treatment group was calculated to yield 94% power to detect a difference of 25% in ACR20 responses between the abatacept 10 mg/k group and the placebo group at the 5% significance level (2-sided), assuming an ACR20 response rate in the placebo group of 25% at 6 months and allowing for a discontinuation rate of 15% in each group.	• ACR responses: all patients who discontinued from the study due to worsening RA disease (lack of efficacy) were considered non-responders from that time point. • However, patients who discontinued for other reasons had their last observations carried forward.

Trial	Alternative hypotheses, objectives	Statistical analysis	Sample size, power calculation	Data management, patient withdrawals
AIM[4]	Alternative hypothesis: 20% difference in ACR20 between the 2 treatment groups Objectives: Study was designed to evaluate the % of patients with 20% improvement in ACR at 6 months, % patients with clinically significant improvement (≥0.3) in HAQ-DI at one year, and radiographic progression of joint erosions at one year	• All efficacy and safety analyses performed on a modified ITT population, defined as all randomly assigned patients who received at least 1 dose of study medication • All statistical tests based on a 2-sided 5% level of significance • Co-primary analyses of ACR20 at 6 months and HAQ-DI response at one year: 2-sided, continuity- corrected chi-square test to compare responses of abatacept group with those of the placebo group • A rank-based analysis of covariance was used to compare CFB in GMS scores between treatment groups at 1 year • HAQ-DI CFB and SF-36: analysis of covariance (ANCOVA) with LOCF using a longitudinal linear mixed-effects model to compare the CFB between treatment groups • DAS28: 2-sided, continuity-corrected chi-square test to compare the responses of abatacept with those of the placebo group • AE: incidence of AE summarised; 95% CIs were used for comparison between groups	• Protocol estimated that 680 patients needed to be enrolled to randomly assign 540 patients. Sample sizes based on a 5% level of significance (2-tailed) • Study had 99% power to detect a difference of 20% in ACR20 between the 2 groups • Sample size allowed detection of an 18% difference in HAQ-DI response rate between the 2 groups, with 98% power • Sample size allowed detection of a 60% reduction in the GMS erosion score relative to placebo (assuming an increase of 1.27 units in placebo for the CFB) with 90% power	• ACR20 and HAQ-DI: missing data for patients who discontinued were imputed as non-responders subsequent to the discontinuation. Additional sensitivity analyses were performed to assess the effect of imputation of missing data. • GMS scores: primary analysis included all observed data at baseline and 12 months. • Missing annual radiographic data was imputed with linear extrapolation for discontinued patients on the basis of the baseline value and the on treatment assessment at the time of discontinuation, if both available. Sensitivity analyses were performed.

Trial	Alternative hypotheses, objectives	Statistical analysis	Sample size, power calculation	Data management, patient withdrawals
ATTEST[26]	Alternative hypothesis: the mean change from baseline in DAS28 (ESR) at 6 months will be different with abatacept and placebo Objective: Study was designed to evaluate the reduction in DAS28 (ESR) with abatacept vs placebo at 6 months	• All efficacy and safety analyses on a modified ITT, defined as all patients randomly assigned who received at least 1 dose of study medication • The abatacept and infliximab groups were compared to Placebo with respect to CFB to Day 197 in DAS28 and in the SF-36 (PCS and MCS) using an analysis of covariance (ANCOVA) model with treatment group as the effect and baseline value as the covariate. • Point estimates, 95% CIs, and p-values were computed for the treatment difference within the framework of the ANCOVA model • The proportion of patients with ACR20, 50 and 70 responses, LDAS, DAS28-defined remission, a good EULAR response, and a clinical meaningful HAQ-DI response was calculated. • The χ^2 test was performed to evaluate the differences (and 95% CIs) between the abatacept or infliximab groups and placebo	• Post-hoc analysis of DAS28 in study Phase 2b demonstrated a 0.88 unit improvement in DAS28 changes at 6 months for 10 mg/kg abatacept compared with placebo, with a 1.25 unit standard deviation. • A total of 150 abatacept-treated subjects and 100 placebo-treated subjects would yield over 99% power to detect a 0.88 unit treatment difference, assuming the same standard deviation and a 20% dropout rate. If the underlying treatment difference was as modest as 0.59 units, the study was still powered at 90% for this endpoint given this sample size. The above calculations were based on a 2-tailed 5% level of significance • Prospectively, this study was not powered for pre-specified comparisons of abatacept with infliximab	• Patients who discontinued the study prematurely were considered as non-responders subsequent to the time of discontinuation for ACR20, 50 and 70 responses, good EULAR responses and clinically meaningful HAQ-DI responses • For all continuous measurements (mean changes in DAS28, SF-36 and the HAQ-DI score), LDAS and DAS28-defined remission, the last observations prior to the discontinuation were carried forward (LOCF) • To assess the effect of antirheumatic medications, a predefined sensitivity analysis was conducted on the data with the last DAS28 (ESR) score just prior to the initiation of the additional DMARD or any increase in MTX or corticosteroid use during days 198-365 carried forward.

Trial	Alternative hypotheses, objectives	Statistical analysis	Sample size, power calculation	Data management, patient withdrawals
IM101-119[27]	Alternative hypothesis: not clear Objective: to evaluate the impact of abatacept or placebo, plus MTX, on MRI changes in wrist synovitis, osteitis, and bone erosion at 4 months	• Comparisons in wrist synovitis change between groups were based on non-parametric ANCOVA • Efficacy was reported for the ITT population • Safety was reported for all randomly assigned patients who received at least 1 dose of study medication	• No data	• One patient on abatacept + MTX discontinued the study during the double-blind period as they no longer met the study criteria due to hyperparathyroidism

Subgroup analyses

The manufacturer's submission states that, in the AIM trial, subgroup analyses were performed by age, gender, race, geographic region, duration of RA, swollen joint count (SJC), tender joint count (TJC), CRP levels, weight, GMS total score, HAQ-DI, and ACR responses at Day 169, wherever applicable, while in the ATTEST trial, subgroup analyses were performed by age, gender, race, geographic region, duration of RA, SJC, TJC, CRP levels and RF status. These analyses were performed to comply with the ethical guidelines for clinical trials. No statistical testing was performed for these subgroups because they were not powered to detect any differences between the treatment groups, but the analyses were held to demonstrate the consistency and robustness of efficacy results across different subpopulations and compared to the entire study population.[6] Further clarification obtained from the manufacturer emphasised that subgroup analyses by baseline disease severity and level of inflammation were performed to check the robustness of the studies, and whether the overall study results were driven by a specific subgroup of patients with greater disease severity at baseline.[2] The results of such analyses were not presented in the manufacturer's submission.

No subgroup analyses were performed in the Kremer Phase 2b trial.

4.1.7 Description and critique of the manufacturer's approach to outcome selection within each relevant trial

The manufacturer's submission listed the outcomes from the included studies which they perceived as relevant to the decision problem follows:

- ACR 20/50/70 responses
- Proportion of patients with a clinically significant improvement in the HAQ-DI (a measure of patient-assessed physical functioning)
- Genant-modified Sharp (GMS) score (a radiographic measurement of joint damage)
- DAS28 change from baseline
- Patient-reported outcomes including morning stiffness, sleep quality, and fatigue
- Patient compliance as measured by the number of missed infusions.

The manufacturer's submission did not list either pain (except inasmuch as it is incorporated into the ACR 20/50/70 responses, and the physical component summary measure of the SF-36) or extra-articular manifestations of disease as outcomes which were perceived to be relevant to the decision problem; they further stated that information on pain was not available in a suitable format for presentation.[2] Despite this, data relating to the pain component of the SF-36 are presented graphically in Figures B 18-22,[6] and published data relating to pain were available for both the Kremer Phase 2b study[3] and the AIM study.[4]

The common primary endpoint in the Kremer Phase 2b and AIM studies was the proportion of patients with an ACR20 response at 6 months. ACR50 and ACR70 responses at 6 months and one year were included as secondary outcomes measures in the Kremer Phase 2b, AIM, and ATTEST studies. However, the primary outcome in the ATTEST study was reduction in disease activity, as measured by the DAS28 (ESR) for abatacept versus placebo at 6 months.

The primary endpoint in study IM101-119 was the change in wrist synovitis score and 4 months, measured by the OMERACT-RAMRIS method.[27] The only other study to assess structural damage was the AIM study, in which change from baseline in bone erosion and joint space narrowing at one year using the GMS system was a primary outcome.[4]

The Kremer Phase 2b, AIM, and ATTEST studies measured physical function using either the HAQ-DI (AIM[4] and ATTEST[26]) or the MHAQ (Kremer Phase 2b[28]). The de novo economic analysis in the manufacturer's submission used change from baseline in an unspecified version of the HAQ as its outcome measure (see Section 6.2).

Other secondary outcome measures included health-related quality of life as measured by the SF-36 scores, global assessment scales, and numbers of adverse events.

4.1.8 Discussion of the extent to which each relevant trial includes the patient population(s), intervention(s), comparator(s) and outcomes as defined in the final scope

The populations included in the relevant trials correspond with that defined in the final scope, ie adults with rheumatoid arthritis who have had an inadequate response to one or more conventional DMARDs, including methotrexate. However, they differ from the population which is eligible for biological agents under NICE guidelines,[14] which is limited to patients with a DAS28 score of ≥5.1. The one study which incorporated a DAS28 score into its inclusion criteria, study IM101-119,[27] recruited patients with moderate as well as severe RA, using a lower threshold than the NICE guidelines (DAS28(CRP) >3.2, or≥6 tender and swollen joints and CRP above the upper limit of normal).

The intervention defined in the final scope is abatacept in combination with methotrexate. Only the AIM study[4] used abatacept at its licensed dosage, ie a fixed dose of 500, 750, or 1000mg for patients weighing <60kg, 60-100kg, and >100kg respectively. The remaining three studies used a broadly similar dose of 10mg/kg. The Kremer Phase 2b study compared this dose with a lower dose of 2mg/kg.[28]

The comparators used in the relevant trials were placebo and also, in the ATTEST study,[26] infliximab. No trial was identified which compared abatacept with the other comparators identified in the final scope – ie a conventional DMARD (eg sulfasalazine or leflunomide) or the biological agents adalimumab, etanercept, certolizumab pegol, or golimumab.

The outcome measures used in the relevant trials addressed some of the outcomes defined in the final scope – in particular, disease activity and physical function. Only two studies, Kremer Phase 2b[28] and study IM101-119[27], reported joint damage. Pain, which is an important outcome, was poorly reported. Moreover, only one study, the AIM study,[4] reported fatigue, although this, together with sleep disturbance (also reported in the AIM study), has been identified by patients with RA as an outcome which is important to them.[24] None of the trials or their LTEs assessed extra-articular manifestations of disease.

4.2 Summary and critique of submitted clinical effectiveness evidence

The manufacturer's systematic review identified four RCTs. These are, in chronological order:

- The Kremer Phase 2b study (IM101100)
- The AIM ('Abatacept in Inadequate responders to Methotrexate') study (IM101102)
- The ATTEST ('Abatacept or infliximab versus placebo, a Trial for Tolerability, Efficacy and Safety in Treating RA') study (IM101043)
- IM101-119.

The submission reviews study IM101-119 separately from the other three studies on the grounds that it used different outcome measures. However, for greater clarity, in this ERG report the four RCTs are discussed together.

The Kremer Phase 2b study was a 12-month randomised, double-blind, placebo-controlled, multicentre study designed to evaluate the safety, efficacy, and immunogenicity of two doses of abatacept (2mg/kg and 10 mg/kg) in adults who had active RA despite methotrexate therapy. Its primary outcome measure was the proportion of patients in each group with a 20% improvement in ACR response criteria at 6 months.[28]

The AIM study was a randomised, double-blind, placebo-controlled, multicentre phase 3 study designed to evaluate the safety and clinical efficacy of abatacept plus methotrexate in adults with persistent and active RA despite methotrexate therapy, and to assess the effects of abatacept on the radiographic progression of structural damage. Abatacept was given at a fixed dose by weight range which approximated to 10 mg/kg (ie 500, 750 or 1000mg for patients weighing <60kg, 60-100kg, and >100kg respectively). Participants were randomised to abatacept or placebo in a 2:1 ratio;[4] the manufacturer's submission states that the purpose of this unequal allocation was to increase safety

information relating to active treatment with abatacept.[6] The primary outcome measures were the proportion of patients in each group with a 20% improvement in ACR response criteria at 6 months, the proportion of patients with a clinically significant improvement in the HAQ-DI at 1 year, and the radiographic progression of joint erosions as measured by comparing GMS scores at baseline and 1 year.[4] The manufacturer clarified that the ACR20 response at 6 months and the HAQ-DI at one year were selected as primary objectives so that the former could be used to evaluate the short-term effect of abatacept on the signs and symptoms of RA and the latter to evaluate its long-term impact on functional disability.[2]

The ATTEST trial was a randomised, triple-blind, double-dummy, placebo- and active (infliximab)-controlled, multicentre phase 3 study designed to evaluate the efficacy and safety of abatacept or infliximab vs placebo in adults who had active RA and an inadequate response to methotrexate therapy. Patients, physicians, and outcome assessors were blinded to treatment group assignment for one year. Treatment with placebo was limited to days 1-197: on day 198, patients in the placebo group were reallocated to abatacept, while patients originally allocated to abatacept or infliximab continued that treatment, but blinding was maintained. The study's primary outcome measure was disease activity in the abatacept and placebo groups as measured by the DAS28 (ESR) at 6 months.[26]

IM101-119 was a multinational, multicentre, randomised, double-blind, placebo-controlled, phase 3b study, of 4 months duration, designed to evaluate the effect of abatacept on changes in wrist synovitis, osteitis, and bone erosion in patients with active RA and an inadequate response to methotrexate. The study's primary outcome measure was the mean change in wrist synovitis score, as measured using the OMERACT-RAMRIS method, at 4 months.[27]

For details of study design, see Table 4. The baseline characteristics of the patients in the included RCTs are presented in Table 9. These show that, in each study, the treatment arms were generally reasonably well balanced in relation to baseline patient and disease characteristics. However, as the manufacturer's submission notes, in the IM101-119 study there was an imbalance in terms of the proportion of participants in each group who tested positive for rheumatoid factor (RF) and anti-CCP2. Little information on baseline characteristics is available for this study, and it is possible that, because of its small size, there may have been other, unreported, imbalances.

IM101-119 differs from the other trials in that it recruited patients with a shorter history of RA and lesser disease severity. Thus, the mean time since first diagnosis was around 2.25 years, whereas in the other three studies it was between 8 to 10 years. Similarly, the mean numbers of swollen and tender joints per treatment group were 8-11 and approximately 13 respectively, whereas in the other three studies participants had a mean of 20-22 swollen joints and 28-32 tender joints at randomisation (for details, see Table 9).

All eligible patients in the Kremer Phase 2b, AIM, and ATTEST studies were stated to have received methotrexate, at a stable dose of 15mg or more per week, for at least 3 months prior to randomisation; this information was not provided for IM101-119 but the inclusion criteria for this study included active RA despite methotrexate therapy.[27] The manufacturer's submission states that the proportion of patients on a biological agent at the time of enrolment was higher in the Kremer Phase 2b study than in the AIM and ATTEST studies. Specific data are not provided for the ATTEST study, but in the other studies the overall numbers were low, between 2.6% and 5.7% in the Kremer Phase 2b study and 0-0.2% in the AIM study (see Table 9).

Although a majority of patients in all four trials tested positive for RF, the proportion of RF-positive patients was highest in the Kremer Phase 2b trial, at 90-99%, and lowest in IM101-119, at 56-83% (see Table 9).

The lower baseline physical function score seen in the Kremer Phase 2b study, when compared with the AIM and ATTEST studies, may perhaps be due to the use of the less sensitive MHAQ rather than the HAQ-DI to assess disability.

Table 9: Baseline characteristics of participants in the included RCTs (data from study publications supplemented where necessary from the manufacturer's submission)

	Kremer Phase 2b[28,3] (n=339)			AIM[4] (n=656)		ATTEST[26] (n=431)			IM101-119 (n=50)[6]	
	Placebo + MTX	Abatacept 2 mg/kg + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Infliximab 3mg/kg + MTX	Placebo + MTX	Abatacept ~10 mg/kg + MTX
Number randomised	119	105	115	219	433	110	156	165	23	27
% female	66%	63%	75%	81.7%	77.8%	87.3%	83.3%	82.4%	69.6%	59.3%
Age in years, mean [range]	54.7 [23-80]	54.4 [23-80]	55.8 [17-83]	50.4 (12.4)	51.5 (12.9)	49.4 (11.5)	49.0 (12.5)	49.1 (12.0)	52.5 (11.5)	51.7 (11.5)
Weight, kg [range]	79.9 [44-140]	78.7 [48-186]	77.8 [40-144]	70.2 (16.1)	72.3 (17.5)	NR	NR	NR		
Years since diagnosis, mean (SD)	8.9 (8.3)	9.7 (8.1)	9.7 (9.8)	8.9 (7.1)	8.5 (7.3)	8.4 (8.6)	7.9 (8.5)	7.3 (6.2)	28.2 months (17.0)	25.7 months (18.0)
No. of prior DMARDs, mean (SD)	NR	NR	NR	1.2 (0.58)	1.3 (0.56)	1.8 (0.91)	1.7 (0.77)	1.7 (0.82)	NR	NR
% patients who had received prior DMARDs	21.0%	18.1%	16.5%	8.7%	22.2%	55.5%	51.3%	52.7%	NR	NR
% patients on corticosteroids at study enrolment	67.2%	67.6%	60.0%	68.50%	72.1%	70.0%	75.6%	71.5%	NR	NR
% on biologics at study enrolment	2.60%	5.70%	2.60%	0%	0.2%	NR	NR	NR	NR	NR
% patients on NSAIDs at study enrolment	NR	NR	NR	82.6%	85.5%	84.5%	85.3%	86.1%	NR	NR
Mean MTX dose, mg/wk (SD)	15.8 (4.1)	15.8 (4.5)	15.0 (4.4)	15.7 (3.5)	16.1 (3.6)	16.6 (3.7)	16.5 (3.7)	16.3 (3.6)	NR	NR
Mean duration of MTX therapy, y (SD)	2.9 (3.5)	2.6 (3.0)	2.5 (2.7)	NR	NR	Months 23.7 (25.6)	Months 18.3 (20.0)	Months 23.6 (26.8)	NR	NR
% patients RF+	90%	90%	99%	78.5%	81.8%	77.3%	87.2%	84.8%	82.6%	55.6%
% patients anti-CCP2+									73.9%	48.1%
Tender joint count, mean (SD)	29.2 (13.0)	28.2 (12.0)	30.8 (12.2)	32.3 (13.6)	31.0 (13.2)	30.3 (11.7)	31.6 (13.9)	31.7 (14.5)	13.3 (7.2)	12.9 (7.1)
Swollen joint count, mean (SD)	21.8 (8.8)	20.2 (8.9)	21.3 (8.4)	22.1 (8.8)	21.4 (8.8)	20.1 (7.0)	21.3 (8.6)	20.3 (8.0)	8.5 (4.1)	11.3 (6.6)
Patient assessment of pain (100-mm VAS), mean (SD)	65.2 (22.1)	64.5 (22.3)	62.1 (21.4)	65.9 (20.6)	63.3 (21.1)	NR	NR	NR	NR	NR
Patient global assessment of disease activity (100-mm VAS) , mean (SD)	62.8 (21.6)	59.4 (23.7)	60.1 (20.7)	62.8 (21.6)	62.7 (21.2)	NR	NR	NR	NR	NR
Physician global assessment of disease activity (100-mm VAS) , mean (SD)	63.3 (15.5)	61.0 (16.7)	62.1 (14.8)	67.4 (17.0)	68.0 (16.0)	NR	NR	NR	NR	NR
Physical function score, mean (SD)	MHAQ 1.0 (0.6)	MHAQ 1.0 (0.5)	MHAQ 1.0 (0.5)	HAQ-DI 1.7 (0.6)	HAQ-DI 1.7 (0.7)	HAQ-DI 1.8 (0.7)	HAQ-DI 1.8 (0.6)	HAQ-DI 1.7 (0.7)	NR	NR
C-reactive protein level (mg/l) , mean (SD)	32 (32)	32 (26)	29 (28)	28 (25)	33 (31)	27 (26)	31 (27)	33 (32)	NR	NR
DAS28(CRP), mean (SD)	NR	NR	NR	NR	NR	NR	NR	NR	5.3 (0.9)	5.3 (1.1)
ESR (mm/h) , mean (SD)	NR	NR	NR	NR	NR	47.0 (32.6)	49.4 (31.2)	47.8 (30.4)	NR	NR

4.2.1 Summary of results

This section presents the main clinical efficacy and safety evidence from the four included RCTs. Data presented in the relevant publications have been supplemented where necessary by data from the manufacturer's submission[6] and supplementary information submitted by the manufacturer.[2] Where data presented in the manufacturer's submission differ from published data, and no reason can be identified, data from the submission have been prioritised on the basis that they may possibly be more recent and more accurate.

The manufacturer's submission did not present relative risks relating either to the individual studies or resulting from the meta-analysis of those studies for all the relevant comparisons. Consequently, for binary outcomes, relative risks for abatacept, at a dose of or approximating to 10 mg/kg, and infliximab have been recalculated by the ERG using a random effects model and Review Manager software.[66] For consistency, the relative risks calculated by the manufacturer for binary outcomes have been replaced by those calculated by the ERG throughout the report. However, because of time constraints, the ERG was not able to recalculate individual or pooled mean differences in continuous outcomes such as changes from baseline.

It should be noted throughout that the ATTEST study was not powered to detect a statistical difference between abatacept and infliximab.

Efficacy

Disease activity

DAS28 scores

Data on DAS28 scores at 6 months and 1 year were available for the AIM and ATTEST studies; IM101-119 presented data at 4 months (see Table 11). Although the manufacturer's submission states that DAS28 scores did not form a reported outcome in the Kremer Phase 2b study, the study investigators did in fact collect DAS28 data, and performed a post-hoc analysis to identify the proportion of patients having low disease activity (DAS28 <3.2) or experiencing remission (DAS28 <2.6) at 6 months and 1 year;[3] moreover, these data are reported in the manufacturer's submission.[6]

The published data from the AIM study[4] relating to the number of participants with DAS28 scores indicating low disease activity (DAS28 $\leq$3.2) or remission (DAS28 <2.6) differ considerably from those presented in the manufacturer's submission (see Table 10). The reason for these differences is not clear, and therefore the data presented in the manufacturer's submission, which are largely more conservative, have been included in Table 11 and used in the meta-analyses performed by the ERG.

Table 10: The AIM study: DAS28 scores – published data[4] and data from the manufacturer's submission[6]

Outcome	Placebo + MTX	Abatacept + MTX
DAS28 ≤3.2 at 6 months		
Published data	10.0%	30.1%
Manufacturer's submission	3.9%	22.4%
DAS28 ≤3.2 at 1 year		
Published data	9.9%	42.5%
Manufacturer's submission	3.8%	27.5%
DAS28 <2.6 at 6 months		
Published data	2.8%	14.8%
Manufacturer's submission	0.6%	9.6%
DAS28 <2.6 at 1 year		
Published data	1.9%	23.8%
Manufacturer's submission	2.2%	17.3%

In the Kremer Phase 2b, AIM and ATTEST studies, relative to placebo, abatacept at a dose of, or approximating to, 10 mg/kg was associated with significantly higher likelihoods of having low disease activity and of achieving remission at 6 months, as was infliximab in the ATTEST study. At 12 months, in Kremer Phase 2b and AIM studies, relative to placebo, abatacept 10 mg/kg was still associated with statistically significantly greater likelihoods of low disease activity or remission (for details, see Table 11). In the ATTEST study, no comparison with placebo was available at 12 months, and the study was not powered to compare abatacept and infliximab; thus, although the point estimates favour abatacept, statistical significance is achieved only in relation to the likelihood of achieving low disease activity.

As only percentages, and not numbers of patients, were presented for study IM101-119, relative risks were not calculated.

Table 11: DAS28 (data from study publications supplemented where necessary from the manufacturer's submission)

Trial	Kremer Phase 2b (n=339)			AIM (n=656)		ATTEST (n=431)			IM101-119 (n=48)	
	Placebo + MTX	Abatacept 2 mg/kg + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Infliximab 3mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX
Number randomised	119	105	115	219	433	110	156	165	23	25
DAS28 at 4 months, number included in analysis										
DAS28 (CRP) mean CFB [95% CI]									-0.55 [-0.95, -0.16]	-1.68 [-2.15, -1.21]
Mean difference vs. placebo [95%CI]										NR
Number with low disease activity (presumably DAS28 ≤3.2) [95% CI]									13.6% [0.0, 28.0]	50.0% [30.8, 69.2]
Number in remission (presumably DAS28 <2.6) [95% CI]									0% [0.0, 0.0]	15.4% [1.5, 29.3]
DAS28 at 6 months, number included in analysis	119	105	115	179	366	102	150	156		
DAS28 (ESR), mean CFB (SE)	NR	NR	NR	-1.33 (0.10)	-2.48 (0.07)	-1.48 (0.15)	-2.53 (0.12)	-2.25 (0.12)		
Mean difference vs. placebo [95%CI]		NR	NR		-1.15 [-1.38, -0.91] p<0.001		-1.04 [-1.42, -0.67] p<0.001	-0.77 [-1.14, -0.39] p<0.001		
Mean difference, abat vs. inflix, [95%CI]							-0.28 [-0.61, 0.06]			
Number with improvement (DAS28 change ≥1.2) (%)				91 (50.8%)	301 (82.2%)	53 (52%)	123 (82%)	113 (72.4%)		
Relative risk vs placebo [95% CI]					1.62 [1.39, 1.88]		1.58 [1.29, 1.93]	1.39 [1.13, 1.72]		
Relative risk vs infliximab [95% CI]							1.13 [1.00, 1.28]			
Number with low disease activity (DAS28 ≤3.2), (%)	23 (19.3%)	32 (30.5%)	46 (40%)	7 (3.9%)	82 (22.4%)	11 (10.8%)	31 (20.7%)	40 (25.6%)		
Relative risk vs placebo [95% CI]			2.07 [1.35, 3.18]		5.73 [2.70, 12.14]		1.92 [1.01, 3.64]	2.38 [1.28, 4.41]		
Relative risk vs infliximab [95% CI]							0.81 [0.53, 1.22]			
Number in remission (DAS28 <2.6), (%)	11 (9.2%)	19 (18%)	30 (26.1%)	1 (0.6%)	35 (9.6%)	3 (2.9%)	17 (11.3%)	20 (12.8%)		
Relative risk vs placebo [95% CI]			2.82 [1.49, 5.36]		17.12 [2.36, 123.94]		3.85 [1.16, 12.81]	4.36 [1.33, 14.29]		
Relative risk vs infliximab							0.88 [0.48, 1.62]			

[95% CI]										
DAS28 at one year, number included in analysis	Assumed 119	Assumed 105	Assumed 115	183	375	PLA-ABA 102	155	155		
DAS28 (ESR), mean CFB (SE)				-1.46 (0.10)	-2.85 (0.07)	-2.68 (0.15)	-2.88 (0.12)	-2.25 (0.12)		
Mean difference vs. placebo [95%CI]					-1.39 [-1.63, -1.16] p<0.001		NA	NA		
Mean difference, abat vs. inflix, [95%CI]							-0.62 [-0.96, -0.29]			
Number with improvement (DAS28 change ≥1.2), (%)				108 (59%)	328 (87.5%)	81 (79.4%)	129 (86%)	117 (75%)		
Relative risk vs placebo [95% CI]					1.48 [1.31, 1.68]		NA	NA		
Relative risk vs infliximab [95% CI]							1.10 [0.98, 1.24]			
Number with low disease activity (DAS28 ≤3.2), (%)	26 (21.9%)	30 (28.6%)	57 (49.6%)	7 (3.8%)	103 (27.5%)	30 (29.4%)	53 (35.3%)	35 (22.4%)		
Relative risk vs placebo [95% CI]			2.27 [1.54, 3.34]		7.18 [3.41, 15.12]		NA	NA		
Relative risk vs infliximab [95% CI]							1.51 [1.05, 2.18]			
Number in remission (DAS28 <2.6), (%)	12 (10.1%)	25 (24%)	40 (34.8%)	4 (2.2%)	65 (17.3%)	16 (15.7%)	28 (18.7%)	19 (12.2%)		
Relative risk vs placebo [95% CI]			3.45 [1.91, 6.23]		7.93 [2.93, 21.43]		NA	NA		
Relative risk vs infliximab [95% CI]							1.47 [0.86, 2.52]			

Meta-analyses undertaken by the ERG indicate that, at 6 months, compared with placebo, abatacept is associated with significantly increased likelihoods of an improved DAS28 score and of achieving DAS28-defined low disease activity and remission (see Table 12). These effects were sustained at 12 months (see Table 13).

Table 12: DAS28 results at 6 months, abatacept 10mg/kg vs placebo (calculations undertaken by the ERG)

Outcome	Relative risk (random effects model) (95% CI)	P value
Improved DAS28	1.60 (1.42, 1.81)	<0.0001
Low disease activity	2.70 (1.44, 5.08)	0.002
In remission	4.12 (1.71, 9.92)	0.002

Table 13: DAS28 results at 1 year, abatacept 10mg/kg vs placebo (calculations undertaken by the ERG)

Outcome	Relative risk (random effects model) (95% CI)	P value
Improved DAS28	1.48 (1.31, 1.68)	<0.00001
Low disease activity	3.89 (1.13, 13.40)	0.03
In remission	4.78 (2.06, 11.09)	0.003

The manufacturer's submission draws attention to the results of the patient-level analysis of the AIM trial by Dougados *et al*.[45] This demonstrates that some patients who did not achieve a clinically meaningful DAS28 response in the first 6 months of abatacept therapy nonetheless achieved low disease activity, as measured by the DAS28, at 1 year. This is consistent with the results tabulated in Table 10 which, while not linked with individual patients, show that, in all three studies which presented results at 6 months and 1 year, the number of patients with an improvement in DAS28 score was higher at 1 year than at 6 months, as were the numbers with DAS28 scores indicating low disease activity and remission.

ACR20, ACR50 and ACR70 responses

The three major RCTs presented data on ACR responses at 6 months and one year. The results reported for the AIM study differ slightly from those included in the Cochrane review of abatacept for rheumatoid arthritis (Analysis 1.1).[15]

In all three studies, at 6 months, both abatacept at a dose of, or approximating to, 10 mg/kg, and infliximab were associated with significantly higher likelihoods of achieving an ACR20, ACR50 or ACR70 response than was placebo. In the ATTEST study, there was no statistically significant difference between the abatacept and infliximab groups in this respect (for details, see Table 14).

At 12 months, in both the Kremer Phase 2b and AIM studies, the likelihood of achieving ACR20, ACR50, and ACR70 responses was still significantly higher in patients randomised to abatacept at a dose of, or approximating to, 10 mg/kg than in those randomised to placebo. In the ATTEST study, at 12 months the difference between abatacept and infliximab in ACR20 response, which favoured abatacept, achieved statistical significance, but there was no statistically significant difference between the groups in relation to ACR50 and ACR70 responses (for details, see Table 14).

Table 14: ACR20/50/70 responses at 6 months and one year (data from study publications supplemented where necessary from the manufacturer's submission[6] and additional data[2]; highlighted data CIC)

Trial	Kremer Phase 2b[28,3] (n=339)			AIM[4] (n=656)		ATTEST[26] (n=431)		
	Placebo + MTX	Abatacept 2 mg/kg + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Infliximab 3mg/kg + MTX
Number randomised	119	105	115	219	433	110	156	165
ACR responses at 6 months, number included in analysis	119	105	115	214	424	110	156	165
ACR20, number of responders (%)	42 (35.3%)	44 (41.9%)	[redacted]	87 (39.7%)	294 (67.9%)	46 (41.8%)	104 (66.7%)	98 (59.4%)
Relative risk vs placebo [95% CI]			1.72 [1.30, 2.29]		1.71 [1.43, 2.03]		1.59 [1.25, 2.04]	1.42 [1.10, 1.83]
Relative risk vs infliximab [95% CI]							1.12 [0.95, 1.33]	
ACR50, number of responders (%)	14 (11.8%)	24 (22.9%)	42 (36.5%)	37 (16.8%)	173 (39.9%)	22 (20.0%)	63 (40.4%)	61 (37.0%)
Relative risk vs placebo [95% CI]			3.10 [1.79, 5.37]		2.36 [1.72, 3.23]		2.02 [1.33, 3.07]	1.85 [1.21, 2.82]
Relative risk vs infliximab [95% CI]							1.09 [0.83, 1.44]	
ACR70, number of responders (%)	2 (1.7%)	11 (10.5%)	19 (16.5%)	14 (6.5%)	86 (19.8%)	10 (9.1%)	32 (20.5%)	40 (24.2%)
Relative risk vs placebo [95% CI]			9.83 [2.34, 41.26]		3.10 [1.81, 5.32]		2.26 [1.16, 4.40]	2.67 [1.39, 5.11]
Relative risk vs infliximab [95% CI]							0.85 [0.56, 1.28]	
ACR responses at one year, number included in analysis	119	105	115	214	424	109 PLA-ABA	156	164
ACR20, number of responders (%)	43 (36.1%)	[redacted]	72 (62.6%)	87 (39.7%)	317 (73.1%)	75 (68.3%)	113 (72.4%)	92 (55.8%)
Relative risk vs placebo [95% CI]			1.73 [1.31, 2.29]		1.84 [1.55, 2.18]		NA	NA
Relative risk vs infliximab [95% CI]							1.29 [1.09, 1.53]	
ACR50, number of responders (%)	24 (20.2%)	24 (22.9%)	48 (41.7%)	40 (18.2%)	209 (48.3%)	56 (50.9%)	71 (45.5%)	60 (36.4%)
Relative risk vs placebo [95% CI]			2.07 [1.36, 3.14]		2.64 [1.96, 3.54]		NA	NA
Relative risk vs infliximab [95% CI]							1.24 [0.95, 1.62]	
ACR70, number of responders (%)	9 (7.6%)	[redacted]	24 (20.9%)	13 (6.1%)	124 (28.8%)	32 (29.1%)	41 (26.3%)	34 (20.6%)
Relative risk vs placebo [95% CI]			2.76 [1.34, 5.68]		4.81 [2.79, 8.32]		NA	NA
Relative risk vs infliximab [95% CI]							1.27 [0.85, 1.89]	

The meta-analyses indicate that abatacept plus methotrexate is more effective than placebo plus methotrexate in relation to ACR response criteria at both 6 months and one year (see Tables 15 and 16).

Table 15: ACR20/50/70 responses at 6 months, abatacept 10mg/kg vs placebo

Outcome	Relative risk (random effects model)	95% CI	P value
ACR20	1.68	1.48, 1.91	<0.00001
ACR50	2.36	1.88, 2.97	<0.00001
ACR70	3.20	1.79, 5.73	<0.0001

Table 16: ACR20/50/70 responses at 1 year, abatacept 10mg/kg vs placebo

Outcome	Relative risk (random effects model)	95% CI	P value
ACR20	1.81	1.56, 2.09	<0.00001
ACR50	2.43	1.91, 3.10	<0.00001
ACR70	3.83	2.22, 6.61	<0.00001

Physical function

All three major studies assessed physical function using measures derived from the HAQ: the Kremer Phase 2b study used the MHAQ,[3] while the AIM[4] and ATTEST[26] studies used the more sensitive HAQ-DI. The threshold for defining a clinically meaningful improvement was set at 0.22 in the Kremer Phase 2b study, and at 0.3 in the AIM and ATTEST studies.

In all three studies, at 6 months, abatacept at a dose of, or approximating to, 10 mg/kg was associated with a statistically significantly greater reduction in mean HAQ score from baseline relative to placebo. Significant differences relative to placebo were also seen at 1 year in the Kremer Phase 2b and AIM studies. In the ATTEST study, at 6 months infliximab was also associated with a statistically significantly greater reduction in mean HAQ score from baseline relative to placebo; there was no significant difference between abatacept and infliximab at 1 year (for details, see Table 17).

In both the Kremer Phase 2b and AIM studies, at both 6 months and 1 year, the likelihood of achieving a clinically meaningful improvement in physical function was significantly higher in the abatacept group than in the placebo group. In the ATTEST study, significantly more patients in both the abatacept and infliximab groups than in the placebo group demonstrated a clinically meaningful improvement in physical function at 6 months; there was no significant difference between the two active interventions. At 1 year, responses were largely maintained in the abatacept and infliximab

groups, and again there was no significant difference between the two groups (for details, see Table 17).

Table 17: HAQ disability score: change from baseline and responders at 6 months and one year (data from study publications supplemented where necessary from the manufacturer's submission)

Trial	**Kremer Phase 2b (n=339)**			**AIM trial (n=656)**		**ATTEST (n=431)**		
	Placebo + MTX	Abatacept 2 mg/kg + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Infliximab 3mg/kg + MTX
Number randomised	119	105	115	219	433	110	156	165
HAQ disability score CFB at 6 months, number included in analysis	119	105	115	211	420	110	156	165
HAQ disability score mean CFB (SE)	-0.14	-0.17	-0.42	-0.40 (0.04)	-0.59 (0.03)	-0.31 (0.06)	-0.69 (0.05)	-0.61 (0.05)
Mean difference vs. placebo [95%CI]			-0.28 [-0.44, -0.12] p<0.05		-0.19 [-0.29, -0.10] p<0.001		-0.38 [-0.53, -0.23] p<0.001	-0.30 [-0.45, -0.15] p<0.001
Clinically meaningful HAQ response (>0.3), number of responders (%)	40 (33.6%) **response >0.22**	NR	67 (58.3%) **response >0.22**	97 (45.3%)	259 (61.1%)	45 (40.9%)	96 (61.5%)	97 (58.8%)
Relative risk vs placebo [95% CI]			1.73 [1.29, 2.33]		1.34 [1.14, 1.58]		1.50 [1.16, 1.94]	1.44 [1.11, 1.86]
Relative risk vs infliximab [95% CI]							1.05 [0.88, 1.25]	
HAQ disability score at one year, number included in analysis	119	105	115	212	422	110 (PLA-ABA)	156	165
HAQ disability score mean CFB (SE)	-0.10	-0.25	-0.47	-0.37 (0.04)	-0.66 (0.03)	-0.56 (0.06)	-0.67 (0.05)	-0.59 (0.05)
Mean difference vs. placebo [95%CI]		NR p<0.087	-0.36 [-0.52, -0.21] p<0.001		-0.29 [-0.38, -0.19] p<0.001		NA	NA
Mean difference, abatacept vs. infliximab [95%CI]							-0.08 [-0.22, 0.06]	
Clinically meaningful HAQ response (>0.3), number of responders (%)	33 (27.7%) **response >0.22**		57 (49.6%) **response >0.22**	84 (39.3%)	270 (63.7%)	63 (57.3%)	90 (57.7%)	87 (52.7%)
Relative risk vs placebo [95% CI]			1.79 [1.27, 2.52]		1.61 [1.35, 1.94]		NA	NA
Relative risk vs infliximab [95% CI]							1.09 [0.90, 1.33]	

Meta-analyses undertaken by the ERG indicate that, at both 6 months and 1 year, relative to placebo, abatacept at a dose of, or approximating to, 10mg/kg is associated with a significantly increased likelihood of achieving a clinically meaningful HAQ response (see Table 18).

Table 18: Clinically meaningful HAQ response, abatacept 10mg/kg vs placebo

	Relative risk (random effects model)	95% CI	p value
6 months	1.46	1.27, 1.67	<0.00001
1 year	1.65	1.41, 1.94	<0.00001

In the manufacturer's meta-analyses, the point estimates suggest that abatacept is associated with mean reductions in the HAQ score at 6 months and 1 year which are just clinically meaningful at the threshold of 0.22 proposed by Bruce *et al*.,[29] rather than the threshold of 0.3 used in the AIM and ATTEST studies (see Table 19). However, the confidence intervals are such that it is impossible to exclude the possibility that, at 6 months, the reduction may not be clinically meaningful, while at 12 months it may be barely so.

Table 19: Reduction in HAQ score, abatacept 10mg/kg vs placebo (data from manufacturer's submission[6]

	Mean change from baseline	95% CI	p value
6 months	-0.2524	-0.3253, -0.1794	NR
1 year	-0.3105	-0.3934, -0.2275	NR

Joint damage

Only two studies, AIM and IM101-119, reported data relating to joint damage.

In the AIM study, radiographic data were available for 391/433 patients (90%) in the abatacept group and 195/219 (89%) in the placebo group.[6] The study authors claimed that, at one year, abatacept was associated with an approximately 50% reduction in change from baseline GMS scores relative to placebo;[4] this claim seems to be based on the median total score in each group (see Table 20). Although the differences in median scores were statistically significant for all three reported outcomes, the Cochrane reviewers questioned their clinical significance.[15]

Table 20: Joint damage: change from baseline at 1 year: data from the AIM study[4]

Outcome	Mean (no variance reported)			Median (25th and 75th percentiles)		
	Placebo	Abatacept	p value	Placebo	Abatacept	p value
Erosion score	1.14	0.63	NR	0.27 (0.0, 1.3)	0.0 (0.0, 1.0)	0.029
Joint-space narrowing score	1.18	0.58	NR	0.0 (0.0, 1.0)	0.0 (0.0, 0.5)	0.009
Total score	2.32	1.21	NR	0.53 (0.0, 2.5)	0.25 (0.0, 1.8)	0.012

In study IM101-119, there was no significant difference between groups in wrist synovitis score at 4 months, the study's primary outcome. Abatacept was associated with a lower mean increase in erosion score than placebo, and with a mean reduction from baseline in both wrist and hand oedema/osteitis score and total RAMRIS score compared with a mean increase from baseline in the placebo group; however, none of these differences was said to be statistically significant (see Table 21, which re-presents the data which were erroneously divided between Tables B 16 and B 17 in the manufacturer's submission). Moreover, in relation to bone erosion, oedema/osteitis, and synovitis, abatacept was not associated with a statistically significant reduction in the relative risk of having one or more newly involved joints, although the point estimates were favourable.

Table 21: Study IM101-119: summary of clinical efficacy results at 4 months (day 113) (data from manufacturer's submission, relative risks calculated by the ERG)

Outcome measure	**Abatacept (N=25)**	**Placebo (N=23)**	**Adjusted mean difference (95% CI)**	**p value**
Synovitis score (non-parametric model): mean change (SD) from baseline to day 113	-0.44 (1.47)	0.52 (1.38)	NR	0.103
Erosion score: adjusted mean change (SE) from baseline to day 113	0.45 (0.43)	0.95 (0.45)	-0.50 (-1.77, 0.76)	NR
Oedema/osteitis score: adjusted mean change (SE) from baseline to day 113	-1.94 (0.86)	1.54 (0.90)	-3.48 (-6.00, -0.96)	NR
RAMRIS score: adjusted mean change (SE) from baseline to day 113	-1.82 (1.13)	2.89 (1.18)	-4.71 (-8.00, -1.42)	NR
Number of newly involved joints			**Relative risk (95% CI)**	
Bone erosion				
0	20/25 (80%)	16/23 (69.6%)		
≥1	5/25 (20%)	7/23 (30.4%)	0.66 (0.24, 1.78)	0.41
Oedema/osteitis				
0	18/25 (72%)	16/23 (69.6%)		
≥1	7/25 (28%)	7/23 (30.4%)	0.92 (0.38, 2.22)	0.85
Synovitis				
0	23/25 (92%)	20/23 (87%)		
≥1	2/25 (8%)	3/23 (13%)	0.61 (0.11, 3.35)	0.57

Pain

The manufacturer's submission did not report data relating to pain; the reason provided in the clarification letter was that such data were not available in a suitable format for presentation within the submission.[2] However, both the Kremer Phase 2b study[3] and the AIM study[4] reported the mean change from baseline in the pain component of the ACR criteria; for the AIM study, these data were only available in an online appendix to the published paper (see http://www.annals.org/content/144/12/865.full.pdf+html).

In the Kremer Phase 2b and AIM studies, both abatacept and placebo were associated with reductions in pain at 6 months and 1 year. However, abatacept, at a dose of, or approximating to, 10 mg/kg, was associated with a significantly greater reduction in pain than placebo (see Table 22).

Table 22: Patient-reported pain (0-100mm VAS)

Trial	**Kremer Phase 2b[28,3] (n=339)**			**AIM trial (n=656)[4]**	
	Placebo + MTX	Abatacept 2 mg/kg + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX
Number randomised	119	105	115	219	433
Pain					
Mean CFB at 6 months	-8.4	-22.7	-46.4	NR	NR
Difference vs. placebo		-14.3 p<0.05	-38.0 p<0.05		
Mean CFB at 1 year (SE)	-12.6	-26.2	-44.9	-23.2 (1.81)	-35.8 (1.17)
Difference vs. placebo [95% CI]		-13.6 p=0.071	-32.3 p<0.001		-12.6 [-16.9, -8.39]

Morning stiffness, sleep quality, and fatigue

Only one study, the AIM study, is known to have collected data relating to morning stiffness, sleep quality, and fatigue. This study used the MOS Sleep measure to assess the impact of abatacept on sleep quality in patients with RA.[40] Data relating to these three outcomes were not reported in the main study publication,[4] but the data relating to sleep quality were reported in depth in a paper by Wells *et al*.,[40] and data relating to all three outcomes were summarised in the manufacturer's submission[6] (for details, see Table 23).

Although improvements in all three outcomes were seen in both the abatacept and placebo groups at 6 months and one year, at both time points abatacept was associated with greater reductions in morning stiffness, fatigue, and an additional parameter described as the SPI score; with the exception of SPI at 6 months, these results appear to be statistically significant (see Table 23). However, the MOS Sleep measure may be used to calculate two sleep problem indices: Sleep Problems Index I and Sleep Problems Index II.[40] It is not clear which of these is reported in the manufacturer's submission, and the data do not appear to correspond to those published by Wells *et al*.,. in relation to either Sleep Problems Index.

Table 23: Morning stiffness, sleep quality, and fatigue: data from the AIM trial reported in the manufacturer's submission[6]

Parameter	Placebo + MTX	Abatacept 10 mg/kg + MTX
Number of randomised patients	219	433
Morning stiffness at 6 months, number included in analysis	176	393
Mean baseline (SD)	84.09 (59.98)	97.48 (61.09)
Mean CFB (SE)	-45.4 (3.29)	-71.7 (2.20)
Mean difference vs. placebo [95%CI]		-26.3 [-34.1, -18.5]
Morning stiffness at one year, number included in analysis	161	382
Mean baseline (SD)	83.45 (59.10)	97.34 (61.36)
Mean CFB (SE)	-55.4 (3.11)	-74.3 (2.01)
Mean difference vs. placebo [95%CI]		-18.9 [-26.2, -11.6]
Sleep quality at 6 months (SPI score), number included in analysis	211	420
Mean baseline (SD)	43.95 (19.06)	43.04 (20.42)
Mean CFB (SE)	-7.80 (1.03)	-10.2 (0.73)
Mean difference vs. placebo [95%CI]		-2.39 [-4.88, 0.09]
Sleep quality at one year (SPI score), number included in analysis	212	423
Mean baseline (SD)	44.05 (19.07)	43.11 (20.51)
Mean CFB (SE)	-6.75 (1.01)	-10.4 (0.72)
Mean difference vs. placebo [95%CI]		-3.60 [-6.04, -1.17]
Reduction of fatigue (VAS) at 6 months, number included in analysis	211	420
Mean baseline (SD)	65.92 (22.81)	63.42 (23.08)
Mean CFB (SE)	-17.2 (1.75)	-25.3 (1.24)
Mean difference vs. placebo [95%CI]		-8.13 [-12.3, -3.91]
Reduction of fatigue (VAS) at one year, number included in analysis	212	423
Mean baseline (SD)	65.87 (22.77)	63.38 (23.06)
Mean CFB (SE)	-16.4 (1.74)	-26.5 (1.23)
Mean difference vs. placebo [95%CI]		-10.1 [-14.3, -5.91]

Health-related quality of life (HRQoL)

The presentation of data relating to HRQoL in the manufacturer's submission is erratic. The submission includes a graph which shows the 6-month results from the Kremer Phase 2b study for all subscales and components of the SF-36 (Figure B 19), but the tabulation of results (Table B 22) does not include the figures underlying that graph. Furthermore, the submission states, curiously, that the Kremer Phase 2b study provided little insight into the HRQoL of the enrolled patients; it does not refer to the detailed one-year HRQoL results from this study which have been published by Emery *et al*.[5] The submission includes additional graphs displaying the results for all subscales and components of the SF-36 at 6 months and one year from the AIM (Figures B 18 and B 21) and ATTEST studies (Figures B 20 and 22), and presents in Table B 22 the numeric results for the physical and mental components of the SF-36 in those studies at 6 months and 1 year.

The available data relating to aspects of HRQoL, as measured by the SF-36 at 6 months and one year, are presented in Table 24. The data relating to the AIM study are taken from the manufacturer's submission; they differ from those presented by Kremer *et al*.,[4] and the reason for this discrepancy is

not clear. Although Russell *et al*.,[36] specifically present data relating to HRQoL from the AIM study, these data are not presented in a format which is compatible with the data from the Kremer Phase 2b and ATTEST studies; they are therefore not utilised in Table 24.

In both the Kremer Phase 2b study and the AIM study, abatacept at a dose of, or approximating to, 10 mg/kg was associated with statistically significant improvements from baseline relative to placebo in the physical and mental components of the SF-36 at both 6 months and one year. In the ATTEST study, at 6 months both abatacept and infliximab were associated with significant improvements in the physical and mental components of the SF-36 relative to placebo. The point estimates suggest that, at 6 months and 1 year, abatacept was associated with greater improvements from baseline than infliximab in both components, but statistical significance was only achieved in relation to the physical component at one year (for details see Table 24).

Table 24: SF-36 physical functioning and mental component at 6 months and 1 year (data from study publications supplemented where necessary from the manufacturer's submission and supplementary data[2], highlighted data CIC)

Trial	Kremer Phase 2b (n=339)			AIM (n=656)		ATTEST (n=431)		
	Placebo + MTX	Abatacept 2 mg/kg + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Placebo + MTX	Abatacept 10 mg/kg + MTX	Infliximab 3mg/kg + MTX
Number randomised	119	105	115	219	433	110	156	165
SF-36 at 6 months (physical component), number included in analysis	119	104	115	207	416	109	154	163
SF-36 PC mean CFB (SE)	[redacted]	[redacted]	[redacted]	4.77 (0.59)	8.82 (0.42)	4.34 (0.82)	8.36 (0.69)	7.66 (0.67)
Mean difference vs. placebo [95%CI]		[redacted]	[redacted]		4.06 [2.64, 5.47] p<0.001		4.02 [1.92, 6.12] p<0.001	3.32 [1.25, 5.40] p=0.002
Mean difference, abatacept vs. infliximab [95%CI]							0.70 [-1.19, 2.58]	
SF-36 at one year (physical component), number included in analysis	NR	NR	NR	207	417	109 (PLA-ABA)	154	163
SF-36 PC mean CFB mean (SE)	2.6 (0.7)	5.2 (0.8)	8.0 (0.8)	4.97 (0.61)	9.12 (0.43)	8 (0.83)	9.52 (0.70)	7.59 (0.68)
Mean difference vs. placebo [95%CI]		2.6 (p=0.05)	5.4 (p<0.001)		4.15 [2.69, 5.62] p<0.001			
Mean difference, abatacept vs. infliximab [95%CI]							1.93 [0.02, 3.84]	
SF-36 at 6 months (mental component), number included in analysis	119	104	115	207*	416	109	154	163
SF-36 MC CFB, mean (SE)	[redacted]	[redacted]	[redacted]	3.83 (0.70)	6.22 (0.49)	1.64 (0.93)	5.14 (0.79)	4.32 (0.76)
Mean difference vs. placebo [95%CI]		[redacted]	[redacted]		2.39 [0.70, 4.07] p=0.005		3.51 [1.10, 5.91] p=0.004	2.68 [0.31, 5.05] p=0.027
Mean difference, abatacept vs. infliximab [95%CI]							0.83 [-1.33, 2.98]	
SF-36 at one year (mental component), number included in analysis	NR	NR	NR	207	417	109 (PLA-ABA)	154	163
SF-36 MC CFB, mean (SE)	2.8 (0.9)	3.5 (1.0)	5.7 (0.9)	4.73 (0.69)	6.86 (0.48)	5.85 (0.97)	5.96 (0.81)	4.03 (0.79)
Mean difference vs. placebo [95%CI]		0.7 (NS)	2.9 (p=0.05)		2.13 [0.48, 3.78] p=0.011			
Mean difference, abatacept vs. infliximab [95%CI]							1.92 [-0.30, 4.15]	

* Adjusted difference based on ANCOVA model with treatment and baseline values as covariates[2]

Critique of reported efficacy data

The manufacturer's submission appears to be complete in that, with the exception of the Japanese and Korean studies mentioned in sections 4.1.3-4.1.4, no apparently relevant clinical studies are known to have been omitted. The Japanese and Korean studies appear to have been excluded because of the nature of their populations. A valid case could probably have been made for their exclusion, but no attempt was made to do so.

However, as indicated throughout the summary of reported efficacy results, the presentation of the relevant data relating to the included studies is generally poor, displaying a number of omissions and inconsistencies in relation to the published data. Thus, the manufacturer's submission failed to refer to some publications relating to the included studies, and did not present all the relevant data which were available in the public domain, at times suggesting that those data did not exist. Moreover, in relation to claims made at various points that the data were not available in the appropriate format, it should be noted that Bristol-Myers Squibb Pharmaceuticals Ltd sponsored all the included studies, and thus should have had access to the study data and could presumably have presented them in an appropriate manner.

Furthermore, the submission states that no studies selected from the systematic review have been excluded from the meta-analysis, despite the fact that study IM101-119, which was identified by the systematic review, was excluded from the meta-analyses. However, the decision to do so, although not appropriately documented, is appropriate, since differs from the other three studies in terms of its population, duration, and many of its outcome measures.

As noted in section 4.1.5, the Cochrane reviewers considered two of the three main studies (Kremer Phase 2b and AIM) to be at high risk of bias. In the Kremer Phase 2b study, the imbalance between treatment groups in discontinuations, combined with concerns about the treatment of missing data, suggest a substantial potential for bias. The AIM study also showed imbalance between treatment groups in discontinuations and, as in the Kremer Phase 2b study, many of these discontinuations were due to lack of efficacy in the placebo arm. However, although the method of analysis used in the AIM study was such as to introduce a risk of bias, fewer than 1% of randomised participants were excluded for nonadherence, suggesting that any consequent bias may not be substantial. Moreover, the fact that, between 6 and 12 months, more patients in the placebo group than in the abatacept group received additional DMARDs is likely to reduce rather than exaggerate the treatment effect at 1 year. Similarly, in the ATTEST study, the fact that, between days 198-365, more patients randomised to infliximab than to abatacept received either an additional DMARD or an increased dose of methotrexate/corticosteroids is likely to disadvantage abatacept relative to infliximab. The details available for the IM101-119 study were so limited as to severely restrict any attempt at quality

assessment, but such information as was available showed an imbalance between treatment groups, with a higher proportion of patients in the placebo group testing positive for RF and anti-CCP2. Although blinding was very important given the largely subjective nature of the outcome measures, none of the studies commented on its success.

The ATTEST study was not powered to compare abatacept with infliximab. However, its results suggest that the two drugs have similar efficacy relative to placebo in relation to all the reported efficacy outcomes.

The reported efficacy data indicate that, over a period of six months to a year, relative to placebo, abatacept, at a dose of, or approximating to, 10 mg/kg, is associated with reduced disease activity, as measured by the DAS28 and ACR response criteria, and with improved physical function, less joint damage, improvements in pain, morning stiffness, sleep quality, and fatigue and improvement in health-related quality of life as measured by the physical and mental summary components of the SF-36. However, although some of the differences (for example, joint damage) are statistically significant, their clinical significance may be less clear. Moreover, the data relate only to a maximum period of 12 months whereas, given the nature of the disease, patients with RA require long-term treatment.

The manufacturer's submission notes on page 221 that the mean baseline DAS28 score in the AIM and ATTEST studies was higher, at 6.4 or above, than the threshold of 5.1 specified in NICE guidance for the initiation of biological therapy, while the mean in the Kremer Phase 2b trial is 5.5. However, although the study results in relation to DAS28 scores differ, the outlier is not the Kremer Phase 2b study, with its lower mean baseline DAS28 score, but the AIM study, in which abatacept is associated with relative risks of low disease activity and remission which are substantially higher than those seen in the Kremer Phase 2b and ATTEST studies (see Table 11). Nonetheless, the results obtained in the included studies may be more favourable to abatacept (and indeed infliximab) than the results which might be obtained in UK clinical practice because, as noted in section 3.1, the populations of the included studies had a shorter duration of RA, and had previously taken fewer conventional DMARDs, than is current standard UK clinical practice before the initiation of biological therapy.

The estimates of abatacept's treatment effects may not be unbiased, for two main reasons. The first, which affects its relative efficacy, relates to the differential discontinuation rates in patients randomised to placebo and active treatment, and the methods used to deal with incomplete data and nonadherence to study therapy. The second, which is more likely to affect the absolute treatment effects, relates to the difference between the study populations and the population likely to receive

abatacept in normal clinical practice in England and Wales. Thus, though the submitted evidence largely reflects the decision problem defined in the final scope, the difference between the two populations is such that the benefit associated with the use of abatacept may be less in UK clinical practice than is seen in the study populations.

Safety and tolerability

The safety and tolerability of a drug may be assessed in a number of ways. In randomised trials, the safety of the study medications is measured by the number of adverse events in each treatment group, while some indication of their tolerability may be gained by studying non-compliance with, or discontinuation from, the study medications. However, RCTs are seldom powered to detect significant differences between treatments with respect to the incidence of specific adverse events, and none of the RCTs included in this submission were so powered. Moreover, even if RCTs have sufficient power to detect significant differences in common adverse events, they cannot form the best source of evidence for rarer adverse events, or those which happen in the intermediate or longer term.

Key adverse event data from the included studies are summarised in Table 25; fuller data may be found in Appendix 1. As may be seen, abatacept at a dose of, or approximating to, 10 mg/kg was not associated with a higher rate of serious AEs than placebo at 6 months; as 12 months, there does not appear to be a significant differenced between abatacept and placebo in this respect. Results from the ATTEST study indicate that, in the short term, abatacept is associated with lower rates of SAEs, lower discontinuation rates due to AEs/SAEs, and lower rates of serious infections and acute infusional events than infliximab.

In the Kremer Phase 2b study, only one serious infection was reported in the first 6 months: this was cellulitis of the left foot requiring hospitalisation, occurring in a patient receiving abatacept 2 mg/kg.[28] Two serious infections were reported at one year: an infected hip joint prosthesis and an opportunistic pulmonary infection (ie an infection caused by a pathogen which does not usually cause infection in a person with a healthy immune system). Both patients were in the placebo group, and both withdrew from the study as a result of those infections. One death was reported in the 2mg abatacept group; this was due to multisystem organ failure, and was not considered to be related to the study drug.[3] The investigators did not identify any pattern in the types of serious AEs which were reported, nor did there appear to be a relationship between serious AEs and the number of infusions of study medication received. The malignancies reported in patients receiving 10 mg/kg abatacept were considered by the investigators to be unrelated to the study drug.[3] No serious AEs were considered to be related to abatacept at a dose of 10mg/kg, although four were reported in the 2 mg/kg group (for details, see Table 25). In relation to total AEs, one-year data were reported only for those AEs reported by at least 5% of patients: these were headache, nasopharyngitis, nausea, cough, and

arthralgia[3] (for details, see Appendix 1). Acute infusional AEs and peri-infusional AEs were not reported.

In the AIM study, the most frequently reported serious AEs were musculoskeletal and connective tissue disorders; these were primarily related to hospitalisation for RA flares or elective surgery for RA.[4] The incidence of common AEs was generally similar in both groups, although headache and nasopharyngitis were more common with abatacept than with placebo. The incidence of infection reported as a serious AE was higher with abatacept than with placebo. However, although both reported deaths were due to pulmonary infections, the patient in the abatacept group who died of bronchopulmonary aspergillosis had a history of TB, asbestos exposure, and pulmonary fibrosis;[4] a patient in the placebo group died of severe bronchopneumonia which was deemed unrelated to the study drug.[6] The incidence of neoplasms of all kinds was similar in both groups (see Appendix 1). Infusion reactions were more common with abatacept than with placebo. Two patients (implicitly in the abatacept group) discontinued treatment because of severe acute infusion reactions: one experienced hypersensitivity (rash and chest pain) after the second infusion, while the second experienced severe hypotension during the fourth infusion, and both AEs resolved shortly after cessation of infusions. Although peri-infusional adverse events were relatively common, affecting almost a quarter of patients in the abatacept group (see Table 25), serious peri-infusional adverse events were said to be rare.[4]

The ATTEST trial reports safety data for all three groups at six months, and for the abatacept and infliximab groups at one year; data relating to AEs other than deaths are not presented for the placebo/abatacept group between days 197-365. During the first six months, the incidence of all adverse events was similar in all three groups, but serious AEs were less common in the abatacept group than in either the placebo or the infliximab groups; the incidence of serious AEs considered to be related to study medication was highest in the infliximab group, due largely to a higher incidence of serious infections (see Table 25). At one year, abatacept was associated with a lower incidence of AEs, serious AEs, related AEs and SAEs, discontinuations due to AEs and SAEs, and acute infusional AEs than was infliximab. Two deaths occurred in the first 6 months: one, in the abatacept group, from a cerebrovascular accident, and one, in the infliximab group, due to fibrosarcoma. Between days 197-365, a patient in the infliximab group with peritoneal TB died of septic shock following surgery, while a patient from the placebo/abatacept group died of pneumonia and sepsis;[26] this last death was considered possibly related to the study medication.[6] Five serious opportunistic infections were reported; all were in the infliximab group.[26]

In IM101-119, two SAEs were reported: both were in the placebo group (atrial fibrillation and overdose), and neither was considered to be related to the study drug (the nature of the overdose is not specified, but presumably does not relate to the study medication). No serious infections,

malignancies, auto-immune events or discontinuation related to an AE or SAE was reported in either group, and most AEs were said to be mild in intensity.[6]

The recent Cochrane review of abatacept in RA found that, while the incidence of total adverse events was significantly higher in patients treated with abatacept than in those treated with placebo, the relative risk was very low (Analysis 1.28, RR 1.05, 95% CI 1.01 to 1.08). The only statistically significant difference between treatment groups in terms of serious adverse events was the greater number of serious infections at 12 months in patients treated with abatacept (Analysis 1.26, Peto odds ratio 1.91, 95% CI 1.07 to 3.42).[15] However, no data relating to this outcome are available for the studies included in this review since it was not an outcome which was reported by the Kremer Phase 2b and Aim studies, while the ATTEST study did not have a placebo arm after 6 months.

Table 25: Numbers of patients suffering adverse events (data from relevant publications supplemented where necessary by the manufacturer's submission and supplementary data, highlighted data CIC)

Study	Kremer Phase 2b[28,3,2]			AIM[4]		ATTEST[26]			IM101-119[6]	
	Placebo	Ab 2mg	Ab 10mg	Placebo	Ab 10mg	Placebo	Ab 10mg	Inflix	Placebo	Ab
Number randomised	119	105	115	656		110	156	165	23	27
Adverse events at 6 months (4 months for study IM101-119) (n)	119	105	115	219	433	110	156	165	23	27
Total AEs	[redacted]	[redacted]	[redacted]	84%	87.3%	92 (83.6%)	129 (82.7%)	140 (84.8%)	14 (60.9%)	20 (74.1%)
Total AEs considered to be related to the study drug	[redacted]	[redacted]	[redacted]	47.5%	49.4%	46 (41.8%)	64 (41.0%)	74 (44.8%)	6 (26.1%)	8 (29.6%)
Deaths	0	0	0	NR	NR	0	1 (0.6%)	1 (0.6%)	0	0
Serious AEs	12 (10.1%)	[redacted]	[redacted]	NR	NR	13 (11.8%)	8 (5.1%)	19 (11.5%)	2 (8.7%)	0
Serious infections	0	1 (1.0%)	0	NR	NR	3 (2.7%)	2 (1.3%)	7 (4.2%)		
Malignant neoplasms				NR	NR	1 (0.9%)	1 (0.6%)	2 (1.2%)		
Serious AEs considered to be related to the study drug	1 (0.8%)	4 (3.8%)	0	NR	NR	3 (2.7%)	3 (1.9%)	8 (4.8%)	0	0
Discontinuation due to AE	[redacted]	[redacted]	[redacted]	NR	NR	1 (0.9%)	3 (1.9%)	8 (4.8%)	0	0
Discontinuation due to serious AE	1 (0.8%)	4 (3.8%)	0	NR	NR	0	2 (1.3%)	4 (2.4%)	0	0
Acute infusional AEs	NR	NR	NR	NR	NR	11 (10.0%)	8 (5.1%)	30 (18.2%)	NR	NR
Adverse events at 1 year (n)	119	105	115	219	433	NR	156	165		
Total AEs	[redacted]	[redacted]	[redacted]	184 (84.0%)	378 (87.3%)	NR	139 (89.1%)	154 (93.3%)		
Total AEs considered related to study drug	[redacted]	[redacted]	[redacted]	104 (47.5%)	214 (49.4%)	NR	72 (46.2%)	96 (58.2%)		
Deaths	0	1	0	1 (0.5%)	1 (0.2%)	1	1 (0.6%)	2 (1.2%)		
Serious AEs	19 (16.0%)	19 (18.1%)	14 (12.2%)	26 (11.9%)	65 (15.0%)	NR	15 (9.6%)	30 (18.2%)		
Infections	NR	NR	NR	5 (2.3%)	17 (3.9%)	NR	NR	NR		
Serious infections	NR	NR	NR	NR	NR	NR	3 (1.9%)	14 (8.5%)		
Malignant neoplasms	3[1]	NR	4[2]	NR	NR	NR	1 (0.6%)	2 (1.2%)		
Serious AEs considered to be related to the study drug	2 (1.7%)	5 (4.8%)	2 (1.7%)	1 (0.5%)	15 (3.5%)	NR	5 (3.2%)	14 (8.5%)		
Acute infusional AEs (ie AEs within 1 hour of start of infusion*)	NR	NR	NR	9 (4.1%)	38 (8.8%)	NR	11 (7.1%)	41 (24.8%)		
Peri-infusional AEs (ie AEs within 24 hours of the start of the infusion)	NR	NR	NR	37 (16.9%)	106 (24.5%)	NR	NR	NR		
Discontinuation due to AEs	[redacted]	[redacted]	[redacted]	4 (1.8%)	18 (4.2%)	NR	5 (3.2%)	12 (7.3%)		
Discontinuation due to serious AEs	[redacted]	[redacted]	[redacted]	3 (1.4%)	10 (2.3%)	NR	4 (2.6%)	6 (3.6%)		

* 3 hours in the ATTEST study

1: 1 patient with endometrial cancer, 1 with squamous cell carcinoma, 1 with malignant melanoma

2: 1 patient with bladder cancer, 2 with basal cell carcinoma, 1 unspecified neoplasm

Compliance with treatment

The manufacturer's submission provides data relating to compliance with treatment, as measured by the number of missed infusions, for the AIM and ATTEST trials (see Table 26). The submission states that, in both studies, the number of missed infusions was low, with a median of 0.2 infusions missed;[6] subsequent clarification has been obtained that the median number of missed infusions was 0, and the mean 0.2.[2] Despite this low figure, between 11% and 16% of patients missed at least one infusion (see Table 26); there was no significant difference in this respect between abatacept and infliximab, or between either active intervention and placebo (see Table 27). It should be noted that patients who discontinued from the studies were not treated as noncompliant.

Table 26: Compliance with treatment during the double-blind period of the abatacept trials, as measured by the number of missed infusions (data from manufacturer's submission)

Missed infusions	**AIM (n=656)**		**ATTEST (n=431)**		
	Placebo + MTX	Abatacept 10 mg/kg + MTX	Placebo + ABA	Abatacept 10 mg/kg + MTX	Infliximab 3mg/kg + MTX
Number of randomised patients	219	433	110	156	165
Mean number of missed infusions (SD)	NR	NR	0.2 (0.44)	0.2 (0.45)	0.2 (0.39)
Median number of missed infusions (range)	NR	NR	0.0 (0.0-2.0)	0.0 (0.0-2.0)	0.0 (0.0-2.0)
Number of subjects who missed 0 infusions (%)	195 (89%)	367 (84.8%)	95 (86.4%)	132 (84.6%)	138 (83.6%)
Number of subjects who missed 1 infusion (%)	20 (9.1%)	61 (14.1%)	12 (10.9%)	20 (12.8%)	26 (15.8%)
Number of subjects who missed 2 infusions (%)	4 (1.8%)	5 (1.2%)	3 (2.7%)	4 (2.6%)	1 (0.6%)
Number of subjects who missed 1 or more infusions	24 (11.0%)	66 (15.2%)	15 (13.6%)	24 (15.4%)	27 (16.4%)
Relative risk of missing ≥1 infusion, vs placebo [95% CI]		1.39 [0.90, 2.16]		1.13 [0.62, 2.05]	1.20 [0.67, 2.15]
Relative risk of missing ≥1 infusion, vs infliximab [95% CI]				0.94 [0.57, 1.56]	

Table 27: Relative risk of missing at least one infusion

Comparison	**Relative risk (random effects model)**	**95% CI**	**p value**
Abatacept vs placebo	1.39	0.90, 2.16	0.14
Infliximab vs placebo	1.20	0.67, 2.15	0.54
Abatacept vs infliximab	0.94	0.57, 1.56	0.81

Critique of reported safety data

It is difficult to compare or combine safety data from the included studies since they do not all report comparable data. Thus, study IM101-199 only reported data at 4 months; the Kremer Phase 2b study presented safety data at both 6 and 12 months, but the AIM study reported very few safety data at 6 months and the ATTEST study, by nature of its design, could not compare active interventions with

placebo at 12 months. Particularly unfortunate omissions in any study of safety and tolerability are the failure to report peri-infusional AEs at 6 months in any study, and at 12 months in the Kremer Phase 2b and ATTEST studies, and to report acute infusional AEs at 6 months in the Kremer Phase 2b and AIM studies, and at 12 months in the Kremer Phase 2b study (see Table 25).

Although adverse reactions to abatacept include allergic and anaphylactic reactions, elevated blood pressure, abnormal liver function tests, headache, dizziness, gastrointestinal disorders, and fatigue,[17] the manufacturer's submission does not include adverse reactions in the cost-effectiveness evaluation on the assumption that there are no clear differences in adverse reactions between treatments, and that adverse events are therefore not expected to be a cost-driver.

Non-RCT evidence

The manufacturer's submission includes under the heading of non-RCT evidence the long-term extension (LTE) phases which followed the Kremer Phase 2b, AIM and ATTEST studies. No independent observational studies of abatacept were either identified in the manufacturer's submission or have come to the attention of the ERG.

Patients who had completed the double-blind phase of the Kremer Phase 2b, AIM and ATTEST were eligible to enter the open-label period in which all participants, regardless of original treatment allocation, received abatacept at a dose of, or approximating to, 10mg/kg. In the manufacturer's submission, data relating to recruitment to, and discontinuation from, the LTE studies were presented in Figures B 26-28. For ease of reference, these data are summarised in Table 28.

As may be seen from Table 28, a high proportion of eligible patients in each study agreed to continue in the LTEs. The proportion who remained in the study varied from 72% at the end of year three in the AIM study LTE and 52% at the end of year 7 in the Kremer Phase 2b study LTE to only 20% at the end of the two-year LTE of the ATTEST study. Comparable data relating to reasons for discontinuation are not presented. This omission is particularly unfortunate in relation to the LTE of the ATTEST study: despite the fact that only 76 of the 372 patients (20%) who had entered the LTE were still ongoing at the end of the two years, reasons for discontinuation are provided for only 43 of the 296 patients who discontinued (see Table 28), and no further explanation is provided.[6]

Table 28: Recruitment to, and discontinuation from, the LTE studies (data from the manufacturer's submission[6])

Study	**Kremer Phase 2b LTE**			**AIM LTE**		**ATTEST LTE**		
	Placebo	Ab 2mg/kg	Ab 10mg/kg	Placebo	Ab 10mg/kg	Placebo/Ab	Ab 10mg/kg	Inflix
Number completing double-blind phase	71	74	90	162	385	104	139	141
Number entering LTE, by original allocation	67 (94%)	68 (92%)	84 (93%)	161 (99%)	378 (98%)	NR	NR	NR
Number entering LTE, overall	219 (93%)			539 (99%)		372 (97%)		
Number discontinuing study in year 1				51				
Primary reasons for discontinuation in year 1								
Adverse events	NR			19		NR		
Lack of efficacy	NR			11		NR		
Withdrawal of consent	NR			12		NR		
Lost to follow-up	NR			4		NR		
Other	NR			4		NR		
Death	NR			1		NR		
Number discontinuing study in year 2	NR			48		NR		
Number discontinuing study in year 3	NR			50		NR		
Number discontinuing over total study period	105 (48%) by end of LTE year 7			149 (28%) by end of LTE year 3		296 (80%) by end of LTE year 2		
Primary reasons for discontinuation over total study period								
Adverse events	42			NR		10		
Lack of efficacy	24			NR		9		
Withdrawal of consent	19			NR		12		
Lost to follow-up	3			NR		NR		
Other	13			NR		NR		
Death	4			NR		NR		

Only the safety data from the LTE studies are considered in the ERG report; the efficacy data are not considered because of the substantial difficulty associated with interpreting efficacy data in the absence of a control arm, particularly in the context of a relapsing remitting disease such as RA.

LTE safety data

As noted earlier, RCTs do not form the best source of evidence relating to adverse events, other than those which are both common and occur within the study timescale, because few RCTs are powered to detect significant differences between treatments in the incidence of specific adverse events. More useful safety data may be obtained from large observational studies of longer duration. In the case of abatacept, the manufacturer's submission claims to form the best source of safety data since it utilises the most recent CSRs, which are said to offer more complete and accurate data than any publication to date.[6]

The manufacturer's submission presents the LTE safety data in a format which is not easy to assimilate. These data are therefore re-presented in Table 29. In the absence of untreated controls, it is difficult to assess the impact of abatacept treatment on AEs, and therefore data from the individual studies are discussed below.

Table 29: Adverse events reported during the open-label LTEs (data from the manufacturer's submission[6])

	Kremer Phase 2b LTE (N=219)	AIM LTE (N=539)	ATTEST LTE (N=372)
Duration of open-label phase	**6 years**	**59 months**	**12 months**
Total patients with AE	Not clear	517 (95.9%)	348 (93.5%)
Patients with AE considered related to study drug	NR	NR	163 (43.8%)
Total patients discontinuing treatment due to AE	42 (10.0%)	54 (10.0%)	9 (2.4%)
Total patients with SAE	Not clear*	211 (39.1%)	82 (22%)
Patients with SAE considered related to study drug	36 (16.4%)	NR	11 (3.0%)
Total patients discontinuing treatment due to SAE	29 (13.2%)		
Total deaths	6 (2.7%)	17 (3.2%)	3 (0.8%)
Deaths considered related to study drug	0	5 (0.9%)	0
Details of SAEs			
Musculoskeletal and connective tissue disorders	43 (19.6%)	78 (14.5%)	31 (8.3%)
Infections and infestations	33 (15.1%)	52 (9.6%)	14 (3.8%)
Neoplasms (benign, malignant and unspecified)	16 (7.3%)	35 (6.5%)	5 (1.3%)
Injury, poisoning and procedural complications	20 (9.1%)	29 (5.4%)	12 (3.2%)
Gastrointestinal disorders	13 (5.9%)	26 (4.8%)	1 (0.3%)
Cardiac disorders	14 (6.4%)	16 (3.0%)	7 (1.9%)
Nervous system disorders	11 (5.0%)	14 (2.6%)	
Hepatobiliary disorders	5 (2.3%)	10 (1.9%)	2 (0.5%)
General disorders and administration site conditions	7 (3.2%)	9 (1.7%)	2 (0.5%)
Renal and urinary disorders	2 (0.9%)	9 (1.7%)	
Respiratory, thoracic and mediastinal disorders	19 (8.7%)	9 (1.7%)	5 (1.3%)
Vascular disorders	13 (5.9%)	8 (1.5%)	2 (0.5%)
Metabolism and nutrition disorders	6 (2.7%)	6 (1.1%)	2 (0.5%)
Reproductive system and breast disorders	6 (2.7%)	6 (1.1%)	5 (1.3%)
Blood and lymphatic system disorders	4 (1.8%)	5 (0.9%)	1 (0.3%)
Eye disorders	1 (0.5%)	5 (0.9%)	1 (0.3%)
Investigations	6 (2.7%)	4 (0.7%)	1 (0.3%)
Psychiatric disorders	1 (0.5%)	2 (0.4%)	NR
Skin and Subcutaneous tissue disorders	3 (1.4%)	2 (0.4%)	1 (0.3%)
Immune system disorders	1 (0.5%)	NR	NR
Ear and labyrinth disorders	1 (0.5%)	NR	NR
Surgical and medical procedures	8 (3.7%)	NR	NR
Endocrine disorders	NR	NR	1 (0.3%)

* This figure cannot be calculated from the tabulated details of SAEs because, as the data from the AIM and ATTEST studies shows, the number of reported SAEs may be greater than the number of patients with SAEs

In the Kremer Phase 2b study, none of the six deaths reported during the open-label LTE were considered to be related to abatacept. The manufacturer's submission states that 113 abatacept-treated patients (51.6%) reported adverse events, most of which were mild or moderate in intensity; the most common were RA (including worsening of RA), nasopharyngitis, upper respiratory tract infection, bronchitis, sinusitis, and urinary tract infection. As Table B 63 in that submission also states that 113 abatacept-treated patients (51.6%) reported serious adverse events, it is not clear which figure is correct; depending upon which is appropriate, the study had either a substantially lower proportion of patients than the AIM and ATTEST LTEs who reported any AE, or a higher proportion who reported an SAE. Thirty-six patients (16.4%) experienced SAEs which were considered to be related to abatacept; the most common of these abatacept-related SAEs were infections and infestations (19, 8.7%) and neoplasms (8, 3.7%). Twenty-nine patients who received abatacept during the open-label LTE (13.2%) discontinued treatment as a result of an SAE. The manufacturer's submission states that the overall incidence of SAEs, infections and infestations reported as both SAEs and AEs, malignant

neoplasms, and auto-immune disorders was no higher in the open-label LTE than was seen in the combined 2 mg/kg and 10 mg/kg abatacept groups during the double-blind period, and no new safety signals were identified during the open-label LTE.[6]

In the AIM study, 517 patients (95.9%) reported adverse events during the open-label LTE phase, the most common being nasopharyngitis, urinary tract infection, and upper respiratory tract infection. Although most AEs were mild or moderate, 31% of patients had at least one AE that was severe or very severe. 211 patients (39.1%) reported serious adverse events, the most common being RA, including worsening of RA in 36 (6.7%) and osteoarthritis in 20 (3.7%), and 54 patients (10.0%) discontinued treatment because of AEs. Five of the 17 deaths which occurred in the LTE were considered to be probably related to abatacept therapy. These were due to pneumonia, septic shock and sinusitis; septic shock and fall; malignant lung neoplasm; lobar pneumonia; and acute lymphocytic leukaemia.

In the LTE of the ATTEST study, 348 patients (93.5%) reported adverse events; these were mostly mild or moderate in intensity, with nasopharyngitis, urinary tract infection, and diarrhoea the most commonly reported. 82 patients (22.0%) reported serious adverse events: of these patients, 30 had originally been randomised to abatacept, 33 to infliximab, and 19 to placebo. The most frequently reported SAE was worsening of RA (18 patients, 4.8%). Eleven patients (3.0%) experienced a total of 12 SAEs which were considered to be related to the study drug: 5 had been originally randomised to abatacept, 4 to placebo, and 2 to infliximab. Nine patients (2.4%) discontinued treatment because of AEs. Three deaths were reported, all in patients who had received abatacept in the double-blind phase, but none was considered to be related to abatacept.

Integrated analyses

In addition to the open-label extensions of the included trials, the manufacturer's submission refers to integrated safety data from 4149 patients with up to 7 years' exposure to abatacept (12,132 person-years of exposure), referencing conference abstracts by Becker *et al.*,[60] Hochberg *et al.*,[61] and Smitten *et al.*[62] These supersede the publications by Simon *et al.*,[67,68] which included only 4134 patients (8,388 patient-years of exposure), and to which the submission makes no reference.

As noted in section 4.1.3 above, the integrated analyses include data from the Kremer Phase 2b, AIM, and ATTEST trials, and their LTEs, together with data from five other studies which did not meet the inclusion criteria for inclusion in the current submission because they included methotrexate-naïve patients or patients who responded inadequately to anti-TNF therapy. Thus the data overlap with those presented for the Kremer Phase 2b, AIM, and ATTEST studies and their LTEs.

Smitten *et al*.,[62] analysed data relating to a total of 3182 patients (7,168 person-years of exposure). They presented separate data for methotrexate-naïve patients, patients who responded inadequately to methotrexate, and patients who responded inadequately to anti-TNF therapy. The data relating to inadequate responders to methotrexate appear to consist of data from the double-blind phases and open-label LTEs of the Kremer Phase 2b, AIM, and ATTEST studies; mean exposure to abatacept in this group was 42.5 months (SD 24.1). Becker *et al*.,[60] and Hochberg *et al*.,[61] both compared short-term and long-term safety events in 4149 patients included in eight studies. Hochberg *et al*., presented data up to December 2009 and relating to 12,132 person-years of abatacept exposure;[61] their data are more recent than those of Becker *et al*., which relate to 11,658 person-years of exposure.[60] In the analysis by Hochberg *et al*., mean exposure to abatacept was 35.6 months (range 1.9 to 104.2 months).

Data from Smitten's analysis are presented in Table 30, alongside the cumulated short- and long-term data from Hochberg *et al*. As may be seen, the point estimates of the incidence rates of the various AEs are very similar, and the confidence intervals overlap, with the sole exception of malignancies, which are more common in inadequate responders to methotrexate. Unfortunately, in the absence of data relating to patients with RA of comparable severity who were not treated with abatacept, it is not possible to assess the extent to which abatacept therapy affects the incidence of adverse events. However, Hochberg *et al*., note that the annual incidence rates for SAEs did not increase with increasing abatacept exposure, and that no new safety events were identified over time.[61] In addition, the manufacturer's submission states that, according to Smitten *et al*., (2010), "malignancies (e.g. colorectal cancer, lung cancer, lymphoma, prostate cancer, or breast cancer) in abatacept patients were not significantly increased compared to that expected based on the general population".[6] The source of the data underlying this statement, which is not in fact contained in Smitten *et al*., 2010,[62] is not clear.

Hochberg *et al*.,'s analysis[61] suggests that the incidence of safety events was generally consistent between short-term and long-term exposure. The one exception is the lower incidence of acute infusion events in the cumulative period than in the short-term period (3.90/100p-y (95% CI 3.52, 4.32) vs. 11.61/100p-y (95% CI 10.14, 13.22). Hochberg *et al*., do not indicate whether this reflects increasing tolerance of abatacept over time, or the withdrawal of patients who suffered acute infusion events relatively early in their treatment with abatacept.

Table 30: Adverse events: incidence rates per 100 person-years

Event	Incidence per 100-person-years (95% CI)	
	Inadequate responders to MTX (data from Smitten *et al.*,[62] n=1280)	**All patients (data from Hochberg *et al.*,[61] and MS,[6] n=4149)**
Overall AEs	260.28 (246.00, 275.17)	NR
Death	0.65 (0.44, 0.93)	0.60 (0.47, 0.76)
Overall serious AEs	14.62 (13.37, 15.97	14.61 (13.85, 15.41)
Serious infections	2.75 (2.28, 3.30)	2.87 (2.57, 3.19)
Opportunistic infections	0.22 (0.11, 0.41)	0.36 (0.27, 0.49)
Hospitalised infections	2.63 (2.17, 3.16)	2.64 (2.35, 2.95)
Serious pneumonia	0.59 (0.38, 0.86)	0.46 (0.34, 0.59)
Malignancies	1.30 (0.98, 1.68)	0.73 (0.58, 0.89)
Autoimmune events	1.78 (1.41, 2.23)	1.99 (1.74, 2.26)
Acute infusional events (ie events occurring within 1 hour of start of abatacept infusion) (n=993 as 1 study did not record acute infusional events)	3.65 (3.00, 4.40)	3.90 (3.52, 4.32)
Peri-infusional events (ie events occurring within 24 hours of start of abatacept infusion)	9.41 (8.42, 10.48)	NR

The data from the integrated analyses appear to support the conclusion drawn in the manufacturer's submission that abatacept is generally well tolerated in both the short-term and the long-term. There is as yet no evidence to indicate the emergence of new clinically important safety issues with extended use. However, as this conclusion is based on an analysis in which the mean exposure to abatacept was only 35.6 months, it cannot be regarded as definitive. If, as the ERG's clinical advisors believe, abatacept is included in the British Society for Rheumatology Biologics Register,[69] then further long-term safety data should eventually become available.

4.2.6 Critique of submitted evidence syntheses

Pairwise meta-analyses

The manufacturer undertook a series of pairwise meta-analyses using data from the three key RCTs to compare the efficacy of abatacept plus methotrexate with that of placebo plus methotrexate in relation to the following outcomes:

- Change from baseline in HAQ score at 24/26 weeks
- Change from baseline in HAQ score at 1 year
- Change from baseline in DAS28 at 24/28 weeks
- DAS28 improvement at 24/28 weeks
- ACR20 at 24/28 weeks
- ACR50 at 24/28 weeks
- ACR70 at 24/28 weeks

Data from these meta-analyses have been incorporate into the summary of clinical effectiveness results in section 4.2.1.

In the pairwise meta-analyses, inverse-variance fixed and random effects methods were used to calculate mean differences for the continuous outcomes (ie changes from baseline). The Mantel-Haenszel fixed effect method and the Der Simonian Laird random effect method were used to calculate odds ratios and relative risks for the binary outcomes.

Network meta-analyses

The manufacturer's submission includes a mixed treatment comparison whose aim is to evaluate the efficacy of abatacept plus methotrexate versus five comparator biologic DMARDs plus methotrexate using a network analysis. The network of treatments includes abatacept, adalimumab, certolizumab, etanercept, golimumab, infliximab, and placebo. The only direct comparison in the network is abatacept versus infliximab; the remaining comparisons are indirect comparisons via placebo.

The manufacturer's submission states that evidence for the mixed treatment comparison was identified using a systematic searching and sifting process similar to that used in the systematic review of clinical effectiveness. Searches to identify all relevant studies were performed in January 201 and updated in October 2010. Four electronic bibliographic databases (Medline, Medline-In-Process, Embase, and the Cochrane Library) were searched, as were the ACR and EULAR conference websites. A handsearch of NICE STA reports was also performed. One NICE submission and 4 relevant abstracts were included in the mixed treatment comparison (MTC) analyses. The strategies used to search the electronic databases, whilst somewhat inelegant, appear to be appropriate. The inclusion and exclusion criteria used to identify relevant studies were not reported in the manufacturer's submission, but were subsequently supplied;[2] they are as follows:

Inclusion criteria

- Population: patients diagnosed with RA with an inadequate response to, or intolerance of, methotrexate
- Interventions: any of the following agents, at their licensed doses and in combination with methotrexate: abatacept, infliximab, etanercept, golimumab, adalimumab or certolizumab pegol
- Comparators: any of the included interventions, or placebo plus methotrexate
- Outcomes: efficacy parameters: CFB in HAQ score; ACR20, ACR50, and ACR70 response rates; and clinical remission response rates (DAS28 <2.6), all at 24/28 and 48/54 weeks; safety parameters: withdrawals due to adverse events at 24/28 weeks
- Study design: RCT

Exclusion criteria

- Interventions: anakinra, tocilizumab, rituximab (these drugs are excluded because they are recommended for patients with an inadequate response to TNF-α inhibitors rather than those with an inadequate response to methotrexate).

Eleven RCTs which compared a biologic DMARD plus methotrexate with placebo plus methotrexate were used to inform the MTC. All reported outcomes at 24/26 weeks. These RCTs included three trials of abatacept, one of which (ATTEST) also compared abatacept with infliximab; two trials of adalimumab, two trials of certolizumab, two trials of etanercept, one trial of golimumab, and one trial of infliximab vs placebo (for details, see Table 31). The manufacturer's submission notes in the text, though not in the supporting table, that the population of the TEMPO study of etanercept may differ from that of the other included studies as it is composed of patients who responded inadequately to conventional DMARDs but not to methotrexate. It therefore suggests that the high placebo response seen in the TEMPO study may be explained by the fact that patients in the placebo group effectively changed their treatment from a conventional DMARD to methotrexate, which is seen as a more efficacious treatment.

Table 31: Summary of trials used to inform the mixed treatment comparison (reproduced from manufacturer's submission, Table B 31)

Trial	Interventions compared (incl. dose, frequency and duration of treatment)				Comparison	Population treated	Primary and secondary study references
	Interventions	Dose	Frequency	Duration			
Abatacept studies							
AIM	Placebo + MTX	N/A	N/A	1 year	Abatacept + MTX vs. Placebo + MTX	Active RA despite MTX treatment	Abatacept CSR, Kremer *et al.*, 2006 and Russell *et al.*, 2007
	Abatacept + MTX	10 mg/kg	Days 1, 15, and 29 and every 28 days thereafter				
Kremer Phase 2b	Placebo + MTX	N/A	N/A	1 year	Abatacept + MTX vs. Placebo + MTX	RA that has remained active despite MTX therapy.	Kremer *et al.*, 2005, Kremer *et al.*, 2003
	Abatacept 2 mg/kg every 4 weeks + MTX	2 mg/kg	Day 1, 15, and 30 and every 30 days thereafter				
	Abatacept 10 mg/kg every 4 weeks + MTX	10 mg/kg					
ATTEST	Placebo + MTX	N/A	N/A	6 months	Abatacept + MTX vs. Placebo + MTX at 6 months only Infliximab + MTX vs. Placebo + MTX at 6 months only Abatacept + MTX vs. Infliximab + MTX	RA and an inadequate response to MTX	Abatacept CSR and Schiff *et al.*, 2007
	Infliximab + MTX	3 mg/kg	Days 1, 15, 43 and 85, and every 56 days thereafter	12 months			
	Abatacept + MTX	10 mg/kg	Days 1, 15 and 29, and every 28 days thereafter	6 and 12 months			
Adalimumab studies							
ARMADA	Placebo + MTX	N/A	N/A	24 weeks	Adalimumab + MTX vs. Placebo + MTX	Active RA despite treatment with MTX	Weinblatt *et al.*, 2003
	Adalimumab + MTX	20 mg	Every other week				
	Adalimumab + MTX	40 mg					
	Adalimumab + MTX	80 mg					
DE019	Placebo + MTX	N/A	N/A	52 weeks	Adalimumab + MTX vs. Placebo + MTX	Active RA receiving with an inadequate response to MTX.	Keystone *et al.*, 2004
	Adalimumab + MTX	20 mg	Weekly				
	Adalimumab + MTX	40 mg	Every other week				
Certolizumab studies							
RAPID 1	Placebo + MTX	N/A	N/A	1 year	Certolizumab + MTX vs. Placebo + MTX	Active RA with an inadequate response to MTX	Keystone *et al.*, 2008 and Strand *et al.*, 2009
	Certolizumab + MTX	200 mg	Every other week				
	Certolizumab + MTX	400 mg	Every other week				
RAPID II	Placebo + MTX			24 weeks	Certolizumab + MTX vs. Placebo + MTX	Active RA despite >= 6 months MTX treatment	Smolen *et al.*, 2009
	Certolizumab + MTX	200 mg	Every other week				
	Certolizumab + MTX	400 mg					
Etanercept studies							
TEMPO	Placebo + MTX	N/A	N/A	52 weeks to 2 years	Etanercept + MTX	Active RA with an	Heijde van der *et al.*,

Trial	Interventions compared (incl. dose, frequency and duration of treatment)				Comparison	Population treated	Primary and secondary study references
	Interventions	Dose	Frequency	Duration			
	Etanercept	25 mg	Twice weekly		vs. Placebo + MTX	inadequate response to MTX	2004, Heijde van der *et al.*, 2006 (PRO), Heijde van der *et al.*, 2006 (2-yr), Heijde van der *et al.*, 2007
	Etanercept + MTX	25 mg					
Weinblatt *et al.*, 1999	Placebo + MTX	N/A	N/A	24 weeks	Etanercept + MTX vs. Placebo + MTX	Active RA despite >= 6 months of MTX	Weinblatt *et al.*, 1999
	Etanercept + MTX	25 mg	Twice weekly				
Golimumab studies							
GO-FORWARD	Placebo + MTX	N/A	N/A	24 weeks	Golimumab + MTX vs. Placebo + MTX	Active RA with an inadequate response to MTX	Keystone *et al.*, 2009
	Golimumab + placebo	100 mg	Every 4 weeks				
	Golimumab + MTX	50 mg					
	Golimumab + MTX	100 mg					
Infliximab studies							
ATTRACT	Placebo + MTX	N/A	N/A	30 weeks to 2 years	Infliximab + MTX vs. Placebo + MTX	Active RA with an inadequate response to MTX	Maini *et al.*, 2004 (2-yr), Lipsky *et al.*, 2000 and Maini *et al.*, 1999
	Infliximab + MTX	3 mg/kg	Every 8 weeks				
	Infliximab + MTX	3 mg/kg	Every 4 weeks				
	Infliximab + MTX	10 mg/kg	Every 8 weeks				
	Infliximab + MTX	10 mg/kg	Every 4 weeks				

The manufacturer's submission also notes that the studies were not completely homogeneous in terms of baseline patient characteristics. Thus, disease duration ranged from a mean of 4.5 years in one arm of the GO-FORWARD study to a mean of 13.1 years in one arm of the ARMADA study. Study inclusion criteria relating to swollen and tender joints also varied: for the abatacept studies, these were ≥10 and ≥12 respectively, compared with ≥9 and ≥6 respectively for the adalimumab studies, ≥6 for each in Weinblatt *et al.*,'s study of etanercept, and ≥4 for each in the GO-FORWARD study (for details, see manufacturer's submission Table B 33). Thus the populations of the abatacept studies may have had more advanced RA than those of the studies of other treatments. The manufacturer's submission notes that this could explain potential differences in the observed relative treatment effects, but states that the random effects approach was used to take into account this heterogeneity across trials. The ERG note that, whilst the random effects model allows an assessment of a population mean and between-study standard deviation, the population mean does not apply to all populations in the presence of heterogeneity. A formal account of heterogeneity would explore the use of meta-regression to explain the heterogeneity such as adjusting for disease duration.

The manufacturer's submission states that no data relating to the outcomes of interest could be identified for the GO-FORWARD study. Nevertheless, it was retained in the MTC, and relevant outcome data are summarised in manufacturer's submission Tables B 35-38.

The analyses used in the MTC analyses are as follows:

- HAQ-DI CFB 24/26 weeks
- ACR20 response 24/28 weeks
- ACR50 response 24/28 weeks
- ACR70 response 24/28 weeks

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

The manufacturer's submission did not present efficacy findings from the MTC relating to the DAS28, arguably the most clinically important outcome measure. Instead, it focused primarily on the change from baseline in the HAQ at 24/26 weeks. The MTC suggested that abatacept plus methotrexate was expected to be more efficacious than placebo plus methotrexate, and was expected to display efficacy comparable to that of most other biologic DMARDs, with numerical differences ranging from -0.11 versus infliximab to 0.09 versus certolizumab pegol (see manufacturer's submission, Section 5.7.6.1, Table B38). The absolute CFB for biological agents in combination with methotrexate was expected to range from -0.46 (infliximab) to 0.65 (certolizumab) (see manufacturer's submission, Section 5.7.6.2, Table B39). It also suggested that all biological agents

considered in the submission would result in comparable proportions of ACR20/50/70 responders, although certolizumab pegol would be expected to have a slightly higher ACR20 response rate than other biological DMARDs (see manufacturer's submission, Section 5.7.6.3 to 5.7.6.8, Tables B40-45).

4.3 Conclusions

The clinical effectiveness evidence submitted by the manufacturer indicated that, relative to placebo, abatacept, at a dose of, or approximating to, 10 mg/kg, reduced disease activity, as measured by the DAS28 and ACR responses, at 6 and 12 months. Abatacept was associated with a relative risk of achieving low disease activity (DAS28 $\leq$3.2) at 12 months of 3.89 (95% CI 1.13, 13.40; p=0.03), and a relative risk of achieving remission (DAS28 <2.6) of 4.78 (95% CI 2.06, 11.09; p=0.003).

Relative to placebo, abatacept also appeared to be associated with improved physical function, as measured using the HAD-DI or MHAQ, at 6 months and 1 year, and with less joint damage at one year; however, the clinical significance of these results was not clear. It also appeared to be associated with improvements at both 6 months and 1 year in pain, morning stiffness, sleep quality, fatigue, and health-related quality of life as assessed by the physical and mental summary components of the SF-36.

As noted above, the ATTEST study was not powered to detect statistical differences between abatacept and infliximab, and a statistically significant difference was identified in only two outcomes: ACR20 response and the physical summary component of the SF-36, both at 1 year. Both results favoured abatacept.

The manufacturer's submission did not present efficacy findings from the MTC relating to the DAS28, arguably the most clinically important outcome measure. Instead, the analysis focused primarily on the change from baseline (CFB) in the HAQ at 24/26 weeks, which suggested that abatacept plus methotrexate was expected to be more efficacious than placebo plus methotrexate, and was expected to display efficacy comparable to that of most other biologic DMARDs; the absolute CFB for biological agents in combination with methotrexate was expected to range from -0.46 for infliximab to 0.65 for certolizumab. The MTC also suggested that all biological agents considered in the submission would result in comparable proportions of ACR20/50/70 responders, although certolizumab pegol would be expected to have a slightly higher ACR20 response rate than other biological DMARDs.

The RCT evidence suggested that abatacept at a dose of, or approximating to, 10 mg/kg was not associated with a higher rate of serious adverse events than placebo at either 6 months or 1 year, and

its adverse event profile appeared to be favourable compared with infliximab. However, the data relating to acute infusional AEs and peri-infusional AEs are incomplete.

Longer-term data incorporated into the integrated safety analyses of abatacept indicated that the incidence of serious AEs did not increase over time, and no new safety events were identified. Thus, abatacept appeared to be generally well tolerated in both the short and the longer term. However, as this conclusion was based on an analysis in which the mean exposure to abatacept was only 3.56 months, it cannot be regarded as definitive. Moreover, the submission indicated an 80% discontinuation rate from the two-year LTE of the ATTEST study, and no explanation was provided for this.

5 COST EFFECTIVENESS

5.1 ERG comment on manufacturer's review of cost-effectiveness evidence

The manufacturer provided a systematic review of previous economic evaluation of biologic and conventional DMARDs (Section 6.1 of the manufacturer's submission[6]). Whilst useful in providing background information and an indication of the likely cost-effectiveness of abatacept, none of these studies adequately addressed the decision problem and the focus of the cost-effectiveness sections of the ERG report will be the *de novo* model submitted by the manufacturer. A brief summary is provided of two studies that explicitly incorporated abatacept although both have limitations. The study by Russell *et al.*,[70] was reported as providing only an estimation of the costs of treatment sequences with no data provided on QALYs. The study by Vera-Llonch *et al.*,[71] compared abatacept treatment in conjunction with methotrexate (MTX) with MTX alone in patients who had an inadequate response to MTX. Following the clarification process and manufacturer's response,[2] it was confirmed that there was a typographical error in the original submission; the incremental cost per QALY of abatacept compared with MTX should have been reported as US$42,348.

5.2 Summary and critique of manufacturer's submitted economic evaluation

The manufacturer supplied a simulation model written within Microsoft Excel. The model was relatively more complex that most seen by the ERG due to the incorporation of user-defined functions and a large number of defined names. The omission of a glossary of names and description of the functions increased the time required to review and validate the model. It is unclear whether the structure of the model contributed to a serious error that was identified by the ERG that will be detailed later.

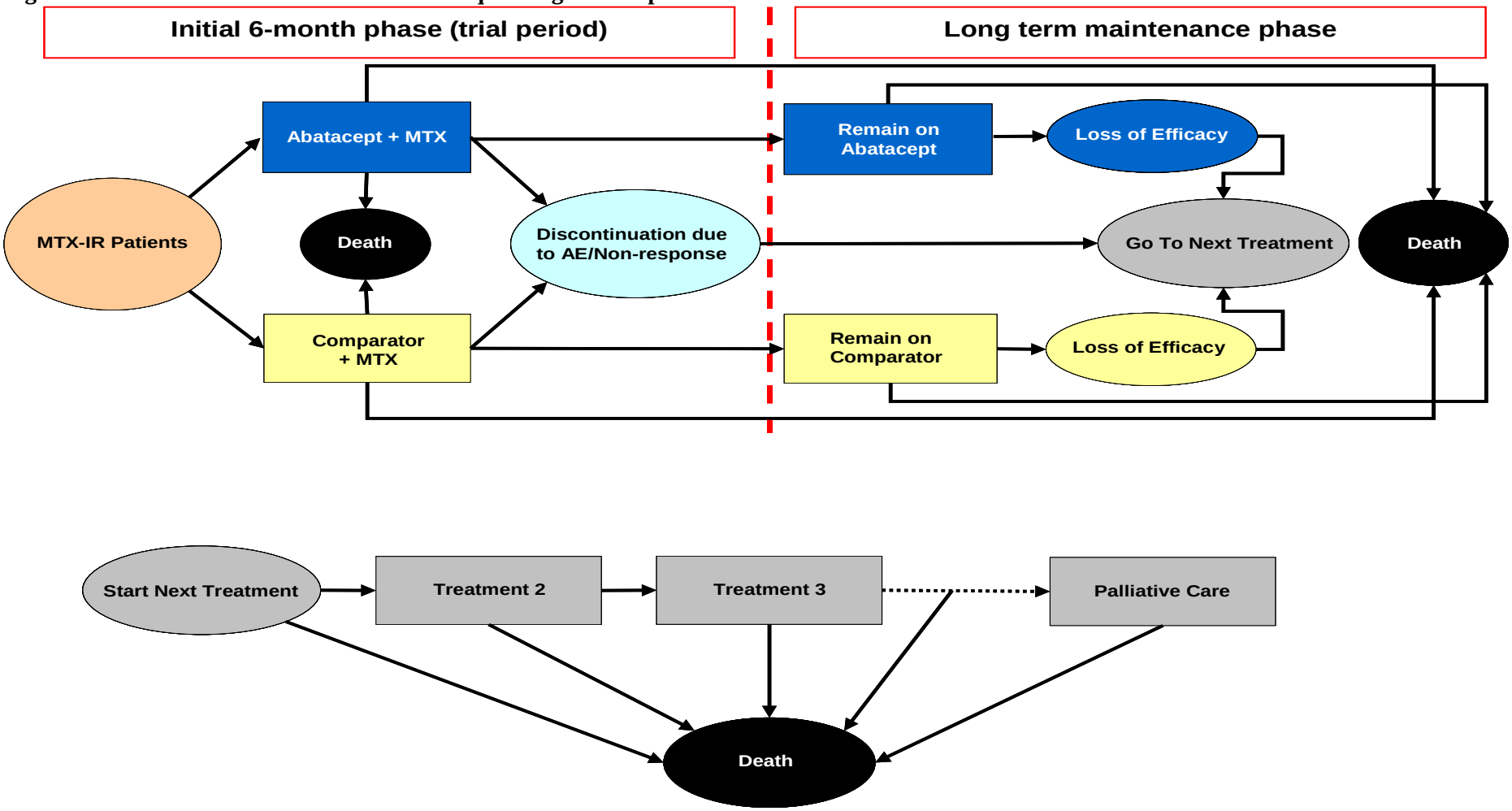
The illustration of the model that was submitted by the manufacturer has been replicated in Figure 1. Further explanation (written by the ERG) has been provided after this figure.

The model attempted to incorporate both individual patient level uncertainty where the starting characteristics (such as age, weight, and HAQ score) and response to treatment and whether a patient suffers an adverse event are simulated for each patient, as well as the parameter uncertainty for values such as the rate of serious adverse events or change in baseline HAQ for an intervention. Exploratory analyses were undertaken by the manufacturer to determine the level of individual patients (8,000) required to achieve a stable answer. Due to the computational time necessary to undertake probabilistic sensitivity analyses, the number of patients in each run was reduced to 1,000 with the manufacturer assuming that 500 sets of parameter values were required to produce stable answers. The ERG comment that there was some variability in results using the same parameter inputs when

the patient numbers were reduced to 1,000 but note that the standard deviation between 10 runs of 1,000 patients was never more than 1% of the mean value and thus the reduction in the number of individual patients simulated would not overly effect the results.

However, as will be later detailed, there was a serious error in the actual implementation of the parameter uncertainty.

Figure 1: Illustration of the treatment sequencing abatacept economic model



MTX: Methotrexate, MTX-IR: Methotrexate-Inadequate Responder, AE: Adverse Event

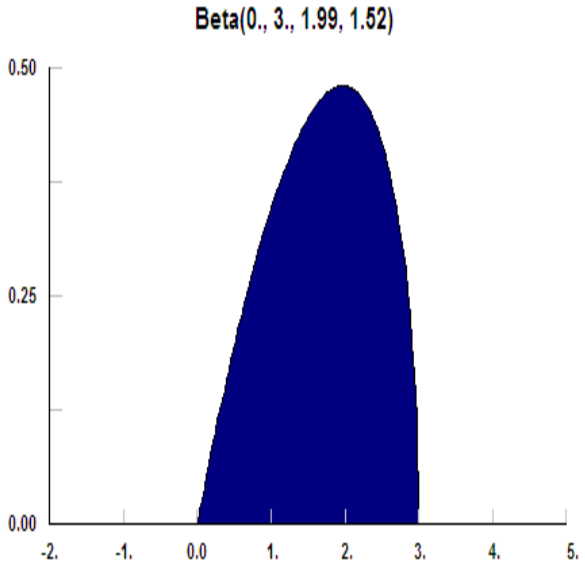
The patients simulated in the model were assumed to have a mean age of 51.5 years with a standard deviation of 12.90; 77.8% of patients were assumed to be female. The mean baseline HAQ was assumed to have a mean of 1.71 with a standard deviation of 0.70. All such data were taken from the clinical study report (abatacept arm) of the AIM trial which was not referenced and does not appear to be publicly available.

The patient weight was sampled from a distribution reported to be data on file at Bristol-Myers Squibb and is given in Table B 75 on page 251 of the manufacturer's submission.

As the HAQ score is bounded between 0 and 3, the normal distribution, which is unbounded, needed to be manipulated. Two parameters were calculated: x (the mean from the normal distribution divided by 3) and y (the standard deviation from the normal distribution divided by 3). These values were used to estimate the alpha and beta for a Beta distribution where alpha = $x*(x*(1-x)/y^2 -1)$ and beta = $x*(x*(1-x)/y^2 -1)$. In the base case, alpha equalled 1.99 and beta equalled 1.52. This Beta distribution was then scaled to cover values between 0 and 3. Note that in the Excel calculation of a Beta distribution the beta parameter equates to the number of failures rather than the number of trials.

The distribution is depicted alongside the proportion within defined ranges in Figure 2. The clinical advisors to the ERG did not believe the distribution was unrepresentative of patients seeking treatment following failure of MTX.

Figure 2: The assumed distribution of HAQ for patients entering the model



HAQ range	Beta
0.0-0.5	5%
0.5-1.0	14%
1.0-1.5	20%
1.5-2.0	23%
2.0-2.5	23%
2.5-3.0	15%

Conceptually, the model submitted by the manufacturer assumes a sequence of treatments, with a patient discontinuing treatment and moving on to the next line of treatment, where appropriate, when one of the following events occurs:

1) The patient experiences a severe adverse event
2) The patient is assumed not to respond to the treatment within the evaluation period
3) The treatment becomes ineffective
4) The patient dies whilst on treatment.

These are discussed in turn, with a focus on the values for three drugs: abatacept (the intervention evaluated within the single technology assessment); infliximab (the only competitor biologic DMARD that is also delivered intravenously); and etanercept (a widely used and relatively inexpensive biologic DMARD). As will be detailed, abatacept was not reported to be a cost-effective use of resources compared to biologic DMARDs[2] but was compared with infliximab in patients who could not, for whatever reason, take a subcutaneous injection. It is this latter comparison that is the focus of the ERG critique of the economic analyses presented and the additional work undertaken by the ERG.

1) The patient experiences a severe adverse event

A serious adverse event is assumed to be able to be experienced only within the initial six months of treatment. Each intervention has a rate of serious adverse events; these are summarised for abatacept, infliximab and etanercept in Table 32. These values are reported to have been taken from a mixed treatment comparison (MTC), although within the model these distributions were assumed to be independent. The values were estimated via a Beta distribution that was calculated using the mean value and a standard deviation estimated from the upper and lower bounds reported from the MTC. The ERG believes that such an approximation is unnecessary, although as later detailed the large amount of evidence available means that the approximation does not unduly influence the results.

Table 32: The assumed rates of serious adverse events for abatacept, etanercept and infliximab

	Mean value	Lower and Upper values*	Assumed Beta Distribution (alpha, beta)
Abatacept	3.00%	2.40% ; 3.60%	93.1; 3011.8
Etanercept	4.76%	3.81%; 5.71%	91.4; 1828.2
Infliximab	10.74%	8.59%; 12.89%	85.6; 711.5

*Although not reported as such in the manufacturer's submission, these were assumed to be the upper and lower 95% confidence intervals

The time at which a serious adverse event would occur is estimated from these data. Should this value be below the six month threshold period, it is assumed that a serious adverse event would not be experienced whilst on that intervention. If a patient is simulated to experience a serious adverse event then the patient immediately discontinues treatment with the HAQ remaining at the value at which the patient began treatment. The model assumes neither costs nor disutility implications of a serious

adverse event, which is likely to be unfavourable to abatacept given the lower mean rate of discontinuation.

2) The patient is assumed not to respond to the treatment within the evaluation period

The model assumes that each intervention will have an evaluation period where the efficacy of an intervention is tested and, if a threshold reduction in HAQ score was not achieved, the intervention would be discontinued and the next line of treatment would be initiated. In the model, the evaluation period was set to be six months with a HAQ reduction of 0.30 required. The rationale for using this threshold was that it was in accordance with the endpoints for the AIM[4] and ATTEST[26] RCTs. The clinical advisors to the ERG were unsure whether this threshold was used in clinical practice as it was more likely that another measure such as the DAS would be used in assessing clinical benefit. The ERG note that the value of 0.30 is only marginally greater than the expected decrease in HAQ that would be associated with MTX treatment alone (0.27). The model assumes that the effects of MTX compared with placebo and the effects of active treatment compared with MTX were additive, and thus on average an intervention would only need an additional 0.03 improvement in HAQ in order to be deemed to have successfully passed the evaluation period.

Table B 39 on page 170 of the manufacturer's submission reports the estimated improvement in HAQ score associated with each treatment. This is replicated for abatacept, etanercept, infliximab and placebo when used in conjunction with MTX in Table 33. It is noted that the confidence intervals were wide for all active interventions, although each was estimated to be markedly better than placebo.

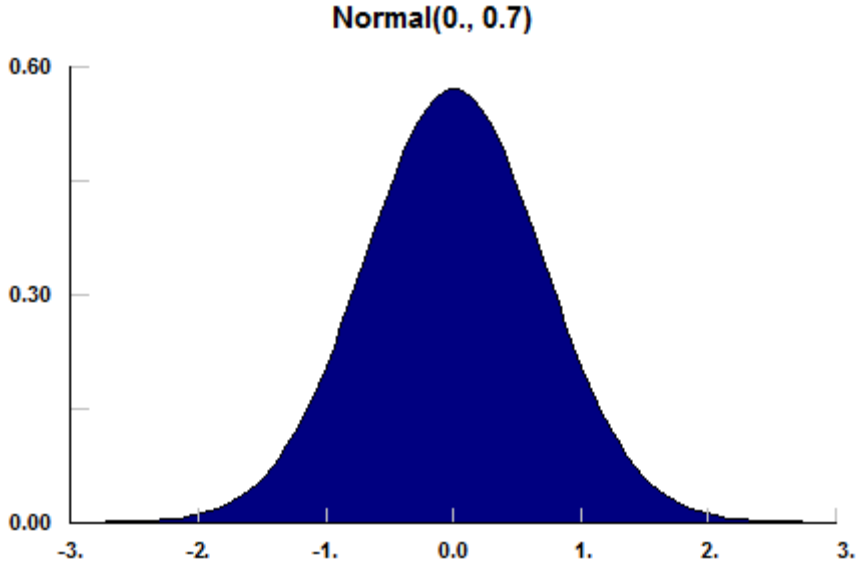
Table 33: The relative efficacy of abatacept, etanercept, infliximab and placebo when used in conjunction with MTX

Treatment	Adjusted mean HAQ CFB	2.5%CrL	97.5%CrL
Placebo + MTX	-0.27	-0.30	-0.24
Etanercept + MTX	-0.55	-0.74	-0.36
Infliximab + MTX	-0.46	-0.62	-0.30
Abatacept + MTX	-0.57	-0.69	-0.43

It was assumed that the change in HAQ would happen gradually with the time to full response of 3 months. It was assumed that there would be a linear change during this period before a steady HAQ was maintained.

The mean improvement in HAQ for the interventions were assumed independent; the ERG notes that using samples from the joint posterior distribution from the MTC would be more appropriate as it maintains the properties of the joint posterior distribution. Individual patient variation in response was incorporated; the manufacturer's model assumes that the standard deviation associated with baseline HAQ (0.70) would equate to the standard deviation in individual response to treatment, although provides no justification for this assumption. The ERG note that this standard deviation is of the same order as the actual treatment effect and may be too large. This value is adjusted within the ERG sensitivity analyses. For information, the assumed inter-patient variability in change in HAQ score has been plotted and is shown in Figure 3. For further clarity it is stated that if the sampled value was zero the patient would have a change in HAQ equal to the mean for that intervention. As HAQ is constrained between zero and three, the manufacturer assumed that any value that exceeded these limits would be curtailed at the appropriate limit.

Figure 3: The variation in change in HAQ score between individuals



Patients in whom the threshold reduction in HAQ of 0.3 was not met at six months were assumed to be a non-responder to that treatment and would discontinue the treatment in favour of the next line of therapy. The HAQ of such patients was assumed to return immediately to the value on treatment initiation. This may be true when a small HAQ improvement was seen, but the ERG question the assumption that the HAQ score would return to its initial value when the patient had deteriorated on treatment. This structural assumption is tested in the ERG sensitivity analyses.

Patients who did meet the threshold requirement continued on treatment and maintained their HAQ score until shortly before discontinuing treatment.

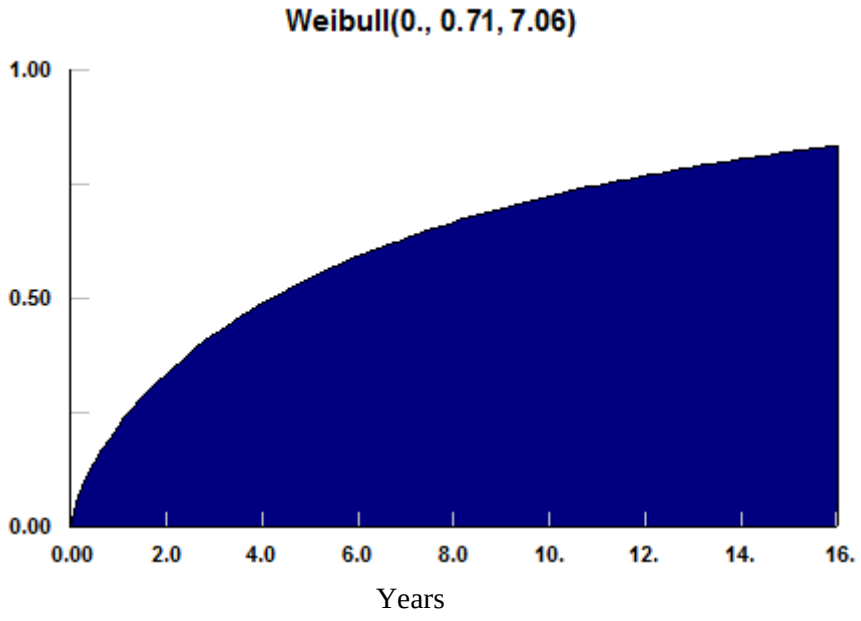
3) The treatment becomes ineffective

The model assumes that, if an intervention is deemed to be effective, then a patient will continue to receive the treatment. For biologic DMARDs, it is assumed that the HAQ score of the patient will remain constant during this period; for conventional DMARDs, it is assumed that there would be a increase in HAQ score of 0.045 per annum and that there would be an increase in HAQ score of 0.060 per annum for those patients receiving palliative care, which is the final line of treatment.

The time to discontinuation for a patient was assumed by the manufacturer to be equal for all biologic DMARDs, and thus the time at which a patient would discontinue abatacept would be identical to the time at which that patient would discontinue infliximab or etanercept. This assumption has been tested by the ERG in the sensitivity analyses. The time to discontinuation for an individual patient on a biologic DMARD was assumed by the manufacturer to be sampled from a weibull distribution with a shape parameter of 0.71 and a scale parameter of 7.06, citing data reported in Malottki *et al*.[72]

For information, a curtailed depiction of the cumulative time to event based on a weibull distribution with a shape of 0.71 and a scale of 7.06 has been shown in Figure 4. The mean value is estimated to be 8.82 years with the median value 4.21 years.

Figure 4: The cumulative discontinuation rate for biologic DMARDs

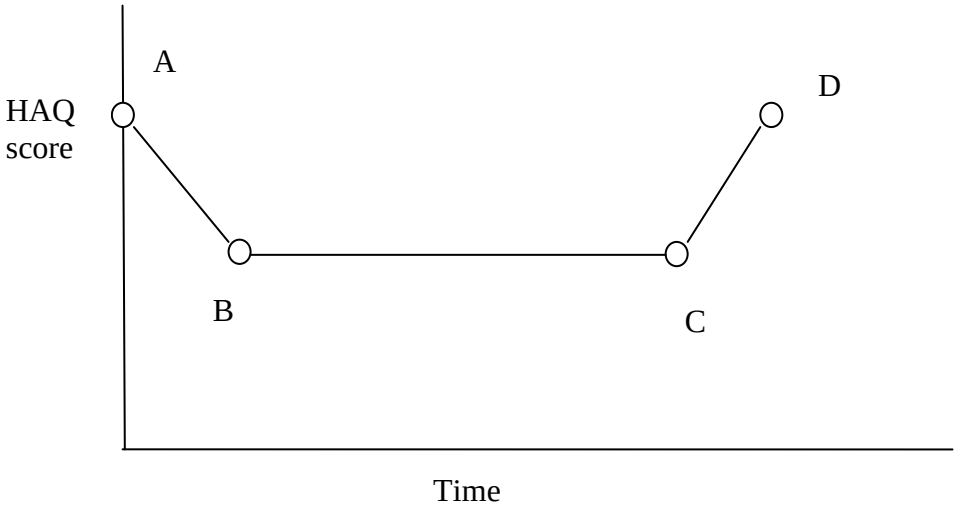


The model assumes for responders that the HAQ score obtained at the end of the evaluation period will be maintained until three months before the sampled time to discontinuation. The HAQ score would then linearly increase until the HAQ at the end of treatment was reached. The HAQ score at the end of treatment was estimated to be the HAQ score on initiation of treatment for biologic DMARDs.

For conventional DMARDs and palliative care the HAQ score on initiation of treatment would have been increased by the specific yearly increase (described above) multiplied by time on treatment.

To demonstrate this point an illustration of the HAQ score for a responder to a biologic DMARD is provided in Figure 5.

Figure 5: An illustration of the change in HAQ score for a responder to a biologic DMARD



Treatment is initiated at Time point A and there is a linear improvement in HAQ score over a three-month period to reach the full effect (Time Point B). The intervention will be discontinued at time point D, although it is assumed that the effects would begin to wane three months earlier (Time Point C). A linear deterioration is assumed over a three-month period. For biologic interventions the HAQ score at time points A and D are assumed identical.

The model has the functionality to include adverse events during the treatment phase that were not severe enough to discontinue treatment, although these have not been incorporated within the manufacturer's submission. It is unclear whether this is favourable or unfavourable to abatacept.

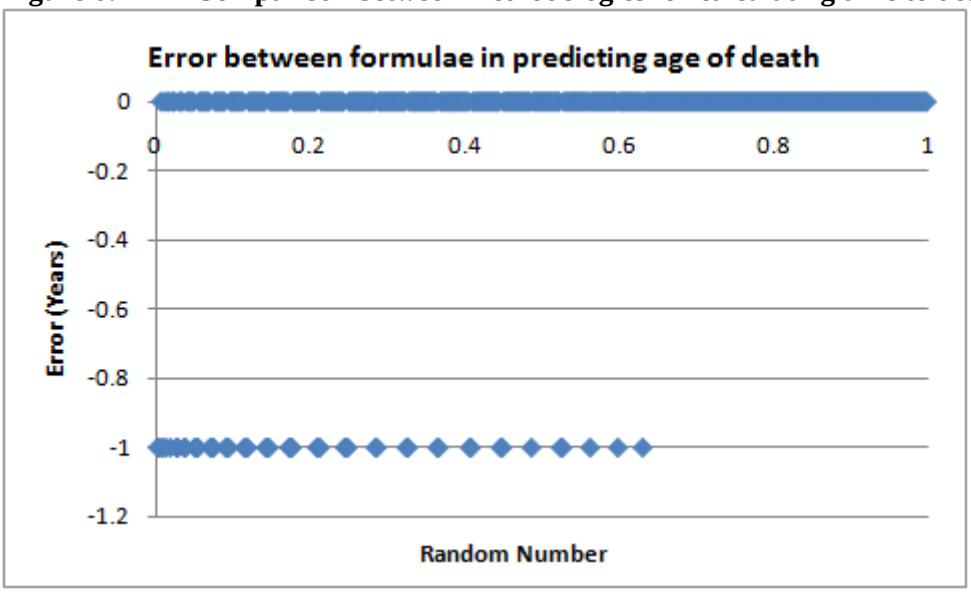
The costs associated with joint replacement have been included within the model with the time to a joint replacement being inversely related to HAQ score. The clinicians providing guidance to the ERG do not consider this assumption to be unreasonable. However, as detailed later, the ERG believe that there is potential for the costs of joint replacement to be double counted and the explicit estimation of the costs of joint replacement were removed in sensitivity analyses.

4) The patient dies whilst on treatment

Death can occur at any time within the model. The HAQ score is assumed to influence the rate of mortality with the mortality hazard ratio increasing by 1.33 (95% confidence interval (CI) 1.10-1.61

for each unit increase in the HAQ score.[73] The model used a non-standard novel approach in estimating the time of death which reduced computational time. Traditionally, the hazard ratio would be used to formulate a new survival curve; however, the manufacturer adjusted the random number used to sample from the original curve. The ERG checked whether this method was valid using a simulated 51 year old woman with a HAQ score of 1.71 and comparing the predicted time of death for each method at chosen random numbers. In the majority of cases, the integer time of death was identical, but in approximately 10% of cases the method adjusting the random number predicted an extra year of survival (as shown in Figure 6). As this change is small, potentially heavily affected by discounting, and applicable to all treatments, the ERG were not concerned by this inaccuracy.

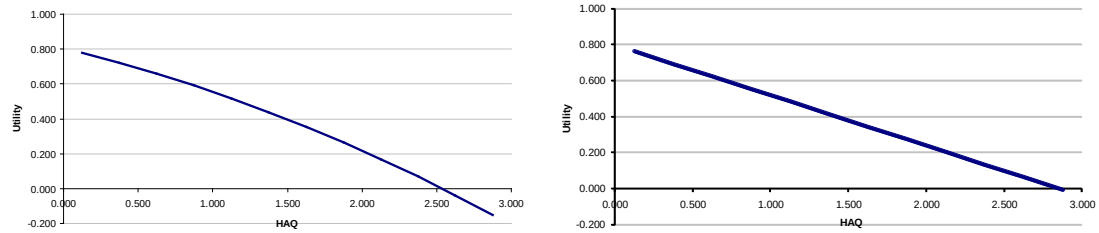
Figure 6: Comparison between methodologies for calculating time to death



The Quality Adjusted Life Years (QALYs) gained during the use of an intervention.

The utility of a patient with RA is assumed to be inversely related to the HAQ score, and the model allows the use of four algorithms. The base case is that of a quadratic approach used in Malottki *et al*.,[72] which is justified as being an algorithm used in a recent review of interventions for RA conducted by NICE.[1] The alternative sources for utility were: algorithms using a linear mapping as used in Barton *et al*;[74] a mapping based on the HUI as reported by Bansback *et al*.,[75] and a table of presumably empirical data relating HAQ score to utility published by Kobelt *et al*.[20] For information, a depiction of the relationship between HAQ score and utility in the base case, and in the sensitivity analysis conducted by the manufacturer (Bansback *et al*.,) is shown in Figure 7.

Figure 7: The relationship between HAQ score and utility



a) Manufacturer Base Case b) Manufacturer Alternative Scenario

The manufacturer calculated the total undiscounted utility on treatment as the summation of three areas represented in Figure 4. These are: the area represented by the triangle between A,B and the line of HAQ score at initiation; the rectangle between B,C and the line of HAQ score at initiation; and the triangle between A,B and the line of HAQ score at initiation.

To improve computational time, the manufacturer attempted to use the characteristics of an exponential distribution that is associated with a constant discount rate. However, the rate of 3.5% per annum, which applies to a discrete value, has been used rather than the continuous rate of 3.44% (1+ln(1.035)). However, the ERG believe that this error will have little impact on the overall results.

The costs associated with treatment

The costs of each individual treatment are described within the manufacturer's submission. The model assumes that there are costs associated with neither conventional DMARDs nor palliative care which were assumed to be contained within the costs associated with disease that is later described. All costs of intervention have been sourced from the British National Formulary Number 60.[18]

Both abatacept and infliximab are weight-based dosages, and thus the costs of treatment are related to the patient's weight. The model incorporates an administration period which initiates the treatment before the standard treatments schedule is begun. Infliximab and abatacept were the only interventions which had an administration phase with both requiring infusions on day 1 and 15, with abatacept then having an infusion at day 29, whereas infliximab had an infusion at day 43.

Patients weighing less than 60kg are assumed to receive 2 vials of abatacept, patients weighing between 60 and 100 kg are assumed to have 3 vials, whereas patients weighing over 100kg require 4 vials. The price of these regimens is £484.34, £726.51 and £968.68 respectively and infusions are required four-weekly.

Infliximab is dosed at 3mg per kg, and is sold in 100mg vials. If vial-sharing is not assumed then patients between 33 and 66kg will require 2 vials, patients between 66 and 100kg will require 3 vials

whilst patients over 100kg will require 4 vials. These regimens are assumed to cost £839.24, £1,258.86 and £1,678.48 respectively with infusions required eight-weekly. If vial sharing is assumed to occur, then the cost of infliximab is £12.59 per kg, which can markedly reduce the price in comparison to when vial-wastage is assumed. For example, a patient weighing 70kg would have an assumed cost of £881.20 when vial sharing is assumed compared with a cost of £1258.86 when vial wastage is assumed.

Etanercept has a twice weekly schedule of 25mg, with an associated cost of £178.76. However, an error in the model assumes that etanercept (and the other biological DMARDs) do not require an integer number of injections and will underestimate the cost of these interventions.

The costs per unit are given in Table 34 which is an abbreviation of Table B80 (page 270) of the manufacturer's submission. The model assumes that all patients receive all infusions although it is noted in the submission that a small number of infusions were missed. The impact of this assumption on the cost-effectiveness of comparisons of biologic DMARDs are unclear, although it is likely to favour conventional DMARDs.

Table 34: Unit costs of abatacept, infliximab and etanercept

Treatment	Unit Cost (2010 £)	Dose per unit	Dose description (SmPc)
Abatacept	£242.17	250 mg	500-1000mg (10mg/kg) week 0,2,4 thereafter every 4 wks
Etanercept	£89	25 mg	25mg twice weekly
Infliximab	£420	100 mg	3mg/kg week 0, 2 and 6 thereafter every 8 weeks

In addition to the acquisition cost of the intervention there are costs associated with administering the medication. These are provided in Table 35 which abbreviates Table B81 on page 271 of the manufacturer's submission.

Table 35: The administration costs of abatacept, infliximab and etanercept

Biologic DMARDs	Route	Cost per Administration (2010 £)	Source
Abatacept	IV (30 min)	£158	Abatacept TNF-IR submission, also referred to in TA195[1] section 4.2.21;price indexed from 2008 to 2010
Etanercept	sc	£30	One time training from nurse specialist (community), thereafter self administration: PSSRU Curtis 2009[76] p.116. Inflated to 2010
Infliximab	IV (2-3hour)	£310	Abatacept TNF-IR submission; price indexed from 2008 to 2010 referred to in TA195[1]

The model has the functionality to incorporate dose escalation for each intervention. Within the base case analysis it is assumed that both infliximab and etanercept are subject to dose escalation. In the base case it is assumed that 29% of patients on infliximab will increase the dose at 12 months to 5mg per kg, and that 1% of patients on etanercept will increase the dose to 37.5 mg per kg at 12 months. It is noted that the manufacturer's submission reported that 31% of patients had dose escalation, which was taken from the DART study[77] which analysed the real life usage of biologic DMARDs. Thus, it appears that the 29% used in the model may be unfavourable to abatacept. The assumption that abatacept does not require a dose escalation may be an artefact of insufficient usage and follow-up and thus may be favourable to this intervention. There may be tentative evidence that the efficacy of abatacept may wane across time in that, in the ATTEST trial, placebo for 6 months, followed by abatacept for 6 months, has a numerically greater percentage of responders at 1 year than patients who have been on abatacept for the entire year, although this is unlikely to be statistically significant (Table B19 of the manufacturer's submission). Sensitivity analyses have been conducted by the ERG which assume that there is no dose escalation for any intervention.

It was assumed that the costs of monitoring were included in the costs associated with disease (that are presented later) and thus were set to zero.

The costs associated with RA

The costs associated with RA used in the model were from Kobelt *et al*.,[20] and inflated to 2010 prices. The costs include: hospitalisations, surgical interventions, ambulatory and community care, and RA cDMARD medication. For information, the cost per HAQ score interval are provided in Table 36.

Table 36: The costs associated with RA by HAQ score interval

Disease related cost (HAQ related)	
HAQ score interval	**Direct costs (2010 £)**
< 0.6	£2,733
0.6 < 1.1	£3,668
1.1 < 1.6	£4,127
1.6 < 2.1	£4,767
2.1 < 2.6	£5,522
>= 2.6	£5,991

The ERG has concerns that these costs include productivity loss which is not part of NICE's reference case.[78] The ERG believe that a more appropriate set of costs would be those detailed in Malottki *et al*., which are reported to be £1,120 per HAQ Score unit. These costs were stated to incorporate joint replacement and hospitalisations.[72]

The treatment sequences evaluated within the submission

Following withdrawal from first-line treatment, the patient is moved to the next line of intervention where the same calculations regarding the costs and QALYs accrued are undertaken. Progression through the interventions continues until the patient is on palliative care.

The sequences evaluated were simplistic and consisted of first-line treatment with a biologic DMARD with a subsequent sequence of leflunomide, gold, azathioprine, ciclosporin, penicillamine and palliative care.

The manufacturer have used this structure as they interpreted the scope as allowing only one biologic treatment before returning to conventional DMARDs. This does not represent current practice in England and Wales where a number of biologic DMARDs can be provided. A comparison of abatacept with a sequence involving multiple biologic DMARDs will be unfavourable to abatacept.

Additionally, a sequence that used both biologic infusions (abatacept followed by infliximab or infliximab followed by abatacept) strategy in patient populations who cannot receive subcutaneous injections has not been evaluated, and it is possible that a multiple infusion strategy is more cost-effective than a single infusion. The submitted model has the functionality to evaluate these sequences if they were deemed appropriate to the decision problem.

The simplification of the model often results in the second and subsequent lines of intervention producing identical costs and QALYs regardless of whether the first-line treatment was infliximab, abatacept or etanercept. This is due to the assumption that, if patients neither have a serious adverse event nor failed to meet the HAQ threshold for a responder, then the time of discontinuation was identical for all biologic treatments. As the effectiveness and duration of treatment was assumed independent of the first line of treatment, identical costs and QALYs were accrued for such patients.

The results reported in the manufacturer's submission.

In the initial submission, the manufacturer did not present the result incrementally, although this was corrected in the response to clarifications. These results are replicated in Tables 37 – 40. In these tables, D refers to dominated (in that another treatment provides greater health at a lower cost); ED refers to extendedly dominated, in that a combination of two other treatments could produce the same health gain at a reduced cost.

The reported deterministic and the probabilistic results were similar and the manufacturer concluded that abatacept (and infliximab) were dominated by adalimumab and certolizumab pegol (depending on rounding) if a patient could receive the intervention subcutaneously. In patients who could not

receive a subcutaneous injection, it was reported that infliximab was extendedly dominated by abatacept and a conventional DMARD and that abatacept had a cost per QALY of £29,888 compared with conventional DMARDs.

Table 37: Deterministic Results - All treatments

Treatment	Cost	QALY	ICER vs cDMARD	ICER (inc. analysis)
cDMARD	£76,276	4.88	Ref	
Certolizumab pegol	£103,976	6.16	£21,592	£21,592
Etanercept	£107,653	6.12	£25,361	D
Infliximab	£109,419	5.96	£30,693	D
Adalimumab	£111,922	6.29	£25,359	£64,732
Abatacept	£114,548	6.16	£29,916	D
Golimumab	£115,372	6.25	£28,592	D

Table 38: Probabilistic Results - All treatments

Treatment	Cost	QALY	ICER vs cDMARD	ICER (inc. analysis)
cDMARD	£75,095	4.75	ref	
Certolizumab pegol	£103,385	6.05	£21,833	£21,833
Etanercept	£107,067	6.02	£25,232	D
Infliximab	£108,456	5.84	£30,565	D
Adalimumab	£111,436	6.15	£25,963	£77,425
Golimumab	£114,105	6.13	£28,332	D
Abatacept	£114,596	6.07	£29,888	D

Table 39: Deterministic Results - cDMARD, abatacept, infliximab only

Treatment	Cost	QALY	ICER vs cDMARD	ICER (inc. analysis)
cDMARD	£76,276	4.88	ref	
Infliximab	£109,419	5.96	£30,693	ED
Abatacept	£114,548	6.16	£29,916	£29,916

Table 40: Probabilistic Results - cDMARD, abatacept, infliximab only

Treatment	Cost	QALY	ICER vs cDMARD	ICER (inc. analysis)
cDMARD	£75,095	4.75	ref	
Infliximab	£108,456	5.84	£30,565	ED
Abatacept	£114,596	6.07	£29,888	£29,888

Sensitivity analyses undertaken by the manufacturer

The manufacturer undertook a range of univariate sensitivity analyses to evaluate the robustness of the cost-effectiveness of abatacept both to conventional DMARDs and to infliximab. These are reported in Tables B91 and B92 in the manufacturer's submission on pages 288 and 289 respectively and are replicated here in Tables 41 and 42. It is noted that the manufacturer does not conduct sensitivity analyses against any biologic DMARD that is given via subcutaneous injection, which in conjunction with their comment that the potential population who could be expected to be treated with abatacept first-line (if approved) would be fewer than 600 per year implies that the manufacturer does

not believe that abatacept is a cost-effective treatment compared with biological DMARDs in patients who can have the intervention delivered subcutaneously.

Table 41: The sensitivity analyses conducted by the manufacturer comparing abatacept and conventional DMARDs

Alternative analyses			Impact on incremental results cDMARD			
Parameters	**Base-case**	**Variation**	**Costs £**	**QALYs**	**ICER (£/QALY)**	**% change**
Base-case	-	-	38,528	1.30	29,646	-
Discount rate effects and costs	3.5%	0% both	48,554	2.09	23,212	21.70%
		1.5% effects and 6% for costs	43,066	0.98	43,853	47.92%
		6% both	34,152	1.00	34,105	15.04%
Analysis time frame	Life time	5 years	23,386	0.28	84,390	184.66%
Utilities	Hurst	Bansback	39,212	1.22	32,047	8.09%
HAQ response rate	0.3	0.22	40,403	1.34	30,095	1.51%

HAQ: Health Assessment Questionnaire, ICER: Incremental Cost Effectiveness Ratio, QALY: Quality Adjusted Life Year, cDMARD: conventional DMARD

It is seen that the incremental cost-effectiveness ratio (ICER) increases greatly when a five-year time horizon is used, although the ERG do not believe that this is an appropriate assumption. The discount rates affect the cost-effectiveness although there appears to be no reason to not use the 3.5% rates specified in the NICE reference case.[78] A reduction in the decrease in HAQ score required to be classified as a responder does not noticeably change the incremental cost-effectiveness ratio (ICER). The use of the alternative utility mapping increases the ICER to £32,047.

The ERG note that no analysis has been undertaken assuming that abatacept is associated with a dose escalation.

Table 42: The sensitivity analyses conducted by the manufacturer comparing abatacept and infliximab

Alternative analyses			Impact on incremental results			
			Infliximab			
Parameters	**Base-case**	**Variation**	**Costs £**	**QALYs**	**ICER (£/QALY)**	**% change**
Base-case	-	-	5,434	0.21	25,355	-
Discount rate effects and costs	3.5%	0% both	8,532	0.33	25,674	1.26%
		1.5% effects and 6% for costs	5,731	0.16	36,065	42.24%
		6% both	4,792	0.18	27,014	6.54%
Analysis time frame	Life time	5 years	3,044	0.06	49,012	93.30%
Utilities	Hurst	Bansback	6,051	0.25	24,390	3.80%
HAQ response rate	0.3	0.22	5,675	0.21	26,884	6.03%
Vial wastage infliximab	Yes	No	10,078	0.17	57,843	128.13%
Dose increase infliximab	Yes	No	9,642	0.26	37,025	46.02%

HAQ: Health Assessment Questionnaire, ICER: Incremental Cost Effectiveness Ratio, QALY: Quality Adjusted Life Year

As infliximab has been recommended for people with severe RA who have failed two DMARDs, including methotrexate it is likely that patients unsuitable for subcutaneous injections would be offered infliximab in preference to starting with conventional DMARDs. Two particularly pertinent sensitivity analyses are those that allow vial sharing for infliximab which increases the ICER of abatacept compared with infliximab to £57,843 and whether a dose increase with infliximab is assumed, which increases the ICER to £37,025. It is currently unclear whether vial sharing happens routinely in practice, which may be influenced by the size of the hospital and the logistics of treating multiple patients with infliximab on the same day. However, the ERG believes that allowing a dose increase for infliximab but not for abatacept is likely to bias the results in favour of abatacept as the shorter duration of follow-up may not have allowed the potential dose increases associated with abatacept to be observed.

The ERG note that a combination of these two sensitivity analyses has not been reported. The ERG has undertaken this sensitivity analysis.

Concerns identified by the ERG

The ERG reviewed the economic model and identified a number of (potential) errors. These have been catalogued and divided into the following broad categories: concerns regarding the conceptual model; concerns regarding the population of the model; concerns regarding the statistical analyses

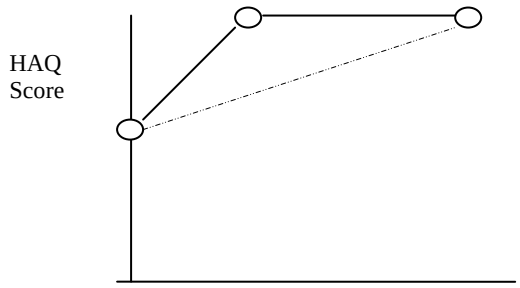
undertaken; concerns regarding the internal validity of the model; and concerns regarding the probabilistic sensitivity analyses (PSA). It is acknowledged that these categories are not mutually exclusive, but each concern will only be detailed once. The five categories are taken in order.

1. Concerns regarding the conceptual model.
   a) As previously stated, the sequences evaluated omit the use of multiple biologic DMARDs which is common practice in England and Wales. In addition, potentially cost-effective strategies for patients who cannot be given a subcutaneous injection that use both abatacept and infliximab have been omitted. It is unclear whether these have been explicitly excluded by the final scope.[79]
   b) Potential strategies that involve switching patients between interventions rather than escalating the dose have been omitted.
   c) The utility of a patient is determined solely by the HAQ score, and is assumed unaffected by the age of the patient. This may lead to implausibly high utility in patients with a low HAQ score. For instance, in the base case an elderly person with a HAQ score of 0.3 is assumed to have a utility of 0.739.
   d) Patient disutility in attending for infusions has not been considered. This is likely to be favourable to abatacept when compared with infliximab as abatacept requires infusions twice as often.
   e) Parameters that should be correlated in the model often are not. For example, the coefficients for parameters within equations for predicting utility should be correlated, as should the shape and scale of Weibull distributions. However the ERG acknowledge that the manufacturer is unlikely to have access to the required data in order to calculate correlation matrices.
   f) The costs of joint replacement appear to be double counted. The mathematical model states that the costs of joint replacements were contained in the underlying disease costs that were sourced from Kobelt *et al*.[20] As such, having an additional calculation to estimate the specific costs of joint replacement will lead to overestimated costs.
   g) No adverse events have been included within the model. If differential rates of adverse events exist between interventions, then this is likely to bias the answer as legitimate costs and disutilities are ignored.
   h) It is unclear whether the threshold selected by the manufacturer to determine whether a patient has a sufficient response to treatment (a reduction of 0.3 or greater in the HAQ score) has relevance to clinical practice, as NICE has defined a responder as a person who has an improvement of 1.2 points or more on the DAS28 score.[1] This threshold should also be considered in conjunction with the estimated average reduction in HAQ score of 0.27 associated with MTX treatment alone.

i) If a patient does not respond to treatment, it is assumed that the HAQ score when initiating the next treatment is equal to that when the previous treatment was initiated. It is unclear why this would be the case should the patient's HAQ have worsened on treatment. The ERG have conducted a sensitivity analyses altering the structure of the model so that the HAQ score on initiating the new treatment was the HAQ score at the end of the six-month evaluation period if the HAQ score had worsened.

j) There is a conceptual error in evaluating the utility of patients when the HAQ score at the end of the treatment period is predicted to be greater than 3. In this circumstance, the HAQ score is set to equal 3 at the end of the treatment period, with a linear increase across the treatment period. This may introduce inaccuracy where the maximum HAQ score of 3 is reached early in the treatment period, with a plateau until end of treatment. This is illustrated in Figure 8. This error is likely to have most influence when a patient reaches palliative care and may remain at a HAQ score of 3 for a considerable time.

Figure 8: An illustration of the error in the model when the maximum HAQ score of 3 is reached



In this example, the patient would be simulated to reach the maximum HAQ score and then remain at this level. The dotted line shows the error introduced by the logic of the model.

2. It is unclear that all biologic interventions would be discontinued at an identical time if a patient neither had an adverse event nor failed to respond to treatment. The ERG has amended the code in order that the time of discontinuation is randomly sampled (from the same distribution) for each intervention for each patient.

3. Concerns regarding the population of the model.

a) As previously detailed, the manufacturer does not explain why the standard deviation associated with baseline HAQ has been assumed to be the standard deviation associated with patient variability in HAQ response to treatment. The assumed patient variation was depicted in Figure 3. Whilst it is unlikely that the manufacturer would have the relevant data, it is expected that the change in HAQ score will be correlated to baseline HAQ score.

b) It is unclear whether there is potential in larger clinical units for the possibility of vial sharing to be achieved in the use of infliximab. As previously described, this assumption is extremely pertinent as it can greatly reduce the acquisition costs of infliximab.

c) As previously detailed, the discount rate used within the user-defined functions in the excel model should be 3.44% rather than 3.5%

d) At present there are no cost of disutility consequences of suffering a serious adverse event that is associated with treatment discontinuation. The manufacturer justifies this assumption stating that this also occurred in a previous appraisal[1] and that it is likely to be unfavourable to abatacept as the manufacturer's estimate of the rate of serious adverse events is lower for abatacept than infliximab.

4. Concerns regarding the statistical analyses undertaken

Estimating the changes in HAQ score for each intervention

The analyses presented in the main report:

- assumed that the population sampling variation is known and equal to the observed value for each treatment arm in each study
- imputed missing standard deviations for the Kremer study[28]
- ignored the fact that one of the studies was a multi-arm study, thereby ignoring correlation between estimates of treatment effect within trials
- estimated the baseline (i.e. placebo) treatment response as a common fixed population mean for each study.

The HAQ data were re-analysed using three incremental analyses:

1. Analysis 1 - to allow for uncertainty in the estimate of the population sampling variation and to estimate the two missing standard deviations by regarding them as uncertain parameters
2. Analysis 2 - to account for the inclusion of a multi-arm study
3. Analysis 3 - to account for between-study variability in the estimate of the baseline (i.e. placebo) response

Analysis 3 is preferred because this fully accounts for the additional sources of uncertainty identified during the review.

The base case results in the manufacturer's submission were reproduced and are presented for comparison alongside the three additional analyses in Table 43.

Table 43: HAQ - Absolute Mean and 95% credible interval using different methodological approaches

Treatment	Base Case	Analysis 1	Analysis 2	Analysis 3
Placebo	-0.270 (-0.302, -0.239)	-0.276 (-0.311, -0.241)	-0.276 (-0.310, -0.242)	-0.272 (-0.386, -0.157)
Adalimumab	-0.600 (-0.775, -0.424)	-0.605 (-0.749, -0.460)	-0.605 (-0.754, -0.460)	-0.603 (-0.789, -0.420)
Certolizumab	-0.656 (-0.814, -0.499)	-0.665 (-0.792, -0.529)	-0.667 (-0.795, -0.538)	-0.660 (-0.830, -0.484)
Etanercept	-0.552 (-0.751, -0.359)	-0.545 (-0.693, -0.399)	-0.547 (-0.694, -0.403)	-0.541 (-0.734, -0.352)
Golimumab	-0.609 (-0.864, -0.361)	-0.615 (-0.830, -0.398)	-0.615 (-0.822, -0.399)	-0.608 (-0.855, -0.357)
Infliximab	-0.461 (-0.620, -0.297)	-0.437 (-0.587, -0.296)	-0.440 (-0.591, -0.297)	-0.432 (-0.622, -0.243)
Abatacept	-0.567 (-0.693, -0.432)	-0.542 (-0.657, -0.439)	-0.541 (-0.656, -0.437)	-0.538 (-0.700, -0.381)
Between study SD – treatment effect	0.09 (0.03, 0.21)	0.05 (0.00, 0.17)	0.05 (0.00, 0.17)	0.05 (0.00, 0.17)
Between study SD – placebo effect	NA	NA	NA	0.17 (0.10, 0.29)

In each case, the estimates of the absolute treatment effects are similar (Analyses 1 and 2), although there is slightly greater uncertainty when the placebo response is assumed to arise from a population of placebo responses (Analysis 3). The similarity between the estimates of treatment effect is most likely a consequence of the fact that the between-study standard deviation for the treatment effect is small so that the treatment effects are essentially common to each study. On the other hand, the between-study standard deviation for the placebo effect is much greater than in the base case so that there is greater uncertainty about the treatment effects. In addition, the population standard deviation, (i.e. the sampling variation) is estimated with relatively little uncertainty (Table 44).

Table 44: Population sample standard deviation and Kramer missing variances: Means and 95% credible intervals

Treatment	Analysis 1	Analysis 2	Analysis 3
Standard deviation	0.625 (0.611, 0.640)	0.625 (0.611, 0.639)	0.625 (0.610, 0.640)
Kramer Placebo variance	0.390 (0.297, 0.498)	0.391 (0.297, 0.499)	0.390 (0.296, 0.499)
Kramer Abatacept variance	0.390 (0.294, 0.499)	0.391 (0.295, 0.500)	0.390 (0.295, 0.500)

The manufacturer used single imputation to impute the missing standard deviations for the Kramer study, which is not necessary in a Bayesian framework (A23). Missing parameters can simply be

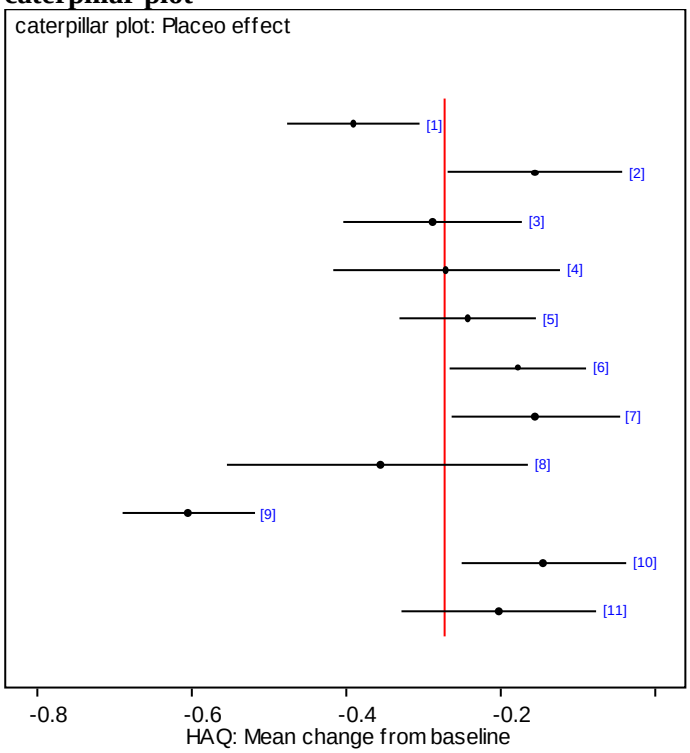
included as uncertain parameters which can be estimated using multiple imputation within WinBUGS (i.e. Markov chain Monte Carlo simulation (MCMC)).

It is not sufficient to state, as the manufacturer did in the clarification responses (A31), that "the assumption that the within study standard deviation is known and equal to the sample standard deviation is commonly used." A method for dealing with this has been described elsewhere,[80] and it is statistically imprecise to make the assumption and not to investigate its validity.

There was some evidence to suggest that the models did not represent the ATTEST[26] and Weinblatt[44] placebo responses particularly well; the deviances from Analysis 3 for these terms were 1.72 and 0.57 respectively. The ATTEST trial has the smallest standard deviation compared with the other trials, and the Weinblatt trial is the oldest trial and also the one with the smallest sample size. The deviance terms are slightly different from those derived by the manufacturer because of the way the sampling variation was accounted for in the analyses.

Figure 9 presents a caterpillar plot of the placebo responses for each study. The smallest placebo response was from the TEMPO study[81] which is study number 9 in the caterpillar plot.

Figure 9: Placebo caterpillar plot



Characterising uncertainty related to the HAQ score within the economic model

The joint posterior distribution of the uncertain parameters in each network meta-analysis, including the base case analysis, cannot be written down analytically. The joint posterior distribution will not follow any standard parametric form and the analysis will induce correlation between parameters. To preserve the underlying joint posterior distribution without having to approximate it as a multivariate normal distribution (or independent normal distributions when correlation is ignored), samples can be drawn from the joint posterior distribution and use these in the economic model. However, in this case, the technically correct approach gave very similar results to the approximate solution.

The results of each analysis are very similar, and the correlation between parameters is small and it is unlikely that these would have a substantial effect on the cost-effectiveness results. Therefore the simplification made by the manufacturer to model the HAQ scores as independent normal variables is unlikely to have introduced a large inaccuracy in the results.

Estimating the rates of serious adverse events (SAEs) for each intervention

The ERG believes that the meta-analysis of SAEs should be undertaken using a random effects rather than a fixed effects model. Six studies should be a sufficient number of studies with which to estimate the between-study standard deviation using a Bayesian framework. If the estimates of treatment effect are implausibly large, then it should be possible to incorporate more informative prior information to exclude implausible effects.

The choice of whether to use a fixed effect or random effects model should be based on the question being asked (i.e. did the treatment work in these studies or can it work) and prior knowledge about the likelihood or not that each study is estimating a common treatment effect. If it is believed that each study is not estimating a common treatment effect then a fixed effect estimate will not only be less variable than a random effects estimate but will also give a different point estimate when studies are of different sizes.

The manufacturer made the arbitrary decision to incorporate uncertainty associated with discontinuations due to SAEs at +/- 20% of the mean because the credible intervals were considered wide. The ERG believes that a better approach would be to consider the inputs to the model and, if necessary, incorporate more informative prior information to exclude implausible values.

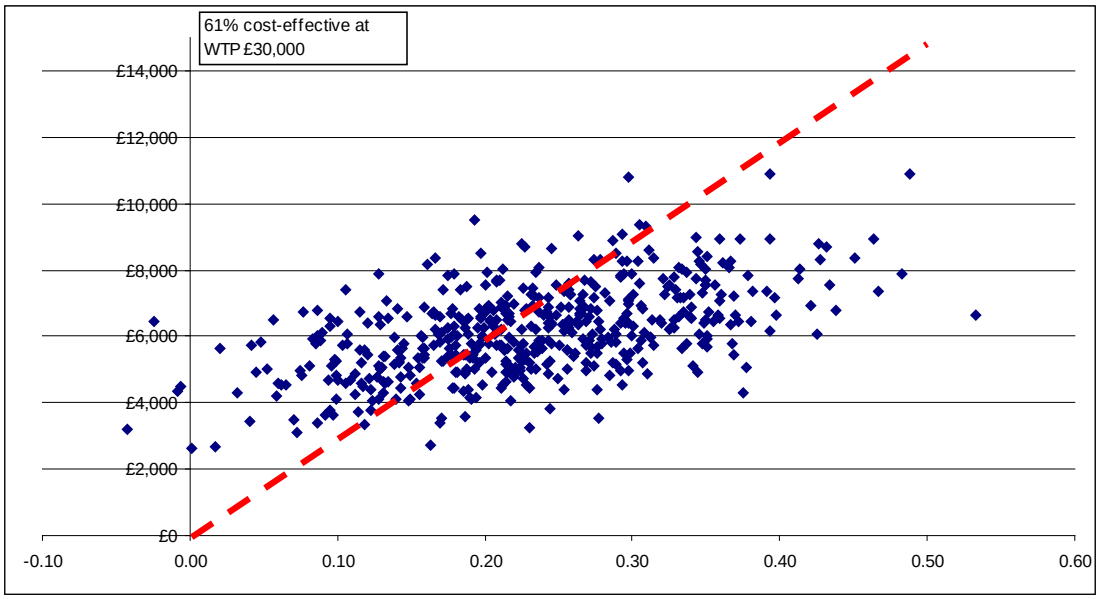
5. Concerns regarding the internal validity of the model.
    a) It is likely that the costs of a nurse training a patient how to administer a subcutaneous injection are strongly correlated. The model assumed that each intervention was sampled independently.
    b) One of the parameters feeding into the eval2disc function is incorrect. For example, in Cell W41 of the 'Model' worksheet the evaldisc2 function the first parameter should be V41 rather than U41. The ERG has amended this error.
    c) The formula used to calculate the costs for biologic DMARDs that are delivered subcutaneously does not round up the dose to an integer number of vials. This will be favourable to such interventions.
    d) The model assumes that patients have an underlying progression in HAQ whilst on conventional DMARDs (0.045 increase in HAQ score per annum). However, this progression is not applied when a patient discontinues a DMARD within 6 months for either lack of efficacy of an adverse event. This will cause some inaccuracy in that, were a conventional DMARD to fail due to lack of efficacy, it would be expected that the HAQ score of the patient would have increased by 0.0225 during this period.
    e) There appears to be an error in the user-defined rxcostdisc function employed in the model as it appears that the cost of the first treatment has been omitted from this calculation. This has been amended by the ERG.
    f) If both the PSA and rndNO flags used within the mathematical model are set to true, then the model does not calculate a valid result as a component of the utility calculation returns a '#Num!' error. It is unclear whether this would also need to be corrected were the manufacturer to correct the logic regarding the PSA that is described later.
    g) As previously detailed, the novel method for adjusting the random number rather than the survival curve adds slight inaccuracy to the predicted time of death (Figure 6). The ERG believes that this error will not have a marked impact on the results.
    h) Inconsistency was noted in the attempted use of probabilistic sensitivity analyses for conventional DMARDs which was incorporated for leflunomide but not for the remaining conventional DMARDs.

6. Concerns regarding the probabilistic analyses.
    a) On inspection of the logic used to perform the PSA, it became apparent that the HAQ score change associated with each treatment was not included within the analyses, with the values erroneously fixed at the midpoint values. This can be seen by inspecting the distributions that should have been used for abatacept and infliximab which are shown in Table 45 and in conjunction with the cost-effectiveness plane reported by the manufacturer comparing the two drugs (replicated in Figure 10).

Table 45: The relative efficacy of abatacept and infliximab when used in conjunction with MTX

Treatment	Adjusted mean HAQ CFB	2.5%CrL	97.5%CrL
Infliximab + MTX	-0.46	-0.62	-0.30
Abatacept + MTX	-0.57	-0.69	-0.43

Figure 10: The cost-effectiveness plane submitted by the manufacturer comparing abatacept with infliximab



Since the relative efficacy of each drug is sampled independently, it would be expected that infliximab would be more efficacious reasonably often as the two confidence intervals overlap. Comparing Monte Carlo samples from the two distributions indicates that this probability is in the region of 14%, ignoring the favourable rates of discontinuation for abatacept due to fewer serious adverse events that cause discontinuation. However, the cost-effectiveness plane submitted by the manufacturer suggests that this probability is very low, and corroborates the opinion of the ERG that changes in the HAQ score were not included in the PSA undertaken by the manufacturer. This error has been corrected by the ERG.

In addition, it is believed that the rates of serious adverse events were not included within the PSA. This has also been amended by the ERG.

5.3 Conclusions

The manufacturer's estimate of cost-effectiveness from the PSA is incorrect due to the omission of key parameters, and will have much smaller confidence intervals than a properly conducted analysis. A number of errors were found within the cost-effectiveness model, which have been detailed previously.

The structural uncertainties undertaken by the manufacturer did not show a large change in the ICER of abatacept compared with conventional DMARDs, with the ICER hovering around £30,000 per QALY gained. The ERG note that no structural uncertainty was undertaken assuming that dose escalation would be required for abatacept.

The change in the ICER was more pronounced for the comparison with infliximab. If an assumption was made that there would be no vial wastage of infliximab, the ICER increases to £57,843; if it is assumed that infliximab is not dose escalated, then the ICER increases to £37,025. The ICER produced when these assumptions were made simultaneously was not reported by the manufacturer.

6 ADDITIONAL WORK UNDERTAKEN BY THE ERG

The ERG re-ran the deterministic analyses comparing abatacept, infliximab and conventional DMARDs to check consistency with the reported results. A discrepancy was found although the cause of this was not known.

The ERG considered the numerous errors within the cost-effectiveness model that have been reported previously. The ones that could be addressed by the ERG within the timescale of the single technology appraisal and that were also deemed by the ERG to potentially have a marked impact on the results have been amended.

Based on the conclusions from the manufacturer's submission, where abatacept was dominated by adalimumab, and given the resources available to the ERG, only evaluations of abatacept followed by conventional DMARDs, infliximab followed by conventional DMARDs and conventional DMARDs will be reported.

The analyses undertaken by the ERG will be detailed in four sections: adjusting the coding of the model to correct for mistakes identified by the ERG; replacing parameter estimates within the model and testing structural uncertainties; additional sensitivity analyses conducted; and errors that were identified but not corrected by the ERG.

Adjusting the coding of the model to correct for mistakes identified by the ERG

1) The model code has been amended so that the PSA now includes the HAQ score change from baseline and the rates of serious adverse events.
2) Errors relating to the cells that form the input parameters for the user-defined eval2disc function have been corrected.
3) The error in the user-defined rxcostdisc function that omitted the cost of the first treatment has been amended by the ERG

Replacing parameter estimates within the model and testing structural uncertainties

1) The ERG believes that the costs associated with RA are better described by the £1,120 per HAQ score unit reported in Malottki *et al.*,[72] as this does not incorporate productivity losses. These costs include the costs of joint replacement which are now set to zero within the model
2) The ERG has removed the assumption that, provided a patient was a responder and did not experience an adverse event, the time of discontinuation for infliximab and abatacept was identical, and instead individually sampled time of discontinuation for abatacept and

infliximab. Whilst this is unlikely to affect the mean results, the uncertainty in the comparison will be greater.

3) The ERG has adapted the model so that, were a patient's HAQ score to increase during the evaluation period of an intervention, then the next line of treatment would begin with the patient's HAQ at this higher score.

4) The manufacturer assumed that the standard deviation associated with baseline HAQ score was equal to the standard deviation in change in baseline HAQ score following treatment. This assumption was not justified. The ERG reduced the value to a standard deviation of 0.3 as at this value 72.5% of patients on etanercept who did not have a serious adverse event would be assumed to respond. This value appeared reasonable to our clinical advisors.

Additional sensitivity analyses conducted

1) The effect on the ICER of increasing the change in HAQ score required to be classed as a responder has been explored. In this analysis, the increase required was changed to 0.5

2) The effect on the ICER for abatacept compared with infliximab of assuming simultaneously that there is no vial wastage for infliximab and that there is no dose escalation required for infliximab has been calculated.

3) The manufacturer used a fixed effects model to estimate the rates of serious adverse events that result in discontinuation within the first six months. A random model was not used as it provided very large uncertainty and the treatments could not be distinguished (Clarification Response A38). It may be the case that the treatments cannot be distinguished, and thus the rates for abatacept and infliximab were set equal at 7% in sensitivity analyses.

Errors that were identified by not corrected by the ERG

1) Note that any errors identified in the model that did not relate to the comparison of abatacept followed by conventional DMARDs, infliximab followed by conventional DMARDs and conventional DMARDs have not been amended.

2) The sequences to be compared have not been altered, and thus the evaluation of infliximab followed by abatacept, or vice versa, compared with the strategies presented by the manufacturer has not been addressed.

3) Potential strategies that involve switching patients at the time of dose escalation have not been evaluated.

4) The utility of patients have been left at values estimated from the HAQ score only and have not been adjusted for patient age.

5) Disutility associated with attending hospital for an infusion (which occurs twice as often for abatacept as infliximab) has not been incorporated.

6) Correlations between shape and scale parameters of weibull distributions and in the utility coefficients have not been added.

7) Costs, disutilities, and rates of occurrences for adverse events by treatment have not been incorporated.

8) The error regarding the underestimation of the HAQ score when a patient reaches a value of 3 (see Figure 7) has not been amended.

9) The discount rates in the excel function have not been changed to 3.44% rather than 3.5%. This is because the costs that do need to be discounted at 3.5% use the same cell in the Excel model, and the impact is likely to be small

10) The meta-analysis of serious adverse events using a random effects model and a more informative prior distribution to exclude implausible results has not been undertaken

11) The omission of underlying HAQ progression in patients on conventional DMARDs that are stopped due to lack of efficacy have not been incorporated.

12) The '#Num!' error within the excel model that occurs when both the PSA and rndNO flags are set to 'on' has not been corrected

13) The inaccuracy in adjusting the random number rather than adjusting the survival curve in calculating the time to death (Figure 6) has not been amended.

14) The assumption that dose escalation is unnecessary for abatacept is a favourable assumption. The ERG has not undertaken a sensitivity analysis to explore how the ICER of abatacept compared with conventional DMARDs would alter were dose escalation to occur as this functionality did not exist in the model.

15) The costs and disutilities associated with adverse events used within the model have been left at zero.

6.1 Impact on the ICER of additional clinical and economic analyses undertaken by the ERG

Verification of the results reported by the manufacturer

The re-run of the model undertaken by the ERG indicated that there was a discrepancy between the results reported by the manufacturer and those provided by the model in terms of the absolute numbers of QALYs accrued. This discrepancy is shown in Table 46.

Table 46: The differences in the results reported by the manufacturer and those produced by the ERG

Source	Treatment	Cost	QALY	ICER vs cDMARD	ICER (inc. analysis)
Manufacturer's Submission	cDMARD	£76,276	4.88	ref	
	Infliximab	£109,419	5.96	£30,693	ED
	Abatacept	£114,548	6.16	£29,916	£29,916
ERG re-run	cDMARD	£76,922	4.73	ref	
	Infliximab	£109,745	5.82	£30,044	ED
	Abatacept	£115,861	6.06	£29,322	£29,322

In order that the results of the sensitivity analyses conducted by the ERG are coherent, the base case results have been set to those produced by the ERG on re-running the model.

Univariate Sensitivity Analyses

The effect of changes made to the manufacturer's model have been explored in a deterministic analyses, each using 8,000 simulated patients. These results are detailed in Table 47. Fuller descriptions of the amendment to the base case are given above.

Table 47: Univariate sensitivity analyses following correction of arithmetic errors

	Treatment	Cost	QALY	ICER vs cDMARD	ICER (inc. analysis)
Manufacturer's Base Case – ERG Re-run	cDMARD	£76,922	4.73	-	
	Infliximab	£109,745	5.82	£30,044	ED
	Abatacept	£115,861	6.06	£29,322	£29,322
Amendment					
Arithmetic Errors corrected	cDMARD	£79,967	4.74	-	
	Infliximab	£112,515	5.82	£30,026	ED
	Abatacept	£119,244	6.06	£29,655	£29,655
Arithmetic Errors corrected and costs from Malottki *et al*., used	cDMARD	£34,475	4.74	-	
	Infliximab	£67,020	5.82	£30,004	ED
	Abatacept	£74,010	6.06	£29,895	£29,895
Arithmetic Errors corrected and time of discontinuation for infliximab and abatacept independently sampled	cDMARD	£79,868	4.74	-	
	Infliximab	£113,061	5.84	£30,328	ED
	Abatacept	£119,099	6.08	£29,402	£29,402
Arithmetic Errors corrected and worsening HAQ score maintained for the next treatment	cDMARD	£85,735	2.54	-	
	Infliximab	£118,034	3.60	£30,315	ED
	Abatacept	£123,948	3.88	£28,464	£28,464
Arithmetic Errors corrected and standard deviation of response to treatment set to 0.3	cDMARD	£81,000	4.49	-	
	Infliximab	£114,768	5.49	£33,628	ED
	Abatacept	£122,088	5.77	£31,969	£26,042
Arithmetic Errors corrected and rate of serious adverse events set equal for abatacept and infliximab	cDMARD	£79,940	4.74	-	
	Infliximab	£112,968	5.82	£30,607	ED
	Abatacept	£118,715	6.02	£30,138	£27,699
Arithmetic Errors corrected and dose escalation for infliximab not assumed	cDMARD	£79,990	4.73	-	
	Infliximab	£109,394	5.80	£27,358	£27,358
	Abatacept	£119,219	6.06	£29,439	£38,113
Arithmetic Errors corrected and vial sharing for infliximab assumed	cDMARD	£79,976	4.74	-	
	Infliximab	£107,589	5.83	£25,276	£25,276
	Abatacept	£119,120	6.04	£29,930	£53,534
Arithmetic Errors corrected and threshold for responders raised to 0.5 HAQ decrease	cDMARD	£80,064	4.73	-	
	Infliximab	£108,607	5.73	£28,699	ED
	Abatacept	£114,811	5.95	£28,642	£32,077
Arithmetic Errors corrected and utility equation from Bansback *et al*., assumed	cDMARD	£79,926	4.74	-	
	Infliximab	£112,082	5.71	£32,833	ED
	Abatacept	£118,820	5.95	£32,077	£32,077

The ICER for abatacept compared with conventional DMARDs is fairly robust ranging from £28,000 - £32,500 dependent on the parameter adjusted. However, the ICER for abatacept compared with

infliximab was shown to be greatly influenced by whether vial sharing for infliximab was assumed and whether dose escalation for infliximab was assumed. In both instances, abatacept and conventional DMARDs no longer extendedly dominated infliximab and the ICER of abatacept compared with infliximab rose to £53,500 and £38,000 respectively.

In order to produce a fuller picture of the uncertainty in the cost-effectiveness of abatacept, the ERG undertook four analyses: an ERG objective analysis; an ERG optimistic analysis; an ERG favourable analysis and an ERG pessimistic analysis. Note that due to errors / lack of functionality identified in the model that has not been corrected by the ERG, these results are only indicative. For example, if abatacept was assumed to require dose escalation then the ICER for abatacept would be less favourable to the intervention; it is unclear in which direction the ICER would move if all of the identified limitations were addressed.

The ERG analyses were defined as
ERG objective analysis: Arithmetic errors corrected; Malottki *et al*., costs used; time of discontinuation for infliximab and abatacept independently sampled; standard deviation of response to treatment set to 0.3; rate of serious adverse events set equal for abatacept and infliximab and dose escalation for infliximab not assumed.

ERG optimistic analysis: ERG objective analyses but the rate of serious adverse events as taken from the manufacturer's submission, the HAQ increase required to be a responder increased to 0.5, and dose escalation assumed for infliximab but not abatacept.

ERG favourable analysis: ERG objective analyses but the rate of serious adverse events as taken from the manufacturer's submission and the HAQ increase required to be a responder increased to 0.5.

ERG pessimistic analysis: ERG objective analyses but vial sharing for infliximab assumed and the utility estimation reported from Bansback *et al*., used.

In addition to these analyses, a further analysis denoted a hybrid analysis was estimated assuming that in 63% of cases infliximab would have no vial wastage as estimated in a previous NICE appraisal.[1] An ICER for this analysis was calculated assuming that the results of the optimistic and pessimistic scenarios were weighted in the ratio of 37:63

The results for the ERG analyses are contained in Table 48.

Table 48: ERG deterministic analyses

Analysis	Treatment	Cost	QALY	ICER vs cDMARD	ICER (inc. analysis)
	cDMARD	£35,545	4.48	-	
	Infliximab	£66,404	5.50	£30,340	£30,340
ERG objective	Abatacept	£76,737	5.76	£32,255	£39,748
	cDMARD	£35,657	4.47	-	
	Infliximab	£66,738	5.51	£30,332	ED
ERG optimistic	Abatacept	£71,499	5.76	£29,661	£29,661
	cDMARD	£35,628	4.48	-	
	Infliximab	£63,604	5.49	£27,615	£27,615
ERG favourable	Abatacept	£73,441	5.76	£29,552	£36,916
	cDMARD	£35,503	4.48	-	
	Infliximab	£61,066	5.37	£28,611	£28,611
ERG pessimistic	Abatacept	£76,525	5.62	£36,045	£63,208
	cDMARD	£35,556	4.48	-	
	Infliximab	£63,016	5.42	£29,294	£29,294
ERG hybrid	Abatacept	£75,322	5.67	£33,519	£49,427

Probabilistic analyses undertaken by the ERG

The analyses undertaken for the deterministic evaluation were also repeated for the probabilistic evaluation, with the ERG having amended the model logic in order that the change in baseline HAQ score and the rate of serious adverse events associated with discontinuation were sampled from the distribution rather than being fixed at the midpoint. Each analysis was undertaken using 1,000 simulated patients and 500 sets of samples from parameter distributions that were indicated by the manufacturer to produce stable results. Each probabilistic run took 10 hours to complete on a computer with dual core 3.16 Gigahertz Intel processors.

The probabilistic results are given in Table 49, with cost-effectiveness planes and cost-effectiveness acceptability curves provided in Figures 11-18. Due to its derivation, cost-effectiveness planes and cost-effectiveness acceptability curves are not presented for the hybrid analyses

Table 49: ERG probabilistic analyses

Analysis	Treatment	Cost	QALY	ICER vs cDMARD	ICER (inc. analysis)
	cDMARD	£35,235	4.41	-	
	Infliximab	£68,087	5.49	£30,423	£30,423
ERG objective	Abatacept	£75,970	5.66	£32,487	£45,299
	cDMARD	£35,410	4.38		
	Infliximab	£65,873	5.32	£32,458	ED
ERG optimistic	Abatacept	£72,783	5.58	£31,328	£31,328
	cDMARD	£35,434	4.39		
	Infliximab	£34,775	5.32	£29,482	£29,482
ERG favourable	Abatacept	£36,187	5.58	£31,282	£37,847
	cDMARD	£35,230	4.41	-	
	Infliximab	£62,544	5.36	£28,620	£28,620
ERG pessimistic	Abatacept	£75,920	5.52	£36,613	£85,209
	cDMARD	£35,297	4.40	-	
	Infliximab	£63,776	5.35	£30,024	£30,024
ERG hybrid	Abatacept	£74,759	5.54	£34,569	£56,896

Comparison of deterministic and probabilistic results

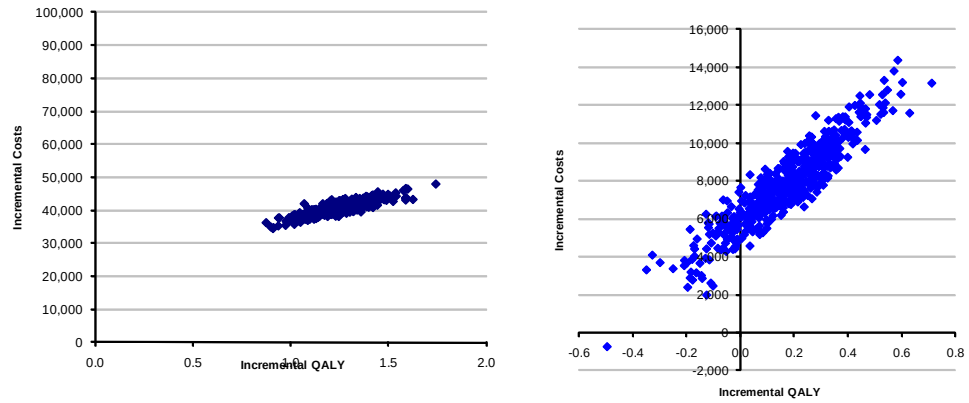
It is seen that the deterministic ICERs are very similar to the probabilistic ICERs for both infliximab and abatacept when compared to conventional DMARDs. The ICER between abatacept and infliximab was more variable and was generally larger in the probabilistic analysis. This is likely to be caused by non-linearities in the model, although the smaller absolute difference in QALYs gained could mean that more patients and sets of parameters needed to be sampled than for the comparison against conventional DMARDs to achieve a stable result.

The analyses undertaken by the ERG indicate that the ICER of abatacept compared with infliximab is in the region of £45,000 although there is uncertainty in this estimate with an optimistic value of approximately £27,000 (and where infliximab is extendedly dominated) and a pessimistic value of approximately £85,000. A hybrid analysis that approximates 63% of patients not having vial wastage estimates an ICER of approximately £57,000.

The analyses undertaken by the ERG indicate that the ICER of abatacept compared with conventional DMARDs is in the region of £32,000 although there is uncertainty in this estimate with an optimistic value of approximately £31,000 and a pessimistic value of approximately £36,500. A hybrid analysis that approximates 63% of patients not having vial wastage estimates an ICER of approximately £34,500.

On the day of submission the ERG noted that the proportion of patients assumed to dose escalate with infliximab used in the model should have been 31% rather than 29%. A deterministic analysis of the optimistic scenario using 29% decreased the ICER of abatacept with conventional DMARDs from £29,661 to £29,550 indicating the extent to which the value of 29% was unfavourable to abatacept. There was insufficient time to re-run the probabilistic analyses, but the impact is expected to be marginal.

Figure 11: The cost-effectiveness planes involving abatacept in the ERG objective analysis



a) abatacept compared with conventional DMARD b) abatacept compared with infliximab

Figure 12: The cost-effectiveness acceptability curve: ERG objective analysis

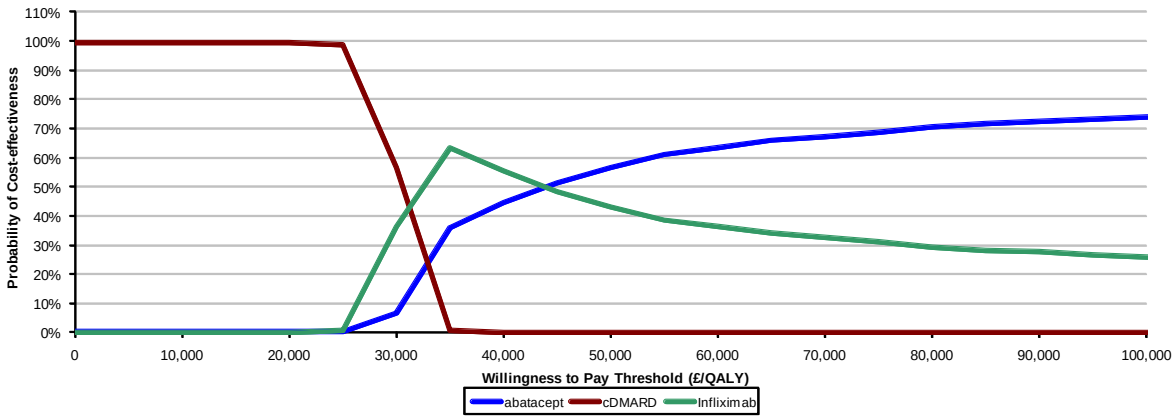
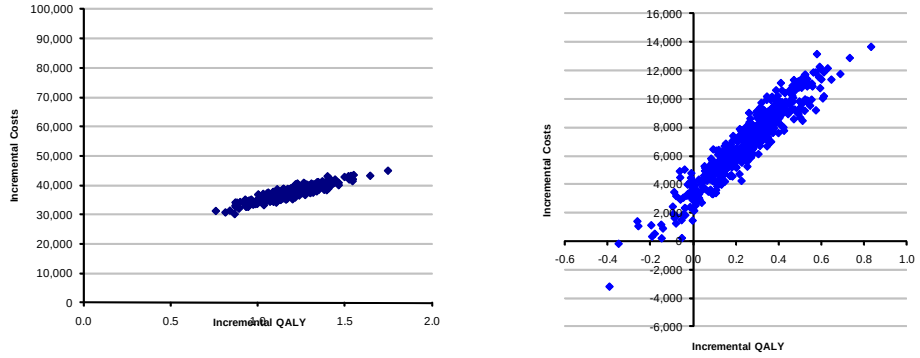


Figure 13: The cost-effectiveness planes involving abatacept in the ERG optimistic analysis



a) abatacept compared with conventional DMARD b) abatacept compared with infliximab

Figure 14: The cost-effectiveness acceptability curve: ERG optimistic analysis

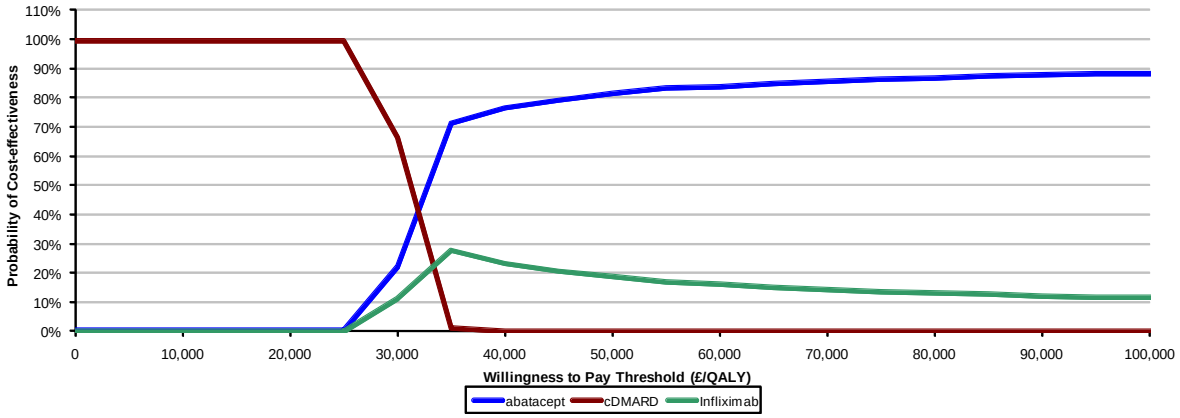
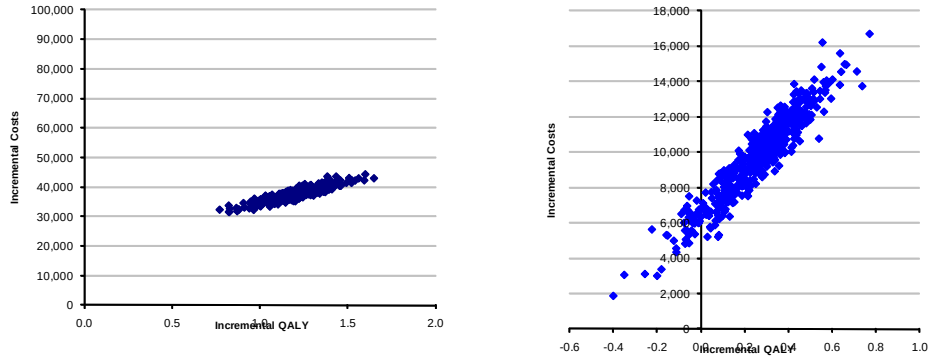


Figure 15: The cost-effectiveness planes involving abatacept in the ERG favourable analysis



a) abatacept compared with conventional DMARD b) abatacept compared with infliximab

Figure 16: The cost-effectiveness acceptability curve: ERG favourable analysis

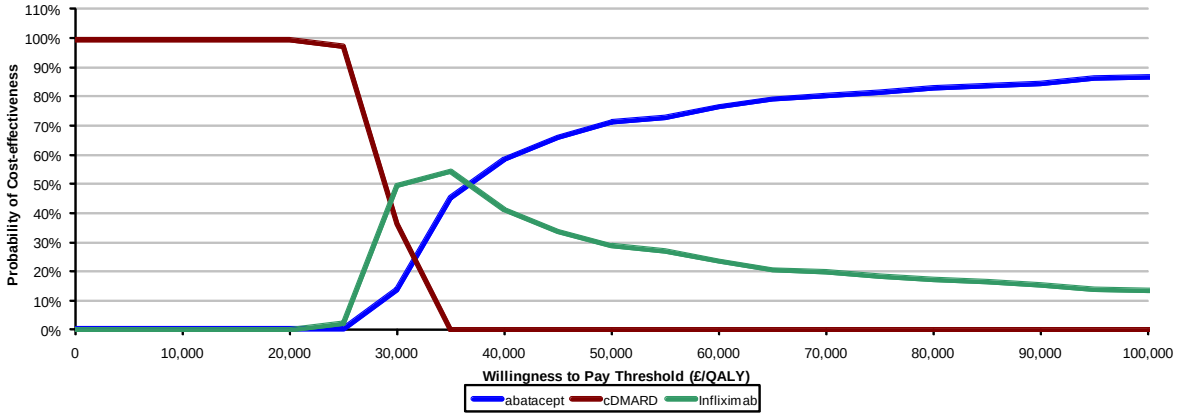
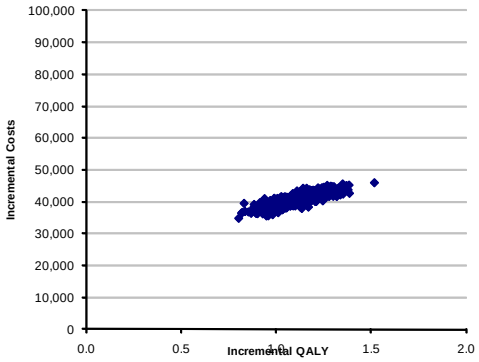
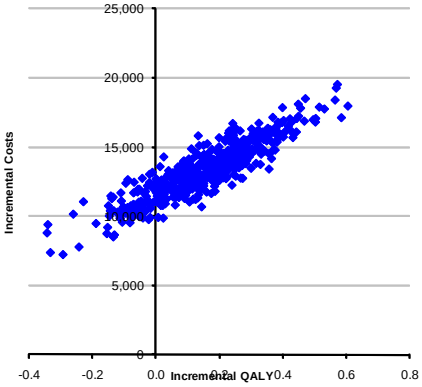


Figure 17: The cost-effectiveness planes involving abatacept in the ERG pessimistic analysis

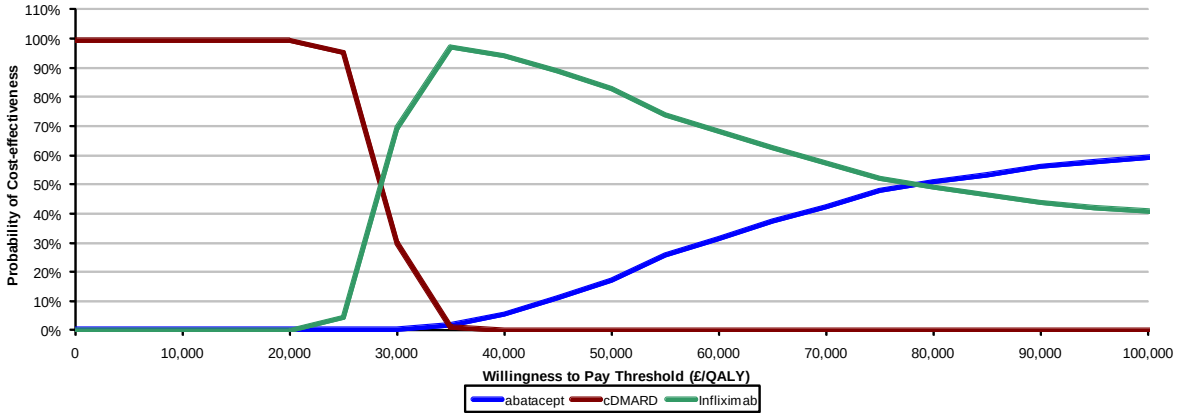


a) abatacept compared with conventional DMARD



b) abatacept compared with infliximab

Figure 18: The cost-effectiveness acceptability curve: ERG pessimistic analysis



7 END OF LIFE

NICE can designate interventions as 'End of Life' medications if they meet a set of criteria which are summarised below.[82]

1 The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;

2 There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment, and;

3 The treatment is licensed or otherwise indicated, for small patient populations.

The ERG has assessed the use of abatacept within RA and does not believe that this intervention would meet the end of life criteria as the first criterion would not be met as, on average, patients with RA are not expected to die within a period of two years. There is also little evidence that abatacept would provide an extension of life compared with infliximab.

8 CONCLUSIONS

8.1 Summary of clinical and cost-effectiveness

The available evidence for clinical effectiveness indicates that, relative to placebo, abatacept, at a dose of, or approximating to, 10 mg/kg, may reduce disease activity in RA, as measured by the DAS28 and ACR responses, at 6 and 12 months. In this context, the most relevant outcome measures would seem to be low disease activity and remission as measured by the DAS28. Relative to placebo, at 12 months abatacept is associated with a relative risk of achieving low disease activity (DAS28 $\leq$3.2) of 3.89 (95% CI 1.13, 13.40; p=0.03), and a relative risk of achieving remission (DAS28 <2.6) of 4.78 (95% CI 2.06, 11.09; p=0.003). Although abatacept also appears to be associated with improved physical function and with less joint damage at one year, the clinical significance of these results is not clear. It may also be associated with improvements at both 6 months and 1 year in pain, morning stiffness, sleep quality, fatigue, and health-related quality of life.

The RCT evidence suggests that, in the short term, abatacept at a dose of, or approximating to, 10 mg/kg is not associated with a higher rate of serious adverse events than placebo, and that its adverse event profile is favourable compared with infliximab. However, the data relating to acute infusional AEs and peri-infusional AEs are incomplete. Longer-term data suggest that the incidence of serious AEs does not increase over time, and no new concerns regarding safety have emerged. However, as this conclusion was based on an analysis in which the mean exposure to abatacept was only 3.56 months, it cannot be regarded as definitive.

The model supplied by the manufacturer had numerous errors the most important of which was the failure of the PSA to incorporate key parameters. The ERG amended some of these errors, and undertook five analyses in order to provide the appraisal committee with additional information in evaluating the cost-effectiveness of abatacept in rheumatoid arthritis after the failure of conventional disease-modifying anti-rheumatic drugs.

Based on previous evaluations of treatments for RA (where in 63% of cases infliximab is vial-shared) the ERG believes that the hybrid analysis would be most pertinent. The hybrid analyses estimate a cost QALY gained of £34,569 when abatacept is compared with placebo and an ICER of £57,896 when abatacept is compared with infliximab although these values would increase if abatacept was associated with dose escalation or if disutilities for more frequent hospital attendance were included.

The optimistic scenario reduces the cost QALY gained to £31,328 when abatacept is compared with placebo and to £27,157 when abatacept is compared with infliximab, although these values are

associated with 0% vial sharing for infliximab, dose escalation for infliximab and not for abatacept and that abatacept had a lesser rate of serious adverse events than infliximab.

The pessimistic scenario increases the a cost QALY gained to £36,613 when abatacept is compared with placebo and to £85,209 when abatacept is compared with infliximab, although these values are associated with 100% vial sharing for infliximab, no dose escalation for either infliximab or abatacept, and that the rates of serious adverse events were equal for infliximab and abatacept.

8.2 Implications for research

A recent abstract by Dougados *et al*., (2010)[45] presents a post-hoc patient-level analysis predicting the likelihood of RA patients achieving a Low Disease Activity State (LDAS) at 1 year following treatment with abatacept in the first 6 months of the AIM trial. This paper also reports that the majority of patients maintained or improved their treatment response or disease status from months 2 to 12, suggesting that patients who had not responded by month 3 may still achieve a clinically meaningful response over time. The sustainability of patient-level responses was also evaluated for the LTE of Kremer Phase 2b (Westhovens *et al*., 2009[52]), revealing that the majority of patients who had achieved LDS, remission or normalised physical function (i.e. HAQ-DI <0.5) by year 1 sustained these outcomes through 5 years.

Schiff *et al*., 2009[26] reported that post-hoc analyses showed that a considerable number of infliximab non-responders (i.e. ACR20 non-responders, or patients with a high disease activity state) who switched to abatacept after 1 year achieved improved clinical responses with abatacept over the second year.

The probability of dose escalation in patients prescribed abatacept, and the increase in dose would need to be assessed through long-term observational studies.

Appendix 1: Adverse events

Table 50: Numbers of patients suffering adverse events (data from relevant publications supplemented where necessary by the manufacturer's submission and supplementary data, highlighted data CIC)

Study	Kremer Phase 2b[28,3,2]			AIM[4]		ATTEST[26]			IM101-119[6]	
	Placebo	Ab 2mg	Ab 10mg	Placebo	Ab 10mg	Placebo	Ab 10mg	Inflix	Placebo	Ab
Number randomised	119	105	115	656		110	156	165	23	27
Adverse events at 6 months (4 months for study IM101-119) (n)	119	105	115	219	433	110	156	165	23	27
Total AEs	[redacted]	[redacted]	[redacted]	84%	87.3%	92 (83.6%)	129 (82.7%)	140 (84.8%)	14 (60.9%)	20 (74.1%)
Total AEs considered to be related to the study drug	[redacted]	[redacted]	[redacted]	47.5%	49.4%	46 (41.8%)	64 (41.0%)	74 (44.8%)	6 (26.1%)	8 (29.6%)
Deaths	0	0	0	NR	NR	0	1 (0.6%)	1 (0.6%)	0	0
Serious AEs	12 (10.1%)	[redacted]	[redacted]	NR	NR	13 (11.8%)	8 (5.1%)	19 (11.5%)	2 (8.7%)	0
Serious infections	0	1 (1.0%)	0	NR	NR	3 (2.7%)	2 (1.3%)	7 (4.2%)		
Malignant neoplasms				NR	NR	1 (0.9%)	1 (0.6%)	2 (1.2%)		
Serious AEs considered to be related to the study drug	1 (0.8%)	4 (3.8%)	0	NR	NR	3 (2.7%)	3 (1.9%)	8 (4.8%)	0	0
Discontinuation due to AE	[redacted]	[redacted]	[redacted]	NR	NR	1 (0.9%)	3 (1.9%)	8 (4.8%)	0	0
Discontinuation due to serious AE	1 (0.8%)	4 (3.8%)	0	NR	NR	0	2 (1.3%)	4 (2.4%)	0	0
Acute infusional AEs	NR	NR	NR	NR	NR	11 (10.0%)	8 (5.1%)	30 (18.2%)	NR	NR
Adverse events at 1 year (n)	119	105	115	219	433	NR	156	165		
Total AEs	[redacted]	[redacted]	[redacted]	184 (84.0%)	378 (87.3%)	NR	139 (89.1%)	154 (93.3%)		
Total AEs considered related to study drug	[redacted]	[redacted]	[redacted]	104 (47.5%)	214 (49.4%)	NR	72 (46.2%)	96 (58.2%)		
Deaths	0	1	0	1 (0.5%)	1 (0.2%)	1	1 (0.6%)	2 (1.2%)		
Serious AEs	19 (16.0%)	19 (18.1%)	14 (12.2%)	26 (11.9%)	65(15.0%)	NR	15 (9.6%)	30 (18.2%)		
Chest pain	0%	0.9%	3.8%	NR	NR	NR	NR	NR		
Myocardial infarction	0%	0.9%	0.8%	NR	NR	NR	NR	NR		
Cardiac disorders				2 (0.9%)	4 (0.9%)	NR	NR	NR		
Gastrointestinal disorder	0%	0%	0.9%			NR	NR	NR		
Musculoskeletal and connective tissue disorders (including RA)	NR	NR	NR	10 (4.6%)	20 (4.6%)	NR	NR	NR		
Infections	NR	NR	NR	5 (2.3%)	17 (3.9%)	NR	NR	NR		
Serious infections	NR	NR	NR	NR	NR	NR	3 (1.9%)	14 (8.5%)		

Nervous system disorders	NR	NR	NR	4 (1.8%)	6 (1.4%)	NR	NR	NR		
Neoplasms (benign, malignant and unspecified)	NR	NR	NR	2 (0.9%)	4 (0.9%)	NR	NR	NR		
Malignant neoplasms	3[1]	NR	4[2]			NR	1 (0.6%)	2 (1.2%)		
Serious AEs considered to be related to the study drug	2 (1.7%)	5 (4.8%)	2 (1.7%)	1 (0.5%)	15 (3.5%)	NR	5 (3.2%)	14 (8.5%)		
Most frequently reported AEs*										
Nervous system disorders	NR	NR	NR	35 (16.0%)	64 (14.8%)	NR	46 (29.5%)	54 (32.7%)		
Headache	18 (15.1%)	17 (16.2%)	17 (14.8%)	26 (11.9%)	76 (17.6%)	NR	23 (14.7%)	32 (19.4%)		
Dizziness	NR	NR	NR	16 (7.3%)	40 (9.2%)	NR	12 (7.7%)	13 (7.9%)		
Infections and infestations	NR	NR	NR	113 (26.1%)	41 (18.7%)	NR	93 (59.6%)	113 (68.5%)		
Nasopharyngitis	11 (9.2%)	19 (18.1%)	17 (14.8%)	25 (11.4%)	66 (15.2%)	NR	20 (12.8%)	26 (15.8%)		
Influenza	NR	NR	NR	12 (5.5%)	31 (7.2%)	NR	13 (8.3%)	11 (6.7%)		
Pharyngitis	NR	NR	NR	10 (4.6%)	26 (6.0%)	NR	12 (7.7%)	17 (10.3%)		
Bronchitis	NR	NR	NR	12 (5.5%)	18 (4.2%)	NR	<5%	<5%		
Upper respiratory tract infection	NR	NR	NR	21 (9.6%)	47 (10.9%)	NR	11 (7.1%)	19 (11.5%)		
Sinusitis	NR	NR	NR	15 (6.8%)	18 (4.2%)	NR	10 (6.4%)	7 (4.2%)		
Gastrointestinal disorders	NR	NR	NR	32 (14.6%)	59 (13.6%)	NR	64 (41.0%)	85 (51.5%)		
Nausea	17 (14.3%)	12 (11.4%)	16 (13.9%)	24 (11.0%)	52 (12.0%)	NR	16 (10.3%)	20 (12.1%)		
Dyspepsia	NR	NR	NR	10 (4.6%)	27 (6.2%)	NR	19 (12.2%)	17 (10.3%)		
Gastritis	NR	NR	NR	<5%	<5%	NR	6 (3.8%)	9 (5.5%)		
Upper abdominal pain	NR	NR	NR	13 (5.9%)	19 (4.4%)	NR	<5%	<5%		
Diarrhoea	NR	NR	NR	21 (9.6%)	47 (10.9%)	NR	21 (13.5%)	21 (12.7%)		
Musculoskeletal and connective tissue disorders	NR	NR	NR	4 (1.8%)	10 (2.3%)	NR	36 (23.1%)	42 (25.5%)		
Back pain	NR	NR	NR	12 (5.5%)	40 (9.2%)	NR	12 (7.7%)	10 (6.1%)		

Arthralgia	NR	16.2%	NR	-	-	NR	NR	NR		
Skin and subcutaneous tissue disorders	NR	NR	NR	14 (6.4%)	34 (7.9%)	NR	28 (17.9%)	50 (30.3%)		
Pruritus	NR	NR	NR	<5%	<5%	NR	5 (3.2%)	10 (6.1%)		
Urticaria	NR	NR	NR	<5%	<5%	NR	3 (1.9%)	11 (6.7%)		
Rash	NR	NR	NR	<5%	<5%	NR	1 (0.6%)	9 (5.5%)		
Respiratory, thoracic and mediastinal disorders	NR	NR	NR	13 (5.9%)	23 (5.3%)	NR	5 (3.2%)	13 (7.9%)		
Cough	15 (12.6%)	NR	NR	13 (5.9%)	29 (6.7%)	NR	<5%	<5%		
General disorders and administration site conditions	NR	NR	NR	25 (11.4%)	43 (9.9%)	NR	25 (16.0%)	36 (21.8%)		
Peripheral oedema	NR	NR	NR	<5%	<5%	NR	8 (5.1%)	6 (3.6%)		
Fatigue	NR	NR	NR	15 (6.8%)	23 (5.3%)	NR	<5%	<5%		
Vascular disorders	NR	NR	NR	4 (1.8%)	19 (4.4%)	NR	23 (14.7%)	37 (22.4%)		
Hypertension	NR	NR	NR	3 (1.4%)	24 (5.5%)	NR	13 (8.3%)	12 (7.3%)		
Hypotension	NR	NR	NR	<5%	<5%	NR	1 (0.6%)	9 (5.5%)		
Psychiatric disorders	NR	NR	NR	3 (1.4%)	8 (1.8%)	NR	19 (12.2%)	23 (13.9%)		
Insomnia	NR	NR	NR	<5%	<5%	NR	5 (3.2%)	12 (7.3%)		
Other	NR	NR	NR			NR				
Urinary tract infection	NR	NR	NR	11 (5.0%)	22 (5.1%)	NR	8 (5.1%)	18 (10.9%)		
Herpes simplex	NR	NR	NR	<5%	<5%	NR	6 (3.8%)	10 (6.1%)		
Gastroenteritis	NR	NR	NR	<5%	<5%	NR	4 (2.6%)	13 (7.9%)		
Most frequently reported AEs considered related to study drug*										
Nasopharyngitis	3.4%	NR	6.1%	NR	NR	NR	NR	NR		
Headache	6.7%	NR	5.2%	NR	NR	NR	NR	NR		
Nausea	5.9%	NR	5.2%	NR	NR	NR	NR	NR		
Acute infusional AEs (ie AEs within 1 hour of the start of the infusion**)	NR	NR	NR	9 (4.1%)	38 (8.8%)	NR	11 (7.1%)	41 (24.8%)		
Peri-infusional AEs (ie AEs within 24 hours of the start of the infusion)	NR	NR	NR	37 (16.9%)	106 (24.5%)	NR	NR	NR		
Discontinuation due to AEs	[redacted]	[redacted]	[redacted]	4 (1.8%)	18 (4.2%)	NR	5 (3.2%)	12 (7.3%)		
Discontinuation due to serious AEs	[redacted]	[redacted]	[redacted]	3 (1.4%)	10 (2.3%)	NR	4 (2.6%)	6 (3.6%)		

* ie AEs seen in at least 5% of patients, and excluding worsening of RA

** 3 hours in the ATTEST study

1: 1 patient with endometrial cancer, 1 with squamous cell carcinoma, 1 with malignant melanoma

2: 1 patient with bladder cancer, 2 with basal cell carcinoma, 1 unspecified neoplasm

9 REFERENCES

1. 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 1. 2010.

2. Lebmeier, M. Re: Single Technology Appraisal: abatacept for the treatment of rheumatoid arthritis after the failure of conventional disease-modifying anti-rheumatic drugs. 2011.

3. Kremer, J. M., Dougados, M., Emery, P., Durez, P., Sibilia, J., Shergy, W., Steinfeld, S., Tindall, E., Becker, J. C., Li, T., Nuamah, I. F., Aranda, R., and Moreland, L. W. Treatment of rheumatoid arthritis with the selective costimulation modulator abatacept: twelve-month results of a phase iib, double-blind, randomized, placebo-controlled trial.[Erratum appears in Arthritis Rheum. 2005 Oct;52(10):3321]. *Arthritis & Rheumatism* 2005; **52** 2263-2271.

4. Kremer, J. M., Genant, H. K., Moreland, L. W., Russell, A. S., Emery, P., Abud-Mendoza, C., Szechinski, J., Li, T., Ge, Z., and Becker, J. C. Effects of abatacept in patients with methotrexate-resistant active rheumatoid arthritis: a randomized trial. *Annals of internal medicine* 2006; **144** 865-876.

5. Emery, P., Kosinski, M., Li, T., Martin, M., Williams, G. R., Becker, J-C., Blaisdell, B., Ware, J. E., Jr., Birbara, C., and Russell, A. S. Treatment of rheumatoid arthritis patients with abatacept and methotrexate significantly improved health-related quality of life. *Journal of Rheumatology* 2006; **33** 681-689.

6. Bristol-Myers Squibb Pharmaceuticals Ltd. Abatacept for the treatment of rheumatoid arthritis after the failure of conventional disease-modifying anti-rheumatic drugs. Manufacturer submission of evidence. 2010.

7. Felson, D. T. Epidemiology of the rheumatoid disease. 2005;1-36. In Koopman, W., and Moreland, L. *Arthritis and allied conditions*. Lippincott, Willilams and Wilkins, Philadelphia.

8. Sokka, T., Abelson, B., and Pincus, T. Mortality in rheumatoid arthritis: 2008 update. *Clinical & Experimental Rheumatology* 2008; **26** S-35-S-61.

9. 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** 793-800.

10. Symmons, D. P. M., Barrett, E. M., Bankhead, C. R., Scott, D. G. I., and Silman, A. J. The incidence of rheumatoid arthritis in the United Kingdom: results from the Norfolk Arthritis Register. *British Journal of Rheumatology* 1994; **33** 735-739.

11. The National Collaborating Centre for Chronic Conditions Rheumatoid arthritis. National clinical guideline for management and treatment in adults. 2009.

12. National Audit Office Services for people with rheumatoid arthritis. 2009.

13. National Audit Office Services for people with rheumatoid arthritis. Methodology. 2009.

14. National Institute for Health and Clinical Excellence Rheumatoid arthritis. The management of rheumatoid arthritis in adults. NICE clinical guideline 79. 2009.

15. Maxwell, L. and Singh, J. A. Abatacept for rheumatoid arthritis. *Cochrane Database of Systematic Reviews* 2009;CD007277.

16. Deighton, C., Hyrich, K., Ding, T., Ledingham, J., Lunt, M., Luqmani, R., and *et al* BSR and BHPR rheumatoid arthritis guidelines on eligibility criteria for the first biological therapy. *Internet* 2010.

17. European Medicines Agency Orencia: EPAR - product information. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000701/WC500048935.pdf 2010.

18. BMJ Group and Pharma Press British national formulary. 2010.

19. European Medicines Agency Summary of opinion (post authorisation). Orencia abatacept. http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion/human/000701/WC500090816.pdf 2010.

20. Kobelt, G., Jonsson, L, Lindgren, P, Young, A., and Eberhardt, K. Modeling the progression of rheumatoid arthritis: a two-country model to estimate costs and consequences of rheumatoid arthritis. *Arthritis & Rheumatism* 2002; **46** 2310-2319.

21. European Medicines Agency Annex 1 Summary of product characteristics. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000165/WC500025821.pdf 2010.

22. European Medicines Agency Assessment report for simponi. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/000992/WC500052372.pdf 2009.

23. Hensor, E. M. A., Emery, P., Bingham, S. J., Conaghan, P. G., and the YEAR Consortium Discrepancies in categorizing rheumatoid arthritis patients by DAS-28(ESR) and DAS-28(CRP): can they be reduced? *Rheumatology* 2010; **49** 1521-1529.

24. Richards, P. and De Wit, M. The OMERACT glossary for patient research partners. *Internet* 2004.

25. van Riel, P. L. DAS-Score.nl. Home of the DAS Disease Activity Score in rheumatoid arthritis. http://www.das-score.nl/www.das-score.nl/ 2010.

26. Schiff, M., Keiserman, M., Codding, C., Songcharoen, S., Berman, A., Nayiager, S., Saldate, C., Li, T., Aranda, R., Becker, J. C., Lin, C., Cornet, P. L., and Dougados, M. Efficacy and safety of abatacept or infliximab vs placebo in ATTEST: a phase III, multi-centre, randomised, double-blind, placebo-controlled study in patients with rheumatoid arthritis and an inadequate response to methotrexate. *Annals of the Rheumatic Diseases* 2008; **67** 1096-1103.

27. Conaghan, P. G., Durez, P., Alten, R., Burmester, G., Tak, P. P., Klareskog, L., Gaillez, C., Le Bars, M., Zhou, X., and Peterfy, C. Impact of abatacept on synovitis and structural damage in methotrexate (MTX)-inadequate responders with active rheumatoid arthritis (RA): a randomized, controlled magnetic resonance imaging (MRAI) exploratory study. 2010.

28. Kremer, J. M., Westhovens, R., Leon, M., Di Giorgio, E., Alten, R., Steinfeld, S., Russell, A., Dougados, M., Emery, P., and Nuamah, I. F. Treatment of rheumatoid arthritis by selective inhibition of T-cell activation with fusion protein CTLA4Ig. *New England Journal of Medicine* 2003; **349** 1907-1915.

29. Bruce, B. and Fries, J. F. The Stanford Health Assessment Questionnaire: dimensions and practical applications. *Health and Quality of Life Outcomes* 2003; **I** 20.

30. Uhlig, T., Haavardsholm, E. A., and Kvien, T. K. Comparison of the Health Assessment Questionnaire (HAQ) and the modified HAD (MHAQ) in patients with rheumatoid arthritis. *Rheumatology* 2006; **45** 454-458.

31. Genant, H. K., Jiang, Y., Peterfy, C., Lu, Y., Redei, J., and Countryman, P. J. Assessment of rheumatoid arthritis using a modified scoring method on digitized and original radiographs. *Arthritis & Rheumatism* 1998; **41** 1583-1590.

32. Khraishi, M., Russell, A., and Olszynski, W. P. Safety profile of abatacept in rheumatoid arthritis: a review. *Clinical Therapeutics* 2010; **32** 1855-1870.

33. Hodgson, R. J., O'Connor, P., and Moots, R. MRI of rheumatoid arthritis - image quantitation for the assessment of disease activity, progression and response to therapy. *Rheumatology* 2008; **47** 13-21.

34. Østergaard, M., Edmonds, J., McQueen, F., Peterfy, C., Lassere, M., Ejbjerg, B., Bird, P., Emery, P., Genant, H., and Conaghan, P. An introduction to the EULAR-OMERACT rheumatoid arthritis MRI reference image atlas. *Annals of the Rheumatic Diseases* 2005; **64** i3-i7.

35. Kalyoncu, U., Dougados, M., Daurès, J-P., and Gossec, L. Reporting patient-reported outcomes in recent trials in rheumatoid arthritis: a systematic literature review. *Annals of the Rheumatic Diseases* 2009; **68** 183-190.

36. Russell, A. S., Wallenstein, G. V., Martin, M. C., Maclean, R., Blaisdell, B., Gajria, K., Cole, J. C., Becker, J-C., and Emery, P. Abatacept improves both the physical and mental health of patients with rheumatoid arthritis who have inadequate response to methotrexate treatment. *Annals of the Rheumatic Diseases* 2007; **66** 189-194.

37. Gossec, L., Dougados, M., Rincheval, N., Balanescu, A., Boumpas, D. T., Canadelo, S., Canadelo, S., and *et al* Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Annals of the Rheumatic Diseases* 2009; **68** 1685.

38. Rand Health Sleep scale from the Medical Outcomes Study. *citation* 1986.

39. Cole, J. C., Dubois, D., and Kosinski, M. Use of patient-reported sleep measures in clinical trials of pain treatment: a literature review and synthesis of current sleep measures and a conceptual model of sleep disturbances in pain. *Clinical Therapeutics* 2007; **29** 2580-2588.

40. Wells, G., Li, T., and Tugwell, P. Investigation into the impact of abatacept on sleep quality in patients with rheumatoid arthritis, and the validity of the MOS-Sleep questionnaire Sleep Disturbance Scale. *Annals of the Rheumatic Diseases* 2010; **69** 1768-1773.

41. 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**

42. Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; **339** 332-339.

43. Westhovens, R., Robles, M., Ximenes, A. C., Nayiager, S., Wollenhaupt, J., Durez, P., Gomez-Reino, J., Grassi, W., Haraoui, B., Shergy, W., Park, S. H., Genant, H., Peterfy, C., Becker, J. C., Covucci, A., Helfrick, R., and Bathon, J. Clinical efficacy and safety of abatacept in methotrexate-naive patients with early rheumatoid arthritis and poor prognostic factors. *Annals of the Rheumatic Diseases* 2009; **68** 1870-1877.

44. Weinblatt, M., Combe, B., Covucci, A., Aranda, R., Becker, J. C., and Keystone, E. Safety of the selective costimulation modulator abatacept in rheumatoid arthritis patients receiving background biologic and nonbiologic disease-modifying antirheumatic drugs: A one-year randomized, placebo-controlled study. *Arthritis & Rheumatism* 2006; **54** 2807-2816.

45. Dougados, M., Kremer, J., Le Bars, M., Poncet, C., Yazici, Y., Emery, P., Westhovens, R., and Schiff, M. Likelihood of achieving LDAS following 12 months of abatacept treatment according to clinical response at months 1-6 in patients with RA and inadequate response to MTX in the AIM trial. *EULAR 2010* 2010.

46. Genovese, M. C., Becker, J. C., Schiff, M., Luggen, M., Sherrer, Y., Kremer, J., Birbara, C., Box, J., Natarajan, K., Nuamah, I., Li, T., Aranda, R., Hagerty, D. T., and Dougados, M. Abatacept for rheumatoid arthritis refractory to tumor necrosis factor alpha inhibition.[Erratum appears in N Engl J Med. 2005 Nov 24;353(21):2311]. *New England Journal of Medicine* 15-9-2005; **353** 1114-1123.

47. Schiff, M., Pritchard, C., Huffstutter, J. E., Rodriguez-Valverde, V., Durez, P., Zhou, X., Li, T., Bahrt, K., Kelly, S., Le, Bars M., and Genovese, M. C. The 6-month safety and efficacy of abatacept in patients with rheumatoid arthritis who underwent a washout after anti-tumour necrosis factor therapy or were directly switched to abatacept: the ARRIVE trial. *Annals of the Rheumatic Diseases* 2009; **68** 1708-1714.

48. Weinblatt, M., Schiff, M., Goldman, A., Kremer, J., Luggen, M., Li, T., Chen, D., and Becker, J. C. Selective costimulation modulation using abatacept in patients with active rheumatoid arthritis while receiving etanercept: a randomised clinical trial. *Annals of the Rheumatic Diseases* 2007; **66** 228-234.

49. Buch, M. H., Boyle, D., Saleem, B., Rosengren, S., Reece, R., Rhodes, L., Radjenovic, A., English, A., Tang, H., Vratsanos, G., O'Connor, P., Firestein, g. S., and Emery, P. Mode of action of abatacept in rheumatoid arthritis patients having failed TNF blockade: a histological, gene expression and dynamic MRI study. *Arthritis & Rheumatism* 2005; **52** S1164.

50. Schiff, M., Keiserman, M., and Codding, C. Two-year efficacy and safety in abatacept-treated patients with RA who received continuous therapy or switched from infliximab to abatacept: the ATTEST trials. *EULAR 2009* 2009;SATO 103.

51. Schiff, M., Keiserman, M. W., Moniz Reed, D., Le Bars, M., Becker, J-C., Zhao, C., and Dougados, M. An increasing proportion of patients achieve a low disease activity state or remission when switched from infliximab to abatacept regardless of initial infliximab treatment response: results from the ATTEST trial. *Arthritis & Rheumatism* 2009; **60** Abstract 1659.

52. Westhovens, R., Kremer, J. M., Moreland, L. W., Emery, P., Russell, A. S., Li, T., Aranda, R., Becker, J. C., Qi, K., and Dougados, M. Safety and efficacy of the selective costimulation modulator abatacept in patients with rheumatoid arthritis receiving background methotrexate: a 5-year extended phase IIB study. *Journal of Rheumatology* 2009; **36** 736-742.

53. Westhovens, R., Kremer, J., Emery, P., Russell, A. S., Li, T., Aranda, R., Becker, J-C., Zhao, C., and Dougados, M. Consistent safety and sustained improvement in disease activity and

treastmetn response over 7 years of abatacept treatment in biologic-naive patients with RA. *EULAR, Copenhagen, 10-13 June 2009* 2009;SAT0108.

54. Genant, H. K., Peterfy, C. G., Westhovens, R., Becker, J-C., Aranda, R., Vratsanos, G., Teng, J., and Kremer, J. M. Abatacept inhibits progression of structural damage in rheumatoid arthritis: results from the long-term extension of the AIM trial. *Annals of the Rheumatic Diseases* 2008; **67** 1084-1089.

55. Kremer, J. M., Genant, H. K., Moreland, L. W., Russell, A. S., Emery, P., Abud-Mendoza, C., Szechinski, J., Li, T., Teng, J., Becker, J. C., and Westhovens, R. Results of a two-year followup study of patients with rheumatoid arthritis who received a combination of abatacept and methotrexate. *Arthritis & Rheumatism* 2008; **58** 953-963.

56. Kremer, J. M., Russell, A. S., Emery, P., and *et al* Long-term safety, efficacy and inhibition of radiographic progression with abatacept treatment in patients with rheumatoid arthritis and an inadequate response to methotrexate: 3-year results from the AIM trial. Unpublished.

57. Genant, H. K., Peterfy, C., Westhovens, R., Becker, J-C., Vratsanos, G., Zhou, X., and Kremer, J. M. Abatacept increases the proportion of patients who remain free of structural damage progression through 5 years in methotrexate inadequate responders with RA. *EULAR, Copenhagen, 10-13 June 2009* 2009;Abstract FR10253.

58. Kremer, J. M., Russell, A. S., Emery, P., Abud-Mendoza, C., Szechinski, J., Becker, J-C., and *et al* Abatacept demonstrates consistent safety and sustained improvements in efficacy through 5 years of treatment in biologic naive patients with rheumatoid arthritis. *EULAR* 2009.

59. Kremer, J., Russell, A. S., Westhovens, R., Teng, J., Rosenblatt, L., and Emery, P. Sustained and clinically meaningful improvements in both day- and night-time aspects of HRQoL are observed with abatacept treatment in patients with rheumatoid arthritis (RA) and previous inadequate response to MTX: 5-year data from the AIM trial. *ACR* 2010;Abstract 1836.

60. Becker, J. C., Westhovens, R., Hochberg, M., Qi, K., Kelly, S., Smitten, A., and Aranda, R. The long-term safety of abatacept in the treatment of rheumatoid arthritis: integrated analyses from the abatacept clinical trial program. *Annals of the Rheumatic Diseases* 2010; **69** 377.

61. Hochberg, M. C., Westhovens, R., Aranda, R., Kelly, S. M., Khan, N., Qi, K., Pappu, R., Delaet, I., Luo, A., Torbeyns, A., Moreland, L. W., Cohen, R. B., Gujrathi, S., and Weinblatt, M. E. Long-term safety of abatacept: integrated analysis of clinical program data of up to 7 years of treatment. *ACR* 2010;S164.

62. Smitten, A., Becker, J. C., Qi, K., Kelly, S., and Aranda, R. Evaluation of the safety profile of abatacept (ABA) by prior therapy in patients with rheumatoid arthritis (RA). *Annals of the Rheumatic Diseases* 2010; **69** 541.

63. Bristol-Myers Squibb Company ABATACEPT. Final Clinical Study Report for Study IM101119. A Phase IIIB Multicenter, Randomized, Double-Blind, Placebo- Controlled Study to Assess Short-Term Changes in Synovitis and Structural Damage Outcomes in Subjects with Active Rheumatoid Arthritis and Inadequate Response to Methotrexate, Treated with Abatacept Versus Placebo on a Background Therapy with Methotrexate. http://ctr.bms.com/pdf//IM101-119%20ST.pdf 2010.

64. US Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research approval package for: application number BLA 125118/000.

Statistical review(s). http://www.accessdata.fda.gov/drugsatfda_docs/nda/2005/125118_S0000_StatR.pdf 2011.

65. Bhandari, M., Busse, J. W., Jackowski, D., Montori, V. M., Schünemann, H., Sprague, S., and *et al* Association between industry funding and statistically significant pro-industry findngs in medical and surgical randomized trials. *Canadian Medical Association Journal* 2004; **170** 477-480.

66. The Cochrane Collaboration Review Manager. The Nordic Cochrane Centre, Copenhagen 2008.

67. Simon, T. A., Smitten, A. L., Franklin, J., Askling, J., Lacaille, D., Wolfe, F., Hochberg, M. C., Qi, K., and Suissa, S. Malignancies in the rheumatoid arthritis abatacept clinical development programme: an epidemiological assessment. *Annals of the Rheumatic Diseases* 2009; **68** 1819-1826.

68. Simon, T. A., Askling, J., Lacaille, D., Franklin, J., Wolfe, F., Covucci, A., Suissa, S., Hochberg, M. C., and Abatacept Epidemiology Study Group. Infections requiring hospitalization in the abatacept clinical development program: an epidemiological assessment. *Arthritis Research & Therapy* 2010; **12** R67.

69. Arthritis Research UK British Society for Rheumatology Biologics Register (BSRBR). http://www.medicine.manchester.ac.uk/musculoskeletal/research/arc/clinicalepidemiology/pharmacoepidemiology/bsrbr/ 2011.

70. Russell, A., Beresniak, A., Bessette, L., Haraoui, B., Rahman, P., Thorne, C., and *et al*. Cost-effectiveness modeling of abatacept versus other biologic agents in DMARDS and anti-TNF inadequate responders for the management of moderate to severe rheumatoid arthritis. *Clin Rheumatol* 2009; **28** 403-412.

71. Vera-Llonch, M., Massarotti, E., Wolfe, F., Shadick, N., Westhovens, R., Sofrygin, O., and *et al*. Cost-effectiveness of abatacept in patients with moderately to severely active rheumatoid arthritis and inadequate response to tumor necrosis factor-(alpha) antagonists. *Journal of Rheumatology* 2008; **35** 1745-1753.

72. Malottki, K., Barton, P., Tsourapas, A., Uthman, A., Liu, Z., Routh, K., Connock, M., Jobanputra, P., Moore, D., Fry-Smith, A., and Chen, Y-F. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor: a systematic review and economic evaluation. http://www.nice.org.uk/nicemedia/live/12135/46677/46677.pdf 2009.

73. Wolfe, F., Mitchell, D. M., Sibley, J. T., and *et al*. The mortality of rheumatoid arthritis. *Arthritis Rheum* 1994; **37** 481-494.

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

75. Bansback, N., Brennan, A., and Ghatnekar, O. Cost effectiveness of adalimumab in the treatment of patients with moderate to severe rheumatoid arthritis in Sweden. *Ann Rheum Dis* 2005; **64** 995-1002.

76. Curtis, L. Unit costs of health and social care. 2009.

77. Blom, M., Kievit, W., Kuper, H. H., Jansen, T. L., Visser, H., den Broeder, A. A., and *et al*. Frequency and effectiveness of dose increase of adalimumab, etanercept, and infliximab in daily clinical practice. *Arthritis Care Res (Hoboken)* 2010; **62** 1335-1341.

78. National Institute for Health and Clinical Excellence Guide to the methods of technology appraisals. 2008.

79. Abatacept for the treatment of rheumatoid arthritis after the failure of conventional disease-modifying anti-rheumatic drugs. Final Scope. *National Institute for Health and Clinical Excellence* 2011.

80. Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley and Sons Ltd 2004.

81. van der Heijde, D. M., Klareskog, L., Singh, A., Tornero, J., Melo, G., Codreanu, C., and *et al*. Patient reported outcomes in a trial of combination therapy with etanercept and methotrexate for rheumatoid arthritis: the TEMPO trial. *Ann Rheum Dis* 2006; **65** 328-334.

82. Appraising life-extending, end of life treatments. *National Institute for Health and Clinical Excellence* 2011.