in collaboration with:

Tocilizumab for the treatment of systemic juvenile idiopathic arthritis

Produced by	Kleijnen Systematic Reviews Ltd. in collaboration with Erasmus University Rotterdam and Maastricht University
Authors	Rob Riemsma, Reviews Manager, KSR Ltd. Maiwenn J. Al, Health Economic Researcher, EUR Stefan K. Lhachimi, Health Economic Researcher, EUR Nigel Armstrong, Health Economist, KSR Ltd. Kate Misso, Information Specialist, KSR Ltd. Nathan Manning, Systematic Reviewer, KSR Ltd. Shona Lang, Systematic Reviewer, KSR Ltd. Johan L. Severens, Professor of Evaluation in Health Care, EUR. Jos Kleijnen, Director, KSR Ltd., Professor of Systematic Reviews in Health Care, Maastricht University.
Correspondence to	Rob Riemsma, Kleijnen Systematic Reviews Unit 6, Escrick Business Park Riccall Road, Escrick York, UK YO19 6FD
Date completed	16/06/2011

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 07/70/01 STA.

Declared competing interests of the authors
None.

Acknowledgements
Dr Mark Wood, Consultant Paediatric Rheumatologist, Paediatric Rheumatology Dept, Leeds General Infirmary.

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:
Riemsma R, Al MJ, Lhachimi SK, Armstrong N, Misso K, Manning N, Lang S, Severens JL, Kleijnen J. Tocilizumab for the treatment of systemic juvenile idiopathic arthritis: a Single Technology Appraisal. York: Kleijnen Systematic Reviews Ltd., 2011.

Contributions of authors
Rob Riemsma 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. Maiwenn Al acted as health economic lead and health economic researcher on this assessment, critiqued the manufacturer's economic evaluation, and contributed to the writing of the report. Stefan K. Lhachimi acted as health economic researcher on this assessment, critiqued the manufacturer's economic evaluation and contributed to the writing of the report. Nigel Armstrong acted as health economist on this assessment, critiqued the manufacturer's economic evaluation and contributed to the writing of the report. Kate Misso critiqued the search methods in the submission and contributed to the writing of the report. Nathan Manning acted as systematic reviewer, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Shona Lang acted as systematic reviewer, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Johan L. Severens critiqued the manufacturer's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the health economic sections of the report. Jos Kleijnen critiqued the manufacturer's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

Abbreviations

AE	Adverse Event
AIC	Akaike Information Criterion
ANK	Anakinra
CAS	Chemical Abstracts Service
CDSR	Cochrane Database of Systematic Reviews
CE	Cost effectiveness
CEA	Cost effectiveness analysis
CHAQ	Childhood Health Assessment Questionnaire
CHAQ-D1	Childhood Health Assessment Questionnaire Disability Index
CHQ	Child Health Questionnaire
CI	Confidence interval
CRD	Centre for Reviews and Dissemination
CRP	C-Reactive Protein
CS	Corticosteroids
DMARDS	Disease Modifying Anti-Rheumatic Drug
ERG	Evidence Review Group
ESR	erythrocyte sedimentation rate
EQ-5D	EuroQol 5 dimensions
HAQ	Health Assessment Questionnaire
Hgb	Haemoglobin
HRG	Healthcare Resource Group
HRQOL	Health Related Quality of Life
HUF	Hungarian Forint (local currency)
HUI	Health Utilities Index
ICER	Incremental Cost Effectiveness Ratio
ILAR	International League of Associations for Rheumatology
INF	Infliximab
IR	Inadequate response
ITT	Intention to treat
IV	Intravenously
JIA	Juvenile idiopathic arthritis
JIA ACR score	Juvenile idiopathic arthritis American College of Rheumatology score
KSR	Kleijnen Systematic Reviews
kg	kilograms
LOM	Loss of movement
LYG	Life Years Gained
mg	milligrams
MAS	Macrophage activation syndrome
MD	Mean Difference
MS	Manufacturer Submission
MTX	Methotrexate
N	Sample Size
N/A	Not applicable
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NR	Not reported
NSAIDS	Non Steroidal Anti-Inflammatory Drug
PAS	Patient Access Scheme
PC	Personal Communication
PSA	Probabilistic sensitivity analysis
PSSRU	Personal Social Services Research Unit
RCT	Randomised Controlled Trial
RA	Rheumatoid Arthritis
RR	Risk Ratio

SAA	Serum Amyloid A
SD	Standard deviation
SE	Standard Error
sJIA	Systemic Juvenile idiopathic arthritis
SMR	Standardized Mortality Ratio
QALY	Quality adjusted life year
STA	Single Technology Assessment
TCZ	Tocilizumab
TNF α	Tumour necrosis factor alpha
UK	United Kingdom
US	United States of America
VAS	Visual analogue scale
WTP	Willingness-to-Pay

Contents

1 SUMMARY

1.1 Scope of the manufacturer submission

The NICE scope for this submission was to assess the clinical effectiveness and cost effectiveness of tocilizumab with or without methotrexate, administered within its license indication, in the treatment of children and young people 2 years and older with systemic Juvenile Idiopathic Arthritis (sJIA) who had not responded adequately to prior NSAID(s) and systemic corticosteroids (population 1) or who had not responded adequately to prior NSAID(s), systemic corticosteroids and methotrexate (population 2). The specified comparators were methotrexate for population 1 and TNF inhibitors (for example, etanercept and infliximab) or anakinra for population 2.

1.2 Summary of clinical effectiveness evidence submitted by the manufacturer

Clearly documented search methods were presented in the submission and clarification response. Absence of comprehensive synonyms and poorly applied study design limits were identified in several of the search strategies, which may have impacted on the recall of the search process. For the most part, the ERG was unable to determine whether any relevant studies were not identified.

The evidence presented by the manufacturer consisted of one small RCT (the TENDER trial) comparing tocilizumab (8 mg/kg, N=37; or 12 mg/kg N=38) with placebo (N=37). Inclusion criteria for the TENDER study population included an inadequate response to previous treatment with NSAIDs and corticosteroids. In TENDER, all patients had evidence of active disease at baseline which was at least 6 months after a definite diagnosis of sJIA. Inadequate response to previous treatment was determined by the treating physician's clinical assessment. A total of 78/112 (70%) patients had been treated with MTX prior to study entry (36 of these entered the study on MTX that had been previously stopped then restarted, 42 of these patients were on their first course of MTX which was ongoing). Twenty-nine patients (26%) had no background MTX at baseline but did receive and stop MTX previously. Five (4%) patients had never received MTX, and could be considered MTX naive.

The ERG has a fundamental problem with the evidence presented in the MS as it is not in accordance with the NICE scope. It is for the Appraisal committee to decide whether it will accept the ERG approach, which means there is no evidence for any comparison in the NICE scope, or accept the MS approach, which means there is some evidence for the second population, but none for the first population.

The main question is: "Which patients in the TENDER trial match which population"? According to the manufacturer 95% of TENDER trial participants match population 2, because "patients are included in the study if they have symptoms of active disease" and "It follows that if patients have tried in the past or are currently administered MTX and continue to have persistent disease then they are inadequate responders" (Response to Clarification Letter, question A2).

The ERG does not agree with this approach. The MS does not provide a clear definition of inadequate responders. It cannot be automatically assumed that all participants in the TENDER trial are inadequate responders to MTX. Because of the lack of information it can only be assumed that the 25% of children in the TENDER trial who stopped using MTX fit this population (population 2). The remaining 75% of children in the TENDER trial should be treated as population 1. Because no data were provided for these two populations, there is no evidence available for any of the comparisons in the NICE scope.

Following the MS approach, no data were provided in the MS for population 1. Therefore, the only comparison left is tocilizumab versus anti-TNFs or anakinra. The manufacturer performed a systematic review to identify trials for the comparators. One trial was identified in children with sJIA, comparing anakinra with placebo. The manufacturer decided to broaden the inclusion criteria to include all trials in juvenile arthritis regardless of subtype, despite advice from their clinical experts to the contrary (see MS, page 116). The ERG agrees with the advice from the clinical experts; therefore, trials in children with other types of juvenile arthritis will be ignored in this report.

In conclusion, following the MS approach, for population 2 (children with sJIA with an inadequate response to NSAIDs, CS and MTX) the MS provided data for an indirect comparison of tocilizumab versus anakinra, using data from the TENDER trial, and a trial of anakinra versus placebo. Strictly speaking, the 5% of participants in the TENDER trial who were MTX naive should be excluded from these analyses. The MS only provided data for all participants in the TENDER trial. However, in response to the clarification letter some data were provided in which MTX naive patients were excluded. These data were not reported for the TENDER trial, but only for the indirect comparison with anakinra. Where possible, the ERG used data for the correct population.

The indirect comparison of tocilizumab versus anakinra shows that ACR30 response favours tocilizumab (RR=2.27, 95% CI: 1.06, 4.85). ACR30 response without fever showed no significant difference between tocilizumab and anakinra.

1.3 Summary of cost effectiveness submitted evidence by the manufacturer

The economic analysis employs a cohort Markov model to evaluate costs and effectiveness of the compared strategies. The model clusters health states into five groups: four groups are different lines of treatment and the fifth group contains 'death' and 'uncontrolled disease'. Each line of treatment consists of five health states: ACR30, 50, 70 and 90 response and "no ACR response". A patient can only move from a particular ACR response in his line to "no ACR response" in the next line or to death. From "no ACR response" the patient can only move to one ACR response level within this line of treatment or to "no ACR response" in the next line. The patient cannot move within a given line to a better or worse health state (say, from ACR 50 to ACR 70). Only after being through all four lines does a patient move to the health state "uncontrolled disease". Transitions to health states are evaluated at 12-week increments (cycles). The model assumes patients start treatment at 2 years and has a time horizon of 16 years.

The probability of a response/non-response within a line of treatment depends on the treatment. The order in which the treatment is applied does not change these transitions. The probability of death is treatment- and health state-independent, whereas the probability of withdrawal is health state independent but higher for MTX while being the same for all other treatment options. All transitions stay constant over time, i.e. are independent of age or disease duration.

The model compares tocilizumab with either MTX or anakinra. For both the intervention and the comparators sequences of four treatments are defined. The transition probabilities for tocilizumab and MTX are directly derived from the TENDER trial, whereas the transition probabilities for anakinra are derived from an indirect comparison. For the transition probabilities of the TNF inhibitors, an indirect comparison was made of tocilizumab versus infliximab. Since the infliximab trial reflects a general JIA population, the indirect comparison results are further adjusted for the differences in the population subtypes. The adjustment factor was derived from long-term efficacy data of patients using etanercept, by assessing the difference in the proportion of responders between the total population

and the systemic JIA patients. The assumption is made that all other TNF inhibitors have the same response rate as infliximab.

In each cycle, the proportion of patients in a given state is calculated. The distribution across states is then used to calculate cycle-specific QALYs and treatment costs.

Each health state leads to an absolute change in the initial CHAQ score. The initial CHAQ score is assumed to be the same for all patients, independent of treatment. Each CHAQ score was mapped onto a utility to allow the calculation of QALYs. This mapping formula was derived in an adult RA population mapping HAQ onto EQ-5D utilities.

The costs depend on the health state for the health-state costs and on the line of treatment for the treatment costs. Resource use for the health state costs was obtained through expert elicitation.
Given the fact that the ERG considers the comparison with MTX invalid (see previous section), the cost effectiveness results for that comparison will not be discussed. The base case result for tocilizumab versus anakinra as derived by the manufacturer is £23,219 per QALY gained.

The ERG considers the ICER presented by the manufacturer biased. This is to a large part related to the problems identified with the indirect comparison which were already mentioned in the previous section; the fact that the indirect comparison with TNF inhibitors was based on one study in a general JIA population leads to biases, even though the manufacturer attempted to correct for this using an adjustment factor.

But also a whole range of other issues identified by the ERG lead to biases in the base case ICER. The main issue was the starting age used in the model. Since the decision problem mentioned children of 2 years and older, the manufacturer used this as the starting age of the model. However, on average, patients will be 7 years before they are eligible for treatment with tocilizumab. It was found that this higher age had a significant impact on the ICER, increasing from £23,000 to £42,500.

Another main issue relates to the model structure. The model is a cohort Markov model assessing different sequences of treatment. The treatment effect is modelled as a relative improvement (ACR response) for each patient. The manufacture claims that this is equal to obtaining an absolute CHAQ score, which is a measure of disease burden that is directly mapped into utilities to calculate QALYs. Considering the wide variation of the initial CHAQ score of patients, the current approach masks the variability in the model outcomes for an individual patient by imposing an undue homogeneity assumption, i.e. only one patient type exists and all patients experience the same absolute utility change from a relative health improvement irrespective of their initial disease manifestation. Thus, the ERG considers the modeling approach not appropriate to inform the decision problem and the outcomes of the model should be interpreted with care.

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

1.4.1 Strengths

Search methods were clearly presented and reported. The manufacturer searched the required databases, with the exception of Medline In-Process for the cost-effectiveness search. Date of searching, date spans and database hosts were, for the most part, well documented. The MS provided sufficient detail for the ERG to appraise the searches.

The industry submission is based on one small high quality randomised controlled trial (TENDER) with a total of 112 participants divided over three treatment arms. Twelve week results show a

consistent significant effect in favour of tocilizumab when compared with placebo. In addition, a systematic review was undertaken with the aim to perform an indirect comparison between tocilizumab and anti-TNFs/ anakinra.

Regarding the cost-effectiveness analyses as submitted by the manufacturer, a strength is the fact that the manufacturer has defined treatment sequences for both intervention and comparator, thus reflecting the use of the intervention in clinical practice.

1.4.2 Weaknesses and areas of uncertainty

All searches would have benefitted from the inclusion of more comprehensive disease and treatment synonyms. Attempts to limit by methodological study design were poorly executed and contained mistakes. Inappropriately applied study design limits were employed in NHS EED, EconLit and CENTRAL.

Although the TENDER trial was a high quality randomised controlled trial, not all statistical analyses were based on Intention-To-Treat (ITT). Many patients, especially in the placebo arm were excluded most likely due to escape medication.

A comparison between tocilizumab and methotrexate was not possible, because no data were presented for population 1 (the methotrexate naïve population). For population 2, the systematic review resulted in one trial (ANAJIS) comparing anakinra with placebo, allowing for an indirect comparison of 12-week data for tocilizumab with 1-month data for anakinra. Both trials (TENDER and ANAJIS) had only two outcomes in common: ACR30 response and ACR30 without fever.

An important weakness of the economic model is that it models the trial, instead of natural disease progression. This is illustrated by the assumption in the model that patients that move to a certain ACR response stay in that state until the patient either withdraws (i.e. moves to the next treatment line) or dies. Given the nature of the disease, this is an unlikely assumption.

The cost estimates for health states, which are a main driver of the cost effectiveness, have been solely defined based on expert opinion. The resulting cost estimates do not seem reasonable, as they present a cost for non-responders (£3,300) that is 6 times higher than the costs for an ACR30 response (£500), whereas an ACR90 response is associated with only a 30% decrease (£350) compared to ACR30. Additionally, due to the wide variation in health status at base line of the patients, patients may be assigned different costs even though at 12 weeks they have the same absolute health status.

The utilities used in the model have been derived using a mapping formula developed in the context of and adult RA population. There is no information available to check the validity of this procedure. Also, after translating the relative ACR response to an absolute (fixed) CHAQ score, and mapping this onto utilities, an additional step is added where for the comparators anakinra and TNF-inhibitors the assumption is made that their ACR responses can be assigned the same utility. It is difficult for the ERG to assess whether this chain of assumptions leads to an over-or underestimation of the cost effectiveness.

Throughout the model, assumptions were made about the statistical uncertainties to be included in the PSA. While some of the assumptions were reasonable, others were not. It is however important to realize that the overall uncertainty about the cost effectiveness of tocilizumab goes far beyond the

statistical uncertainty, as it is related more to fundamental problems in model structure and availability and use of effectiveness evidence.

1.5 Summary of additional work undertaken by the ERG

Based on several shortcomings identified in the manufactures economic evaluation, the ERG defined a new base case analysis with the main changes made:

- The starting age is 7 years, with a time horizon of 11 years.
- A correction of the withdrawal rate (increase)
- Adjust the ACR response probabilities for tocilizumab to reflect the MTX non-responder population (95% of whole populations)\
- Adjust the relative risk of anakinra to reflect the non-MTX-naïve population in the indirect comparison

This ERG defined base case analysis leads to an ICER of £42,552 per QALY gained. As a result of the various changes made the ICER has increased substantially. This is almost entirely explained by the higher starting age.
To assess the uncertainty around this estimate, the ERG performed a PSA. Based on these outcomes, we find that the probability that the ICER is below £20,000 and £30,000 is 5% and 22%, respectively.

Based on this ERG base case, a few additional scenarios were explored.

In the first, we varied the withdrawal probabilities in such a way that high-responders would have a lower probability of withdrawing than low responders. This was implemented by assuming withdrawal of 5% for ACR30, 3.5% for ACR50, 2.7% for ACR70 and 1.5% for ACR90. Note that there is no evidence base for the specific values used; our goal was to use realistic values such that the base case withdrawal risk of 3.13% would be between the ACR50 and ACR70 response. The resulting ICER was £40,916 per QALY gained, slightly lower than the base case ICER.

In the second scenario, we explored the effect of the assumption that after the initial response, patients either stay in their current health state, withdraw and move to next line or die. We assumed that patients would move between all health states with a probability of 10% per transition, that is, patients in the ACR30 state had (per cycle) a 10% chance of moving to ACR50, a 10% chance of moving to ACR70 and a 10% chance of moving to ACR90. We assumed that both improvements and deteriorations would occur. The resulting ICER was £53,051 per QALY gained, 24% higher than the base case ICER. This indicates that the assumption that patients who do not move to the next treatment line stay in the same health state indefinitely is a rather strong assumption.

In the third scenario, the ERG used the starting age of 9.7 that was observed in the TENDER trial data across all patients (Table 8 in the MS) in an analysis; the ICER changes to £46,611 per QALY gained.

Finally, the ERG explored the effect of variation in the treatment sequences, and a slightly lower ICER can be achieved by using anakinra as second line treatment for tocilizumab instead of etanercept. When infliximab is added as a treatment option, sequences with infliximab dominate sequences with etanercept in the same treatment line.

2 BACKGROUND

2.1 Critique of manufacturer's description of underlying health problem.

Does the ERG believe that the manufacturer's description of the underlying health problem is appropriate and relevant to the decision problem under consideration?

The manufacturer's description of the underlying health problem is in line with NICE guidance [1], and hence seems reasonable and relevant to the decision problem. For completeness the following is reproduced from the MS:

"Juvenile idiopathic arthritis (JIA) is a term that covers a heterogeneous group of syndromes in which the onset of inflammatory arthritis occurs before the age of 16 years and lasts for more than 6 weeks. JIA is characterised by persistent joint swelling, pain and limitation of movement. The cause of JIA is poorly understood, but may relate to genetic and environmental factors" (MS, page 20).

"A classification system for JIA has been developed by the International League of Associations for Rheumatology (ILAR). There are seven categories of JIA: systemic, oligo arthritis (formerly pauciarticular), polyarthritis rheumatoid factor positive, polyarthritis rheumatoid factor negative, enthesitis related arthritis, psoriatic arthritis and unclassified (types that do not correspond to any, or to more than one, category) (Petty et al., 2004). The clinical manifestations and severity of the different sub-types varies considerably. sJIA is a multiorgan disease characterised by arthritic symptoms, fever, transient rash, liver and spleen enlargement. It is distinct from other subtypes and is often resistant to treatment. The overall outcome of the disease is poor with a high risk of long-term functional impairment. Macrophage activation syndrome (MAS) is a severe, life threatening complication to sJIA which affects around 7% of children, which is associated with serious morbidity and sometimes death (Yokota et al., 2010)" (MS, page 20).

"JIA is a relatively rare disease, with an estimated incidence in the UK of 0.1 per 1000 children per year, equivalent to 1000 children diagnosed per year. The prevalence is in the order of 1 per 1000 children, and about 10,000 children in the UK are affected. Approximately 10% of children diagnosed with JIA have systemic disease. Of these patients, those who have had an inadequate response to NSAIDs and corticosteroids and are 2 years of age and older will be eligible for Tocilizumab treatment." (MS, page 21).

2.2 Critique of manufacturer's overview of current service provision

Does the ERG believe that the manufacturer's overview of current service provision is appropriate and relevant to the decision problem under consideration?

The ERG broadly agrees with the manufacturer's description of current service provision. The MS states that there are no specific NICE guidance documents or national protocols for the treatment of sJIA and in addition there are no licensed therapies for the treatment of sJIA.

The British Society for Paediatric and Adolescent Rheumatology recommends treatment for JIA within multidisciplinary teams including paediatric rheumatologist, paediatric rheumatology clinical nurse specialist, ophthalmologist, general practitioner, paediatric physiotherapist, paediatric clinical psychologist, paediatric occupational therapist, podiatrist (Davies et al 2010). Drug therapy for sJIA typically begins with systemic corticosteroids to treat systemic symptoms. Later in the disease, the systemic features can be mild / absent and at that stage steroid joint injections are often helpful (and reduce steroid toxicity). Synthetic disease modifying anti-rheumatic drugs (DMARDs), like

methotrexate, sulfasalazine or cyclosporine A, constitute second line therapy. Failure on these treatments may then lead to anti-TNF α therapies, including etanercept, adalimumab, infliximab, anakinra and other biologic DMARDs.[2]

There are several viable therapies not mentioned by the manufacturers, including: anakinra,[3-5] operative treatments, and autologous stem cell transplantation.[6]

3 CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM

Table 3.1: Statement of the decision problem (as presented by the manufacturer)

	Final scope issued by NICE	**Decision problem addressed in the submission**	**Rationale if different from the scope**
Population	1. Children and young people 2 years and older with systemic JIA which has not responded adequately to prior NSAID(s) and systemic corticosteroids. 2. Children and young people 2 years and older with systemic JIA which has not responded adequately to prior NSAID(s), systemic corticosteroids and methotrexate.	Yes Yes	N/A N/A
Intervention	Tocilizumab with or without methotrexate	Yes	N/A
Comparator(s)	1. For children and young people 2 years and older with systemic JIA which has not responded adequately to prior NSAID(s) and systemic corticosteroids: ▪ **methotrexate**	This is addressed to some extent in the submission. The pivotal trial WA18221 (TENDER) (De Benedetti et al. 2010), was designed to compare tocilizumab versus placebo.	The pivotal trial WA18221 (TENDER) (De Benedetti et al. 2010) was a randomised, double-blind, placebo controlled trial. This compared tocilizumab + standard of care versus placebo + standard of care This design is the preferred choice for the demonstration of efficacy because there are no licensed therapies in sJIA and an actively controlled study would be difficult to compare due to ethical issues in this patient population. No studies currently exist directly comparing tocilizumab with TNFα

	2. For children and young people 2 years and older with systemic JIA which has not responded adequately to prior NSAID(s), systemic corticosteroids and methotrexate: ▪ **TNFα inhibitors (for example, etanercept and infliximab)** ▪ **anakinra**	This is addressed to some extent in the submission, but there is very limited appropriate comparison that can be made in the absence of any head-to-head studies involving tocilizumab, and limited RCTs involving other biologics in sJIA.	inhibitors and anakinra, hence no/very limited appropriate comparisons can be made.
Outcomes	Outcomes to be considered include: ▪ disease activity ▪ physical function ▪ joint damage ▪ pain ▪ steroid sparing ▪ mortality ▪ adverse effects of treatment ▪ health-related quality of life	Yes, although joint damage as assessed by radiographic progression is not currently available from the 12 week data from TENDER (De Benedetti et al. 2010) due to the short timeframe. 'Fever' is also an outcome which will be addressed	N/A
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 time horizon for the economic evaluation should reflect the chronic nature of the condition. Costs will be considered from an NHS and Personal Social Services perspective.	Yes	N/A
Subgroups to be considered		N/A	N/A
Special considerations, including issues related to equity or equality		N/A	

3.1 Population

To what extent does the clinical evidence submitted by the manufacturer match the patient population described in the final scope? Where there is a mismatch, provide further details. Does the clinical

evidence submitted by the manufacturer reflect the characteristics of the patient population in England and Wales eligible for treatment? If not, provide further comment.

There was no mismatch between the two populations defined in the scope and the decision problem. However, the included trials had inclusion criteria broader than that defined in the scope. In particular, they were not limited to form the two relevant populations of inadequate response to NSAID(s) and corticosteroids or NSAID(s), corticosteroids and methotrexate. Whilst the inclusion criteria for TENDER would suggest population 1 (inadequate response to NSAIDs and CS) the MS argues that the TENDER trial population is actually equivalent to population 2 (inadequate response to NSAIDs, CS and MTX). According to the MS, "70% of patients (all with prior inadequate response to NSAIDs and corticosteroids) at baseline were still receiving methotrexate, yet had active disease (a further entry requirement for the study), thus could be considered to be failing on methotrexate alone." (p. 39)

At best this inference means TENDER does not address population 1, at worst the inference made about population 2 is unreliable and neither population is addressed.

3.2 Intervention

Does the intervention described in the MS match the intervention described in the final scope? What is the technology and what is its relevant or proposed marketing authorisation/ CE mark?

There was no mismatch between scope and decision problem.
Tocilizumab is a solution for infusion with a dosing frequency of once every two weeks. The recommended posology is 8 mg/kg in patients weighing greater than or equal to 30 kg or 12 mg/kg in patients weighing less than 30 kg.

Tocilizumab is indicated for the treatment of active sJIA in patients 2 years of age and older, who have responded inadequately to previous therapy with NSAIDs and systemic corticosteroids. Tocilizumab can be given as monotherapy (in case of intolerance to MTX or where treatment with MTX is inappropriate) or in combination with MTX. (RoActemra draft Summary of Product Characteristics – Anticipated June 2011).

3.3 Comparators

Do the comparators described in the MS match the comparators described in the final scope? If not, provide further details. Where evidence is limited or not available for relevant comparators has the manufacturer asked an unbiased clinical panel, or carried out its own survey, and do the views elicited agree with what the clinical advisors to the ERG advocate?

There is a mismatch between scope and the decision problem. For population 1 the comparator in the scope is methotrexate, yet the trial the MS relies upon is the TENDER trial which compares tocilizumab plus standard care versus placebo plus standard care. Whilst the comparator in this study was placebo, 70% of patients at baseline were also receiving methotrexate. The MS uses a post-hoc analysis to compare tocilizumab with those patients in the placebo group also receiving methotrexate. However, this is not an acceptable comparison because trial participants were not randomised in this way. Therefore this comparison is based on observational data from two different populations, those using MTX and those not using MTX.

There is also a mismatch between scope and the decision problem for population 2. For those patients with an inadequate response to NSAIDs, CS and methotrexate the comparators should be TNF inhibitors (e.g. etanercept, adalumimab, infliximab) and anakinra. There is no direct evidence presented for these comparators and hence indirect comparisons are made. For tocilizumab evidence of the TENDER trial is used, but instead of using data from those patients described by the manufacturer as relevant to the decision problem (95% of included patients), data for all patients were used in the indirect comparison. For the comparators, the manufacturer decided to broaden the inclusion criteria (see MS, page 116):

> "Due to the dearth of clinical evidence in systemic JIA, Roche augmented the dataset with evidence from a rapid review performed with objective to identify all pivotal trials in juvenile arthritis regardless of subtype."

Despite advise from their clinical experts to the contrary (see MS, page 116):

> "Clinical experts [PC Westhovens R 02/03/2011, Wright S 16/03/2011], stressed the differences between a systemic JIA population and other subtypes and advised against comparing evidence from different populations."

3.4 Outcomes

Do the outcomes in the MS match the outcomes described in the final scope? If not, provide further details. Consider clinical effectiveness, adverse events, quality of life and health economic outcomes and a discussion of appropriate mechanisms for measuring these outcomes. Is the focus of the submission on appropriate outcomes or has it been limited to non-ideal outcomes?

The clinical characteristics of systemic JIA are arthritis and daily fever, plus rash or lymph node enlargement or hepatomegaly or splenomegaly or serositis. However the disease is associated with low health related quality of life, with the disease affecting physical, emotional and social well being.
The outcomes to be considered for the final scope were; disease activity, physical function, joint damage, pain, steroid sparing, mortality, adverse effects of treatment, health-related quality of life.
The primary endpoint presented in the MS for the included TENDER trial was the percentage of responding patients to treatment at week 12. Responders were patients who had an improved JIA ACR30 score at Week 12 and absence of fever (temperatures <37.5°C) in the 7 days preceding the Week 12 assessment day. JIA ACR30 response is defined as three of any six core outcome variables improved by at least 30% from the baseline assessments, with no more than one of the remaining variables worsened by more than 30%. Patients who withdrew, received escape medication, or for whom the endpoint could not be determined were classified as non-responders. The JIA Core Outcome Variables consist of:

1. Physician Global Assessment of Disease Activity (100 mm VAS)
2. Parent/patient global assessment of overall well-being (100 mm VAS)
3. Number of Joints with Active Arthritis
4. Number of Joints with Limitation of Movement
5. ESR (erythrocyte sedimentation rate)
6. Functional Ability (Childhood Health Assessment Questionnaire, measures 8 everyday functional activities)

The secondary outcomes presented in the MS for the TENDER trial were; the individual results for each JIA ACR component at 12 weeks, JIA ACR 50/70/90 responses at 12 weeks, corticosteroid

reduction, fever, rash, pain and laboratory outcomes (CRP levels, anaemia and Hgb levels, thrombocytosis, leucocytosis).

Adverse events presented in the MS for the TENDER trial were: infections, infusion reactions, immunogenicity, neutrophil count, platelet count, hepatic transaminase elevations, and lipid parameters. There were further tables in this section that provided information on AE including death, and macrophage inactivation syndrome (MAS) but not hepatomegaly, splenomegaly, serositis or lymph node enlargement. Although death was reported, it was only reported within adverse events and the 12 week follow-up period is not sufficient to measure this outcome.

From these data the scoping outcomes of disease activity, physical function, pain, adverse events and steroid sparing have been matched by the MS. There appear to be no outcomes for 'joint damage' or 'HRQOL' in the clinical effectiveness. The manufacturer states that joint damage is not currently available. The parent/patient global assessment of overall well-being (100 mm VAS) and CHAQ (functional ability), see page 196 do not measure HRQOL.

Consideration of the clinical characteristics of sJIA would suggest it could be important to consider outcomes that define lymph node enlargement, hepatomegaly, splenomegaly and serositis. MAS was mentioned within the decision problem therefore it would have been advantageous to present this more clearly within adverse events.

Overall the presented clinical effectiveness outcomes did not match the scoped outcomes. There was a general lack of clarity regarding which trial outcomes were being used to match the scope outcomes. There were no appropriate outcomes for joint damage, mortality and HRQOL. In addition, further outcomes may have been useful especially for adverse events relating to sJIA.

3.5 Other relevant factors

For example: Does the MS include a section on equity considerations? Is there an ongoing Patient Access Scheme application?

According to the MS, [redacted] (MS, page 17). No Patient Access Scheme was reported for sJIA.

No equity and equalities issues were mentioned in the MS (MS, page 25).

COMMENT

The ERG has a fundamental problem with the evidence presented in the MS as it is not in accordance with the NICE scope. It is for the Appraisal committee to decide whether it will accept the ERG approach, which means there is no evidence for any comparison in the NICE scope, or accept the MS approach, which means there is some evidence for the second population, but none for the first population. Both approaches and the evidence for each are now discussed.

ERG approach:

In the ERG approach, we adhered to the NICE scope. This means there are two populations:

1: Children with sJIA with an inadequate response to NSAIDs and CS
2: Children with sJIA with an inadequate response to NSAIDs, CS and MTX

For population 1, we have two relevant comparisons:

1A. Tocilizumab alone versus MTX
1B. Tocilizumab plus MTX versus MTX

For population 2, we have one relevant comparison:

2A. Tocilizumab alone versus TNF inhibitors or Anakinra

For comparison 1A, in theory data from the TENDER trial could have been used in an indirect comparison (not directly from trial data as attempted by the manufacturer) using data from TENDER together with data from trials comparing MTX versus placebo. Actually, such a comparison was not possible, because no data were reported for the relevant population in the TENDER trial, children who were MTX naive (only 5% in the TENDER trial), and no data were sought in the MS for MTX versus placebo.

For comparison 1B, in theory data from the TENDER trial could have been used directly. However, despite our request, no data from the TENDER trial were provided for the relevant population: children who used MTX (70% of the TENDER trial participants). The manufacturer ignored our request for separate data because in the MS approach, these children fit population 2 together with children who previously used MTX (i.e. 95% of the TENDER trial participants). Therefore, according to the MS approach the 70% of children in the TENDER trial who used MTX are not a relevant population on its own.

For comparison 2A, data from the TENDER trial could have been used in an indirect comparison, using data from TENDER together with data from trials comparing anti-TNFs or anakinra versus placebo. Such a comparison was not possible, because no data were provided for the relevant population despite our request. The relevant population for this comparison is, according to the ERG approach, children who have previously failed on MTX (25% of participants in the TENDER trial). The manufacturer ignored our request for separate data for the same reason as described above. According to the manufacturer children who have previously failed on MTX are only part of population 2 and therefore not of interest on their own.

In conclusion, according to the ERG interpretation of the scope, no data have been provided for any of the relevant comparisons.

MS approach:

In the MS approach, the population is defined in a different way to the ERG approach. According to the MS approximately 95% of participants are population 2, which was explained as follows (Response to Clarification Letter, Question A2):

> "With regards to the TENDER trial, patients are included in the study if they have symptoms of active disease. It follows that if patients have tried in the past or are currently administered MTX and continue to have persistent disease then they are inadequate responders. Therefore, the majority of patients in the TENDER trial are children and young people with inadequate response to NSAID(s), systemic corticosteroids and MTX. In essence, the study design, as

> presented from the clinical study report, is not stratifying patients based on MTX inadequate response (MTX-IR). Nevertheless, the TENDER study included those patients de facto.
>
> As explained in section 6.2.1 of the MS, there is a small group of patients that are MTX naive (N=5, ~5%), for whom we cannot infer their response to MTX. As stated in the submission, 70% of patients in TENDER were on tocilizumab with MTX and had active disease and these patients can be classified as MTX inadequate responders. The remaining 25% of patients who had MTX in the past, we can assume are also MTX inadequate responders.
> In conclusion, the proportion of participants in TENDER that are inadequate responders to MTX is ~95%."

The ERG does not agree with this approach. The MS does not provide a clear definition of inadequate responders. In a previous NICE submission by Roche (tocilizumab for RA), the manufacturer explained in the response to the clarification letter: "Inadequate response was defined in the clinical trials and in the submission as a response less than ACR20 at 6 months (ie, patients with an ACRn of < 20 at week 24)." No such information is provided for children included in the TENDER trial. It cannot be automatically assumed that all participants in the TENDER trial are inadequate responders to MTX. Because of the lack of information it can only be assumed that the 25% of children in the TENDER trial who stopped using MTX fit this population (population 2).

Following the MS approach, it automatically followed that only ~5% of participants in the TENDER trial fit population 1. Nevertheless, the manufacturer presented a post-hoc analysis to address the comparator for population 1: methotrexate.

> "Population 1 compares to methotrexate and population 2 compares to TNFα inhibitors or anakinra. Whilst the comparator in this study was placebo, 70% of the placebo patients (and tocilizumab arm) at baseline were also receiving methotrexate. Therefore a post-hoc analysis comparing the tocilizumab treatment arm to the 70% of patients in the placebo group receiving methotrexate will be carried out to address the comparator for population 1."

As described before, this analysis is flawed: participants were not randomised based on whether or not they received MTX. Therefore, the manufacturer compared two groups of patients whose heterogeneity will influence the treatment effects. More importantly, following the TENDER inclusion criteria for active disease and the MS approach which claims active disease status despite MTX therapy equates to inadequate response, means the effects of tocilizumab were compared with MTX in a population that was specifically selected as being not responsive to MTX. In addition, the manufacturer previously stated that all children receiving MTX were population 2; and therefore not relevant for this comparison.

In conclusion, following the MS approach no data were provided in the MS for population 1.

Therefore, the only comparison left is tocilizumab versus anti-TNFs or anakinra. The manufacturer performed a systematic review to identify trials for the comparators. One trial was identified in children with sJIA, comparing anakinra with placebo.

As described above (see section 3.3), the manufacturer decided to broaden the inclusion criteria to include all trials in juvenile arthritis regardless of subtype, despite advice from their clinical experts to

the contrary (see MS, page 116). The ERG agrees with the advice from the clinical experts; therefore, trials in children with other types of juvenile arthritis will be ignored in this report.

In conclusion, following the MS approach, for population 2 (children with sJIA with an inadequate response to NSAIDs, CS and MTX) the MS provided data for an indirect comparison of tocilizumab versus anakinra, using data from the TENDER trial, and a trial of anakinra versus placebo. Strictly speaking, the 5% of participants in the TENDER trial who were MTX naive should be excluded from these analyses. The MS only provided data for all participants in the TENDER trial. However, in response to the clarification letter some data were provided in which MTX naive patients were excluded. These data were not reported for the TENDER trial, but only for the indirect comparison with anakinra. Where possible, the ERG will use data for the correct population.

ERG conclusion regarding the decision problem

According to the ERG interpretation of the scope, no data have been provided in the MS for any of the relevant comparisons. Following the MS approach, no data were provided for population 1; for population 2, data from the TENDER trial can be used for an indirect comparison with anakinra.

As far as the ERG is concerned, this should be the end of the ERG report. However, to help the Appraisal Committee interpret the available evidence, we have followed the manufacturer's approach to the decision problem as described above and the evidence presented in the MS is discussed in the remainder of this report.

4 CLINICAL EFFECTIVENESS

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

The MS reports a systematic review of the comparative clinical effectiveness of tocilizumab, etanercept, anakinra, adalimumab, and infliximab for the treatment of systemic juvenile idiopathic arthritis. The MS also reports a rapid review of juvenile arthritis undertaken to augment the very limited results of the systematic review.

4.1.1 State objective of systematic review. Provide description of manufacturers search strategy and comment on whether the search strategy was appropriate. If the manufacturer did not perform a systematic review, was this appropriate?

List databases and other sources of information including unpublished sources, describe any restrictions.

Overall the searches were well-documented. A detailed commentary on the individual search strategies is provided below. There were some potential weaknesses in the strategies provided. These problems may have led to potentially studies being missed.

Searches were carried out on all the required databases, Medline, Medline In-Process and Embase, using the Datastar Web search interface to identify clinical studies on the use of tocilizumab for the treatment of systemic juvenile idiopathic arthritis (sJIA). The Cochrane Library was also searched. In addition, abstracts of conference proceedings and Roche's in-house databases were also checked. Handsearching of reference lists and publication e-alert was undertaken. Searches were conducted on 15 March 2011. The MS presented full search strategies for all the required databases. Additional details of the conference abstract searches were provided in the clarification response.[7, 8]

An amended version of the PRESS evidence-based checklist[9] was used to inform the assessment of the quality of search strategies presented in the manufacturer's submission.[10] The submission was checked against the Single Technology Appraisal (STA) specification for manufacturer/sponsor submission of evidence.[1]

4.1.1.1 Search for clinical evidence

Search strategy for section 5.1, clinical evidence

All the required databases were searched for clinical effectiveness studies. In addition Biosis Previews, Biosis Previews Update, Embase Alert and Roche's internal databases were also searched. Supplementary handsearching of reference lists and publication e-alert was combined with an internet search of the EULAR and ACR conference abstracts. Medline, Embase and Biosis Previews were searched via the Datastar host and limited from 1993 to date searched. The database provider for Cochrane Library access was not given; from the search strategy it appeared that the database was accessed via Wiley. Date spans, dates of searching and search strategies were reported for the majority of resources. Complete information about the Cochrane Library and conference abstract searching was provided in the clarification response.[7, 8] The ERG was unable to replicate the manufacturer's Datastar searches due to lack of access to that database provider. Exploratory searches were undertaken by the ERG using the OvidSP host.

For the databases searched via Datastar, all searches consisted of a Population facet (juvenile idiopathic arthritis) and an Intervention facet (Tocilizumab). Once combined, the searches were limited by the term 'systemic'. Although basic in structure, the facets were combined using appropriate logic.

All the Datastar searches would have benefitted from inclusion of more comprehensive text terms and synonyms for both the population and intervention facet; for instance the strategies could have included sJIA as an abbreviation for 'systemic juvenile idiopathic arthritis'. The search could have included several additional intervention terms, such as Roactemra, Atlizumab, Actemra, r-1569, r1569 and the CAS Registry Number (375823-41-9).

Despite searching all the Datastar resources in one continuous session, there was no consistency between the search strategies for each database. Embase and Embase Alert were searched with different strategies, as were Medline and Medline In-Process, and Biosis Previews and Biosis Previews Update. Only the Medline search strategy incorporated a term for Interleukin 6 Receptors. If 'Interleukin 6 receptors' was a relevant synonym for the intervention facet, this term should have been reproduced consistently in all the search strategies, including the Cochrane Library. Furthermore, additional synonyms for interleukin 6 could have been included to optimise recall e.g. il-6 or anti-IL-6 or anti-interleukin-6. The Embase Alert search consisted of a single keyword term for Tocilizumab; incorporation of additional synonyms expanded to all fields would have retrieved more records. The Biosis Previews Update search was restricted to a single juvenile idiopathic arthritis term; once again, expansion to include synonyms searching all fields could have picked up more records. Line 41 of the Datastar strategy combined the final sets for all the databases. Curiously, this line also contained several redundant lines which were already included in subsequent combination lines. Although inconsistent, this would not have affected the recall of the strategy.

Although the Medline, Embase and Biosis Previews searches combined juvenile idiopathic arthritis, systemic and Tocilizumab, the Cochrane strategy differed in that it did not include the 'systemic' limit but did include the free text term '***interleukin 6'***. The Cochrane search was fit for purpose with no major omissions or errors.

The manufacturer undertook internet searches of conference proceedings for EULAR and ACR. The MS reported the date range for the conference abstract searches as 2005-2010. Following the ERG's clarification letter, the manufacturer presented further detail regarding the conference abstract searches, including the website address for the internet searches and the search terms employed.[7] There appeared to be a discrepancy in the date span of these searches with the MS reporting the date range as 2005-2010, but the detailed methods described the ACR search as 2006-2010. A total number of conference abstracts retrieved for both sources was presented in the MS. The ERG considered these conference abstract searches fit for purpose and a useful addition the clinical effectiveness searches, as searches to identify unpublished and ongoing studies, and conference abstracts can aid retrieval of relevant references.[11-13]

The MS did not include details of the search terms used to search Roche's internal databases, therefore the ERG was unable to comment on these searches.

Search strategy for section 5.8, Indirect and mixed treatment comparisons

The MS reported searches of all the required databases: Medline, Medline In-Process, Embase and the Cochrane Library. Medline, In-Process and Embase searches were undertaken using the Datastar host, appropriate date spans, the date of searching and the full search strategies were reported in the MS.[10] Details of the Cochrane Library strategy were absent from the MS, and were requested by the ERG as part of the clarification process.[10] The clarification response[7] from the manufacturer stated the original Cochrane search had not been recorded on 28.3.11. The manufacturer ran additional searches of the Cochrane Library on 13.5.11 in order to provide that search strategy in the clarification response.[8]

The Medline, In-Process and Embase searches were presented as individual Datastar strategies which were clearly structured into population and intervention facets with the addition of a study design filter. The comparator interventions were identified as etanercept, anakinra, adalimumab and infliximab. Methotrexate (MTX) was not included in the indirect comparison searches.

The ERG considered methotrexate an important comparator intervention and was concerned that searches were not undertaken for studies of MTX in sJIA to allow for an indirect comparison of tocilizumab with MTX, therefore the ERG carried out additional searches which are described later in this section.

The same intervention search terms were applied to all searches, and consisted of the comparator drug's generic name in combination with the brand name, limited to the title and abstract fields. In order to make the searches more sensitive, additional synonyms, brand names and the CAS registry number could have been included, together with thesaurus index terms where available. As with the clinical effectiveness searches, all strategies would have benefit from the inclusion of more comprehensive synonyms for the intervention and population.

The Embase search strategy was presented first in the MS, and contained comprehensive variations of the disease terms. The first line of the search contained a potential typographical error which did not appear relevant to the topic i.e. ***(juvenile adj arthritis adj c adj '12').ab.*** This error would have impaired retrieval of records with 'juvenile arthritis' in abstract. The Embase search incorporated an RCT search filter and attempted to remove references to books, conference papers, editorials, letters and reviews from the retrieved results. The ERG noted a few areas of weakness in the RCT filter, the most important of which being the inclusion of '***retracted article'*** as a synonym for randomised controlled trial. The ERG was unclear why this term was included, as it did not appear to relate to any aspect of controlled trials or randomisation. The Embase RCT filter employed appeared to be a pragmatic collection of terms, limited solely to the title and abstract fields. The ERG felt that an objectively derived filter which incorporated relevant Emtree terms would have increased the sensitivity and relevance of the search results. Line 15 of the RCT filter attempted to remove various publication types combined Emtree terms from the results, by means of the Boolean operator 'NOT'. Unfortunately this attempt was not entirely successful as it appeared that line 15 was intended to search the Emtree ***Exp randomised controlled trial***, for example:

(book or conference adj paper or editorial or letter or review).p.t. not (exp adj randomised or randomized) adj controlled adj trial

OvidSP syntax was used, which failed to work in Datastar and resulted incorrect parentheses. The correct Datastar syntax should have applied, e.g.

(book or editorial or letter or review).pt. not Randomized-Controlled-Trial.DE.

The incorrect syntax was repeated in line 16. Both lines 15 and 16 would have impacted on the effectiveness of the RCT filter and impaired recall of relevant randomised controlled trials.

The Medline search strategy included appropriate MeSH and free text terms for the population facet, and these lines were combined using appropriate Boolean logic.

The Medline RCT filter appeared similar to the Embase filter, and included many of the same irregularities and mistakes. As before, ***'retraction of publication'*** and ***'retracted publication'*** were employed as alternate terms to pick up randomised controlled trials. The ERG was unclear why the manufacturer felt this was appropriate, as no explanation accompanied the strategy. The RCT filter included several terms limited to the publication type field which were not Medline publication types (***'journal correspondence'*** and ***'randomised'***). As this field contains controlled vocabulary, these terms would not have retrieved any records, and would not add anything relevant to the results.

The intervention facet lacked CAS registry numbers, additional synonyms, appropriate MeSH, and as before, MTX was omitted from the search strategy. The ERG concluded that a more comprehensive, objectively-derived RCT filter and the inclusion of MTX would have aided retrieval of relevant comparator studies.

The Medline In-Process strategy replicated the Medline search and included all the same problems.
The Cochrane Library search strategy was supplied in the clarification response. The strategy was structured into population and intervention facets, in combination with an RCT filter. As previously, MTX was not included as a comparator intervention. As the Cochrane Library resource is comprised of several databases of specific study designs, the inclusion of an RCT was incorrect and detrimental. In order to limit a subject search to trials, the searcher can simply view the results of Central (Cochrane Central Register of Controlled Trials), a database made up solely of reports of controlled trials. Application of any RCT filter would impair recall of results from the Cochrane Library. Furthermore the flawed RCT filter employed in the Medline search was used in Cochrane with the same errors, which would have exacerbated this misjudgement.

In response to the lack of evidence identified by the MS searching, the manufacturer carried out an additional rapid review. It appears the rapid review involved handsearching the reference lists of published systematic reviews to identify additional trials. Checking the reference lists of relevant systematic reviews and included studies in combination with searches of electronic searches is recommended for comprehensive information retrieval.[14, 15]

ERG Indirect and mixed treatment comparison searches

As the MS did not include methotrexate as a comparator invention, the ERG undertook additional searches for the relevant indirect and mixed treatment comparisons, which include adalimumab, etanercept, infliximab, anakinra, abatacept and methotrexate. The searches were limited with objectively-derived RCT search filters appropriate for each database.[16, 17] Searches were undertaken between 24.5.11-26.5.11, and were limited to omit animals studies, and incorporated suitable terms for the population and comparator interventions, including generic and product names. In order to optimise sensitivity, a combination of database fields were searched, including the CAS registry number. The ERG searched Medline, Medline In-Process and Daily Update, Embase, CDSR, Central, Clinicaltrials.gov, mRCT, WHO ICTRP and EUCTR, which retrieved 2007 references. Full ERG search strategies are presented in Appendix 1. Two ERG reviewers independently screened the

additional search results and identified 1 relevant study which met the inclusion criteria (Woo, 2000[18]). This study will be discussed further in section 4.2.6.

Search strategy for section 5.9, Non-RCT evidence

The MS stated that the only non-RCT evidence used was the TENDER trial (page 35[10]). On page 352 in the Search Strategy Section for 5.8 (Non-RCT evidence), the MS stated that "Non-RCT evidence was searched for in-parallel with the randomised clinical trial evidence outlined in appendix 9.2". As the information for the TENDER trial forms part of a submitted 2011 EULAR abstract, the ERG assumed it was identified through searches of Roche's in-house databases. The searches for clinical effectiveness (Appendix 9.2) were not limited by study design, therefore non-RCT references would have been retrieved. A full commentary and critique of the search strategies can be found in 4.1.1.1 Search for clinical evidence.

Search strategy for section 5.10 Adverse Events

All the required databases were searched for adverse events studies. In addition Biosis Previews, Biosis Previews Update, Embase Alert and Roche's internal databases were also searched. Supplementary handsearching of reference lists and publication e-alert was combined with an internet search of the EULAR and ACR conference abstracts. Medline, Embase and Biosis Previews were searched via the Datastar host and limited from 1993 to date searched. The database provider for Cochrane Library access was not given; from the search strategy it appeared that the database was accessed via Wiley. Date spans, dates of searching and search strategies were reported for the majority of resources. Complete information about the Cochrane Library and conference abstract searching was provided in the clarification response.[7, 8] The ERG was unable to replicate the manufacturer's Datastar searches due to lack of access to that database provider. Exploratory searches were undertaken by the ERG using the OvidSP host.

For the databases searched via Datastar, all searches consisted of a Population facet (juvenile idiopathic arthritis), adverse events facet and an Intervention facet (Tocilizumab, plus comparators). Methotrexate was included in the intervention facet, along with etanercept, adalimumab, infliximab, anakinra and interleukin 6. Once combined, the searches were limited by the term 'systemic'. Although basic in structure, the facets were combined using appropriate logic.

There appears to be several redundant lines which are omitted from the final combination set for the Embase search. Lines 1, 10, 12, 14, 16, 18 and 20 were not incorporated into the final search results. This resulted in reduced sensitivity for both the population and intervention facets. Furthermore the intervention facet was made up of Emtree terms limited to major focus; this may have meant that references discussing multiple comparators might not have been identified. The ERG considered the Embase strategy unnecessarily restrictive for this reason. Although free-text searches for the population and interventions were undertaken, only the EMTREE index terms were combined in the final results. The CRD Guidance [19] stresses the importance of using a combination of text-terms and subject indexing to ensure comprehensive recall of references. Relying solely on subject indexing terms can impair recall and may miss relevant references which would have been retrieved by the inclusion of synonyms and variations in product names.

The adverse events facet includes a combination a free-text terms and an Emtree term. The Embase strategy would have benefited from incorporation of specified adverse effects terms, and additional

text words for 'adverse events', as recommended by the CRD Guidance [19] and Golder.[20] The ERG would suggest using a more comprehensive approach, including text terms for known adverse reactions, such as infection, lymphoma, leukopenia, neutropenia, drug-induced lupus, blood disorders, abdominal pain, fatigue, fever or ulcerative stomatitis. It would also be beneficial to remove the restriction to major focus which was applied to the Emtree term.

The Medline search resembled the Embase strategy in structure and content, therefore the ERG felt many of the same recommendations for improvement of the Embase search were applicable for Medline. The MeSH terms were again restricted to major focus, and the population and intervention terms lacked synonyms. As before, the Medline search contained redundant lines (lines 25 and 38). The adverse events facet replicated the Embase terms, therefore the Medline strategy would have benefited from incorporation of specified adverse effects terms, floating subheadings and additional text words for 'adverse events', as recommended by the CRD Guidance [19] and Golder.[20]

The Biosis Previews followed the same basic search strategy and reproduced the same weakness with redundant lines and absence of drug synonyms. The Embase Alert, Biosis Previews Update and Medline In-Process searches were much simplified; surprisingly Embase Alert and Biosis Previews Update were only searched for sJIA references, rather than adverse effects of interventions.

The Cochrane Library search investigated drug toxicity in sJIA, with no additional terms for adverse events, side effects or named AEs, such as infection or neutropenia. The Cochrane Library interface enables searching with the same complexity as Medline, requiring the searcher to make very minor syntactical adaptations to an existing Medline strategy. Therefore the manufacturer could have carried out a detailed search of the Cochrane with multiple synonyms for all three facets: population, intervention/comparators and adverse events. The ERG did not consider the AE Cochrane strategy adequate.

In addition, abstracts of conference proceedings and Roche's in-house databases were also checked. Handsearching of reference lists and publication e-alerts was undertaken.

4.1.1.2 Search strategy for cost-effectiveness

Search strategy for section 6.1, cost-effectiveness

Most of the required databases were searched for cost-effectiveness studies. In addition HEED (Health Economic Evaluations Database) was also searched. Although the MS reported that Medline In-Process was searched, no strategies were provided therefore the ERG was unable to comment on this search.

Medline and Embase were searched via the Datastar host and limited from inception to date searched. NHS EED was searched via the Cochrane Library (Wiley). The EconLit strategy provided in the clarification response[7, 8] gave the host as EBSCO, and HEED was accessed directly. Date spans, dates of searching and search strategies were reported for the majority of resources. Complete information about the EconLIT, HEED and NHS EED searching was provided in the clarification response.[7, 8] The ERG was unable to replicate the manufacturer's Datastar searches due to lack of access to that database provider. Exploratory searches were undertaken by the ERG using the OvidSP host.

For the databases searched via Datastar, all searches consisted of a Population facet (juvenile idiopathic arthritis) and a study design filter (cost-effectiveness). Once combined, the searches were limited to human studies, and to remove letters, editorials and note publication types. Although basic in structure, the facets were combined using appropriate logic.

The Population facet for the Datastar searches was different from that used in all previous Datastar searches, and included much broader population terms, such as paediatric or juvenile arthritis. Both the Medline and Embase searches employed similar cost effectiveness study design filters, which included terms for cost, economics, health status, quality of life and burden of illness.

For both searches the human limit was incorrect applied. Correct syntax for the Embase search line 38 should have read:
38 AND HUMAN=YES
Correct syntax for the Medline search line 36 should have read:
36 AND HUMAN=YES

Although both the Medline and Embase cost-effectiveness filters included a comprehensive combination of free-text and subject indexing terms, there were some issues with the way that Emtree/MeSH terms were included. For example, explosion of the Emtree term 'health economics' (line 14) in the Embase strategy, would also retrieve references indexed with narrower terms, such as 'economic evaluation' and 'cost-effectiveness analysis'. Although application of this level of explosion meant that the narrower terms did not also need including individually, many of the narrower terms were repeated later within the strategy. Although unnecessary, this would not have affected recall of the search strategies. The MeSH index term *Cost of illness/* was repeated twice in the Medline cost-effectiveness filter, while erroneous this would not have affected recall. The ERG was unable to replicate the Datastar searches due to lack of access to this search provider, however the Medline and Embase searches were considered adequate and acceptable recall.

The MS reported that the NHS EED and HEED databases were searched for cost-effectiveness information. Both strategies were absent from the MS, and were requested by the ERG as part of the clarification process[10]. The clarification response [7] from the manufacturer stated the original HEED & NHS EED searches had not been recorded on 3.12.10. The manufacturer ran additional searches of HEED & NHS EED on 14.5.11 and provided that search strategies in the clarification response.[8]

The full NHS EED strategy was provided in the clarification response, and showed that NHS EED was searched via the Cochrane Library. As with the Datastar searches, the population terms were broad and included juvenile arthritis. Although the NHS EED search was limited to NHS EED records, a cost-effectiveness filter was also applied to limit the search strategy. The ERG felt that the inclusion of an economics filter was incorrect and detrimental. Restriction of the results to NHS EED was adequate but the addition of the filter would impair recall of results from the NHS EED.

The HEED search involved a single population facet, combining arthritis terms, restricted to child*. The ERG felt the HEED strategy was clearly presented and appropriate.

Although EconLit is a required database[1], the database was not searched for the MS. Following a clarification query by the ERG, the manufacturer's clarification response presented details of an EconLit search carried out on 15.5.11. The EconLit strategy was structured into a population facet and a cost-effectiveness facet. The population terms resembled those used for Medline and NHS EED, and

included broader terms, such as juvenile arthritis. The search also incorporated a long and comprehensive economics filter, which include HRQL terminology. Whilst thorough in the inclusion of economics synonyms, this facet was unnecessary when searching an economics database. The population component of the EconLit strategy retrieved a total of 34 references. It would have been quicker and safer to take the cautious approach of screening all 34 records. Following application of the economics filter, the EconLit search retrieved 23 records.

Search strategy for section 6.4.5, Measurement and valuation of health effects
The MS stated that the search for HRQL evidence was undertaken as part of the cost-effectiveness searches. The ERG queried the absence of HRQL terminology from the cost-effectiveness search strategy, and suggested inclusion of terms, such as HRQOL, HRQL, EuroQOL, EQ5D and SF36. The clarification response[7] stated that the MS cost-effectiveness searches incorporated a Centre for Reviews and Dissemination (CRD) economic evaluation filter. The ERG was unable to verify the source of the cost-effectiveness filter, despite checking the CRD website[22], CRD Guidance[19], NHS EED Handbook[23], ISSG Search Filter Resource [24], CRD Report 4[25] and contacting the CRD Information Service.

Although the cost-effectiveness terms included phrases, such as quality of life, the ERG did not consider this approach sufficient to identify all HRQL terminology. The clarification response[7] described how the updated EconLit and NHS EED searches included some additional HRQOL terms, which failed to identify any missing studies. The ERG noted that the Medline, In-Process and Embase searches were not repeated with the additional terms included. As discussed in section 4.1.1.2, it was not appropriate to limit economics databases, such as NHS EED and Econlit, by study design. As the Medline, In-Process and Embase searches were not re-run with the additional HRQL terms, the ERG concluded that there was a possibility that relevant HRQL references might have been missed by these strategies.

Search strategy for section 6.5.3, Resource identification, measurement and valuation
The MS reported that the searching for this section formed part of the economic evidence search, therefore the ERG comments for that section also apply to this one.

SUMMARY OF SEARCHES
Clinical evidence: For the most part, searches were well reported and clearly structured into population and intervention facets. All the required databases were searched, in addition to a comprehensive range of other sources, such as in-house databases, conference abstracts and handsearching. Aside from inconsequential errors in combination of redundant lines in the final sets, no obvious errors were detected in the search strategies. The ERG felt that the strategies would have benefited from additional terms for sJIA and Roactemra which may have resulted in identification of additional studies, however the ERG was unable to re-run and screen all the searches due to time constraints, and was therefore unable to determine whether any relevant studies might have been overlooked.

Indirect and mixed treatment comparison: The indirect and MTC comparison searches were inadequate and not fit for purpose, due to mistakes and the omission of MTX, an important comparator. Searches were clearly documented, and missing strategies were provided in the clarification response. The searches were structured into population and comparator interventions. Strategies contained errors in syntax and subject indexing which would have affected recall. Study

design filters were poorly structured with incorrect and irrelevant terms which would have impacted on the effectiveness of the search. Application of an RCT filter to the CENTRAL search was irrelevant and detrimental. The ERG undertook and screened additional searches, resulting in identification of one potentially relevant reference.[18]

Adverse events: Searches were clearly documented, and missing strategies were provided in the clarification response. The Medline, Embase and Biosis Previews searches were structured appropriately, with relevant facets and Boolean logic. All search strategies would have benefitted from addition population, intervention/comparator and adverse events terms. The Biosis Previews Update, Embase Alert, Medline In-Process and Cochrane Library searches lacked structure and consistency. Several strategies contained redundant lines which were not incorporated into the final results. On the whole, the adverse events searches were basic and would have benefitted from a more comprehensive appropriate, as per current AE searching advice.[19, 20]

Cost-effectiveness: Searches were clearly documented, and missing NHS EED, EconLit and HEED strategies were provided in the clarification response. The ERG was unclear whether the required database Medline In-Process was searched, as no strategy was presented. The Medline and Embase strategies were appropriately structured with correct Boolean logic. The economics study design filters for both databases contained redundant indexing which did not impair recall. The ERG concluded the Medline, Embase and HEED searches were acceptable. The ERG was concerned that the application of economics filters to both the EconLit and NHS EED searches was unnecessary and potentially detrimental to recall.

HRQL: The search to identify HRQL studies were undertaken as part of the cost-effectiveness searching. The ERG felt that inclusion of additional HRQL terminology in the Medline and Embase searches would have been beneficial. EconLit and HEED were inappropriately limited by study design which may have impaired sensitivity of the searches. The ERG concluded that relevant HRQL references may not have been retrieved.

Resource identification, measurement and valuation: The MS reported that the cost-effectiveness searches in section 6.1.1 were assumed to identify all the relevant literature for this section. Therefore the same ERG comments about errors, synonyms and inappropriate inclusion of economics filters for the EconLit and HEED searches applied to the searches for section 6.1, applied to this section.

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

The inclusion and exclusion criteria used in the selection of evidence for the systematic review were presented in the MS[10] (MS, page 33). The table in the MS was labelled 'eligibility criteria used in search strategy' but was presented within the description of the study selection process (Section 5.2.1). It was not clear from the MS how many reviewers were involved in the study selection process. Best practice specifies that two reviewers be involved in the application of inclusion and exclusion criteria in order to limit bias in study selection. Details of the inclusion and exclusion criteria applied in the MS are presented in Table 4.1.

Table 4.1: Inclusion/exclusion criteria used in the study selection (as presented by the manufacturer (MS, page 33))

	Clinical effectiveness
Inclusion criteria	**Population** Patients with systemic juvenile idiopathic arthritis (sJIA) or systemic juvenile rheumatoid arthritis **Interventions** Tocilizumab, interleukin-6 receptor inhibitor **Outcomes** Disease activity, physical function, joint damage, pain, steroid sparing, mortality, adverse effects of treatment, health-related quality of life, fever **Study design** No restrictions **Language restrictions** No restrictions
Exclusion criteria	No exclusion criteria were used at database level searches. The following exclusions were used during hand screening of results **Population** Patients with JIA subtypes other than systemic ie, oligo arthritis (formerly pauciarticular), polyarthritis rheumatoid factor positive, polyarthritis rheumatoid factor negative, enthesitis related arthritis, psoriatic arthritis and unclassified **Interventions** Studies that do not include tocilizumab **Outcomes** None excluded **Study design** None excluded **Language restrictions** Languages other than English

The population, interventions and outcomes are the same as those mentioned in the decision problem and the NICE scope. This search is limited to intervention trials, and does not include comparators. The search for comparators is described in chapter 5.7 (MS, page 113), and includes the same population, RCTs only and the comparators mentioned in the NICE scope: adalimumab, anakinra, etanercept and infliximab. So far everything is in accordance with the NICE scope. However, when this search retrieves one trial only, the manufacturer conducts a rapid review, which was not conducted in a systematic way to identify "all pivotal trials in juvenile arthritis regardless of subtype" (MS p. 116). Four more trials were identified, one for each comparator. As these trials do not include children with systemic JIA, they will not be further discussed in this report.

4.1.3 What studies were included in the clinical effectiveness review and what were excluded? Provide a table of identified studies. Please identify the <u>most important</u> clinical effectiveness studies.

The flow diagram in the MS (p. 34) shows that 2 trials were included. Details of the studies and their populations as presented in the MS (Table 8 p. 36) are reproduced below in Table 4.2.

The Phase III TENDER study (WA18221) (De Benedetti et al. 2010) is identified as the trial most relevant to the decision problem. Somewhat ironically, the MRA316JP study (Yokota et al. 2008) is excluded from further discussion as the comparator is placebo and the MS states, "as such this study does not address either population in the Decision Problem." (MS, page 40).

In addition, the search for trials including comparators identified one trial (Quartier et al. 2010). The ANAJIS trial had 1 month follow-up after which all respondents received anakinra for another 11 months in an open label phase. Previous response to NSAIDs, CS and/or MTX is not reported. The mean duration of steroid treatment (predniso(lo)ne) was 3.3 years for all respondents, and 19 out of 24 had used MTX, 13 had used etanercept and 19 had used DMARDs and/or biologic agents.

Table 4.2: List of relevant RCTs

Trial no. (acronym)	Intervention	Comparator	Population	Primary study ref.
WA18221 TENDER (De Benedetti et al. 2010) Phase III	Tocilizumab 12 mg/kg < 30 kg or 8 mg/kg ≥ 30 kg every 2 weeks for 12 weeks (69% of patients received concomitant methotrexate)	Placebo (70% of patients received concomitant methotrexate)	112 patients with active sJIA and a previous inadequate response to NSAIDs and corticosteroids (ages 2-17) 37 received placebo 75 received tocilizumab	De Benedetti et al. Arthritis Rheum 2010;62(10 Supple 1):S596 (Presented at ACR congress 2010) [redacted]
MRA316JP (Yokota et al. 2008) Phase III	Tocilizumab 8mg/kg every 2 weeks 6 week open-label lead-in phase followed by a 12 week randomised double-blind, placebo controlled, withdrawal, phase III trial (Concomitant methotrexate therapy not permitted)	Placebo (Concomitant methotrexate therapy not permitted)	56 patients with active sJIA and a previous inadequate response to conventional treatments (ages 2-19) received tocilizumab, in the open-label lead-in phase 23 received placebo 20 received tocilizumab, in double-blind phase	Yokota et al. Lancet 2008;371:998-1006 [redacted]
ANAJIS (Quartier et al. 2010) Phase III	Anakinra 2 mg/kg subcutaneous daily, max. 100 mg 1 month randomised double-blind, placebo controlled, phase III trial (8 out of 12 received concomitant methotrexate therapy)	Placebo (11 out of 12 patients received concomitant methotrexate therapy)	12 patients with active sJIA (ages 2-20) received anakinra, in the double-blind phase 12 received placebo Previous response to conventional treatments not reported.	Quartier et al. Ann Rheum Dis 2011; 70(5): 747–754

Baseline characteristics of children included in the three trials are described in Table 4.3.

Table 4.3: Characteristics of participants in the included studies.

Trial	Intervention/ Duration	% Female	Mean Age ± SD	Mean Disease Duration± SD	Physician Global Assessment Mean ± SD	Concomitant Therapy (I/C) %
TENDER	Tocilizumab, 8 or 12 mg (n=75) Placebo (n=37) 12 weeks	T: 52% P: 46%	T: 10.0 ± 4.64 P: 9.1 ± 4.43	T: 5.17 ± 3.98 P: 5.06 ± 4.45	T: 69.6 ± 15.65 P: 61.4 ± 21.12	DMARD: 71% Biologics: 82% MTX: 95% Etanercept: NR CS: NR
Yokota	Tocilizumab 8 mg/kg (n=23) Placebo (n=20) 12 weeks	T: 65% P: 65%	T: 8.0 ± 4.3 P: 9.3 ± 4.5	T: 4.6 ± 3.5 P: 4.7 ± 4.0	T: 51 (21-96)** P: 51 (18-95)	DMARD/biol*: NR MTX:NR Etanercept: NR CS: 100%
ANAJIS	Anakinra 2 mg/ kg (n=12) Placebo (n=12) 1 month	A: 58% P: 67%	A: 9.5 ± 5.19 P: 7.5 ± 3.73	A: 4.2 ± 3.33 P: 3.2 ± 1.95	A: 63 ± 20.57 P: 57 ± 29.74	DMARD/biol*: 79% MTX: 79% Etanercept: 54% CS: NR

*) DMARD and/or biologic agent; **) Range; CS=Corticosteroids

4.1.4 Provide details of any relevant studies not discussed in the submission? Why were these studies excluded and how were these studies identified by the ERG?

Absence of comprehensive synonyms and poorly applied study design limits were identified in several of the search strategies, which may have impacted on the recall of the search process. For the most part, the ERG was unable to determine whether any relevant studies were not identified.

The indirect and mixed treatment comparison searches were inadequate due to mistakes in the strategies, incorrect study design limits and omission of methotrexate, which is an important comparator. Additional searches by the ERG identified one potentially relevant study (Woo, 2000 [18]). However, because no data in the relevant population were available from the TENDER trial, an indirect comparison of tocilizumab with methotrexate was not possible.

4.2 Summary and critique of submitted clinical effectiveness evidence

If there is more than one RCT described in the MS, it may be appropriate to discuss each trial individually using the headings described.

4.2.1 Summary of submitted clinical evidence for each relevant trial.

Results of the TENDER trial were described in section 5.5 to 5.9 (MS, pages 80-194). We have summarised the available evidence for each outcome and population of the NICE scope in Table 4.4. In addition we have provided calculations of effect size. We did not summarise evidence that does not relate to the outcomes in the scope. The TENDER trial did not present results for joint damage or health related quality of life. However, tocilizumab significantly improved steroid sparing, mortality, adverse events, physical function, disease activity and pain, in comparison to placebo. The statistical analysis for the TENDER trial was not always ITT, as stated in table 18 (MS, page 89). Many patients, especially in placebo were excluded most likely due to escape medication.

Table 4.4: 12 week outcomes for the TENDER trial

Outcome\population	**Population 1[+]:**	**Population 2:**		
		Tocilizumab* (Responders/ Patients analysed)	**Placebo (Responders/ patients analysed)**	**Effect size**
Primary endpoint: JIA ACR 30 **	NR	68/75	9/37	RR = 3.73 (95%CI: 2.1, 6.61) ǂ
JIA ACR 50 JIA ACR 70 JIA ACR 90		64/75 53/75 28/75	4/37 3/37 2/37	RR=7.89 (95%CI: 3.11, 20.11) ǂ RR= 8.72 (95%CI: 2.92, 26.0) ǂ RR= 6.91 (95%CI: 1.74, 27.4) ǂ
Steroid sparing*	NR	17/70	1/31	RR=17.57 (95%CI: 2.49, 123.89) ǂ
Mortality	NR	0/75	0/37	
Adverse events of treatment: Patients with ≥1 AE Patients with ≥1 serious AE Patients with ≥1 infection Patients with ≥1 serious infection	 NR NR NR NR	[redacted]	[redacted]	[redacted]
		Tocilizumab* Adjusted mean change from baseline (patients analysed)	**Placebo Adjusted mean change from baseline (patients analysed)**	**Effect size**
Disease activity: Physician Global Assessment of Disease Activity (100 mm VAS) No. Active joints	 NR NR	 -69.6 (73) -70.6 (73)	 -41.1 (17) -37.2 (17)	 MD = -64.4 (95%CI: -87.5, -41.3)ǂ MD = -33.4

				(95%CI: -53.2, -13.6)‡
Physical function No. of Joints with Limitation of Movement CHAQ-D1 score	 NR NR	 -51.6 (72) -45.6 (72)	 -22.5 (17) -10.3 (17)	 MD =-29.1 (95%CI: -53.4,-4.9)‡ MD =-35.3 (95%CI: -63.5, -7.1)‡
Joint damage	NR	NR		
Pain Vas (0-100mm)	NR	-41.0 (73)	-1.1 (17)	Difference = -39.8 (85%CI: -55.1, -24.6)‡
Health related Quality of Life:	NR	NR		

+Population 1= Children with sJIA who have not responded adequately to prior NSAID and systemic corticosteroids; Population 2= Children with sJIA who have not responded adequately to prior NSAID, systemic corticosteroids and MTX.
‡ results are significant..
* Tocilizumab is a combination of 8mg/kg and 12mg/kg doses
** analysis was reported as ITT, but patients who withdrew, received escape medication, or for whom the endpoint could not be determined were classified as non-responders.
MD = Adjusted mean difference. Analysis of variance adjusted for the randomization stratification factors applied at baseline.
*** Patients Receiving Oral CS with JIA ACR70 Response at Week 6/8 who Reduced Oral CS dose by ≥ 20% Without Subsequent JIA ACR30 Flare or Occurrence of Systemic Symptoms

The MS identified two studies from its searches; TENDER and Yokota 2008. Yokota was subsequently excluded from further analysis by the MS, due to the reasons below:

> 'This is due to the study design and population. The study was initiated with a 6 week open-label led-in phase. Patients with an ACR Pedi 30 response and CRP levels below 5mg/L were then randomised in a double-blind manner to receive either placebo or tocilizumab for a further 12 weeks, with rescue therapy available if necessary. This was followed by an open-label extension period for at least 48 weeks. Methotrexate treatment was not permitted throughout the duration of the study. The comparator was placebo, and as such this study does not address either population in the Decision Problem.' (MS, page 40).

However the ERG thought that the Yokota trial should remain in the analysis for the following reasons; study design was not an inclusion/-exclusion criteria, a Japanese population was not an exclusion criteria, placebo was used as a comparator by the TENDER trial and so should not be an exclusion criteria here (following the MS approach), MTX treatment was not an inclusion/exclusion criteria and MTX treatment is mixed in the TENDER trial. The Yokota trial could not be used in further analysis because it was a withdrawal trial and was therefore too dissimilar to the other trials for further synthesis. Therefore the results for Yokota are presented in Table 4.5 as stand-alone results. No evidence was presented for joint damage, physical function, steroid sparing or health related quality of life.

Table 4.5: 12 week double blind phase outcomes for Yokota (withdrawal trial)

Outcome\population	Population 1[+]:	Population 2:		
		Tocilizumab*	**Placebo**	**Effect size**
Primary endpoint: Rate of maintained response** (Responders/total patients analysed)	NR	16/20	4/23	RR=4.6 (95% CI: 1.84, 11.51)ǂ
Disease activity: JIA ACR 30 (Responders/total patients analysed) - completers - ITT Mean Physician Global Assessment of Disease Activity, 100 mm VAS (patients analysed) - completers - ITT		 16/16 17/20 4.9 ±4.4 (16) 7.5± 10.3 (20)	 4/4 8/23 4.5 ± 7 (4) 24.7 ± 25.8 (23)	 RR=1 RR=2.44 (95% CI: 1.36, 4.40)ǂ MD=4 (-5.39, 6.19) MD= -17.2 (-29.7, -4.75) ǂ
Mean Pain Vas, 0-100mm (patients analysed) - completers - ITT		 7.6±8.3 (16) 7.9±9.3 (20)	 19.5 ± 17.3 (4) 34.5 ± 25.4 (23)	 MD=-11.9 (-24.07, 0.27) MD=-26.6 (-38.7, -14.5) ǂ
Steroid sparing	NR	NR	NR	
Mortality	NR	0/21	0/23	
Adverse events : Patients with ≥1 AE Patients with ≥1 serious AE Patients with ≥1 infection Patients with ≥1 serious infection	 NR NR NR NR	 18/21 0/21 14/21 0/21	 19/23 0/23 11/23 0/23	 RR=1.04 (95%CI:0.8,1.3) RR=1.39 (95%CI:0.83,2.35)
Physical function	NR	NR	NR	
Joint damage	NR	NR	NR	
Health related Quality of Life:	NR	NR	NR	

+Population 1= Children with sJIA who have not responded adequately to prior NSAID and systemic corticosteroids; Population 2= Children with sJIA who have not responded adequately to prior NSAID, systemic corticosteroids and MTX.
Outcomes were measured as baseline at the start of the blinded part of the study to 12 weeks.
ǂ results are significant.
* Tocilizumab is a dose of 8mg/kg
** percentage of patients completeing the study to whom withdrawal criteia and rescue criteria did not apply.
ITT, carried out by last observation day

The Yokota trial was a withdrawal trial, hence after 6 weeks of drug in an open phase design, those patients intolerant to the drug would have dropped out from the trial (12 patients), before commencement of the 12 week double blind phase. This design will particularly affect the numbers of patients subsequently suffering from adverse events and positively influence the response to drug during the following blinded phase. Indeed no serious adverse events were observed within this phase

whilst 2 were observed during the open phase. No deaths were recorder in either phase. To inform the results more fully we have provided ITT and patients completing analyses.
The primary endpoint of the Yokota trial was the rate of maintained response, which was found to significantly improve with tocilizumab. JIA ACR 30 outcomes were less useful since all patients responded to tocilizumab during the open label lead in phase, therefore there was a high rate of response despite large numbers of patients dropping out of the placebo arm due to rescue medication. ITT-analysis by the last observation day indicated that JIA ACR 30 improvement was significantly higher for tocilizumab than placebo. Tocilizumab also significantly improved the pain scores and disease activity scores.

The ERG identified the ANAJIS trial as the only trial eligible for the indirect comparison of tocilizumab versus anakinra (or other treatments, such as TNF alpha inhibitors). In Table 4.6 we have summarised the outcomes which fit those scoped by NICE. There were no reported outcomes for joint damage, mortality, health related quality of life or steroid sparing. Anakinra was found to significantly inhibit disease activity, measured by four different outcomes (ACR-pedi 30 either with or without fever, physician assessment and number of active joints). Anakinra was not found to significantly affect pain, physical function nor adverse events. No significant differences for adverse events were observed between anakinra or placebo treatment.

Table 4.6: 1 month outcomes for ANAJIS trial

Outcome\population	**Population 1[+]:**	**Population 2:**		
		Anakinra* (n=12)	**Placebo (n=12)**	**Effect size**
Primary endpoint: Modified ACR-pedi 30** (Responders)	NR	8	1	RR=8 (95% CI:1.17, 54.5)ǂ
Disease activity: ACR-pedi 30 and no fever*** (Responders)	NR	11	6	RR= 1.83 (95% CI:1.02, 3.31)ǂ
Physician Assessment of Disease Activity, §	NR	-63	-20	MD= -43 (-9.4, -76.6) ǂ
No. Active joints §	NR	-46	-18	MD= -28 (95%CI: -2.9, -53.1) ǂ
Patients' assessment of pain§	NR	-29	-21	MD=-8 (95% CI: 4.4, -20.4)
Steroid sparing	NR	NR	NR	
Mortality	NR	NR	NR	
Adverse events :				
No. AE	NR	14	13	RR= 1.08 (95% CI:)
No. serious AE	NR	0	0	
No. Infections	NR	2	2	RR= 1 (95%CI: 0.17, 5.98)
Physical function No. of Joints with Limitation of Movement§	NR	-36	-20	MD=-16 (95%CI: 4.9,-36.9)
CHAQ §	NR	-37	-9	MD= -28 (95%CI: 17.0, -73.0)
Joint damage	NR	NR	NR	
Health related Quality of Life:	NR	NR	NR	

+Population 1= Children with sJIA who have not responded adequately to prior NSAID and systemic corticosteroids; Population 2= Children with sJIA who have not responded adequately to prior NSAID, systemic corticosteroids and MTX.
ǂ results are significant.
* Anakinra is a dose of 2mg/kg
**Body temperature <38°C for more than 7 days, CRP and ESR normalised or decreased by at least 50% (=systemic symptoms responders) and also, in responders to the trial primary objective, ACR-pedi 30, 50, 70 or 100 (whichever level is indicated) response

compared with D1.
***Body temperature <38°C for more than 7 days and ACR-pedi 30 response compared with D1.
§mean change from baseline to 1 month

4.2.2 Describe and critique the manufacturer's approach to validity assessment for each relevant trial.

Formal appraisals of the validity of the TENDER trial and the ANAJIS trial were clearly presented in the MS (table 11, pages 75-9; table 34, pages 134-135). All the criteria listed under Section 5.4.1 (MS, page 75 - as specified in the NICE STA Specification for manufacturer/sponsor submission of evidence) were addressed in the quality assessment findings. These finding are reproduced in Table 4.7 alongside a validity assessment provided by the ERG for the Yokota trial.

The ERG checked the quality assessment findings against the provided trial data or original study publications and any additional points are discussed within this Section. Responses that the ERG have highlighted and discussed in further detail are marked with an asterisk (*). Overall the ERG was in agreement with the MS quality assessment findings. However, the ERG would disagree that the TENDER trial used ITT analysis. Several of the relevant outcomes in the TENDER trial were not analysed on an ITT basis, many were per protocol (see Table 4.4). The Yokota trial had a 6 week open phase design before commencing the 12 week blinded trial, therefore those patients intolerant to the drug would have dropped out from the trial. This design will bias the results of the outcomes – particularly adverse effects, therefore caution should be used in the analysis of the results.

It was not clear if any of the procedures for searching, screening, assessing validity, extraction and synthesis were undertaken by a single reviewer and independently checked by a second reviewer or using a consensus of multiple reviewers.

Table 4.7: Quality assessment results for RCTs

Trial no. (acronym)	TENDER	Yokota (ERG)	ANAJIS
Was randomisation carried out appropriately?	Yes	Unclear	Yes
Was the concealment of treatment allocation adequate?	Yes	Unclear	Yes
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Unclear	Yes
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Yes	Yes
Were there any unexpected imbalances in drop-outs between groups?	No	Yes, more patients withdrawn for escape medication	Not clear
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	Not clear
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes, no missing data at end of 12 week randomised phase*	Yes, last observation carried forward	Yes
Adapted from Centre for Reviews and Dissemination (2008) Systematic reviews. CRD's guidance for undertaking reviews in health care. York: Centre for Reviews and Dissemination			

4.2.3 Describe and critique the statistical approach used within each relevant trial.

The statistical analyses in the included TENDER and ANAJIS studies are described on pages 65-66 and Table 33 (MS) and copied into Table 4.8 below. The analyses for the Yokota trial were copied from the trial data.

As mentioned previously not all the outcomes in the TENDER trial were analysed by ITT in table 18 (MS, page 89). Many patients, especially in placebo were excluded, most likely due to escape medication.

As the major trial was TENDER there must be some concern that the population was heterogeneous with regards to whether the patients respond to MTX or not and whether the patients were receiving MTX or not. We have already discussed that the post-hoc analysis performed by the MS to address this problem is invalid and the manufacturers did not provide individual data for these populations to allow a valid analysis. In addition, the ERG also concluded that the indirect analysis performed by the MS is invalid due to the inclusion of several trials in juvenile arthritis regardless of subtype.

The statistical analyses in the ANAJIS trial and Yokota 2008 trial seemed appropriate.

Table 4.8: Statistical analyses in the relevant RCTs

Trial no. (acronym)	Hypothesis objective	Statistical analysis	Data management, patient withdrawals
TENDER (De Benedetti et al. 2010)	To evaluate the efficacy (JIA ACR 30 + absence of fever) of tocilizumab versus placebo in combination with stable ongoing	[redacted]	[redacted]

	therapy at 12 weeks, in sJIA patients with persistent disease activity and an inadequate response to NSAIDs and systemic corticosteroids.	[redacted]	[redacted]
ANAJIS (Quartier et al. 2010) (Anakinra)	The primary objective was to compare the efficacy after 1 month's treatment with anakinra (2mg/kg subcutaneously daily, maximum 100 mg) or placebo in the two groups of patients. To be responders to a modified American College of Rheumatology Pediatric (ACRpedi) 30 score built for the purpose of the trial. No null hypothesis specified	Qualitative and quantitative data were compared using Wilcoxon test and Fisher exact test, respectively. The R statistical software was used for statistical analysis.	Not Specified
Yokota et al 2008	To investigate the efficacy, maintenance of response, safety and pharmacokinetics of tocilizumab.	The rate of maintained response was compared using the $\chi 2$ test (exact). The period of maintained response was determined using the Kaplan-Meier method and the groups were compared using the log-rank test.	Patients who dropped out of the blind study because withdrawal criteria applied or patients in whom response was maintained on the last observation day were treated as censored.

4.2.4 Describe and critique the manufacturer's approach to outcome selection within each relevant trial.

According to the NICE scope the outcomes required were:

- Disease activity;
- physical function;
- joint damage;
- pain;
- steroid sparing;
- mortality;
- adverse effects of treatment;
- health-related quality of life; and

Appendix 9.4.6 (MS, page 349) does not include outcomes in the inclusion criteria while it is stated that studies will be excluded if ACR outcomes are not reported (MS, page 350).

Table 4.9 below compares the outcomes identified in the scope with those reported in the relevant trials.

Table 4.9: Scope outcomes reported in relevant trials

	Outcomes identified in NICE scope							
Trial	Disease activity	Physical function	Joint damage	Pain	Steroid sparing	Mortality	Adverse effects of treatment	Health related quality of life

TENDER	✔	✔	X	✔	✔	✔	✔	X
Yokota*	✔	X	X	✔	X	✔	✔	X
ANAJIS	✔	✔	X	✔	X	X	✔	X

*Yokota is a withdrawal trial and not directly comparable with the two effectiveness trials

For the indirect comparison only ACR 30, ACR 30 plus absence of fever, ACR 50 and ACR 70 are used. While both studies also report:

- Physician assessment of disease activity (100mm VAS);
- Number of active joints;
- Patients' assessment of pain (100mm VAS);
- Number of joints with Limitation of Movement;
- CHAQ score;
- Number of adverse events;
- Number of serious adverse events; and
- Number of infections.

4.2.5 To what extent does each relevant trial include the patient population(s), intervention(s), comparator(s) and outcomes as defined in the final scope?

Populations

The included trials had inclusion criteria broader than that defined in the scope. In particular, they were not limited to form the two relevant populations of inadequate response to NSAID(s) and corticosteroids or NSAID(s), corticosteroids and methotrexate. Whilst the inclusion criteria for TENDER would suggest population 1 (inadequate response to NSAIDs and CS) the MS argues that the TENDER trial population is actually equivalent to population 2 (inadequate response to NSAIDs, CS and MTX). According to the MS, "70% of patients (all with prior inadequate response to NSAIDs and corticosteroids) at baseline were still receiving methotrexate, yet had active disease (a further entry requirement for the study), thus could be considered to be failing on methotrexate alone." (MS, p. 39)

At best this inference means TENDER does not address population 1, at worst the inference made about population 2 is unreliable and neither population is addressed.

Intervention

Tocilizumab is a solution for infusion with a dosing frequency of once every two weeks. The recommended posology is 8 mg/kg in patients weighing greater than or equal to 30 kg or 12 mg/kg in patients weighing less than 30 kg.

Tocilizumab is indicated for the treatment of active sJIA in patients 2 years of age and older, who have responded inadequately to previous therapy with NSAIDs and systemic corticosteroids. Tocilizumab can be given as monotherapy (in case of intolerance to MTX or where treatment with MTX is inappropriate) or in combination with MTX. (RoActemra draft Summary of Product Characteristics – Anticipated June 2011).

Comparators

For population 1 the comparator in the scope is methotrexate, yet the trial the MS relies upon is the TENDER trial which compares tocilizumab plus standard care versus placebo plus standard care. Whilst the comparator in this study was placebo, 70% of patients at baseline were also receiving methotrexate. The MS uses a post-hoc analysis to compare tocilizumab with those patients in the placebo group also receiving methotrexate. However, this is not an acceptable comparison because trial participants were not randomised in this way. Therefore this comparison is based on observational data from two different populations, those using MTX and those not using MTX.

There is also a mismatch between scope and the decision problem for population 2. For those patients with an inadequate response to NSAIDs, CS and methotrexate the comparators should be TNF inhibitors (e.g. etanercept, adalumimab, infliximab) and anakinra. There is no direct evidence presented for these comparators and hence indirect comparisons are made. For tocilizumab evidence of the TENDER trial is used, but instead of using data from those patients described by the manufacturer as relevant to the decision problem (95% of included patients), data for all patients were used in the indirect comparison. For the comparators, the manufacturer decided to broaden the inclusion criteria (see MS, page 116):

> "Due to the dearth of clinical evidence in systemic JIA, Roche augmented the dataset with evidence from a rapid review performed with objective to identify all pivotal trials in juvenile arthritis regardless of subtype."

Despite advise from their clinical experts to the contrary (see MS, page 116):

> "Clinical experts [PC Westhovens R 02/03/2011, Wright S 16/03/2011], stressed the differences between a systemic JIA population and other subtypes and advised against comparing evidence from different populations."

The ERG agrees with the advice from the clinical experts; therefore, trials in children with other types of juvenile arthritis will be ignored in this report.

Outcomes

The outcomes listed in the NICE scope of disease activity, physical function, pain, adverse events and steroid sparing have been matched by the MS. However, there appear to be no outcomes for 'joint damage' or 'HRQOL' in the clinical effectiveness section. The manufacturer states that joint damage is not currently available. The parent/patient global assessment of overall well-being (100 mm VAS) and CHAQ (functional ability), see MS, page 196 do not measure HRQOL.

Consideration of the clinical characteristics of sJIA would suggest it could be important to consider outcomes that define lymph node enlargement, hepatomegaly, splenomegaly and serositis. MAS was mentioned within the decision problem therefore it would have been advantageous to present this more clearly within adverse events.

Overall the presented clinical effectiveness outcomes did not match the scoped outcomes. There was a general lack of clarity regarding which trial outcomes were being used to match the scope outcomes. There were no appropriate outcomes for joint damage, mortality and HRQOL. In addition, further outcomes may have been useful especially for adverse events relating to sJIA.

4.2.6 Where appropriate, describe and critique any meta-analysis, indirect comparisons and/ or mixed treatment analysis carried out by the manufacturer.

This section should include a summary of the manufacturer's methods and results as described in the MS. The ERG should critique the methods used and interpret the results in light of the methods used by the manufacturer and generalisability to patients in England and Wales.

The MS describes an indirect comparison starting on page 117 with a list of the identified studies. The methodology and eligibility criteria for included studies is presented in Tables 26 and 27 (pages 119-123) and Tables 28-32 (pages 124-128) provide summaries of patient characteristics. Table 33 (pages 129-133) summarises the statistical methods of the included studies, Table 34 (pages 134-135) presents the quality assessment and Table 35 (pages 136-137) presents relevant results. Table 38 (page 141) summarises the data used in the indirect comparison and Table 39 (page 143) presents the results of the indirect comparison.

In the MS the methods are described as follows (MS, page 142):

> The summary measure selected for this analysis is the relative risk (RR). The RR and its precision are calculated for each study and each outcome using the n/N data presented above.

The efficacy of tocilizumab, anakinra and infliximab is indirectly compared using placebo as a common comparator, following the method developed by Bucher at al. [1997]. This indirect comparison is carried out on RRs. Given data are only available in one study for each treatment there is no need for meta-analysis.

The indirect effect of tocilizumab compared to its comparators and its associated 95% bilateral confidence interval are calculated using the formulas below. Due to the mathematical characteristics and distribution of the RRs, it is necessary to perform the analysis on the logarithmic scale and then back-transform (exponentiate) the results.

(1) ln(RR)A vs B= ln(RR)A vs P – ln(RR)B vs P
(2) SE(ln(RR)A vs B) = [Var(ln(RR)A vs P)+ Var(ln(RR)B vs P)]1/2 the 95% CI around the logarithm of the indirect effect is calculated as:
(3) ln (RR)A vs B ± 1.96*(ln(RR)A vs B)
Values for the RRs and confidence intervals were calculated using Stata SE version 8.2. The indirect comparisons were carried out in Excel.

The table below reproduces the results of the indirect comparison supplied by the manufacturer in response to the clarification letter (page 5-6). These results exclude the 5% of MTX naive patients.

Table 4.10: Results of the indirect comparison analysis

Comparison	Outcome	Base-case analysis (TENDER*)		Sensitivity analysis (excl. MTX naive)	
		RR	95% CI	RR	95% CI
TCZ vs. ANK	ACR30	2.37	1.10, 5.10	2.27	1.06, 4.85
TCZ vs. INF	ACR30	2.87	1.49, 5.55	2.75	1.44, 5.26
	ACR50	5.35	1.91, 14.97	5.04	1.81, 14.04
	ACR70	4.61	1.16, 18.38	4.33	1.09, 17.20

* analysis was reported as ITT, but patients who withdrew, received escape medication, or for whom the endpoint could not be determined were classified as non-responders.

COMMENT

As noted above, the ERG does not think the comparison with infliximab is valid as the infliximab trial (Ruperto et al. 2007) does not include children with systemic JIA.

Regarding MTX, the MS uses a post-hoc analysis to compare tocilizumab with those patients in the placebo group receiving methotrexate. However, this analysis is flawed: participants were not randomised based on whether or not they received MTX. Therefore, the manufacturer compared two groups of patients whose heterogeneity will influence the treatment effects. More importantly, following the TENDER inclusion criteria for active disease and the MS approach which claims active disease status despite MTX therapy equates to inadequate response, means the effects of tocilizumab were compared with MTX in a population that was specifically selected as being not responsive to MTX. In addition, the manufacturer previously stated that all children receiving MTX were population 2; and therefore not relevant for this comparison. It would have been possible to produce an indirect comparison of tocilizumab vs methotrexate using data from the TENDER trial and from

Woo et al 2000.[18] In the absence of the requested data from the manufacturer (individual data for tocilizumab without methotrexate and placebo without methotrexate from the TENDER trial) this was not possible. It should also be noted that data from Woo et al.[18] are probably minimal as most data are reported for children with sJIA and extended oligoarticular arthritis combined; in addition, outcomes from both trials may not be comparable.

The ERG have investigated heterogeneity within and across TENDER and ANAJIS trails. Inclusion criteria are similar for both trials. Table 4.11 presents baseline characteristics for TENDER and ANAJIS.

Table 4.11: Patient characteristics at baseline for TENDER and ANAJIS trials

	TENDER				ANAJIS	
Characteristics	**Placebo n=37**	**TCZ 8mg/kg n=37**	**TCZ 12mg/kg n=38**	**All TCZ n=75**	**Placebo n=12**	**Anakinra n=12**
Female, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	8(67)	7(58)
Age, mean, years (SD)	[redacted]	[redacted]	[redacted]	[redacted]	7.5 (3.73)	9.5 (5.19)
Disease mean duration (SD)	[redacted]	[redacted]	[redacted]	[redacted]	3.2 (1.95)	4.2 (3.33)
CRP, mg/l, mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]	84 (65.74)	66 (64.4)
ESR, mean, (SD)	[redacted]	[redacted]	[redacted]	[redacted]	57 (27.85)	44 (23.37)
Active joints, mean no. (SD)	[redacted]	[redacted]	[redacted]	[redacted]	16 (15.84)	16 (13.12)
Joints with LOM, mean no. (SD)	[redacted]	[redacted]	[redacted]	[redacted]	17 (14.91)	16 (14.88)
Physician's global VAS, mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]	57 (29.74)	63 (20.57)
Parent/patient global VAS, mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]	55 (26.51)	50 (24.39)
CHAQ, mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]	1.44 (0.625)	1.67 (0.845)
DMARD and/or biologic agent, no. Patients	[redacted]	[redacted]	[redacted]	[redacted]	11 (91.6)	8 (66.7)
Methotrexate, no. Patients (%)	[redacted]	[redacted]	[redacted]	[redacted]	11 (91.6)	8 (66.7)

Of particular note are the scores for disease duration and CRP. In the Tender trial the TCZ 8mg group had a disease duration of 6.33 years compared to 4.03 years for TCZ 12mg resulting in a difference of 2.3 years. Patients in the ANAJIS trial had shorter disease duration, up to 3.13 years less. The CRP scores for the TCZ groups in the Tender trial are considerably higher than the placebo group and both placebo and anakinra groups in the ANAJIS trial. The Tender trail notes these scores were driven by three patients with very high CRPs which distorted the median/mean values. Nonetheless, these scores indicate a spike in disease activity at baseline which would likely resolve during the trial thus making improvement within these groups relatively easier.

Moreover, 4/12 (30%) of patients in the anakinra group were MTX naive compared to only 5% in the TENDER trial.

However, the most important difference between the two trials is the length of follow-up: 12 weeks for TENDER and 1 month for ANAJIS. Therefore, results of this indirect comparison need to be interpreted with caution.

When we look at the results from both trials, only two outcomes are found to be comparable across trials: ACR30 response and ACR30 without fever (see table 4.12). The indirect comparison of tocilizumab versus anakinra shows that both outcomes favour tocilizumab, but only ACR30 response shows a significant difference (RR=2.27, 95% CI: 1.06, 4.85).

Table 4.12: Indirect comparison results based on the TENDER and ANAJIS trials

	TENDER (12 weeks)			ANAJIS (1 month)			Toc vs Ana
Outcome	Toc	Pla	RR	Ana	Pla	RR	RR
ACR30	68/75	9/37	3.73 (2.10, 6.61)	11/12	7/12	1.57 (0.95, 2.61)	2.37 (1.10, 5.10)
ACR30, using data from 95% of TENDER trial participants							2.27 (1.06, 4.85)
ACR30, no fever	64/75	9/37	3.51 (1.97, 6.24)	11/12	6/12	1.83 (1.02, 3.31)	1.91 (0.84, 4.37)
ACR30, no fever, using data from 95% of TENDER trial participants							NR

4.2.7 Additional clinical work conducted by the ERG

Provide details of any additional work conducted by the ERG in relation to clinical effectiveness. If the results of any of the additional work affect the size of the ICER, refer the reader to the summary table in Section 6.

No further additional work was undertaken by the ERG in relation to clinical effectiveness.

4.3 Conclusions

Describe the completeness of the MS with regard to relevant clinical studies and relevant data within those studies. Does the submission contain an unbiased estimate of the technology's (relative and absolute) treatment effects in relation to relevant populations, interventions, comparators and outcomes? Are there any remaining uncertainties about the reliability of the clinical effectiveness evidence? Reference should also be made concerning the extent to which the submitted evidence reflects the decision problem defined in the final scope.

The ERG has a fundamental problem with the evidence presented in the MS as it is not in accordance with the NICE scope. It is for the Appraisal committee to decide whether it will accept the ERG approach, which means there is no evidence for any comparison in the NICE scope, or accept the MS approach, which means there is some evidence for the second population, but none for the first population.

The main question is: "Which patients in the TENDER trial match which population"? According to the manufacturer 95% of TENDER trial participants match population 2, because "patients are included in the study if they have symptoms of active disease" and "It follows that if patients have tried in the past or are currently administered MTX and continue to have persistent disease then they are inadequate responders".

The ERG does not agree with this approach. The MS does not provide a clear definition of inadequate responders. It cannot be automatically assumed that all participants in the TENDER trial are inadequate responders to MTX. Because of the lack of information it can only be assumed that the 25% of children in the TENDER trial who stopped using MTX fit this population (population 2). The remaining 75% of children in the TENDER trial should be treated as population 1. Because no data were provided for these two populations, there is no evidence available for any of the comparisons in the NICE scope.

Following the MS approach, no data were provided in the MS for population 1. Therefore, the only comparison left is tocilizumab versus anti-TNFs or anakinra. The manufacturer performed a systematic review to identify trials for the comparators. One trial was identified in children with sJIA, comparing anakinra with placebo. The manufacturer decided to broaden the inclusion criteria to include all trials in juvenile arthritis regardless of subtype, despite advice from their clinical experts to the contrary (see MS, page 116). The ERG agrees with the advice from the clinical experts; therefore, trials in children with other types of juvenile arthritis will be ignored in this report.

In conclusion, following the MS approach, for population 2 (children with sJIA with an inadequate response to NSAIDs, CS and MTX) the MS provided data for an indirect comparison of tocilizumab versus anakinra, using data from the TENDER trial, and a trial of anakinra versus placebo. Strictly speaking, the 5% of participants in the TENDER trial who were MTX naive should be excluded from these analyses. The MS only provided data for all participants in the TENDER trial. However, in response to the clarification letter some data were provided in which MTX naive patients were excluded. These data were not reported for the TENDER trial, but only for the indirect comparison with anakinra. Where possible, the ERG will use data for the correct population.

The indirect comparison of tocilizumab versus anakinra shows that ACR30 response favours tocilizumab (RR=2.27, 95% CI: 1.06, 4.85). ACR30 response without fever showed no significant difference between tocilizumab and anakinra.

5 COST EFFECTIVENESS

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

5.1.1 State objective of cost effectiveness review. Provide description of manufacturers search strategy and comment on whether the search strategy was appropriate. If the manufacturer did not perform a systematic review, was this appropriate?

The objective of the cost-effectiveness review in the MS was to identify existing economic evaluations relevant to the STA Decision Problem. The search strategies for the review are discussed in detail in section 4.1.1.2.

5.1.2 State the inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate.

The manufacturer used the following inclusion criteria for the search (MS, page 203):

- **Study design** to include economic evaluations, pharmacoeconomic studies, cost effectiveness and cost utility analysis
- **Disease area** to include all juvenile arthritis (no restrictions by subtype)
- **Population** (no restrictions by age or disease severity)
- **Treatments** (no restrictions by treatment all interventions included)

Included citations were indicated by "Inc". Excluded citations were indicated by "Exc" and the reason for exclusion provided as follows:

- "Not study" (not an economic or cost/utility study)
- "Not population" (not JIA population)
- "Not disease" (not arthritis related, other disease area)
- "Not outcome" (no cost or utility data reported)
- "Duplicate" (duplicate reference in the database)
- "Not English" (language of publication is other than English)

COMMENT

It was not clear from the MS how many reviewers were involved in the study selection process. Best practice specifies that two reviewers be involved in the application of inclusion and exclusion criteria in order to limit bias in study selection.

5.1.3 What studies were included in the cost effectiveness review and what were excluded? Where appropriate, provide a table of identified studies. Please identify the most important cost effectiveness studies.

The manufacturer's literature search identified a total of 6 studies that present a synthesis of cost and effects. Of the total of 6 studies that were identified there were 3 cost-utility studies[26-28], 1 willingness to pay (WTP) study[29], 1 decision analysis regarding treatment options for knee monoarthritis[30] and 1 record that was not clear with regard to many domains including authors and study type.[31] See table below.

Table 5.1: Study characteristics and results (Economic Evaluations)

Study	Population	Treatment considered	Form of analysis	Type of model	Time-horizon	Main results	Relevance to the Decision Problem
Cummins et al. 2002 (review of Wyeth submission)	JIA	Etanercept vs. Placebo	Cost-utility	Not explicitly determined. Believed to be Markov model.	Lifetime	Base case ICER: £16,082. Sensitivity analysis ICER range: £3,900 - £34,000 Amended model ICER: £24,000	Not very relevant. Some information on resource use in the UK can be used for the C/E model. The model is an RA model adapted for JIA. The model has uncertain validity.
Barron et al. 2004	JIA	Hypothetical medications	WTP	No model considered	N/A	WTP $395 ± $329	Not relevant. The analysis is based on hypothetical interventions and elicits WTP.
Beukelman 2008	JIA	Optimal treatment of knee monarthritis	Decision analysis	Markov decision analysis comparing 3 treatment strategies	6 months	Of 3 most common treatment strategies compared: The number of patients that need to be treated with intraarticular corticosteroid injections is 3.8 with additional cost of 6.7 months of active arthritis	Not relevant.
Budapest University: Etanercept study 2006	JIA	Not clear	Not clear	Not clear	Not clear	Incremental QALY gain per patients is 0,72 QALY. ICER is 9.7million HUF per QALY (265HUF=1€)	No information is provided.
Epps 2005	JIA	Hydrotherapy	Cost utility	Model not defined	6 months	Patients in the combined group (+hydro) had slightly higher mean costs (£20.90) and lower mean QALYs (–0.0478).	Cost data and resource use may be relevant. QALYs specific for hydrotherapy not JIA population.
Ungar et al. 2010	JIA (poly articular)	Biologics in DMARD IR	Cost utility	Decision model	2 consecutive 6 month intervals	The additional costs per additional ACR Ped 30 responder at one year (95% CI) were $26,061 (17,070, 41,834), $46,711 (30,042, 75,787), $16,204 (11,393, 22,608) and $31,209 (16,659, 66,220), for etanercept, adalimumab, abatacept and infliximab, respectively.	Costs specific for Canada setting. Effectiveness combined from several RCTs, response data used not QALYs therefore may not be useful.

DMARD IR: Disease modifying anti rheumatic drug inadequate response, ICER: Incremental cost effectiveness ratio, JIA: Juvenile idiopathic arthritis, QALY: *quality-adjusted life year*, WTP: willingness to pay

5.1.4 What does the review conclude from the data available? Does the ERG agree with the conclusions of the cost effectiveness review? If not, provide details.

No specific conclusions from the economic review were provided in the MS. However, from the remarks in table 5.1 it may be concluded that the manufacturer considered none of the studies useful for the current analysis. The ERG agrees with that (implicit) conclusion.

Regarding the HRQL review, the manufacturer states that of the 34 studies identified in the review reporting on quality of life, none reported any utility values that were used in the economic evaluation presented here. The ERG agrees with that decision.

Regarding the review of resource use, no specific conclusions were provided in the MS. However, from the remarks in table 95 in the MS it may be concluded that the manufacturer considered two studies[32, 33] useful for the current analysis. The ERG agrees with that (implicit) conclusion.

5.2 Summary and critique of manufacturer's submitted economic evaluation by the ERG

Summarise and critique the cost effectiveness evidence submitted by the manufacturer (headings 5.2.1 to 5.2.11 are suggested headings). It is noted that the ERGs may prefer NOT to combine the summary and critique of the submitted economic evidence and instead report summary and critique sections separately.

Table 5.2 presents a summary of the *de novo* economic model developed by the manufacturer.

Table 5.2: Summary of the manufacturer's economic evaluation

	Approach	**Source / Justification**	**Signpost (location in MS)**
Model	Markov model with constant transition probabilities of relative improvement of health states.		Section 6.2.2 (p.213)
States and events	Health states are the relative improvement (ACR30, 50, 70 and 90 response). Events are a particular ACR response, non-response, withdrawal, or death.	The health states defined mirrored the endpoints of the TENDER trial. MS states that the structure of the model is developed to closely reflect real practice.	Section 6.2.2 (p.213), 6.2.4 (p.216), 6.2.5 (p. 217)
Comparators	A sequence of treatments where only the first-line medication is changed (TCZ vs MTX and TCZ vs anakinra) followed by (anakinra,) eternacept, adalimumab and abatacept	First line of the sequence was based on NICE scope. The subsequent treatments were based on prevailing clinical practice (clinical expert opinion).	Section 6.2.3 (page 214,2115)
Natural History	Based on the Markov model	Response rates of the MTX users in the placebo group of TENDER No deterioration in uncontrolled disease is assumed.	Section 6.3 (p. 222)
Treatment effectiveness	Treatment influences the probabilities of each ACR response and the withdrawal probability	Based on TENDER and indirect comparison	Section 6.3 (p. 222)

Adverse events	Not modelled	Manufacturer claims no differences in adverse events between treatments	Section 6.4.8 (p. 264) and 6.5.7 (p. 286)
Health related QoL	A CHAQ score is assigned to each of the four health state. The CHAQ score is mapped into utilities using a mapping formula derived in adult RA patients mapping HAQ onto EQ-5D utilities.	Manufacturer acknowledges that assumptions that CHAQ is equal to HAQ and adult EQ-5D is equal to HRQL of child has no evidence basis. It is solely due to lack of other available data that this mapping method is preferred for the analysis in order to derive QALYs for the economic model.	Sections 6.4.3 (p. 249) and 6.4.9 (p. 265)
Resource utilisation and costs	Treatment cost (depending on medication, age, administration, and wastage due to package size) and health state cost (accounting for hospital stays and specialist treatments, depending on the share of patients affected given current health state)	Based on UK reference costs, literature and expert opinion	Section 6.5.5 (p. 275) and 6.5.6 (p. 279)
Discount rates	A 3.5% discount rate was used for both costs and effects	According to NICE reference case	Section 6.2.6 (p.218)
Sub groups	No subgroup analysis undertaken	No justification given	Section 6.9 (p. 312)
Sensitivity analysis	One-way deterministic sensitivity analysis, scenario analyses and probabilistic sensitivity analysis	Ranges based on observed confidence intervals, expert opinion and assumptions	Section 6.6 (p. 287)

The ERG has assessed the manufacturer's economic evaluation using the Philips et al. checklist for quality assessing decision analytic models.[34] This is shown in Appendix 2 and is used to assist the narrative critique in the following sections.

5.2.1 NICE reference case checklist

Table 5.3: Comparison of the MS model with the NICE reference case

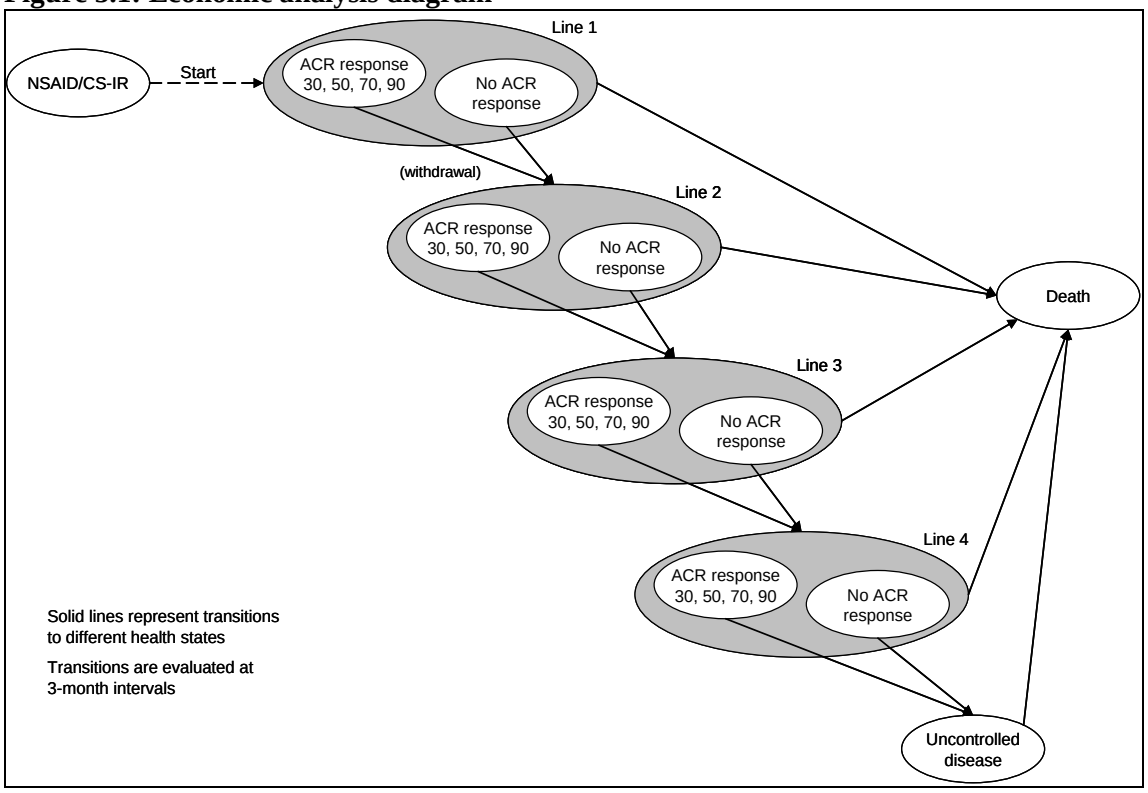
Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether de-novo evaluation meets requirements of NICE reference case
Comparator(s)	Therapies routinely used in the NHS, including technologies regarded as current best practice	Yes	
Type of economic evaluation	Cost-effectiveness analysis	Yes	
Perspective on costs	NHS and PSS	Yes	
Perspective on outcomes	All health effects on individuals	No	Adverse events are not modelled
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	Time horizon 16 years, i.e. age 2 to 18

Synthesis of evidence on outcomes	Systematic review	Unclear	A systematic review was performed, augmented with rapid reviews. Most model parameters were based only on TENDER trial.
Measure of health effects	QALYs	Yes	
Source of data for measurement of HRQL	Reported directly by patients and/or carers	Yes	CHAQ was used to map onto utilities
Source of preference data for valuation of changes in HRQL	Sample of public	Yes	
Discount rate	Annual rate of 3.5% on costs and health effects	Yes	
Equity weighting	No special weighting	Yes	
Sensitivity analysis	Probabilistic sensitivity analysis	Yes	

5.2.2 Model structure

The economic analysis employs a Markov model to evaluate costs and effectiveness of the compared strategies. Transitions to health states are evaluated at 12-week increments (cycles). The model outcomes are evaluated by cohort analysis. Half-cycle correction is applied to the model. A diagrammatic representation of the model is presented in the Figure below.

Figure 5.1: Economic analysis diagram



The Markov chain has in total 25 states. The model clusters the states into five groups: four groups are different lines of treatment and the fifth group contains the states death and uncontrolled disease. Each line of treatment consists of five health states. Those five states are ACR response at the 30, 50, 70, or

90 level and “no ACR response”. A patient can only move from a particular ACR response in his line to “no ACR response” in the next line or to death. From “no ACR response” the patient can only move to one ACR response level within this line of treatment or to “no ACR response” in the next line. The patient cannot move within a given line to a better or worse health state (say, from ACR 50 to ACR 70). Only after being through all four lines does a patient move to the health state “uncontrolled disease”.

The manufacturer states in response to the clarification letter that there is no evidence about transitions between ACR states within one medication line, except that for tocilizumab the proportion of ACR70 and ACR90 responders increases following the first 12 weeks.

The probability of a response/non-response within a line of treatment depends on the treatment. The order in which the treatment is applied does not change these transitions. The probability of death is treatment- and health state-independent, whereas the probability of withdrawal is health state independent but higher for MTX while being the same for all other treatment options. All transitions stay constant over time, i.e. are independent of age or disease duration.

In each cycle, the proportion of patients in a given state is calculated. The distribution across states is then used to calculate cycle-specific QALYs and treatment costs. Those are then discounted and summed up over the length of treatment.

Each health state leads to an absolute change in the initial CHAQ score. The initial CHAQ score is assumed to be the same for all patients, independent of treatment. For a given CHAQ score a utility is assigned to calculate QALYs.

The costs depend on the health state for the health-state costs and on the line of treatment for the treatment costs.

COMMENT

The ERG is of the opinion that the current economic model does not adhere to conventions in Markov modelling. In a Markov cohort model the health states defined should comprise the full range of conditions that are relevant to a patient population, and the states should be mutually exclusive. In the submission, the health states defined reflect a change in a patients’ condition (change in CHAQ based on ACR response) instead of the absolute condition of the patient. Change in a patients’ condition should be included in a Markov model as a health state transition and not as a health state as such. The consequence of using a change in a patients’ condition as a health state is that the Markov states are heterogeneous and not mutually exclusive regarding a patients’ condition, depending on the disease variation of the cohort at the start of the model. In the TENDER trial, there is substantial disease variation at baseline, as becomes clear from pages 90 to 96 in the MS. Table 5.4 shows this variation. From this, it is clear that, patients with exactly the same absolute health profile at 12 weeks based on the six components of the JIA core set can have an ACR30 response, an ACR70 response or be non-responsive, depending on the initial health profile. As a result, it makes no sense to assign one utility value and one health state cost estimate to an ACR30 response or an ACR50 response, given that the patient in ACR30 might actually be in a better absolute health state.

Table 5.4 Baseline values JIA core set

JIA Core Set component	Baseline value tocilizumab, all patients Mean ± SD	Baseline value placebo Mean ± SD
No. of active joints	21.3 ± 15.9	16.9 ± 12.9
No. of joints with limitation of movement	20.7 ± 15.9	17.9 ± 15.9
Patient/parent global assessment VAS	60.3 ± 23.8	56.3 ± 21.2
Physician global assessment VAS	69.6 ± 15.7	61.4 ± 21.1
CHAQ-DI score	1.7 ± 0.8	1.7 ± 0.8
ESR	57.6 ± 31.2	54.1 ± 35.4

The ERG requested in the clarification letter that the manufacturer would provide a new model using absolute health states defined by for example CHAQ values (a measure of disease manifestation). In their response, the manufacturer stated that:

> *"The economic model uses absolute CHAQ scores as recommended above. The methods of the model adhere to the Markov modelling conventions of health state cohort homogeneity. This is explained in sections 6.2.4, and 6.3.8 of the MS, where Roche presents that ACR is a relative measure and not considered as an absolute measure of health status. Utility values are assigned to the cohort, based on state membership and according to CHAQ score–an absolute score derived after response to treatment."*

However, while it is true that each health state was assigned an absolute CHAQ score, this score was simply the average of the patients in each ACR response group. Thus, the wide variation in CHAQ score at base line, i.e. 1.7 ± 0.8, was reduced to a single value of 1.7, and the same was done for each ACR response groups. But this means that patients with very different observed CHAQ scores are now forced to be (or better, appear as) homogeneous by assigning an average value for the CHAQ score. The correct approach, if one wants to define health states based on CHAQ score, is to define ranges of the score, with for instance four ranges: <0.75, 0.75-1, 1-1.5, >1.5. (Note that this is a purely hypothetical definition of ranges) Now each patient is classified at baseline and at 12 weeks into a health state based on the observed CHAQ of that patient. Other options are also possible, for instance using some absolute ACR score based on a function of the six components of the core set. The purpose of this is to generate homogeneous health states where patients truly have approximately the same health status, the same quality of life and the same resource use.

The manufacturer assumed for the model, due to lack of data, that patients that move to a certain ACR response stay in that state until the patient either withdraws (i.e. moves to the next treatment line) or dies. Given the nature of the disease, this is an unlikely assumption. The ERG has tested the impact of the assumption through a scenario analysis (see section 5.3)

In the model, no adverse events have been modelled. However, from the MS (page 180, table 53) it is clear that there are differences between treatments in terms of adverse events. The TENDER trial showed more infection and serious infections in the tocilizumab group compared to the placebo

group. Literature review showed that the various anti-TNF alphas are also associated with serious infections. Inclusion of adverse events would have impact on the comparison of tocilizumab versus MTX, as it would decrease the effects of tocilizumab and increase the costs. However, given that the ERG has indicated that the comparison of tocilizumab versus MTX is not appropriate for addressing the first population of the decision problem, the ERG has not explored the impact of adverse events in this comparison. (See Chapter 3 and section 5.2.3 for more details on the population considered) For tocilizumab versus biologics (the second population) the ERG accepts the exclusion of adverse events, as they seem to occur approximately equally between the various treatments.

Another important assumption made in the model is that the effectiveness of the second, third and fourth line treatments are independent of treatment history. However, especially for the class of anti-TNF alphas, it seems reasonable to assume that patients that did not respond to one treatment have a smaller probability of response to the next treatment than patients who had not used the first treatment before. However, given that patients receiving tocilizumab stay longer on their first line treatment, the assumption of independent effectiveness is in favour of the anakinra treatment sequence.

The model further assumes that the risk of withdrawal is constant for all ACR states. This assumption may be less realistic, as it seems more likely that a patient with an ACR30 response will withdraw than a patient with a ACR90 response. However, the ERG acknowledges that there is not enough data available for health state specific withdrawal rates. Additionally, this assumption is in favour of the control group, since more patients in the tocilizumab group achieve higher ACR responses.

A minor issue concerns the cycle length in the model. In the electronic model, a cycle length of 3 months is used. However, the MS states that the cycle length is 12 weeks, in line with the duration of the TENDER trial, and the transition probabilities used in the model are all based on the 12-week cycle length. The ERG has corrected this (see section 5.3).

5.2.3 Population

The manufacturer states (MS p.211) that there are two populations for which tocilizumab is licensed:

1. Children and young people 2 years and older with systemic JIA which has not responded adequately to prior NSAID(s) and systemic corticosteroids.
2. Children and young people 2 years and older with systemic JIA which has not responded adequately to prior NSAID(s), systemic corticosteroids and methotrexate.

They then state:
'The first group of patients reflects all patients in the TENDER trial. The second population is represented by the majority of patients in the TENDER trial.'

They go on to state: '…the second population is a subgroup of the first…'. They explain that 95% of TENDER trial patients had either tried MTX or were still on MTX and that, due to presence of active disease, could be presumed to belong to population 2. On this basis they have two sets of comparisons, one for population 1, which allows comparison of tocilizumab with MTX and one for population 2, which does not allow comparison with MTX.

The age range in the TENDER trial is from 2 to 17 and the median age is 10 years. In the economic evaluation, all patients have a starting age of 2 years.

COMMENT

The ERG believes that the definition of population 1, which came from the scope was meant to mean those who had not responded adequately to only NSAIDs and corticosteroids and not MTX, as opposed to those who had not responded adequately to NSAIDs and corticosteroids, some of whom who had and some of whom who had not responded adequately to MTX. Another way of expressing this is that population 1 are all MTX naïve and population 2 have all tried MTX and inadequately responded to MTX. Therefore, populations 1 and 2 are mutually exclusive and population 2 is not a sub-set of population 1. It is therefore misleading to state that 'the first group of patients [population 1] reflects all patients in the TENDER trial.' Since 95% are deemed to have not responded adequately to MTX then, although they might also have not responded adequately to NSAIDs and corticosteroids, they would in fact belong only to population 2. Indeed, only the 5% that are MTX naïve would belong to population 1.

Thus, according to the ERG (see also Chapter 3), even when following the MS approach to the decision problem, no data were provided for population 1; for population 2, data from the TENDER trial (minus the 5% MTX-naïve patients) can be used for an indirect comparison with anakinra. Consequently, the ERG will, in the remainder of this chapter, only discuss the economic evaluation for population 2.

Additionally, the ERG notices that the modelled patient population does not take into account the heterogeneity in the patient population of the clinical trial by starting the whole cohort at age 2. Scenario analyses show that a higher initial patient age has a substantial and negative effect on cost-effectiveness. Based on literature, the ERG considers an average starting age of 7 more valid. The details of this can be found in sections 5.2.10 and 5.3.

5.2.4 Interventions and comparators

The manufacturer submission aims to compare the intervention tocilizumab with two comparators, that is, one comparator for each of the two populations mentioned in the previous section. Both intervention and comparators are sequence of treatments. This sequence approach accounts for the fact that some patients may not show any response or treatments lose efficacy and, hence, a treatment is always part of a line of treatments.

Table 5.5 Sequence of treatments for population 1

	Strategy A	**Strategy B**
Line 1	Tocilizumab	Methotrexate
Line 2	Anakinra	Anakinra
Line 3	Etanercept	Etanercept
Line 4	Adalimumab	Adalimumab
Uncontrolled disease		

Table 5.6 Sequence of treatments for population 2

	Strategy A	**Strategy B**
Line 1	Tocilizumab	Anakinra
Line 2	Etanercept	Etanercept
Line 3	Adalimumab	Adalimumab
Line 4	Abatacept	Abatacept
Uncontrolled disease		

The intervention, tocilizumab, is indicated for the treatment of active sJIA in patients 2 years of age and older, who have responded inadequately to previous therapy with NSAIDs and systemic corticosteroids. Tocilizumab can be given as monotherapy (in case of intolerance to MTX or where treatment with MTX is inappropriate) or in combination with MTX. (RoActemra draft Summary of Product Characteristics – Anticipated June 2011). Tocilizumab is a solution for infusion with a dosing frequency of once every two weeks. The recommended posology is 8 mg/kg in patients weighing greater than or equal to 30 kg or 12 mg/kg in patients weighing less than 30 kg.

The order in which the various treatment options after the first line failed appear is mainly based on expert opinion (see MS page 214)

COMMENT

As mentioned in section 4.2.5, for population 1 the comparator in the scope is methotrexate, yet the trial the MS relies upon is the TENDER trial which compares tocilizumab plus standard care versus placebo plus standard care. Whilst the comparator in this study was placebo, 70% of patients at baseline were also receiving methotrexate. The MS uses a post-hoc analysis to compare tocilizumab with those patients in the placebo group also receiving methotrexate. However, this is not an acceptable comparison because trial participants were not randomised in this way. Therefore this comparison is based on observational data from two different populations, those using MTX and those not using MTX (see Section 3 for details). More importantly, following the TENDER inclusion criteria for active disease and the MS approach which claims active disease status despite MTX therapy equates to inadequate response, means the effects of tocilizumab were compared with MTX in a population that was specifically selected as being not responsive to MTX. This makes a valid modelling of the cost-effectiveness impossible.

For the second population the comparators are TNF-inhibitors/anakinra. There is also a mismatch between scope and the decision problem for population 2. For those patients with an inadequate response to NSAIDs, CS and methotrexate the comparators should be TNF inhibitors (e.g. etanercept, adalumimab, abatacept, infliximab) and anakinra. There is no direct evidence presented for these comparators and hence indirect comparisons are made. For tocilizumab, evidence from the TENDER trial is used, but instead of using data from those patients described by the manufacturer as relevant to the decision problem (95% of included patients), data for all patients were used in the indirect comparison and the economic evaluation (see Chapter 3 for details).

The ERG, for their additional analyses, therefore used data from the 95% of patients in the TENDER trial who were not MTX naïve, as provided by the manufacturer in their response to the clarification letter for the ERG defined base case which will be presented in section 5.3.

For the comparators, the manufacturer decided to broaden the inclusion criteria (see MS, page 116) to identify all pivotal trials in juvenile arthritis regardless of subtype. This was despite advice from their clinical experts to the contrary (see MS, page 116) based on the differences between a systemic JIA population and other subtypes. Thus, in the review of the clinical evidence the ERG has, in line with the advice from the clinical experts, chosen to ignore the trials in children with other types of juvenile arthritis in this report.

For the economic evaluation, the manufacturer has attempted to correct for the mismatch between the TNF-inhibitors' population (any JIA) and the population relevant in this evaluation (sJIA) by

adjusting the ACR response rates for the TNF-inhibitors based on literature (see also section 5.2.6.2, table 5.9).[35]

Given that the scope and decision problem include comparison with any TNF inhibitor and sequences are appropriate, one would have expected a comparison of potentially many combinations of treatments starting with tocilizumab versus starting with some other treatment (anakinra, infliximab, adalimumab, abatacept and etanercept). However, the MS only includes an incremental analysis of one comparison for population 2 i.e. tocilizumab versus anakinra each followed by etanercept, adalimumab and abatacept. A comparison with a sequence where anakinra and etanercept were switched was also made, but only with tocilizumab first instead of a full incremental analysis of all three options.

Some explanation was given for excluding some sequences, but there remained much arbitrariness and some lack of logic e.g. infliximab was excluded as a treatment option in the sequences, but a trial comparing it to placebo was used as the source of effect for all other TNF inhibitors. Given that there was only one source of effect and therefore TNF inhibitors only differed in terms of cost, if other sequences were allowed, then a comparison with the cheapest would be useful. Therefore, the ERG performed a full incremental analysis, assuming infliximab could not be used i.e. comparing starting with tocilizumab vs. anakinra (followed by etanercept, adalimumab and abatacept) vs. etanercept (followed by anakinra, adalimumab and abatacept). Comparison was also made with sequences that include infliximab, i.e. starting with infliximab (followed by etanercept, adalimumab and abatacept) and starting with anakinra (followed by infliximab, adalimumab and abatacept). These results will be discussed in section 5.3.

5.2.5 Perspective, time horizon and discounting

The MS model has a time horizon of 16 years. That means that a patient in the model starting treatment at age 2, turns 18 and can be considered an adult at the end of the simulation period. The model allows shorter and longer time durations for sensitivity analysis (up to 30 years). The discount rates applied were 3.5% for utilities and costs and costs are considered from an NHS and Personal Social Services perspective.

COMMENT

The ERG concludes that the discount rates and perspectives are in line with the NICE reference case. The time horizon used (until patients become adult) is adequate. Given the uncertainty in how the disease will develop in adulthood any extrapolation beyond childhood is uncertain. However, lengthening the time horizon, assuming continued use of the same treatment as used in childhood, did not alter the conclusion.

5.2.6 Treatment effectiveness

The transitions in the model can be divided into those that are treatment independent and those that are treatment dependent. We will first discuss the treatment independent set of parameters and then the treatment dependent set.

5.2.6.1 Treatment independent parameters

Baseline patient characteristics

The starting age in the model is 2 years. The patients started all with the same weight of 13.25kg that increase over time to 62.5kg until the age of 18.

Mortality Risk

Due to lack of data to determine elevated mortality risk dependent on the patient condition, the model assumes a constant mortality risk based on evidence from Hashkes et al.[36] Hashkes and colleagues studied records of a paediatric registry in the US (49,000 patients) and provide evidence on the observed survival of patients by subtype of JIA. The study reports that systemic disease is a predictor of higher mortality risk. Nevertheless, the standardised mortality ratio (SMR) reported on 962 patients (14 years follow-up) with systemic JIA is not statistically significant (p=0.15), as there were only 21 events. Evidence from a figure (see figure 15 in MS) of the observed mortality in sJIA patients, as reported by Hashkes et al.,[36] was extracted and a constant annual risk of mortality was estimated; 0.07% (R^2=0.8656). This risk is assumed to apply across all treatments and all health states and is included in the analysis for completion, without having an impact on the incremental cost-effectiveness results.

The mortality risk could be expressed as an age-dependent probability based on UK life tables. However, given the availability of evidence on mortality of the analysis population from a recent study on systemic JIA patients it was deemed more appropriate by the manufacturer to use the disease-specific risk.

To assess the impact of the mortality rate on the outcomes, the annual mortality risk is assumed to have a range between 0.000076 (lowest risk in UK population for 2-18 year old patients)[37] to 0.001324 (apply equal difference over 0.0007).

The one-year mortality risk was then adjusted for the duration of the cycle, using the following formula: $P_j = 1 - [1 - P(t_0, t_j))]^{1/j}$ where j represents the number of equal time intervals[38]. The 12-week mortality rate thus estimated is 0.000162 per 12 weeks.

In the clarification letter, the ERG requested information on other models considered to fit the observed data. In their response, the manufacturer stated that both exponential and linear parameterizations were fitted to the mortality data.

Linear model
Survival= -0.0007*years + 1
SE around the time (-0.0007) parameter=0.000015
R^2=0.959

Exponential model
Survival= exp(-0.0007*years)
SE around the time (-0.0007) parameter=0.000015
R^2=0.958

The two models have very similar results, and the manufacturer stated that the linear model is preferred for the economic analysis.

5.2.6.2 Treatment dependent parameters

ACR Response

The ACR response is a measure of relative improvement in the current health state attributed to the medication. The model includes 4 different levels of ACR responses (30, 50, 70, 90). When a new treatment is started the simulated cohort is distributed among the four possible ACR responses or to the health state ACR-non-response (in order to start the next line of treatment) according to the transition probabilities given for each treatment.

For tocilizumab, evidence from the TENDER clinical trial is used for the transitions probabilities for ACR 30, ACR 50, ACR 70, ACR 90 response and no-ACR 30 response. The proportion of the cohort that falls within each response category is obtained by adjusting the reported response rates (table 5.7) in order to ensure the categories are non-overlapping. The transformation involves a simple re-expression of these proportions in non-overlapping categories. To obtain the final ACR probability input (table 5.8) and avoid the double-counting of patients, the following method was utilised:

1. the proportion of ACR 30 includes ACR 50, ACR 70, and ACR 90
2. subtract from ACR 30 the sum of the other 3 categories
3. continue the same with ACR 50, and ACR 70

For anakinra, the result of the indirect comparison is used for the transition probability for ACR 30. The RR of ACR 30 for tocilizumab versus anakinra is used to derive the probability of ACR 30 with anakinra treatment, given the probability of ACR 30 with tocilizumab treatment (see Table below).
The ANAJIS study[39] reports evidence for 1 month. The ACR 30 response probability from the indirect comparison could be further adjusted to reflect projected response to anakinra at 12 weeks. The manufacturer remarks that a degradation of ACR 30 response probability for anakinra for longer than the trial duration is supported by the reported evidence: around 50% of responders to the modified ACR 30 lost response after 1 month. However, in this model, due to lack of evidence to perform this adjustment, the base case analysis assumes no degradation for the anakinra ACR 30 response. This assumption is conservative against tocilizumab since further degradation of response for anakinra would improve the cost-effectiveness results.

The probabilities of ACR 50, ACR 70 and ACR 90 are derived by applying to the probability of ACR 30 with anakinra the relative difference of ACR 50 to ACR 30, ACR 70 to ACR 30, and ACR 90 to ACR 30 with tocilizumab (see Table below).

Table 5.7: ACR evidence comparison TCZ vs. ANK

	Tocilizumab	Anakinra	Anakinra rates adjustment
ACR 30	0.907	0.3827	Based on RR= 2.37 (section 5.7)
ACR 50	0.853	0.3599	Adjusted based on difference ACR 50-30 TCZ
ACR 70	0.707	0.2983	Adjusted based on difference ACR 70-30 TCZ
ACR 90	0.373	0.1574	Adjusted based on difference ACR 90-30 TCZ

ACR response: American college rheumatology response criteria, ANK: anakinra, RR: Relative risk TCZ, tocilizumab

Table 5.8: ACR probabilities: comparison TCZ vs. ANK

	Tocilizumab	Anakinra	PSA distribution
pACR NR	0.093	0.617	Dirichlet
pACR 30	0.054	0.023	
pACR 50	0.146	0.062	
pACR 70	0.334	0.141	
pACR 90	0.373	0.157	

ACR response: American college rheumatology response criteria,
ANK: anakinra, RR: Relative risk TCZ, tocilizumab

The ACR response of the other biologics (abatacept, infliximab, etanercept, adalimumab) is derived by the indirect comparison of TENDER versus Ruperto et al.[40] Ruperto et al. concerns a clinical study of infliximab versus placebo. The assumption is made that all other biologics have the same response rate as infliximab. Similarly to the above, the relative risk is used to derive the probability of ACR 30, ACR 50 and ACR 70 for the other biologics. The probability of ACR 90 is derived by applying to the probability of ACR 50 with biologic the relative difference of ACR 90 to ACR 50 with tocilizumab (see Table below).

However, the ACR response in Ruperto et al.[40] reflects a JIA population of which only 16% are systemic JIA patients. Thus, the indirect comparison results are further adjusted for the differences in the population subtypes. Data from an observational study by Prince et al.[35] are used to correct ACR response rates of biologics. Prince and colleagues report long-term efficacy data of patients using etanercept from a Dutch registry. The study reports evidence from 146 patients, out of which 27% were systemic JIA. The adjustment factor consists of the difference in the proportion of responders between the total population and the systemic JIA patients (see Table below).

Table 5.9: Prince et al. 2009 adjustment

	Total patients	sJIA	Adjustment factor	Reference
ACR 30	79%	59%	0.75	Prince et al. 2009
ACR 50	67%	43%	0.65	Prince et al. 2009
ACR 70	51%	27%	0.53	Prince et al. 2009

ACR response: American College Rheumatology response criteria

Table 5.10: ACR evidence comparison TCZ vs. biologics

	Tocilizumab	Biologics	Biologics rates adjustment
ACR 30	0.907	0.238	Based on RR= 2.87 (section 5.7) with Prince 2009 adjustment
ACR 50	0.853	0.103	Based on RR= 5.35 (section 5.7) with Prince 2009 adjustment
ACR 70	0.707	0.082	Based on RR= 4.61 (section 5.7) with Prince 2009 adjustment
ACR 90	0.373	0.045	Adjusted based on difference ACR 90-50 TCZ

ACR response: American College Rheumatology response criteria, RR: relative risk

Table 5.11: ACR probabilities: comparison TCZ vs. biologics

	Tocilizumab	Biologics	PSA distribution
pACR NR	0.093	0.762	Dirichlet
pACR 30	0.054	0.135	
pACR 50	0.146	0.021	
pACR 70	0.334	0.037	
pACR 90	0.373	0.045	

ACR response: American College Rheumatology response criteria, PSA: Probabilistic sensitivity analysis

Probabilities of withdrawal from treatment

Withdrawal risk given a response to treatment, i.e. from an ACR response health state, does not dependent on treatment except for methotrexate (which has a three times higher withdrawal rate than all the other treatments)

For MTX treatment, the most recent evidence is from Ruperto et al.[41] and Woo et al.[18] Evidence from Woo et al.[18] is selected for the base case analysis as it reflects data for the systemic subtype populations. Woo et al.[18] is a crossover study, in which it is reported that over a period of 4 months 4 patients out of a sample of 45 withdrew due to exacerbation of systemic condition. The withdrawal risk assumed is 9% for 4 month duration.

For biologic treatments the review identified a number of sources with evidence on withdrawal; the majority of them based on treatment with etanercept. Due to insufficient evidence to differentiate between treatments, the economic analysis assumes the same withdrawal risk across all biologics. The base case analysis assumes the annual risk of withdrawal is based on Lovell et al.[42] Lovell et al.[42] was selected as the most relevant evidence given duration of the observational data (8 years). Figure 5.2 presents the observed data. A constant risk of withdrawal over a year is assumed to be 9.48% (R2=0.70). An annual risk of withdrawal around 10% is also supported by clinical expert opinion [PC: Wright S. 16/03/2011, Woo P. 21/03/2011, Baildam E 28/03/2011].

However, in the clarification letter, the ERG requested information on other models considered to fit the observed data. In their response, the manufacturer stated that both exponential and Weibull parameterizations were fitted to the withdrawal data. The AIC criterion showed that the exponential model was the most appropriate (AIC 177.05 versus 178.74). Using a standard linear regression approach, the exponential model was then compared to a simple linear model using the R2 criterion which measures the goodness of fit of the model to the data. Results are shown below:

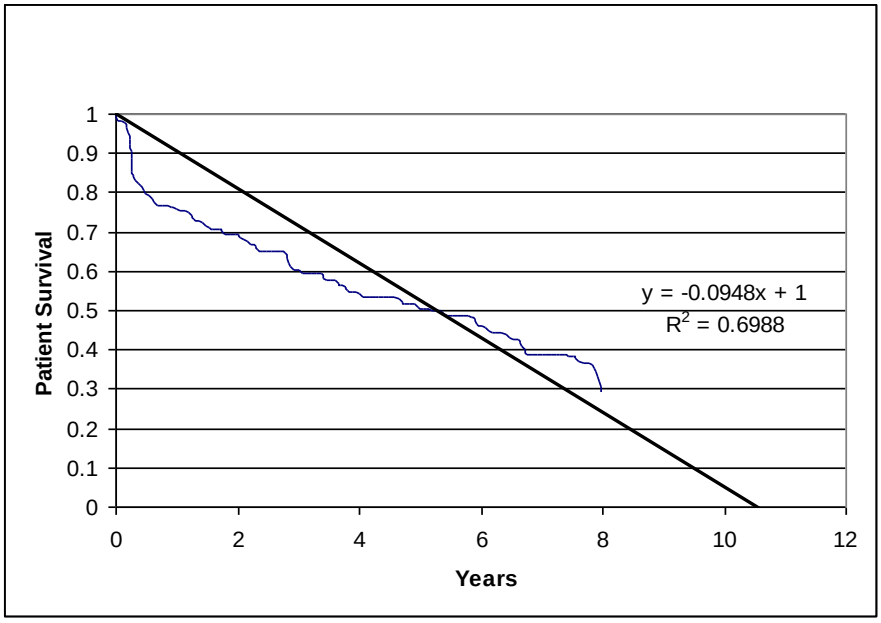
Linear model
Equation: survival= -0.0948*years + 1
SE around the time parameter: 0.0018
R^2=0.954
The intercept term is set to 1 to ensure that survival is 1 at the time origin and therefore does not have a precision associated with it.

Exponential model
Equation : survival= exp(-0.138*years)
SE around the time parameter: 0.0016

R^2=0.982

From this, it follows that the performances of the two models are very similar. Roche identified in the MS (sections 6.3.3 and 6.10.3) that a time-dependent probability (exponential in this case) would be accurate for the withdrawal risk. However, given the evidence on withdrawal risk with MTX and other biologics, it was not possible to differentiate greatly across treatments on this parameter. It was deemed more appropriate from a model parsimony view to use a constant risk (R2=0.954).

Figure 5.2: Observed withdrawal risk for biologic treatments plus linear model



The one-year withdrawal risk was then adjusted for the duration of the cycle, using the following formula: $P_j = 1 - [1 - P(t_0, t_j))]^{1/j}$ where j represents the number of equal time intervals.[38]

Table 5.12 Withdrawal probabilities

	Reported risk	Adjusted risk
MTX	0.0889 per 4 months	0.0674 per 12 weeks
All biologics	0.0948 per year	0.0227 per 12 weeks

MTX: methotrexate

COMMENT

As mentioned in earlier sections, an important issue is the population on which the ACR responses are based. The ERG does not agree with the approach in the MS to compare the whole tocilizumab group (regardless of MTX use) with only the MTX users in the placebo group. Therefore all of the analysis and results for population 1, for which MTX is a comparator, is considered by the ERG to be invalid (i.e. tocilizumab cannot be compared to MTX) and therefore analyses based on that population will not be discussed here. The ERG only considers that the analysis and results for population 2 i.e. those with an inadequate response to MTX to be valid and only given the assumption that the 95% of

patients in the TENDER trial who were not MTX-naïve can be considered to have had an inadequate response to MTX.
Thus, for the comparison of tocilizumab versus anakinra, the ERG does not agree with the use of the whole tocilizumab group, as this group included 5% MTX naïve patients. In their response to the clarification letter, the manufacturer provided ACR response data for the 95% patients in the tocilizumab group that were not MTX-naïve. These numbers should be considered as the base case values.

Table 5.13 12 week ACR responses for the TENDER trial

	Tocilizumab, all patients	**Tocilizumab, non MTX naïve (95%)**
ACR 30	0.907	0.907
ACR 50	0.853	0.840
ACR 70	0.707	0.694
ACR 90	0.373	0.360

Additionally, the manufacturer provided in their response to the clarification letter the results of the indirect comparisons based on only non-MTX-naïve patients (see table 5.14).

Table 5.14: Results of the indirect comparison analysis

Comparison	**Outcome**	**Base-case analysis (TENDER)**		**Sensitivity analysis (excl. MTX naive)**	
		RR	**95% CI**	**RR**	**95% CI**
TCZ vs. ANK	ACR30	2.37	1.10, 5.10	2.27	1.06, 4.85
TCZ vs. INF	ACR30	2.87	1.49, 5.55	2.75	1.44, 5.26
	ACR50	5.35	1.91, 14.97	5.04	1.81, 14.04
	ACR70	4.61	1.16, 18.38	4.33	1.09, 17.20

Based on this, the following transition probabilities for tocilizumab, anakinra and biologic were derived by the ERG:

Table 5.15 ACR transition probabilities

	Tocilizumab non MTX naïve	**Anakinra**	**Biologics**
pACR NR	0.093	0.600	0.753
pACR 30	0.067	0.030	0.139
pACR 50	0.146	0.064	0.023
pACR 70	0.334	0.147	0.039
pACR 90	0.36	0.159	0.046

It is important to realize that the ACR transition probabilities are fully reliant on rather strong assumptions. First it is assumed that infliximab, etanercept, adalimumab and abatacept are all equally effective in reaching ACR responses. It is surprising that the manufacturer is making this assumption, given the fact that they are excluding infliximab as a treatment option because "it is not

recommended for JIA patients due to insufficient evidence, nor is it widely used by clinicians as a treatment on systemic JIA patients after NSAID-IR or CS-IR and/or MTX-IR" (see MS page 215). The ERG considers this an inconsistent approach.

Second, it is assumed that the relationship between response in all JIA and sJIA in patients using etanercept is also applicable to infliximab, adalimumab and abatacept. Combined with the first assumption this leads to the assumption that all biologics have the same response rate in sJIA patients. The question arises how realistic this is, given that in the recently publised NICE report on biologics in psoriatic arthritis a clear difference between biologics was seen.[43]

The manufacturer uses a constant risk of withdrawal in the model, based on a linear survival function. However, a linear survival function does not imply a constant risk over time. With the curve presented by the manufacturer, the probability of withdrawal in year 1 would be 9.48% but would increase in year 10 to 65%. To arrive at a constant risk, an exponential survival curve would have to be assumed. As the manufacturer states in their response to the clarification letter, the exponential survival function actually had a (slightly) better fit to the observed curve than the linear function. Thus, the ERG proposes to use that function: $S(t)=e^{(-0.138*years)}$

From this, we derive a constant withdrawal risk of 12.9%, which is then transformed into a 12-week probability of 3.13%. This withdrawal rate will be part of the new ERG base case. The same error is also made for the mortality risk. However, in this instance the linear approach and the exponential approach lead to approximately the same 12-week transition probability.

5.2.7 Health related quality of life

The TENDER study included the Child Health Questionnaire (CHQ) as an instrument eliciting patient HRQL. The CHQ assesses a child's physical, emotional, and social well-being from the perspective of a parent or guardian (CHQ-PF50). The questionnaire was completed twice during the randomised period of the study: at baseline (visit 1) and at week 12 (visit 7).

During the design of the economic evaluation a number of methods were attempted to translate CHQ scores to QALYs for the model. However, due to lack of robust data and many other limitations that are briefly described below, an alternative method to provide QALYs was selected.

The limitations of the use of CHQ are:

- Missing data; when evaluating the CHQ scores of patients in the dataset a number of missing values were identified: 23 (21%) at baseline and 29 (26%) at week 12. This data is not missing at random. Since missing values were not at random, none of the established methods of value imputation or using a method of available cases only were deemed appropriate here.
- Lack of mapping formula in the literature; a literature review did not identify any method that could provide mapping of CHQ to QALY.
- Lack of mapping formula from the trial; in previous cases in adult RA data from the trial was used to derive a mapping formula between EQ-5D and HAQ scores. The TENDER trial does not include any data on HUI or EQ-5D to perform a similar analysis for sJIA affected patients.
- Separate physiological and psychosocial scores; the inclusion of CHQ in the database would need to differentiate between physiological and psychosocial data.

Instead a new mapping approach was developed, using CHAQ. This mapping formula is derived from an adult RA population, and therefore, the use of this formula implies the following assumptions;

- CHAQ of child is equal to HAQ score of adult
- Adult EQ-5D is equal to the HRQL of a child

The manufacturer acknowledges that the above assumptions have no evidence basis. It is solely due to lack of other available data that this mapping method is preferred for the analysis in order to derive QALYs for the economic model.

The mapping formula for the base case analysis is using data from a pool of two tocilizumab trials of adult RA patients (OPTION and LITHE, N=1800). The method is briefly described below.

HAQ scores were regressed on EQ-5D utility data using a linear mixed model. The significance of coefficients for HAQ and the square of HAQ were tested and the fit of strictly linear and non-linear models were compared.

Results showed that a linear model generated coefficient estimates similar to those reported by Boggs and colleagues[44]:

EQ5D =0.89 – 0.28*HAQ (p<.0001)

Consistent with Boggs and colleagues[44], inclusion of a model term for the square of the HAQ score resulted in an improved fit and a significant coefficient for the non-linear term.

Table 5.16: HAQ / EQ-5D mapping formula

Coefficient	Estimate	Standard Error	Pr>t
Intercept	0.8229	0.008621	<0.0001
HAQ	-0.1125	0.01360	<0.0001
HAQ_SQ	-0.06874	0.005200	<0.0001

HAQ: Health assessment questionnaire

Analysis suggested that the model with the squared term model has a better 'fit' and hence was selected to inform the base-case model.

This non-linear formula reflects the assumption that decreases in CHAQ level are more valuable (as measured by change in utility) for severely disabled patients than for patients who are less disabled. Sensitivity analysis tested alternative utility mapping scenarios assuming linear relationship between the two model parameters:

1. Roche quadratic (base case): HRQL=0.82-0.11*CHAQ-0.07*(CHAQ^2)
2. NICE quadratic [45]: HRQL=0.804-0.203*CHAQ-0.045*(CHAQ^2)
3. Roche linear: HRQL=0.89 – 0.28*CHAQ
4. Boggs et al.[44] linear: HRQL= 0.76 - 0.28*CHAQ+0.05*Female

In the MS (section 6.2.4), it is stated that health states reflect the condition of patients dependent on ACR response after a 12 week period on treatment. The utility of the health state is characterised by the resulting CHAQ triggered by the ACR response. Based on the average CHAQ score per ACR

response state, combined with the above presented mapping formula, utilities for each health state are presented in Table 5.17.

Table 5.17: Summary of quality-of-life values for cost-effectiveness analysis

Health state name	CHAQ	Assumed QoL	Assumed SE	Adult RA values (for reference)
No response or uncontrolled disease	1.7442	0.4152	30% of the mean	0.4651
ACR 30	1.2699	0.5674		0.5660*
ACR 50	1.1351	0.6050		0.6084
ACR 70	0.8601	0.6736		0.6289
ACR 90	0.6692	0.7150		N/A

*refers to ACR 20 and not ACR 30 in adult RA

COMMENT

The ERG acknowledges that, due to lack of data, both in the trial and in the literature, very strong assumptions are required in order to assign a utility to each health state in the model, and the ERG considers the approach taken by the manufacturer reasonable and acceptable.

The ERG explored how reasonable the assumed standard error of 30% of the mean is for use in the PSA. We found that combining the uncertainty in baseline CHAQ with the uncertainty around the parameter estimates of the mapping formula led to a standard error of less than 10% of the mean. The ERG considers the 30% uncertainty used in the model reasonable, since it also takes the additional uncertainty due to the assumptions required for the mapping procedure into account.

It should be noticed here that, without explicitly mentioning this in the MS, the manufacturer assumes that the CHAQ scores per ACR response found in the TENDER trial are also valid for the anakinra and biologic. Whether this is a reasonable assumption is difficult to judge. The baseline CHAQ score in the anakinra study is 1.55 (SD 0.74) and in the infliximab study (used for the indirect comparison of biologic) the baseline CHAQ score was 1.5 (SD 0.7).[39, 40]

The baseline score in the anakinra group is slightly lower than that observed in the TENDER trial, while the baseline score in the infliximab study is much lower. The latter reflects the fact that the infliximab study was done in the whole JIA population instead of the sJIA population.

5.2.8 Resources and costs

5.2.8.1 Costs of treatment

Treatment costs are a composite of the cost of the medication and the cost for administering the medication. In some cases the required dosage depends on the body weight of the patient. Table 5.18 presents data on drug acquisition cost, cost of administration, and dosage. Table 5.19 presents the resulting treatment costs per year for each treatment.

Table 5.18: Overview of treatment administration for selected ages

		Abatacept	Adalimumab	Anakinra	Etanercept	Infliximab	Tocilizumab
	Maintenance dose	Every 4 weeks: 10 mg/kg when less than 75 kg and 6-17 years; 750mg over 75kgs and <100kgs if over 17 years	40 mg every other week	2mg/kg	0.4mg/kg twice a week (max 25mg)	3mg/kg every 8 weeks	12 mg/kg for patients < 30 kg; 8 mg/kg for patients ≥ 30 kg) and administered intravenously (IV) every two weeks
	Number of administrations per year	13	26	364	104	6.5	26
	Drug unit cost	£ 242.17 per 250 mg	£357.5 per 40 mg	£26.23 per 100 mg	£83.38 per 25mg	£83.38 per 25mg	£102.4 per 80mg, £256 per 200mg and £512 per 400mg
	Adminstration cost	£150	Self administration. 20% < 10 year and 10%>10 year require nurse visit. £ 13 per home visit				£150
Age	**Weight**	**Dosage per administration (in mg)**					
2	13.25	132.5	40	26.5	5.3	39.75	159
5	18.5	185	40	37	7.4	55.5	222
8	25	250	40	50	10	75	300
10	31	310	40	62	12.4	93	248
12	39.25	392.5	40	78.5	15.7	117.75	314
14	49	490	40	98	19.6	147	392
16	55.75	557.5	40	111.5	22.3	167.25	446
18	62.5	750	40	125	25	187.5	500

Table 5.19 Cost per year for a treatment assume wastage due to varying package size (mean values)

Age	Abatacept	Adalimumab	Anakinra	Etanercept	Infliximab	Tocilizumab
2	£ 5,124	£ 9,383	£ 10,494	£ 9,566	£ 3,701	£ 10,570
5	£ 5,124	£ 9,383	£ 10,494	£ 9,566	£ 3,701	£ 13,233
8	£ 8,273	£ 9,383	£ 10,494	£ 9,566	£ 3,701	£ 17,226
10	£ 8,273	£ 9,349	£ 10,021	£ 9,431	£ 3,701	£ 13,233
12	£ 8,273	£ 9,349	£ 10,021	£ 9,431	£ 6,428	£ 17,226
14	£ 11,421	£ 9,349	£ 19,569	£ 9,431	£ 6,428	£ 19,889
16	£ 11,421	£ 9,349	£ 19,569	£ 9,431	£ 6,428	£ 19,889

18	£	11,421	£	9,349	£	19,569	£	9,431	£	6,428	£	23,882

5.2.8.2 Health-state costs

A resource use schedule for a JIA patient was identified by Epps et al.[27] and modified for the current economic analysis. The list of resources was augmented by a similar schedule by Thornton et al.[32, 33] and Barton et al.[46] Evidence from Thornton and colleagues [2008a and 2008b] was reviewed but disposed in favour of expert opinion.

In order to establish the differences between the resource use for each health state, items from the combined cost schedule were presented to clinical experts in structured interviews to determine (see also section 6.3.5 MS):

- the proportion of patients that make use of a resource item
- the frequency of use.

Several items were excluded from the Epps et al.[27] and Thornton et al.[32, 33] list. An outpatient visit for ear, nose and throat check, as well as a visit to an ophthalmologist were excluded as not relevant for systemic patients based on clinical opinion [PC: Westhovens R 02/03/2011, Wright S 16/03/2011, Woo P 21/03/2011, Baildam E 28/03/2011]. A visit to orthopaedic surgeon was excluded because it was considered very rare (once in a lifetime for ACR 30 responders and every 2-3 years for patients with uncontrolled disease). The cost of a nephrology visit was suggested that should be included only as urinalysis tests and no outpatient visit [PC: Baildam E 28/03/2011]. The social worker cost was deemed to be outside the perspective of the NHS, as patients in the UK might more often receive disability living allowance rather than employ a social worker.

Diagnostic tests were limited to full blood count, liver function, erythrocyte sedimentation rate, C-reactive protein, urea, electrolytes and creatinine, assuming that the remaining tests would be included in relevant HRGs for paediatric arthritis visits or inpatient stay. To avoid double-counting diagnostics tests were excluded from the proportion of patients who are hospitalised.

The health state-cost only depend on the ACR response level and are independent from any other health outcomes.

Table 5.20 Unit costs

Item	Unit cost	Reference
Inpatient stay (per day)	£310 inflated to 2010 = £428.32	Epps et al. 2005; inflation indices by PSSRU/Curtis 2010
GP visit (per visit)	32	PSSRU/Curtis 2010 p.167
Haematological (per visit)	91	NHS reference cost: 253 Consultant Led: Follow up Attendance Non-Admitted Face to Face
Radiological (per visit)	£101 inflated to 2010 = £139.55	Epps et al. 2005; inflation indices by PSSRU/Curtis 2010
Podiatrist (per visit)	11	PSSRU/Curtis 2010 p.154
Opthalmologist (per visit)	£51 inflated to 2010 = £70.47	Epps et al. 2005; inflation indices by PSSRU/Curtis 2010
Rheumatology paediatric (per visit)	£193 inflated to 2010 = £266.66	Epps et al. 2005; inflation indices by PSSRU/Curtis 2010
Psychologist paediatric (per visit)	89	PSSRU/Curtis 2010 p.181 (assume 1 hour visit)
Orthodontist (per visit)	101	NHS reference cost:143 Consultant Led: Follow up Attendance Non-Admitted Face to Face
Occupational therapist (per visit)	15	PSSRU/Curtis 2010 p.152
Full blood count	£11.15 inflated to 2010 = £15.41	Barton et al. 2004; ; inflation indices by PSSRU/Curtis 2010
Liver function test	£6.19 inflated to 2010 = £8.55	Barton et al. 2004; ; inflation indices by PSSRU/Curtis 2010
Erythrocyte sedimentation rate	£11.15 inflated to 2010 = £15.41	Barton et al. 2004; ; inflation indices by PSSRU/Curtis 2010
C-reactive protein	£11.15 inflated to 2010 = £15.41	Barton et al. 2004; ; inflation indices by PSSRU/Curtis 2010
Urea, electrolytes and creatinine	£0.08 inflated to 2010 = £0.11	Barton et al. 2004; ; inflation indices by PSSRU/Curtis 2010

Table 5.21 Resource use responders

Item	Value	PSA	Reference
Inpatient stay (annual units)			
Number of days	7.5	Assume range 5-10	PC: Westhovens R 02/03/2011, Wright S 16/03/2011, Baildam E 28/03/2011
Proportion of patients ACR 30	0.225	Assume range 20%-25%	PC: Wright S 16/03/2011
Proportion of patients ACR 50	0.1475	N/A	Assume linear reduction from 25% (ACR 30) to 0% (ACR 90)
Proportion of patients ACR 70	0.0715	N/A	Assume linear reduction from 25% (ACR 30) to 0% (ACR 90)
Proportion of patients ACR 90	0	N/A	Assume linear reduction from 25% (ACR 30) to 0% (ACR 90)
Outpatient diagnostic tests (annual units)			
Number of tests	3	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011; assume 3 weeks of flare and tests performed once a week
Proportion of patients ACR 30	0.775	N/A	1-% patients with inpatient stay
Proportion of patients ACR 50	0.8525	N/A	1-% patients with inpatient stay
Proportion of patients ACR 70	0.9285	N/A	1-% patients with inpatient stay
Proportion of patients ACR 90	1	N/A	1-% patients with inpatient stay
GP visit (annual visits)	3.5	Assume range 3-4	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Haematological (annual visits)	2	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011, Baildam E 28/03/2011
Radiology (annual visits)	2	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011, Baildam E 28/03/2011
Radiology(proportion of patients)	0.2	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Podiatrist (annual visits)	1	Assume SE 30% of mean	PC: Wright S 16/03/2011
Podiatrist (proportion of patients)	0.025	Assume range 2-3	PC Wright S 16/03/2011
Opthalmologist (annual visits)	2	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011, Baildam E 28/03/2011
Opthalmologist(proportion of patients)	1		PC: Baildam E 28/03/2011
Rheumatology paediatric (annual visits)	3	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Rheumatology paediatric (proportion of patients)	1		PC: Westhovens R 02/03/2011
Psychologist paediatric (annual visits)	1	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Psychologist paediatric (proportion of patients)	0.2	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Orthodontist (annual visits)	1	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Orthodontist (proportion of patients)	0.2	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Occupational therapist (annual visits)	1	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Occupational therapist(proportion of	0.2	Assume SE 30% of mean	PC: Westhovens R 02/03/2011,

patients)			Wright S 16/03/2011

Table 5.22 Resource use non-responders

Item	Value	PSA	Reference
Inpatient stay (annual units)			
Number of days	24.5	Assume range 21-28	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Proportion of patients	0.9	Assume range 85%-95%	PC: Wright S 16/03/2011
Outpatient diagnostic tests (annual units)			
Number of tests	18	Assume range 12-24	PC: Westhovens R 02/03/2011, Wright S 16/03/2011; assume flare 3 times a year, 4-8 weeks per flare and tests performed once a week
Proportion of patients	0.1	N/A	1-% patients with inpatient stay
GP visit (annual visits)	20.8	Assume range every 2-3 weeks	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Haematological (annual visits)	12	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011, Baildam E 28/03/2011
Radiology (annual visits)	2	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011, Baildam E 28/03/2011
Radiology(proportion of patients)	0.9	Assume range 85%-95%	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Podiatrist (annual visits)	1	Assume SE 30% of mean	PC: Wright S 16/03/2011
Podiatrist (proportion of patients)	0.1	Assume SE 30% of mean	PC Wright S 16/03/2011
Opthalmologist (annual visits)	2	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011, Baildam E 28/03/2011
Opthalmologist(proportion of patients)	1		PC: Baildam E 28/03/2011
Rheumatology paediatric (annual visits)	10	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Rheumatology paediatric (proportion of patients)	1		PC: Westhovens R 02/03/2011
Psychologist paediatric (annual visits)	1.5	Assume range 1-2	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Psychologist paediatric (proportion of patients)	0.85	Assume range 75%-95%	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Orthodontist (annual visits)	1	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Orthodontist (proportion of patients)	0.35	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Occupational therapist (annual visits)	3.5	Assume SE 30% of mean	PC: Westhovens R 02/03/2011, Wright S 16/03/2011
Occupational therapist(proportion of patients)	1		PC: Westhovens R 02/03/2011

Combining the data on resource use with the unit costs presented in table 5.20 yields the following costs per health state.

Table 5.23 Health state costs

Health state	Costs per year
Uncontrolled disease	£3,360
Response ACR 30	£504
Response ACR 50	£449
Response ACR 70	£396
Response ACR 90	£345

5.2.8.3 Adverse Events Costs

In the comparison of tocilizumab vs. biologics, as discussed in section 5.9 of MS, a review of comparator safety did not identify any notable differences in serious adverse events with high incidence (over 5%).

In their submission, the manufacturer states that in all comparisons the identified adverse events are of minor severity, lasting a short duration, and have a minuscule cost impact for their management. Therefore, it can be assumed that they do not have a considerable bearing on the incremental cost of the two model arms.

COMMENT

The ERG has not identified any issue with the manufacturer's approach to estimating drug administration costs.

Regarding the health state costs, the ERG considers the difference in costs between uncontrolled disease and ACR response very large. In the clarification letter, the ERG requested the manufacturer to justify these differences. In their reply, the manufacturer confirms that length of hospitalisation is the main driver of the health state cost estimates. According to the manufacturer, all clinical experts suggested that if the analysis considers a week of hospitalisation for the average patient who responds to treatment then assuming a month for non-responders is an underestimate. Given the severe symptoms of the patient's condition, namely fever and skin rash, and also considering that all patients are of young age, it seems plausible to suggest that extensive hospitalisation would occur while patients experience disease flare. It is also expected that non-responders would experience a number of disease flares in a given year. Clinical experts suggested that for non-responders the length of hospitalisation could far exceed three weeks and that the average patient could stay in hospital for as much as three months in a year. The analysis has taken a conservative approach on this estimate and considers the lowest value suggested by clinical experts (3-4 weeks a year).

Given the fact that, due to the relative definition of the health states, patients in the ACR30 health state may have the same absolute health state (as measured by the CHAQ score) as a non-responder, the ERG still considers the difference in health care costs between non-responders and responders as high. Of course, the manufacturer claims that, by linking ACR response to an average CHAQ score, in fact absolute health states have been used. As a matter of consistency, it ERG considers that it

would have been preferable if the experts had not been asked about resource use of non-responders versus responders, but about the resource use typical of patients with a certain CHAQ profile. This would probably have led to a more gradual decline of the health care costs with better health states.

5.2.9 Cost effectiveness results

Base Case Analysis

The base case analysis compares tocilizumab with two different treatment strategies as comparator: a) MTX as first line treatment and b) ANK as first line treatment. As discussed the former analysis, i.e. comparison with MTX, is invalid and will not be further discussed (details of those model outcomes can be found in MS on p 249). Table 5.24 presents the base case results of the cost-effectiveness of tocilizumab compared with anakinra. The additional cost per QALY amount to £23,219.

Table 5.24: Base-case results: tocilizumab versus anakinra

Technologies	Total costs (£)	Total LYG in response	Total QALYs	Incr. costs (£)	Incr. LYG in response	Incr. QALYs	ICER (£) (QALYs)
Strategy TCZ	£138,927	6.1284	5.3223	£11,697	1.7797	0.5038	£23,219
Strategy ANK	£127,230	4.3486	4.8185				

Table 5.25 shows both the additional treatment costs and the cost savings for health care costs of the tocilizumab treatment strategy.

Table 5.25: Summary of costs by strategy: comparison vs. ANK

	Strategy TCZ	Strategy ANK	Incremental
Treatment cost	£82,620	£47,808	£34,812
Health state cost	£56,307	£79,422	-£23,114
Total cost	£138,927	£127,230	£11,697

Note that due to rounding, not all sums add up

Subgroup analysis

No subgroup analysis was conducted in the base case scenario.

COMMENT

Based on the original submission of the manufacturer, the findings of the ERG, and the response of the manufacturer to the clarification letter, the ERG had to conclude that the baseline and probabilistic sensitivity analyses performed so far were not optimal. For this reason the ERG ran the manufacturers cost-effectiveness model using the following choices:

- All response transitions based on the TENDER population that was not MTX-naïve (95% of the trial population).
- Start age of treatment is 7 years
- Withdrawal rate based on exponential time-to-event curve
- Cycle length adjusted to 12 weeks

The results of these ERG analyses are shown in section 5.3

5.2.10 Sensitivity analyses

The manufacturer assessed the various uncertainties in the economic evaluation through deterministic sensitivity analysis, scenario analysis and probabilistic sensitivity analysis. While the first two show which parameters and assumption have the largest impact on the model outcomes, the latter shows the overall uncertainty around the ICER. Unfortunately, all sensitivity analyses and scenario analyses provided by the manufacturer are based on the original model for tocilizumab versus anakinra, for which the analysis of the trial data was done for all patients, instead of only on patients not MTX-naïve. Consequently, the main relevance of these analyses is not the absolute ICERs they present, but rather the order of magnitude of the impact on the ICERs, since we assume that this relative impact will hold in the ERG base case analysis. All three type of sensitivity analyses are discussed in the next paragraphs.

5.2.10.1 Deterministic sensitivity analyses

In the manufacturer's submission, an extensive deterministic sensitivity analysis was done to explore the impact of the input parameters on the outcomes one by one. Details on the parameters varied and the rationale for the deterministic sensitivity can be found in table 97 of the MS (page 288). Many input parameters had little effect on the outcomes, so in the table below (a summary of table 111 in the MS) we have included only those analyses where the ICER changed by more than 10%.

From this analysis, the manufacturer concluded that of the clinical parameters, two had most influence on the model, the anakinra response rates and the withdrawal probability. Using the anakinra response rates for all biologics had the effect of increasing the ICER. When tested an assumed low limit value for the withdrawal risk the ICER is increased. This is anticipated as a greater proportion of patients receive treatment accruing increasing cost. In this case, the increase in costs does not offset the increase in QALYs. The opposite is observed when applying the higher limit for withdrawal risk; less patients on treatment.

According to the manufacturer, overall the model results are not sensitive to input changes surrounding utilities. However, the model results were very sensitive to input changes regarding costs for both comparisons. When assuming double the values for health-state unit costs the tocilizumab strategy is dominating anakinra. Moreover, given clinical expert opinion, one of the analyses tested the result of increasing the length of stay in hospital for non-responder patients. This results in a lower ICER from the base-case. The infusion administration cost is also found to have a great impact in the model results for both comparisons. The impact is estimated higher on the tocilizumab strategy as there are no infusion costs for the biologic comparator strategies.

Table 5.26 Selected results of the univariate sensitivity analysis

	Strategy TCZ		Strategy Anakinra		Incremental results		
	Cost	**QALYs**	**Cost**	**QALYs**	**Cost**	**QALYs**	**ICER £/QALY**
Base case	£138,927	5.3223	£127,230	4.8185	£11,697	0.5038	£23,219
Clinical parameters							
Remove Prince adjustment from ACR response rates	£137,224	5.4438	£124,678	4.998	£12,547	0.4458	£28,142
Adjust the ANK ACR response rates, so it degraded over time	£138,927	5.3223	£127,509	4.6959	£11,419	0.6264	£18,230
Use the ANK ACR response rates for all biologics	£135,472	5.6323	£122,068	5.2766	£13,404	0.3557	£37,683
Use ACR+fever definition	£138,608	5.2518	£127,538	4.8548	£11,070	0.397	£27,884
Do not use ACR 90	£139,393	5.2338	£127,499	4.7673	£11,894	0.4664	£25,499
Withdrawal risk low limit	£146,369	5.7122	£125,528	5.0938	£20,841	0.6183	£33,706
Withdrawal risk high limit	£135,571	5.0316	£128,974	4.6185	£6,597	0.4131	£15,970
Utilities							
NICE formula for mapping	£138,927	4.5201	£127,230	3.9625	£11,697	0.5576	£20,976
Baseline CHAQ CHAQ =2	£138,927	4.4994	£127,230	3.9218	£11,697	0.5776	£20,252
Costs							
Half cost of inpatient stay	£123,307	5.3223	£103,279	4.8185	£20,029	0.5038	£39,756
Half all health-state unit costs	£110,774	5.3223	£87,519	4.8185	£23,255	0.5038	£46,160
Increase duration of patients in hospital by 50% for non-responders	£153,677	5.3223	£150,384	4.8185	£3,293	0.5038	£6,537
All visits to low limit	£126,187	5.3223	£111,097	4.8185	£15,091	0.5038	£29,954
All visits to high limit	£151,657	5.3223	£143,347	4.8185	£8,310	0.5038	£16,495
Half administration cost of infusion	£129,525	5.3223	£126,425	4.8185	£3,100	0.5038	£6,153
Double administration cost of infusion	£157,732	5.3223	£128,839	4.8185	£28,892	0.5038	£57,350
Double nurse visit cost	£139,112	5.3223	£129,107	4.8185	£10,005	0.5038	£19,859
No wastage for costs	£129,098	5.3223	£118,881	4.8185	£10,217	0.5038	£20,281

5.2.10.2 Scenario analyses

The manufacturer considers two structural assumptions in scenario analyses, i.e. the age at which the patients starts treatment (2, 5 or 10 years) combined with various time horizons (up to the age of 32) and the number of treatment lines considered in the sequence. Additionally, etanercept is considered as a comparator for tocilizumab instead of anakinra. Table 5.27 presents the results.

Table 5.27 Scenario analyses tocilizumab versus anakinra

	Strategy TCZ		Strategy ANK		Incremental results		
	Cost	QALYs	Cost	QALYs	Cost	QALYs	ICER £/QALY
Base case	£138,927	5.322	£127,230	4.819	£11,697	0.504	£23,219
Scenarios for age/duration							
Patients 2-32	£162,134	5.660	£150,341	5.225	£11,793	0.435	£27,089
Patients 2-22	£153,098	5.640	£140,043	5.137	£13,055	0.504	£25,922
Patients 5-18	£134,575	4.903	£116,326	4.419	£18,249	0.485	£37,660
Patients 5-32	£170,057	5.739	£151,871	5.277	£18,186	0.462	£39,391
Patients 10-18	£108,444	3.735	£90,044	3.338	£18,401	0.397	£46,369
Scenarios for line of treatments							
One treatment only	£142,311	4.977	£133,382	4.320	£8,930	0.656	£13,603
Two treatments	£141,281	5.112	£131,698	4.510	£9,583	0.602	£15,923
Three treatments	£140,367	5.226	£130,153	4.675	£10,214	0.551	£18,546
Use Etanercept as comparator	£141,047	5.447	£127,335	4.805	£13,712	0.641	£21,379

In the scenario analyses the duration of the model in combination with the starting age of patients has a notable effect to the model results. The model estimates that the older the patient and the longer the duration of treatment the higher the incremental cost and the ICER.

When evaluating different scenarios of treatment sequences (one, two, and three agents) it is estimated that the ICER increases with more treatments in the strategy. The manufacturer states that this signifies that the base case analysis has taken a conservative approach to evaluate four treatments in the sequence.

5.2.10.3 Probabilistic sensitivity analyses

In the manufacturer's submission, a probabilistic sensitivity analysis (PSA) was performed to study the impact of all input uncertainty simultaneously. To this end, probability distributions were specified for all input parameters. We refer the reader to table 99 in the MS for all details on distributions and their parameters used for the PSA.

In the comparison with anakinra, out of 1,000 samples 38% to 63% were below a cost-effectiveness threshold of £20,000 and £30,000 per QALY respectively (See figure 5.3 and figure 5.4 below).

Figure 5.3: Scatter plot: comparison versus anakinra

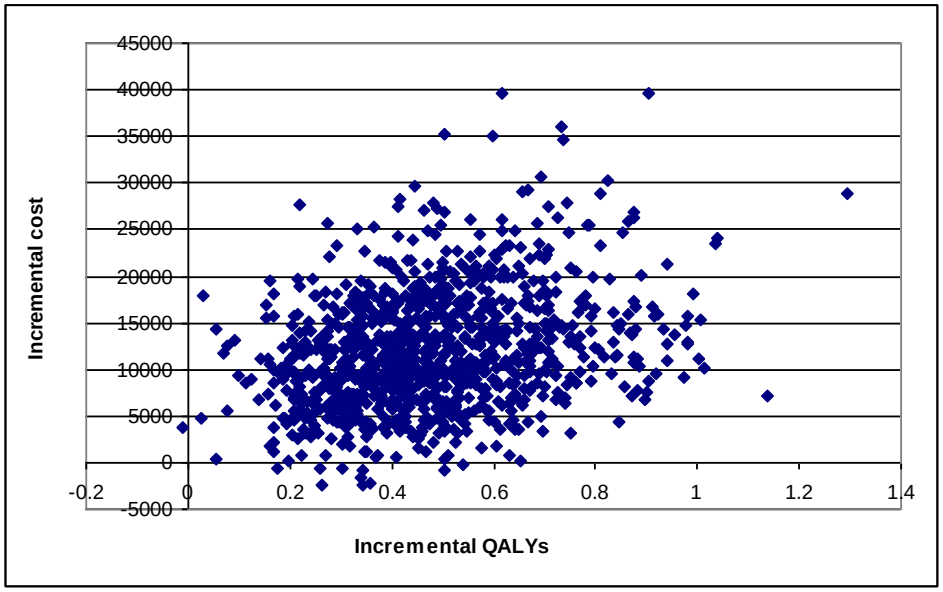
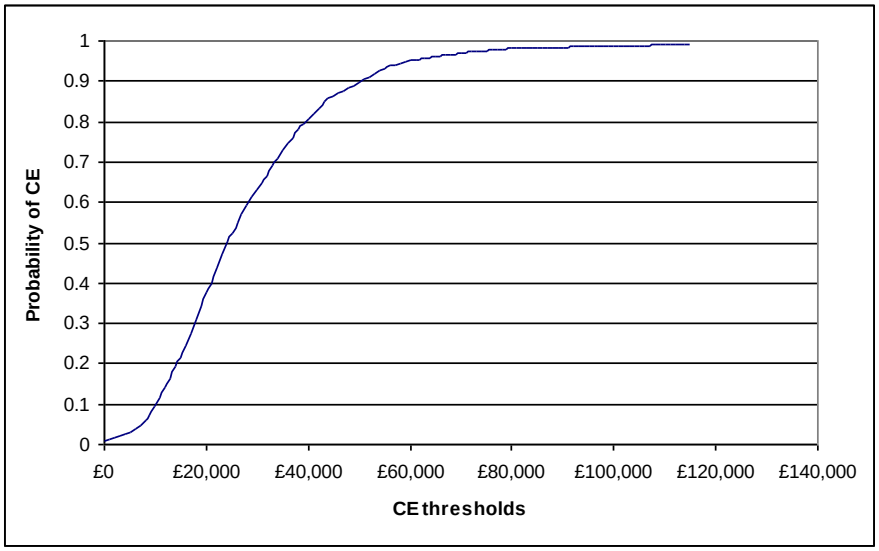


Figure 5.4: CEAC: comparison versus anakinra



In the clarification letter, the ERG asked the manufacturer for an additional PSA. In table 87, page 236 of MS a summary of variable values is presented, with the distributions used in the PSA. The text in the table for the various transition probabilities mentions 'assume N=alpha of the one parameter Gamma distribution for each ACR category'. This suggests that the true N (number of patients in the trial) is used in deriving random draws from the Dirichlet distribution. However, in the electronic model, N=100 is assumed for all treatments and all transitions. This, the ERG requested to change the model input to reflect the true N on which the transition probabilities are based, and provide the PSA outcomes.

In response, the manufacturer changed the model input to reflect the true N on which the transition probabilities were based. For the input related to indirect comparison the N of the original trial of the comparator was assumed. In the comparison with anakinra, out of 1,000 samples 35.1% to 60.5% were below a cost-effectiveness threshold of £20,000 and £30,000 per QALY respectively.

Furthermore, the ERG requested an additional PSA allowing for more variation in the duration of the hospitalization and the percentage of patients requiring hospitalization. The manufacturer addressed this by assuming a margin of 80% of the mean as standard error. Results were presented for the cumulative effect of both ERG requests. Now, in the comparison with anakinra, out of 1,000 samples 30.2% to 43.4% were below a cost-effectiveness threshold of £20,000 and £30,000 per QALY respectively.

COMMENT

In general, the ERG agrees with the conclusions of the manufacturer about the key drivers of the cost-effectiveness results. Both the use of the anakinra response rates for all biologics and variation in the withdrawal risk had a large impact on the ICER. It is interesting to observe that an increased withdrawal rate leads to a more favourable ICER. Given the very strong, and somewhat unrealistic, assumptions on which the response rates for the biologics are based, the manufacturer base case results should be interpreted with caution. The analysis with the anakinra response rates used for all biologics might be seen as an upper limit of the response rates, and in that case, the ICER increases with 60%.

Additionally, the model outcomes were importantly influenced by changes in the health state costs (which is mainly driven by the hospitalization rate and length) and in the infusion administration cost. The ERG considers it unfortunate that the only source of information about the resource use in the various health states was expert opinion. More detailed (observational) data about hospitalisation in sJIA would greatly enhance the robustness of the outcomes.

The scenario analyses show that patient age has a substantial effect on the cost-effectiveness of the intervention; for example, a starting age of 5 leads to an ICER of approximately £36,000 and a starting age of 10 lead to an ICER of approximately £47,000 when comparing TCZ with anakinra.
These scenario analyses are particularly relevant, given the wide range of ages for the onset of disease. Some evidence suggests that the peak age of onset of sJIA is between 18 months and 2 years [47]. In a UK cohort the peak age was 2 years with a mean of 6 years.[48] In the CAPS study, another UK prospective study, the median age of onset is reported to be 6.4 years (interquartile range 4.2, 9.8).[49]
Combined with the fact that the average age in the TENDER trial was close to 10 years old, the ERG has included an alternative start age of 7 years in the ERG base case (section 5.3). This is based on the observed average of 6 years,[48] combined with 1 year of diagnosis and (failed) treatments with NSAIDs, CS and MTX.

The ERG considers the sensitivity analyses performed on the utilities limited. The CHAQ scores, which are directly mapped into utilities, are only varied slightly. Only the starting CHAQ values (base case: 1.7 ± 0.8) is varied slightly (to 1.63, 1.73, 2, respectively) to reflect the mean starting CHAQ when accounting for different subgroups. As described in the section of the model structure, the manufacture claims to model relative improvement which should be avoided in Markov modelling. In the response to the clarification letter the manufacturer states that, in effect, absolute CHAQ scores are modelled. But how the model is set-up, this leads a) to the assumption that all patients have initially the same CHAQ score and b) all relative improvement leads to the same absolute (improved) CHAQ score. A clarifying example can be given by the initial CHAQ distribution which has mean of 1.7 ± 0.8. Assuming normality, this translates into some 16% of all patients having an initial CHAQ of less than .9 and some 16% of all patients having a CHAQ score

higher than 2.4. As a relative increase is modelled, the heterogeneity in the treatment health states is of a similar magnitude. The ERG is of the opinion that the manufacturer should have addressed all these heterogeneities.

Furthermore, the incremental change in CHAQ score/utilities between health states when having a treatment response, which is also affected by uncertainty, is neither part of the deterministic sensitivity analysis nor the probabilistic sensitivity analysis.

As already pointed out by the ERG in the clarification phase, the statistical uncertainty (variability) around the ACR responses is currently based on the assumption that they were all based on a sample size of 100. On request, the manufacturer provided PSA output based on corrected sample sizes, but unfortunately did not provide an updated version of the electronic model. It appears from the explanation given by the manufacturer that only the sample sizes were adjusted to reflect those observed in the clinical studies with tocilizumab, anakinra and infliximab. However, it also appears that no account has been taken of the fact that the transition probabilities for the biologics are in fact a multiplication of the transition probability for tocilizumab, a relative risk and an adjustment factor, and that they are all associated with uncertainty. This assumption of the ERG is for example based on the fact that nowhere in the MS the uncertainty around the adjustment factors are reported.[35]

In the ERG base case (see section 5.3) the statistical uncertainties regarding the transition probabilities will be reflected.

5.2.11 Model validation

The manufactures states that the model assumptions were validated by clinical experts (PC: Westhovens R 02/03/2011, Wright S 16/03/2011, P Woo 21/3/2011, E Baildam 28/03/2011). An independent analyst verified the model calculations. A report was produced with comments on the model. All comments were acknowledged and considered for the final version of the model. No further details were provided in the MS.

In the clarification letter, the ERG indicates that internal and external model validation are lacking. It was requested that the model outcomes be verified with real life data (external validity), for instance using the trial data on follow up CHAQ-score. Additionally, it was requested that if the pivotal trials collected utility data, to validate the model against these data as well. Furthermore, it was requested to show the internal validity of the model for instance by performing extreme values analyses.

In their response, the manufacturer indicated that the TENDER clinical trial and the literature provide limited evidence for comparison with the economic model. The patient CHAQ from the TENDER trial extension at 72 weeks were provided in the table below. The manufacturer cautions that the number of patients in the non-response, ACR30, and ACR 50 groups at week 72 are very small and that it is not appropriate to draw conclusions on the CHAQ score of those patients. With regards to the ACR 70 and ACR 90 categories, the model assumption underestimates improvement in CHAQ. Neither the TENDER study nor other clinical trials report utility data for comparison with the model outcomes. In terms of internal validation, a number of extreme values were tested in the model and a table with results was provided.

Table 5.28 CHAQ score comparison between TENDER clinical trial and economic model taken from the response to the clarification letter

	TENDER (N)	TENDER outcomes at week 72	Model assumed values
No-response	4	0.90625	1.7442
ACR 30	3	1.291667	1.2699
ACR 50	8	1.296875	1.1351
ACR 70	20	0.63125	0.8601
ACR 90	23	0.418478	0.6692

COMMENT

From the manufacturer's description it is not clear how extensive the expert reviewed the model. Also, no insight is given into the findings of these experts. Apparently neither the personal communications with the external experts nor the mentioned report are part of the submission.

The comments of experts, however, are very important to properly assess the model. A number of influential parameters, such as costs, are - at least partly – justified by expert opinion.

The extreme values analysis indicates that the model behaves as required.

5.3 Additional work undertaken by the ERG

Provide details of any additional work conducted by the ERG in relation to cost effectiveness. If the results of any of the additional work affect the size of the ICER, refer the reader to the summary table in Section 6.

New base case analysis

Based on several remarks made in section 5.2 of this report, the ERG defined a new base case analysis:

- The starting age is 7 years, with a time horizon of 11 years.
- The cycle length is adjusted to 12 weeks instead of the current 3 months
- The withdrawal rate is based on the exponential distribution
- Adjust the ACR response probabilities for tocilizumab to reflect the MTX non-responder population (95% of whole populations)
- Adjust the relative risk of anakinra to reflect the non-MTX-naïve population in the indirect comparison
- Adjusted parameters for the distribution of treatment response for anakinra and other biologics for the PSA, to include additional uncertainty around the relative risks and around the adjustment factor.

For the withdrawal rate, we used an SE of 30% of the mean (as it was in the manufacturer's model) since this is more conservative than the standard error based on the standard error derived from the exponential model fit.

For the uncertainty around the response probabilities for anakinra and biologic, we attempted to correctly consider the standard errors around the relative risks and adjustment factors, whilst maintaining a Dirichlet distribution. Thus, as a proxy, for anakinra a sample size of 12 was assumed (based on the clinical study) [39] and for the other treatments, a sample size of 25 was assumed, based on the trial sample size of 58,[40] and downward adjusted to correct for the additional uncertainty due to the adjustment factor used.

The results of this ERG defined base case analysis are presented in table 5.29 and 5.30. As a result of the various changes made the ICER has increased substantially (see also table 5.24). This is almost entirely explained by the higher starting age.

Table 5.29: ERG Base-case results: tocilizumab versus anakinra

Technologies	Total costs (£)	Total LYG in response	Total QALYs	Incr. costs (£)	Incr. LYG in response	Incr. QALYs	ICER (£) (QALYs)
Strategy TCZ	£121,952	4.9668	4.3065	£16,318	1.3630	0.3835	£42,552
Strategy ANK	£105,634	3.6038	3.9230				

Table 5.30: Summary of costs by strategy: ERG comparison vs. ANK

	Strategy TCZ	Strategy ANK	Incremental
Treatment cost	£76,193	£42,183	£34,010
Health state cost	£45,760	£63,451	-£17,692
Total cost	£121,952	£105,634	£16,318

To assess the uncertainty around this estimate, we have performed a PSA. Figure 5.5 presents the outcomes of the PSA on the CE-plane, while figure 5.6 presents the acceptability curve. Based on these outcomes, we find that the probability that the ICER is below £20,000 and £30,000 is 5% and 22%, respectively.

Figure 5.5 PSA outcomes on CE-plane

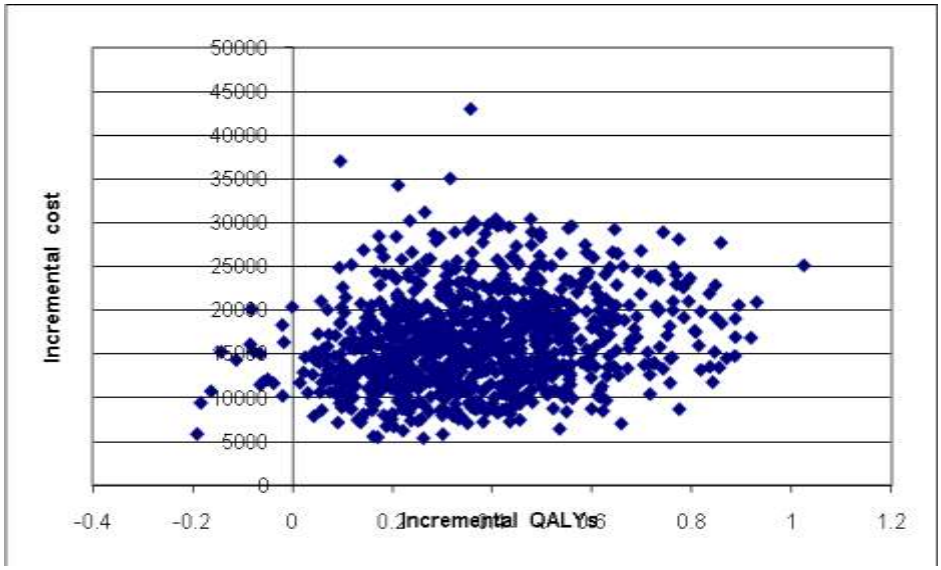
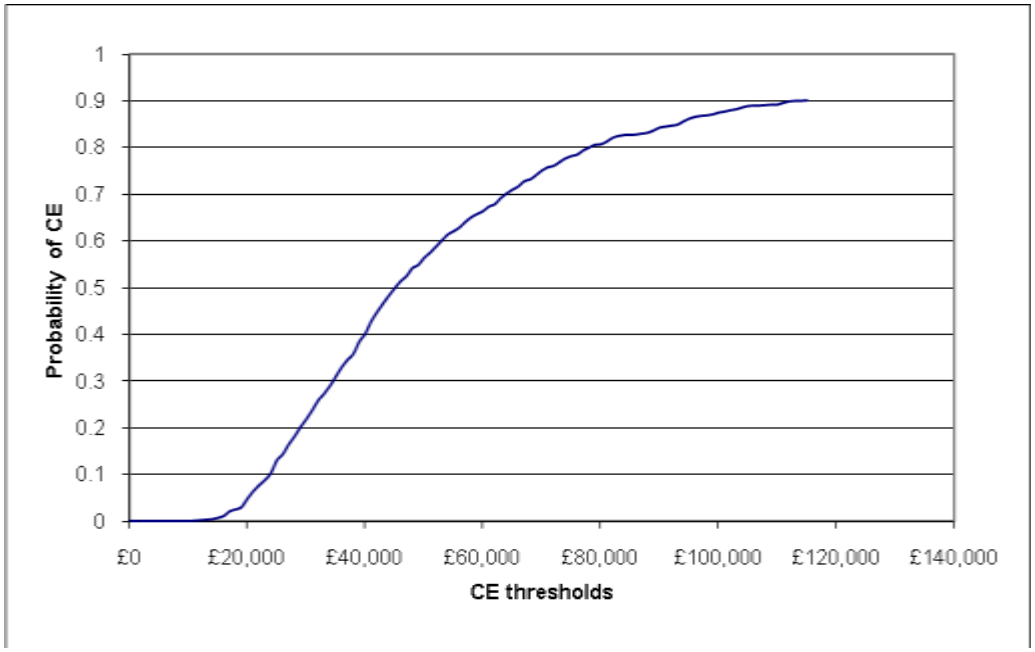


Figure 5.6 Acceptability curve



Additional scenarios based on ERG base case

Based on the ERG base case, a few additional scenarios were explored.

In the first, we varied the withdrawal probabilities in such a way that high-responders would have a lower probability of withdrawing than low responders. This was implemented by assuming withdrawal of 5% for ACR30, 3.5% for ACR50, 2.7% for ACR70 and 1.5% for ACR90. Note that there is no evidence base for the specific values used; our goal was to use realistic values such that the base case withdrawal risk of 3.13% would be between the ACR50 and ACR70 response. The resulting ICER was £40,916 per QALY gained, slightly lower than the base case ICER.

In the second scenario, we explored the effect of the assumption that after the initial response, patients either stay in their current health state, withdraw and move to next line or die. We assumed that patients would move between all health states with a probability of 10% per transition, that is, patients in the ACR30 state had (per cycle) a 10% chance of moving to ACR50, a 10% chance of moving to ACR70 and a 10% chance of moving to ACR90. We assumed that both improvements and deteriorations would occur.

The resulting ICER was £53,051 per QALY gained, 24% higher than the base case ICER. This indicates that the assumption that patients who do not move to the next treatment line stay in the same health state indefinitely is a rather strong assumption.

As an alternative to the ERG starting age of 7, which was derived from literature, the ERG also explored the starting age of 9.7 that is observed in the TENDER trial data across all patients (Table 8 in the MS) in an analysis; the ICER changes to £46,611 per QALY gained.

Finally, the ERG explored the costs and effects from various alternative sequences for treatment. The decision problem states that tocilizumab should be compared to anakinra and TNF inhibitors. The

main focus of the MS is on anakinra as comparator, and etanercept as comparator is explored in a scenario analysis. However, only a pair wise comparison is done, instead of the full incremental analysis of the three treatment options. Additionally, the ERG considers anakinra as second line treatment after tocilizumab also a viable option. In the table 5.31, we present the results of this full incremental analysis. It shows that starting with etanercept followed by anakinra is dominated by anakinra followed by etanercept. Interestingly, the strategy of tocilizumab followed by etanercept is extendedly dominated by tocilizumab followed by anakinra. Thus, the ICER of interest becomes that of tocilizumab – anakinra versus anakinra - etanercept, which is £39,026, slightly lower than our base case ICER.

Table 5.31 Cost-effectiveness results for various treatment sequences, excluding infliximab

Strategy	QALY	Costs	Incr. QALY	Incr. costs	ICER
ETA-ANK-ADA-ABA	3.9113	£105,819			
ANK-ETA-ADA-ABA	3.9230	£105,634	0.0118	-£185	dominates
TCZ-ETA-ADA-ABA	4.3065	£121,952	0.3835	£16,318	£42,552
TCZ-ANK-ETA-ADA	4.4082	£124,569	0.1017	£2,617	£25,730

ETA: etanercept, ANK: anakinra, ADA: adalimumab, ABA: abatacept, TCZ: tocilizumab
Incr: incremental

In section 5.2.4 it was mentioned that exclusion of infliximab as a treatment option in the sequences showed a lack of logic since a trial comparing it to placebo was used as the source of effect for all other TNF inhibitors. Thus, we also explored sequences in which infliximab is also considered. Various sequences were explored and the table below shows the most relevant options (others are all dominated or extendedly dominated).

From this, we see that starting with anakinra compared to infliximab is a cost-effective strategy, with an ICER of £18,287 per QALY gained. The ICER of tocilizumab – anakinra versus anakinra – infliximab is £43,607 per QALY gained.

Table 5.32 Cost-effectiveness results for various treatment sequences, including infliximab

Strategy	QALY	Costs	Incr. QALY	Incr. costs	ICER
INF-ETA-ADA-ABA	3.7545	£98,250			
ANK-INF-ADA-ABA	3.9230	£101,332	0.1685	£3,082	£18,287
TCZ-ANK-INF-ADA	4.4082	£122,490	0.4852	£21,158	£43,607

INF: infliximab, ETA: etanercept, ANK: anakinra, ADA: adalimumab, ABA: abatacept, TCZ: tocilizumab, Incr: incremental

5.4 Conclusions

Describe the completeness of the MS with regard to relevant cost effectiveness studies and data described in any de novo economic evaluations. Does the submission contain an unbiased estimate of the technology's ICERs in relation to relevant populations, interventions comparators and outcomes? Are there any remaining uncertainties about the reliability of the cost effectiveness evidence? Reference should also be made concerning the extent to which the submitted evidence reflects the decision problem defined in the final scope.

The systematic review in the MS did not identify any relevant cost-effectiveness studies, thus making a *de novo* economic analysis necessary. The *de novo* economic analysis is based on data that does not allow drawing inference for population 1 as defined in the scope and has only limited evidence for population 2 as defined in the scope.

No reliable data has been presented by the manufacturer to inform population1, i.e. data pertaining to MTX naïve patients, since 95% of the patients in the TENDER trial have used MTX or are currently using it and the MS comparison is based on a non-randomised comparison of all patients in the tocilizumab arm versus only those who do use MTX in the placebo arm.

For population 2, i.e. patients who have not responded to MTX, exactly the same tocilizumab population is used, based on the argument that even though patients do use MTX, they have active disease and do thus not respond to the MTX. Note that also the 5% MTX naïve patients are included. There is no direct evidence presented for the comparators in this population, i.e. anakinra and TNF-inhibitors, and hence indirect comparisons are made.

For the indirect comparison, the manufacturer decided to broaden the inclusion criteria to identify all pivotal trials in juvenile arthritis regardless of subtype. This was despite advice from their clinical experts to the contrary based on the differences between a systemic JIA population and other subtypes. The only relevant study compared infliximab to placebo. All ACR responses in the model are based on this one study, while infliximab itself is not considered a comparator by the manufacturer. The non-systematic approach to the search for studies of TNF-inhibitors introduces uncertainty about the completeness of the search.

For the economic evaluation, the manufacturer has attempted to correct for the mismatch between the TNF-inhibitors' population (any JIA) and the population relevant in this evaluation (sJIA) by adjusting the ACR response rates for the TNF-inhibitors based on an observational study in patients using etanercept.

Based on the above, the ICER for population 1 is clearly biased and has therefore not been reproduced. The ICER for population 2 is also likely to be biased, but here the argument could be made that the TENDER population (minus the 5% MTX naïve patients) matches the decision problem.

The remaining biases in the latter ICER are partially related to the problems identified with the indirect comparison but also to a wide range of other issues identified by the ERG. The main issue was the starting age used in the model. Since the decision problem mentioned children of 2 years and older, the manufacturer used this as the starting age of the model. However, on average, patients will

be 7 years before they are eligible for treatment with tocilizumab. It was found that this higher age had a significant impact on the ICER, increasing from £23,000 to £42,500.
Another main issue relates to the model structure. The model is a cohort Markov model assessing different sequences of treatment. The treatment effect is modelled as a relative improvement (ACR response) for each patient. The manufacture claims that this is equal to obtaining an absolute CHAQ score, which is a measure of disease burden that is directly mapped into utilities to calculate QALYs. Considering the wide variation of the initial CHAQ score of patients, the current approach masks the variability in the model outcomes for an individual patient by imposing an undue homogeneity assumption, i.e. only one patient type exists and all patients experience the same absolute utility change from a relative health improvement irrespective of their initial disease manifestation. Thus, the ERG considers the modeling approach not appropriate to inform the decision problem. Consequently, the outcomes of the model should be interpreted with care.

Remaining uncertainties about the reliability of the cost effectiveness evidence

Important uncertainties remain, beyond the major issues mentioned above.

The cost estimates for health states have been solely defined based on expert opinion. The resulting cost estimates do not seem reasonable, as they present a cost for non-responders (£ 3,300) that is 6 times higher than the costs for an ACR30 response (£500), whereas an ACR90 response is associated with only a 30% decrease (£350) compared to ACR30. Additionally, due to the wide variation in health status at base line of the patients, patients may be assigned different costs even though at 12 weeks they have the same absolute health status.

Since the cost estimates are a main driver of the cost effectiveness, it might be of value to find more reliable cost estimates. The problems could be overcome if observational data would be obtained on resource use in the sJIA population, related to CHAQ score of the patient.

The utilities used in the model have been derived using a mapping formula developed in the context of and adult RA population. There is no information available to check the validity of this procedure. Also, after translating the relative ACR response to an absolute (fixed) CHAQ score, and mapping this onto utilities, an additional step is added where for the comparators anakinra and TNF-inhibitors the assumption is made that their ACR responses can be assigned the same utility. It is difficult for the ERG to assess whether this chain of assumptions leads to an over-or underestimation of the cost effectiveness.

An important shortcoming of the model is that it models the trial, instead of natural disease progression. This is illustrated by the assumption in the model that patients that move to a certain ACR response stay in that state until the patient either withdraws (i.e. moves to the next treatment line) or dies. Given the nature of the disease, this is an unlikely assumption.

Throughout the model, assumptions were made about the statistical uncertainties to be included in the PSA. While some of the assumptions were reasonable, others were not. It is however important to realize that the overall uncertainty about the cost effectiveness of tocilizumab goes far beyond the statistical uncertainty, as it is related more to fundamental problems in model structure and availability and use of effectiveness evidence.

6 IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

Where appropriate, this section should include a table which shows (i) the effect of any major clinical or cost parameter change or structural change on the size of the base-case ICER and (ii) the effect of making all changes simultaneously on the size of the base-case ICER.

The impact of the additional analyses the ERG has undertaken on the ICER is presented in Table 6.1. The ERG base case analysis leads to a much higher ICER, which is almost entirely explained by the higher starting age used. This emphasizes that the variability in the patient population is an important driver of the cost effectiveness of tocilizumab versus anakinra.

Additionally, the ERG explored the effect of 2 modelling assumptions on the ICER, using plausible but fictitious transition rates between ACR response states and withdrawal risks dependent on health state. These analyses showed that the effect of these modelling assumptions are not negligible. Also, the ERG explored the effect of changing the starting age of treatment to the average age in the TENDER trial. This increased the ICER somewhat more.

Finally, the ERG explored the effect of variation in the treatment sequences, and a slightly lower ICER can be achieved by using anakinra as second line treatment for tocilizumab instead of etanercept. When infliximab is added as a treatment option, sequences with infliximab dominate sequences with etanercept in the same treatment line.

Table 6.1. The effect of any major clinical or cost parameter change or structural change on the size of the base-case ICER

Technologies	ICER (£))
Base case manufacturer's submission, analyses based on analysis total tocilizumab population	
Tocilizumab vs. anakinra	23,219
Alternative ERG base case Starting age 7, exclusion MTX-naïve patients tocilizumab populations, withdrawal risk corrected	
Tocilizumab vs. anakinra	42,552
Alternative ERG base case plus allow transitions between ACR response states	
Tocilizumab vs. anakinra	53,051
Alternative ERG base case plus allow withdrawal dependent on ACR response state	
Tocilizumab vs. anakinra	40,916
Alternative ERG base case plus starting age 9.7 (time horizon 8.3 years)	
Tocilizumab vs. anakinra	46,611
Alternative ERG base case various treatment sequences, excluding infliximab	
Tocilizumab - anakinra vs. anakinra - etanercept	39,026
Alternative ERG base case various treatment sequences, including infliximab	
Anakinra – infliximab vs. infliximab - etanercept	18,287

Tocilizumab – anakinra vs. anakinra - infliximab	43,607

7 END OF LIFE

Where appropriate, this section should summarise the manufacturer's case for using the NICE end of life treatment criteria and discuss to what extent the manufacturer's argument is valid.

Not relevant.

8 CONCLUSIONS

The section should focus on any difference(s) of opinion between the manufacturer and the ERG that might influence the size of the ICER. Priority should be focussed on discussing information that will be useful to the Appraisal Committee including strengths, weaknesses and remaining uncertainties. Further summary of evidence is not required in this section.

The evidence presented by the manufacturer consisted of one RCT (the TENDER trial) comparing tocilizumab (8 mg/kg, N=37; or 12 mg/kg N=38) with placebo (N=37). Inclusion criteria for the TENDER study population included an inadequate response to previous treatment with NSAIDs and corticosteroids. In TENDER, all patients had evidence of active disease at baseline which was at least 6 months after a definite diagnosis of sJIA. Inadequate response to previous treatment was determined by the treating physician's clinical assessment. A total of 78/112 (70%) patients had been treated with MTX prior to study entry (36 of these entered the study on MTX that had been previously stopped then restarted, 42 of these patients were on their first course of MTX which was ongoing). Twenty-nine patients (26%) had no background MTX at baseline but did receive and stop MTX previously. Five (4%) patients had never received MTX, and could be considered MTX naive.

The ERG has a fundamental problem with the evidence presented in the MS as it is not in accordance with the NICE scope. It is for the Appraisal committee to decide whether it will accept the ERG approach, which means there is no evidence for any comparison in the NICE scope, or accept the MS approach, which means there is some evidence for the second population, but none for the first population.

The main question is: "Which patients in the TENDER trial match which population"? According to the manufacturer 95% of TENDER trial participants match population 2, because "patients are included in the study if they have symptoms of active disease" and "It follows that if patients have tried in the past or are currently administered MTX and continue to have persistent disease then they are inadequate responders" (Response to Clarification Letter, question A2).

The ERG does not agree with this approach. The MS does not provide a clear definition of inadequate responders. It cannot be automatically assumed that all participants in the TENDER trial are inadequate responders to MTX. Because of the lack of information it can only be assumed that the 25% of children in the TENDER trial who stopped using MTX fit this population (population 2). The remaining 75% of children in the TENDER trial should be treated as population 1. Because no data were provided for these two populations, there is no evidence available for any of the comparisons in the NICE scope.

Following the MS approach, no data were provided in the MS for population 1. Therefore, the only comparison left is tocilizumab versus anti-TNFs or anakinra. The manufacturer performed a systematic review to identify trials for the comparators. One trial was identified in children with sJIA, comparing anakinra with placebo. The manufacturer decided to broaden the inclusion criteria to include all trials in juvenile arthritis regardless of subtype, despite advice from their clinical experts to the contrary (see MS, page 116). The ERG agrees with the advice from the clinical experts; therefore, trials in children with other types of juvenile arthritis have been ignored in this report.

In conclusion, following the MS approach, for population 2 (children with sJIA with an inadequate response to NSAIDs, CS and MTX) the MS provided data for an indirect comparison of tocilizumab versus anakinra, using data from the TENDER trial, and a trial of anakinra versus placebo. Strictly speaking, the 5% of participants in the TENDER trial who were MTX naive should be excluded from these analyses. The MS only provided data for all participants in the TENDER trial. However, in response to the clarification letter some data were provided in which MTX naive patients were excluded. These data were not reported for the TENDER trial, but only for the indirect comparison with anakinra. Where possible, the ERG will use data for the correct population.

The indirect comparison of tocilizumab versus anakinra shows that ACR30 response favours tocilizumab (RR=2.27, 95% CI: 1.06, 4.85). ACR30 response without fever showed no significant difference between tocilizumab and anakinra.

For the economic evaluation, the manufacturer has attempted to correct for the mismatch between the TNF-inhibitors' population (any JIA) and the population relevant in this evaluation (sJIA) by adjusting the ACR response rates for the TNF-inhibitors based on an observational study in patients using etanercept.

Based on the above, the ICER for population 1 is clearly biased and has therefore not been reproduced. The ICER for population 2 is also likely to be biased, but here the argument could be made that the TENDER population (minus the 5% MTX naïve patients) matches the decision problem.

The biases in the ICER of population 2 are partially related to the problems identified with the indirect comparison but also to a wide range of other issues identified by the ERG. The main issue was the starting age used in the model. Since the decision problem mentioned children of 2 years and older, the manufacturer used this as the starting age of the model. However, on average, patients will be 7 years before they are eligible for treatment with tocilizumab. It was found that this higher age had a significant impact on the ICER, increasing from £23,000 to £42,500.

The ERG does not agree with the model structure used in the MS. The model is a cohort Markov model where the treatment effect is modelled as a relative improvement (ACR response) for each patient. Considering the wide variation of the initial health status of patients, the current approach leads to health states that are not mutually exclusive; patients with the same health status may be in different health states, depending on their baseline health status. This leads to problems when assigning utilities and costs. Thus, the ERG considers the modeling approach not appropriate to inform the decision problem and consequently, the outcomes of the model should be interpreted with care.

Throughout the model, assumptions were made about the statistical uncertainties to be included in the PSA. While some of the assumptions were reasonable, others were not. It is however important to

realize that the overall uncertainty about the cost effectiveness of tocilizumab goes far beyond the statistical uncertainty, as it is related more to fundamental problems in model structure and availability and use of effectiveness evidence.

8.1 Implications for research

For children with systemic Juvenile Idiopathic Arthritis who are MTX-naive, the relative effect of tocilizumab should be established in comparison with MTX.

For children with systemic Juvenile Idiopathic Arthritis who have shown to be non-responsive to MTX, good quality trials are necessary for anakinra and relevant anti-TNFs. These interventions should also be compared with tocilizumab in head-to-head trials.

Long term monitoring of tocilizumab, anakinra and anti-TNF treatments is required to address questions relating to occurrence of rare and serious adverse events, the prolonged maintenance of clinically meaningful response, rate and reason for patient withdrawal from treatments, and the success of sequential treatments when one treatment has failed.

In order to inform the cost-effectiveness of tocilizumab an improved health economic model should be developed that either allows for patients level simulation, to account for patient heterogeneity, or that uses health states that are defined such that patients within each health state all have approximately the same health status. Such models exist for instance for COPD, with health states defined based on FEV1 % predicted, or in rheumatoid arthritis, with health states defined based on DAR score.

To attach utilities and costs to health states in such model additional research is required. Ideally a quality of life questionnaire would be administered in sJIA patients that allows for utilities but is also suitable for children. Additionally, resource use in sJIA patients should be investigated in such a way that it allows for health state specific costs.

A full systematic review to inform the effectiveness of all relevant anti TNF alphas would also be warranted. This would improve the ability to perform a full incremental analysis comparing all possible treatment sequences.

References

[1] National Institute for Health and Clinical Excellence. *Single Technology Appraisal: specification for manufacturer/sponsor submission of evidence*. London: NICE, October 2009, 2009. 76p.

[2] Prince FH, Otten MH, van Suijlekom-Smit LW. Diagnosis and management of juvenile idiopathic arthritis. *Br Med J* 2010;341:c6434.

[3] Lequerre T, Quartier P, Rosellini D, Alaoui F, De Bandt M, Mejjad O, et al. Interleukin-1 receptor antagonist (anakinra) treatment in patients with systemic-onset juvenile idiopathic arthritis or adult onset Still disease: preliminary experience in France. *Ann Rheum Dis* 2008;67(3):302-8.

[4] Woo P. Anakinra treatment for systemic juvenile idiopathic arthritis and adult onset Still disease. *Ann Rheum Dis* 2008;67(3):281-2.

[5] Zeft A, Hollister R, LaFleur B, Sampath P, Soep J, McNally B, et al. Anakinra for systemic juvenile arthritis: the Rocky Mountain experience. *J Clin Rheumatol* 2009;15(4):161-4.

[6] Wulffraat NM, Vastert B, Tyndall A. Treatment of refractory autoimmune diseases with autologous stem cell transplantation: focus on juvenile idiopathic arthritis. *Bone Marrow Transplant* 2005;35 Suppl 1:S27-9.

[7] Roche Products Ltd. Tocilizumab for the treatment of systemic juvenile idiopathic arthritis – Roche response to clarification questions. Welwyn Garden City, Hertfordshire: Roche Products Ltd, 2011: 28p.

[8] Roche Products Ltd. Tocilizumab for the treatment of systemic juvenile idiopathic arthritis – Roche response to clarification questions. Appendix of literature reviews. Welwyn Garden City, Hertfordshire: Roche Products Ltd, 2011: 15p.

[9] McGowan J, Sampson M, Lefebvre C. An evidence based checklist for the peer review of electronic search strategies (PRESS EBC). *Evidence Based Library and Information Practice* 2010;5(1):1-6.

[10] Roche Products Ltd. RoActemra (tocilizumab) for the treatment of systemic juvenile idiopathic arthritis (sJIA): submission to National Institute of Health and Clinical Excellence. Single technology appraisal (STA). Welwyn Garden City, Hertfordshire: Roche Products Ltd, 2011: 395p.

[11] Sterne J, Egger M, Moher D, Cochrane Bias Methods Group. *Chapter 10: Addressing reporting biases [Internet]. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration*, 2011 [cited 9.6.11] Available from: http://www.cochrane-handbook.org

[12] Lefebvre C, Manheimer E, Glanville J. *Chapter 6.2.3 Unpublished and ongoing studies [Internet]. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration*, 2011 [cited 9.6.11] Available from: http://www.cochrane-handbook.org

[13] Lefebvre C, Manheimer E, Glanville J. *Chapter 6.2.2.4 Conference abstracts or proceedings [Internet]. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of*

Interventions Version 5.1.0 [updated March 2011]. *The Cochrane Collaboration*, 2011 [cited 9.6.11] Available from: http://www.cochrane-handbook.org

[14] Centre for Reviews and Dissemination. *1.3.1 Identifying research evidence for systematic reviews [Internet]*. *In: Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York: University of York*, 2009 [cited 9.6.11] Available from: http://www.york.ac.uk/inst/crd/systematic_reviews_book.htm

[15] Lefebvre C, Manheimer E, Glanville J. *Chapter 6.2.2.1 Handsearching [Internet]*. *In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. *The Cochrane Collaboration*, 2011 [cited 9.6.11] Available from: http://www.cochrane-handbook.org

[16] Lefebvre C, Manheimer E, Glanville J. *Chapter 6: searching for studies. Box 6.4.c: Cochrane Highly sensitive search strategy for identifying randomized controlled trials in Medline: Sensitivity-maximizing version (2008 version); OVID format [Internet]*. *In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. *The Cochrane Collaboration*, 2011 [cited 5.4.11]. Available from: http://www.cochrane-handbook.org

[17] Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. *J Med Libr Assoc* 2006;94(1):41-7.

[18] Woo P, Southwood TR, Prieur AM, Dore CJ, Grainger J, David J, et al. Randomized, placebo-controlled, crossover trial of low-dose oral methotrexate in children with extended oligoarticular or systemic arthritis. *Arthritis Rheum* 2000;43(8):1849-57.

[19] Centre for Reviews and Dissemination. *Systematic reviews: CRD's guidance for undertaking reviews in health care [Internet]*. York: University of York, 2009 [cited 7.6.11] Available from: http://www.york.ac.uk/inst/crd/systematic_reviews_book.htm

[20] Golder S, McIntosh H, Duffy S, Glanville J. Developing efficient search strategies to identify reports of adverse effects in Medline and Embase. *Health Information and Libraries Journal* 2006;23(1):3-12.

[21] Riemsma R, Al M, Armstrong N, Misso K, Manning N, Lang S, et al. Clarification Letter – Tocilizumab for systemic juvenile idiopathic arthritis. York: Kleijnen Systematic Reviews Ltd, 2011.

[22] Centre for Reviews and Dissemination. *CRD homepage [Internet]*. York: University of York, 2011 [cited 7.6.11] Available from: http://www.york.ac.uk/inst/crd/index.htm

[23] Centre for Reviews and Dissemination. *NHS Economic Evaluation Database Handbook [Internet]*. York: University of York, 2007 [cited 7.6.11]. 116p. Available from: http://www.york.ac.uk/inst/crd/pdf/nhseed-handb07.pdf

[24] InterTASC Information Specialists' Sub-Group (ISSG). *InterTASC Information Specialists' Sub-Group Search Filter Resource [Internet]*. York: Centre for Reviews and Dissemination, 2011 [cited 7.6.11] Available from: http://www.york.ac.uk/inst/crd/intertasc/

[25] Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews (2nd edition)*. York: University of York, 2001

[26] Cummins C, Connock M, Fry-Smith A, Burls A. A systematic review of effectiveness and economic evaluation of new drug treatments for juvenile idiopathic arthritis: etanercept. *Health Technol Assess* 2002;6(17):1-43.

[27] Epps H, Ginnelly L, Utley M, Southwood T, Gallivan S, Sculpher M, et al. Is hydrotherapy cost-effective? A randomised controlled trial of combined hydrotherapy programmes compared with physiotherapy land techniques in children with juvenile idiopathic arthritis. *Health Technol Assess* 2005;9(39):iii-iv, ix-x, 1-59.

[28] Ungar WJ, Costa V, Hancock-Howard R, Feldman BM, Laxer RM. Cost-effectiveness of biologics in polyarticular-course juvenile idiopathic arthritis patients unresponsive to disease-modifying antirheumatic drugs. *Arthritis Care Res* 2011;63(1):111-9.

[29] Barron AC, Lee TL, Taylor J, Moore T, Passo MH, Graham TB, et al. Feasibility and construct validity of the parent willingness-to-pay technique for children with juvenile idiopathic arthritis. *Arthritis Rheum* 2004;51(6):899-908.

[30] Beukelman T, Guevara JP, Albert DA. Optimal treatment of knee monarthritis in juvenile idiopathic arthritis: a decision analysis. *Arthritis Rheum* 2008;59(11):1580-8.

[31] Brodszky V, Pentek M, Majer I, Karpati K, Gulacsi L. *[Etanercept in patients with juvenile idiopathic arthritis: systematic review and economic evaluation] [Internet]*. Budapest: Unit of Health Economics and Technology Assessment in Health Care (HUNHTA), 2006 [cited 15.6.11] Available from: http://hecon.uni-corvinus.hu/download/english/publ/hecon_research_20_eng.pdf

[32] Thornton J, Ashcroft D, O'Neill T, Elliott R, Adams J, Roberts C, et al. A systematic review of the effectiveness of strategies for reducing fracture risk in children with juvenile idiopathic arthritis with additional data on long-term risk of fracture and cost of disease management. *Health Technol Assess* 2008;12(3):iii-ix, xi-xiv, 1-208.

[33] Thornton J, Lunt M, Ashcroft DM, Baildam E, Foster H, Davidson J, et al. Costing juvenile idiopathic arthritis: examining patient-based costs during the first year after diagnosis. *Rheumatology* 2008;47(7):985-90.

[34] Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;8(36):iii-iv, ix-xi, 1-158.

[35] Prince FH, Twilt M, ten Cate R, van Rossum MA, Armbrust W, Hoppenreijs EP, et al. Long-term follow-up on effectiveness and safety of etanercept in juvenile idiopathic arthritis: the Dutch national register. *Ann Rheum Dis* 2009;68(5):635-41.

[36] Hashkes PJ, Wright BM, Lauer MS, Worley SE, Tang AS, Roettcher PA, et al. Mortality outcomes in pediatric rheumatology in the US. *Arthritis Rheum* 2010;62(2):599-608.

[37] Office of National Statistics. Office of National Statistics homepage [Internet]. 2011 [cited 16.3.11]. Available from: http://www.statistics.gov.uk/default.asp

[38] Miller DK, Homan SM. Determining transition probabilities: confusion and suggestions. *Med Decis Making* 1994;14(1):52-8.

[39] Quartier P, Allantaz F, Cimaz R, Pillet P, Messiaen C, Bardin C, et al. A multicentre, randomised, double-blind, placebo-controlled trial with the interleukin-1 receptor antagonist anakinra in patients with systemic-onset juvenile idiopathic arthritis (ANAJIS trial) [Epub]. *Ann Rheum Dis* 2010 [cited 22.12.10];doi: 10.1136/ard.2010.134254.

[40] Ruperto N, Lovell DJ, Cuttica R, Wilkinson N, Woo P, Espada G, et al. A randomized, placebo-controlled trial of infliximab plus methotrexate for the treatment of polyarticular-course juvenile rheumatoid arthritis. *Arthritis Rheum* 2007;56(9):3096-106.

[41] Ruperto N, Murray KJ, Gerloni V, Wulffraat N, de Oliveira SK, Falcini F, et al. A randomized trial of parenteral methotrexate comparing an intermediate dose with a higher dose in children with juvenile idiopathic arthritis who failed to respond to standard doses of methotrexate. *Arthritis Rheum* 2004;50(7):2191-201.

[42] Lovell DJ, Reiff A, Ilowite NT, Wallace CA, Chon Y, Lin SL, et al. Safety and efficacy of up to eight years of continuous etanercept therapy in patients with juvenile rheumatoid arthritis. *Arthritis Rheum* 2008;58(5):1496-504.

[43] Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, et al. *Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation [Internet]*. York: CRD/CHE Technology Assessment Group (Centre for Reviews and Dissemination/Centre for Health Economics), University of York, 4th December, 2009 [cited 15.6.11]. 68p. Available from: http://www.nice.org.uk/nicemedia/live/11966/47853/47853.pdf

[44] Boggs R, Sengupta N, Ashraf T. Estimating health utility from a physical function assessment in rheumatoid arthritis (RA) patients treated with adalimumab (HUMIRA) (D2E7). Presented at ISPOR Fifth Annual European Congress, 2002 3-5 Nov; Rotterdam, The Netherlands. 2002.

[45] National Institute for Health and Clinical Excellence. *Adalimumab, eternacept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF-alpha inhibitor. Addendum to TA 195 [Internet]*. London: NICE, 2010 [cited 14.6.11] Available from: http://guidance.nice.org.uk/TA195

[46] Barton P, Jobanputra P, Wilson J, Bryan S, Burls A. The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. *Health Technol Assess* 2004;8(11):iii, 1-91.

[47] Woo P. Systemic juvenile idiopathic arthritis: diagnosis, management, and outcome. *Nat Clin Pract Rheumatol* 2006;2(1):28-34.

[48] Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S, et al. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *J Clin Invest* 1998;102(7):1369-76.

[49] Adib N, Hyrich K, Thornton J, Lunt M, Davidson J, Gardner-Medwin J, et al. Association between duration of symptoms and severity of disease at first presentation to paediatric rheumatology: results from the Childhood Arthritis Prospective Study. *Rheumatology* 2008;47(7):991-5.

Appendix 1: ERG Search Strategies

ERG Indirect/MTC trials search

Medline (OvidSP): 1948-2011/05/wk2
Searched 24.5.11

1 randomized controlled trial.pt. (305892)
2 controlled clinical trial.pt. (82328)
3 randomized.ab. (212836)
4 placebo.ab. (124063)
5 drug therapy.fs. (1449020)
6 randomly.ab. (154440)
7 trial.ab. (219764)
8 groups.ab. (1027548)
9 or/1-8 (2678337)
10 animals/ not (animals/ and humans/) (3501643)
11 9 not 10 (2271276)
12 Arthritis, Juvenile Rheumatoid/ (7486)
13 systemic$.ti,ab,ot. (268295)
14 12 and 13 (1306)
15 (sJIA or SoJIA or (Systemic adj3 JIA)).ti,ab,ot,hw. (237)
16 (Systemic adj3 (Child$ or juvenile$ or p?ediatr$) adj3 arthriti$).ti,ab,ot,hw. (540)
17 ((Child$ or juvenile$ or p?ediatr$) adj4 Idiopathic Arthrit$).ti,ab,ot,hw. (1709)
18 (Systemic$ adj4 (JRA or JCA)).ti,ab,ot,hw. (230)
19 ((Child$ or juvenile$ or p?ediatr$) adj4 Still$ disease).ti,ab,ot,hw. (109)
20 (systemic$ adj3 arthrit$).ti,ab,ot,hw. (2184)
21 exp Child/ or Adolescent/ (2121749)
22 (child$ or kid or kids or preschool$ or pre school$ or toddler$ or preteen$ or pre teen$ or preadolescen$ or adolescen$ or p?ediatric$ or prepubescent or prepubert$ or pre pubescent or pre puberty or school age$ or schoolage$ or young person$ or young people or p?ediatr$ or juvenile$ or teenage$ or teen or teens).ti,ab,ot. (1015111)
23 21 or 22 (2378925)
24 20 and 23 (674)
25 or/14-19,24 (2940)
26 (Methotrexate or MTX or mexate or amethopterin or 59-05-2).af. (37708)
27 (Abatacept or Orencia or 332348-12-6 or bms-188667 or bms188667 or CTLA4-Ig or CTLA4Ig).af. (2202)
28 (Anakinra or Kineret or 143090-92-0).af. (483)
29 (infliximab or remicade or 170277-31-3 or avakine or revellex).af. (6077)
30 (Etanercept or Enbrel or 185243-69-0 or 200013-86-1 or tnr001 or tnr-001).af. (3709)
31 (Adalimumab or humira or trudexa or 331731-18-1 or D2E7 or d2-e7).af. (1873)
32 or/26-31 (47312)
33 11 and 25 and 32 (421)

Trials filter:
Lefebvre C, Manheimer E, Glanville J. Chapter 6: searching for studies. Box 6.4.c: Cochrane Highly sensitive search strategy for identifying randomized controlled trials in Medline: Sensitivity-maximizing version (2008 version); OVID format. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org

Medline In-Process Citations (OvidSP): up to 2011/05/23
Medline Daily Update (OvidSP): up to 2011/05/23
Searched 24.5.11

1 randomized controlled trial.pt. (851)
2 controlled clinical trial.pt. (50)
3 randomized.ab. (9982)
4 placebo.ab. (4056)
5 drug therapy.fs. (1441)
6 randomly.ab. (9827)
7 trial.ab. (10621)
8 groups.ab. (57961)
9 or/1-8 (77459)
10 animals/ not (animals/ and humans/) (2050)
11 9 not 10 (77062)
12 Arthritis, Juvenile Rheumatoid/ (7)
13 systemic$.ti,ab,ot. (9812)
14 12 and 13 (1)
15 (sJIA or SoJIA or (Systemic adj3 JIA)).ti,ab,ot,hw. (16)
16 (Systemic adj3 (Child$ or juvenile$ or p?ediatr$) adj3 arthriti$).ti,ab,ot,hw. (26)
17 ((Child$ or juvenile$ or p?ediatr$) adj4 Idiopathic Arthrit$).ti,ab,ot,hw. (154)
18 (Systemic$ adj4 (JRA or JCA)).ti,ab,ot,hw. (3)
19 ((Child$ or juvenile$ or p?ediatr$) adj4 Still$ disease).ti,ab,ot,hw. (6)
20 (systemic$ adj3 arthrit$).ti,ab,ot,hw. (92)
21 exp Child/ or Adolescent/ (1509)
22 (child$ or kid or kids or preschool$ or pre school$ or toddler$ or preteen$ or pre teen$ or preadolescen$ or adolescen$ or p?ediatric$ or prepubescent or prepubert$ or pre pubescent or pre puberty or school age$ or schoolage$ or young person$ or young people or p?ediatr$ or juvenile$ or teenage$ or teen or teens).ti,ab,ot. (38362)
23 21 or 22 (39301)
24 20 and 23 (28)
25 or/14-19,24 (171)
26 (Methotrexate or MTX or mexate or amethopterin or 59-05-2).af. (846)
27 (Abatacept or Orencia or 332348-12-6 or bms-188667 or bms188667 or CTLA4-Ig or CTLA4Ig).af. (57)
28 (Anakinra or Kineret or 143090-92-0).af. (49)
29 (infliximab or remicade or 170277-31-3 or avakine or revellex).af. (448)
30 (Etanercept or Enbrel or 185243-69-0 or 200013-86-1 or tnr001 or tnr-001).af. (228)
31 (Adalimumab or humira or trudexa or 331731-18-1 or D2E7 or d2-e7).af. (191)
32 or/26-31 (1457)
33 11 and 25 and 32 (17)

Trials filter:
Lefebvre C, Manheimer E, Glanville J. Chapter 6: searching for studies. Box 6.4.c: Cochrane Highly sensitive search strategy for identifying randomized controlled trials in Medline: Sensitivity-maximizing version (2008 version); OVID format. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org

Embase (OvidSP): 1980-2011/wk20
Searched 24.5.11

1 Random$.tw. or clinical trial$.mp. or exp treatment-outcome/ (1838287)
2 animal/ (1657862)
3 animal experiment/ (1437511)
4 (rat or rats or mouse or mice or murine or rodent or rodents or hamster or hamsters or pig or pigs or porcine or rabbit or rabbits or animal or animals or dogs or dog or cats or cow or bovine or sheep or ovine or monkey or monkeys).mp. (4733080)
5 or/2-4 (4733080)
6 exp human/ (12389443)
7 human experiment/ (289060)
8 or/6-7 (12390826)
9 5 not (5 and 8) (3798193)
10 1 not 9 (1727986)
11 juvenile rheumatoid arthritis/ (10736)
12 systemic$.ti,ab,ot. (320228)
13 11 and 12 (2057)
14 (sJIA or SoJIA or (Systemic adj3 JIA)).mp. (340)
15 (Systemic adj3 (Child$ or juvenile$ or p?ediatr$) adj3 arthriti$).mp. (687)
16 ((Child$ or juvenile$ or p?ediatr$) adj4 Idiopathic Arthrit$).mp. (2521)
17 (Systemic$ adj4 (JRA or JCA)).mp. (252)
18 ((Child$ or juvenile$ or p?ediatr$) adj4 Still$ disease).mp. (117)
19 (systemic$ adj3 arthrit$).mp. (4586)
20 Child/ or boy/ or girl/ or hospitalized child/ or preschool child/ or school child/ or toddler/ or exp Adolescent/ or handicapped child/ (1862442)
21 (child$ or kid or kids or preschool$ or pre school$ or toddler$ or preteen$ or pre teen$ or preadolescen$ or adolescen$ or p?ediatric$ or prepubescent or prepubert$ or pre pubescent or pre puberty or school age$ or schoolage$ or young person$ or young people or p?ediatr$ or juvenile$ or teenage$ or teen or teens).ti,ab,ot. (1190042)
22 or/20-21 (2262235)
23 19 and 22 (1113)
24 or/13-18,23 (4416)
25 methotrexate/ (100846)
26 (Methotrexate or MTX or mexate or amethopterin or 59-05-2).af. (104887)
27 (Abatacept or Orencia or 332348-12-6 or bms-188667 or bms188667 or CTLA4-Ig or CTLA4Ig).af. (2934)
28 recombinant interleukin 1 receptor blocking agent/ (2709)
29 (Anakinra or Kineret or 143090-92-0).af. (1923)
30 (infliximab or remicade or 170277-31-3 or avakine or revellex).af. (19177)
31 (Etanercept or Enbrel or 185243-69-0 or 200013-86-1 or tnr001 or tnr-001).af. (12590)
32 (Adalimumab or humira or trudexa or 331731-18-1 or D2E7 or d2-e7).af. (7946)
33 or/25-32 (121454)
34 10 and 24 and 33 (589)
35 limit 34 to embase (556)

Trials filter:
Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE: increased specificity filter. Journal of the Medical Library Association 2006;94(1):41-7.

Cochrane Central Register of Controlled Trials (CENTRAL Issue 2:2011) (Internet): all dates Searched 24.5.11

#1 MeSH descriptor Arthritis, Juvenile Rheumatoid, this term only 168
#2 systemic*:ti,ab 15894
#3 (#1 AND #2) 18
#4 (sJIA or SoJIA):ti,ab,kw 0
#5 (Systemic near/3 JIA):ti,ab,kw 2
#6 (Systemic near/3 (Child* or juvenile* or pediatr* or paediatr*) near/3 arthriti*):ti,ab,kw 6
#7 ((Child* or juvenile* or pediatr* or paediatr*) near/4 Idiopathic Arthrit*):ti,ab,kw 82
#8 (Systemic* near/4 (JRA or JCA)):ti,ab,kw 5
#9 ((Child* or juvenile* or pediatr* or paediatr*) near/4 Still* disease):ti,ab,kw 19
#10 (systemic* near/3 arthrit*):ti,ab,kw 39
#11 MeSH descriptor Child explode all trees 13
#12 MeSH descriptor Adolescent, this term only 69451
#13 (child* or kid or kids or preschool* or pre school* or toddler* or preteen* or pre teen* or preadolescen* or adolescen* or p?ediatric* or prepubescent or prepubert* or pre pubescent or pre puberty or school age* or schoolage* or young person* or young people or pediatr* or paediatr* or juvenile* or teenage* or teen or teens):ti,ab 53540
#14 (#11 OR #12 OR #13) 108508
#15 (#10 AND #14) 13
#16 (#3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #15) 117
#17 Methotrexate or MTX or mexate or amethopterin or 59-05-2 4633
#18 Abatacept or Orencia or 332348-12-6 or bms-188667 or bms188667 or CTLA4-Ig or CTLA4Ig 71
#19 Anakinra or Kineret or 143090-92-0 49
#20 infliximab or remicade or 170277-31-3 or avakine or revellex 588
#21 Etanercept or Enbrel or 185243-69-0 or 200013-86-1 or tnr001 or tnr-001 477
#22 Adalimumab or humira or trudexa or 331731-18-1 or D2E7 or d2-e7 242
#23 (#17 OR #18 OR #19 OR #20 OR #21 OR #22) 5455
#24 (#16 AND #23) 35
#25 (#16 AND #23) 22 *limited to CENTRAL

Clinical trials.gov (Internet)
Searched 25.5.11
http://www.clinicaltrials.gov
Advanced search option

Search terms	Condition	Results
(Methotrexate OR MTX OR mexate OR amethopterin OR 59-05-2) AND (Arthritis AND systemic)	-	15
(Abatacept OR Orencia OR 332348-12-6 OR bms-188667 OR bms188667 OR CTLA4-Ig OR CTLA4Ig) AND (Arthritis AND systemic)	-	4
(Anakinra OR Kineret OR 143090-92-0) AND (Arthritis AND systemic)	-	4
(Etanercept OR Enbrel OR 185243-69-0 OR 200013-86-1 OR tnr001 OR tnr-001) AND (Arthritis AND systemic)	-	4
(Adalimumab OR humira OR trudexa OR 331731-18-1 OR D2E7 OR d2-e7) AND (Arthritis AND systemic)	-	7
Sjia OR sojia	-	11
Systemic* AND jia	-	14
Systemic AND (Child* OR juvenile* OR pediatr* OR paediatr*) AND arthritis	-	75
idiopathic AND (Child* OR juvenile* OR pediatr* OR paediatr*) AND arthritis	-	65

Child* OR juvenile* OR pediatr* OR paediatr*	Still* AND disease	9
Total		208

metaRegister of Controlled Trials (Internet)
Searched 26.5.11
http://www.controlled-trials.com/
Advanced search option

Search terms	**Results**
(Abatacept OR Orencia OR 332348-12-6 OR bms-188667 OR bms188667 OR CTLA4-Ig OR CTLA4Ig) AND (Arthritis AND systemic)	28
(Anakinra OR Kineret OR 143090-92-0) AND (Arthritis AND systemic)	31
(Etanercept OR Enbrel OR 185243-69-0 OR 200013-86-1 OR tnr001 OR tnr-001) AND (Arthritis AND systemic)	87
(Adalimumab OR humira OR trudexa OR 331731-18-1 OR D2E7 OR d2-e7) AND (Arthritis AND systemic)	66
Sjia OR sojia	6
Systemic* AND jia	16
Systemic AND (Child* OR juvenile* OR pediatr* OR paediatr*) AND arthritis AND idiopathic	43
idiopathic AND (Child* OR juvenile* OR pediatr* OR paediatr*) AND arthritis	90
(Child* OR juvenile* OR pediatr* OR paediatr*) AND (Still* AND disease)	0
Total	**367**

WHO International Clinical Trials Registry Platform (WHO ICTRP) (Internet)
Searched 26.5.11
http://www.who.int/ictrp/en/
Advanced search option
– Recruitment status = ALL

Title search	**Results**
(Abatacept OR Orencia OR 332348-12-6 OR bms-188667 OR bms188667 OR CTLA4-Ig OR CTLA4Ig) AND (Arthritis AND systemic)	36
(Anakinra OR Kineret OR 143090-92-0) AND (Arthritis AND systemic)	12
(Etanercept OR Enbrel OR 185243-69-0 OR 200013-86-1 OR tnr001 OR tnr-001) AND (Arthritis AND systemic)	77
(Adalimumab OR humira OR trudexa OR 331731-18-1 OR D2E7 OR d2-e7) AND (Arthritis AND systemic)	64
Sjia OR sojia	8
Systemic* AND jia	9
Condition search	**-**
Systemic AND (Child* OR juvenile* OR pediatr* OR paediatr*) AND arthritis AND idiopathic	9
idiopathic AND (Child* OR juvenile* OR pediatr* OR paediatr*) AND arthritis	53
(Child* OR juvenile* OR pediatr* OR paediatr*) AND (Stills AND disease)	0
Total	**268**

EU Clinical Trials Register (EUCTR) (Internet)
Searched 26.5.11
https://www.clinicaltrialsregister.eu/
Advanced search option

Search terms	**Results**
(Abatacept OR Orencia OR 332348-12-6 OR bms-188667 OR bms188667 OR CTLA4-Ig OR CTLA4Ig) AND (Arthritis AND systemic)	44
(Anakinra OR Kineret OR 143090-92-0) AND (Arthritis AND systemic)	50
(Etanercept OR Enbrel OR 185243-69-0 OR 200013-86-1 OR tnr001 OR tnr-001) AND (Arthritis AND systemic AND idiopathic)	11
(Adalimumab OR humira OR trudexa OR 331731-18-1 OR D2E7 OR d2-e7) AND (Arthritis AND systemic AND idiopathic)	11
Sjia OR sojia	6
Systemic* AND jia	14
idiopathic AND systemic AND arthritis ***limited to Under 18**	12
(Child* OR juvenile* OR pediatr* OR paediatr*) AND (Stills AND disease)	0
Total	**148**

Appendix 2: Phillips et al. Checklist

Results of assessing the manufacturers report based on the checklist by Phillips et al.

1. **Is there a clear statement of the decision problem?**
   Yes, the decision problem is clearly stated.
2. **Is the objective of the evaluation and model specified and consistent with the stated decision problem?**
   In the MS the included trials had inclusion criteria broader than that defined in the decision problem. In particular, they were not limited to form the two relevant populations of inadequate response to NSAID(s) and corticosteroids or NSAID(s), corticosteroids and methotrexate. Whilst the inclusion criteria for TENDER would suggest population 1 (inadequate response to NSAIDs and CS) the MS argues that the TENDER trial population is actually equivalent to population 2 (inadequate response to NSAIDs, CS and MTX). According to the MS, "70% of patients (all with prior inadequate response to NSAIDs and corticosteroids) at baseline were still receiving methotrexate, yet had active disease (a further entry requirement for the study), thus could be considered to be failing on methotrexate alone." (p. 39)
   At best this inference means TENDER does not address population 1, at worst the inference made about population 2 is unreliable and neither population is addressed.
3. **Is the primary decision-maker specified?**
   The term is not used, but implicitly the NHS is assumed
4. **Is the perspective of the model stated clearly?**
   Yes, it is the perspective NHS.
5. **Are the model inputs consistent with the stated perspective?**
   Yes.
6. **Has the scope of the model been stated and justified?**
   No
7. **Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?**
   No, it does not include adverse events and it does not present sufficient evidence for all the populations described in the decision problem.
8. **Is the structure of the model consistent with a coherent theory of the health condition under evaluation?**
   No, the model assumes no patient heterogeneity in disease burden/disease manifestation within a health state. It does also does not account for adverse events
9. **Are the sources of data used to develop the structure of the model specified?**
   Yes
10. **Are the causal relationships described by the model structure justified appropriately?**
    Yes
11. **Are the structural assumptions transparent and justified?**
    Yes, the model structure is transparent. But the use of four ACR response health states (which are relative health states) instead of modelling the CHAQ score, which is the central parameter of the utility calculations, is not sufficiently justified. The answer in the clarification letter response did not address this issue sufficiently. The model is validated by expert opinion, however, their opinions are not included in the submission.

12. **Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?**
    No, the patient heterogeneity is not sufficiently modelled. Health states are not mutually exclusion, causing problems with assigning meaningful cost and utility estimates to the health states. Adverse events cannot be included when more evidence becomes available.
13. **Is there a clear definition of the options under evaluation?**
    Yes, there is an explanation for the sequence of treatments. But it is not discussed if the sequence could be varied and how it affects the model outcomes.
14. **Have all feasible and practical options been evaluated?**
    No, not all possible sequences of the treatments have been varied.
15. **Is there justification for the exclusion of feasible options?**
    Yes, it is justified with prevailing clinical practice and licensing of the other treatment options.
16. **Is the chosen model type appropriate given the decision problem and specified causal relationships within the model?**
    No, the main problem is that the model uses relative response as health states. This leads to health states that are not mutually exclusive, i.e. depending on the health status at baseline, patients with the same health status at 12 weeks may be considered non-response, ACR30 response or better. Thus, it does not account sufficiently for heterogeneity in the patient population. Furthermore, it assumes constant transition probabilities, i.e. independent of age, time, or prior treatment, leading to a narrow and rigid model.
17. **Is the time horizon of the model sufficient to reflect all important differences between options?**
    Yes, it is until patients become adults, but the model allows up to 30years simulation duration.
18. **Are the time horizon of the model, the duration of treatment and the duration of treatment effect described and justified?**
    Yes, although there is a mismatch between duration of treatment (12weeks) and the corresponding cycle length (13weeks). The effects of the treatment are to be constant. However, this mitigated because no constant efficacy is assumed, that is, treatments lose their efficacy over time.
19. **Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?**
    Only to some extend; they reflect the biological process but only as disease progression. The disease states are based on improvements and not on the actual health states, which is preferable in Markov models, in particular in this case as patient heterogeneity is large. Adverse events are not modelled.
20. **Is the cycle length defined and justified in terms of the natural history of disease?**
    No, it is justified by the length of the clinical trial.
21. **Are the data identification methods transparent and appropriate given the objectives of the model?**
    Yes
22. **Where choices have been made between data sources, are these justified appropriately?**
    Yes
23. **Has particular attention been paid to identifying data for the important parameters in the model?**
    Only to some extent. For the indirect comparisons, only a rapid review was done of the literature instead of a full systematic review.

24. **Has the quality of the data been assessed appropriately?**
    Yes
25. **Where expert opinion has been used, are the methods described and justified?**
    To some extent. Some information is given about the elicitation process, but more details would have been welcome.
26. **Is the data modelling methodology based on justifiable statistical and epidemiological techniques?**
    No, the initial CHAQ score which is directly mapped into utilities show strong variation; this has not been accounted for properly.
27. **Is the choice of baseline data described and justified?**
    Yes. However, for the comparator, data from an infliximab study has been applied to all TNF-inhibitors which may not be realistic. Additionally, correction factors have been derived from an etanercept trial, and again applied to all TNF-inhibitors. Given the fact that according to the manufacturer infliximab itself is not a viable treatment option because of lack of evidence, the quality of the comparator data is somewhat questionable.
28. **Are transition probabilities calculated appropriately?**
    No, transition probabilities are unadjusted for personal characteristics and considered constant over time. This might be due to lack of data.
29. **Has a half-cycle correction been applied to both cost and outcome?**
    Yes.
30. **If not, has this omission been justified?**
    N/A.
31. **If relative treatment effects have been derived from trial data, have they been synthesised using appropriate techniques?**
    Yes
32. **Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified?**
    Yes
33. **Have alternative extrapolation assumptions been explored through sensitivity analysis?**
    Yes.
34. **Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified?**
    Yes
35. **Have alternative assumptions regarding the continuing effect of treatment been explored through sensitivity analysis?**
    Yes, the withdrawal risk has been altered.
36. **Are the costs incorporated into the model justified?**
    No, the treatment costs are justified, the health state costs are based on expert opinion, and show a remarkable difference between costs of non-responders and the costs of responders. The ERG is sceptical about these values.
37. **Has the source for all costs been described?**
    Yes, partly based on expert opinion.
38. **Have discount rates been described and justified given the target decision-maker?**
    Yes
39. **Are the utilities incorporated into the model appropriate?**
    No, the utilities have been assigned to a health state. Only five health states are modelled. The variation within a health state is ignored. Furthermore, the (dis)utility of adverse events cannot be modelled as adverse events are not modelled.

40. **Is the source for the utility weights referenced?**
    Yes, it is based on the CHAQ score of the clinical trial and then mapped into HRQL scores. This is because of lack of better data.
41. **Are the methods of derivation for the utility weights justified?**
    Yes, the mapping procedure is clearly described and under these circumstances justified.
42. **Have all data incorporated into the model been described and referenced in sufficient detail?**
    Yes.
43. **Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)?**
    N/A
44. **Is the process of data incorporation transparent?**
    Yes
45. **If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?**
    No, they have been described but are not justified sufficiently.
46. **If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?**
    Yes
47. **Have the four principal types of uncertainty been addressed?**
    No
    Methodological uncertainty is not discussed at all.
    Structural uncertainty is only explored superficially, by varying number and sequence of selected treatments.
    Heterogeneity: insufficient analysis of sub-groups.
    Parameter uncertainty has been assessed in the PSA but distributions are not sufficiently discussed.
48. **If not, has the omission of particular forms of uncertainty been justified?**
    No
49. **Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?**
    No. Methodological uncertainty, i.e. gauging the importance or uncertainty of particular analytical steps, has not been assessed sufficiently.
50. **Is there evidence that structural uncertainties have been addressed via sensitivity analysis?**
    Yes, some alternative scenarios have been run for different time horizons and ages. But this has not been explored sufficiently.
51. **Has heterogeneity been dealt with by running the model separately for different subgroups?**
    No, only starting age has been varied.
52. **Are the methods of assessment of parameter uncertainty appropriate?**
    Yes.
53. **If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?**
    No. Clearly stated, but not justified, in most cases based on expert opinion. A key parameter is calculated wrongly, clearly biasing the results.

54. **Is there evidence that the mathematical logic of the model has been tested thoroughly before use?**
    Some, the model was reviewed by a modelling expert and some extreme value analysis was done.
55. **Are any counterintuitive results from the model explained and justified?**
    N/A
56. **If the model has been calibrated against independent data, have any differences been explained and justified?**
    N/A
57. **Have the results of the model been compared with those of previous models and any differences in results explained?**
    No prior models have been discussed.

ERRATUM

Following the Factual Error Check from the manufacturer, the following pages have been revised:

Page 12: The following sentence has been removed: "However, it should be noted that NICE guidance on etanercept for JIA is for all subtypes of JIA."

Page 24: In section 4.1.1.1 (p. 24), the ERG states that 'conference paper' is not an Embase publication type, hence there is an error in the search term. However, Roche applied the RCT filter used for BMJ clinical evidence, which includes 'conference paper' as an Embase publication type. Therefore, page 24 has been revised by removing the statement.

Page 44: In Table 4.11 the 11th line which reads: 'DMARD and/or biologic agent no. patients', is incorrect for the all tocilizumab group. [redacted] This has been corrected.

2 BACKGROUND

2.1 Critique of manufacturer's description of underlying health problem.

Does the ERG believe that the manufacturer's description of the underlying health problem is appropriate and relevant to the decision problem under consideration?

The manufacturer's description of the underlying health problem is in line with NICE guidance [1], and hence seems reasonable and relevant to the decision problem. For completeness the following is reproduced from the MS:

"Juvenile idiopathic arthritis (JIA) is a term that covers a heterogeneous group of syndromes in which the onset of inflammatory arthritis occurs before the age of 16 years and lasts for more than 6 weeks. JIA is characterised by persistent joint swelling, pain and limitation of movement. The cause of JIA is poorly understood, but may relate to genetic and environmental factors" (MS, page 20).

"A classification system for JIA has been developed by the International League of Associations for Rheumatology (ILAR). There are seven categories of JIA: systemic, oligo arthritis (formerly pauciarticular), polyarthritis rheumatoid factor positive, polyarthritis rheumatoid factor negative, enthesitis related arthritis, psoriatic arthritis and unclassified (types that do not correspond to any, or to more than one, category) (Petty et al., 2004). The clinical manifestations and severity of the different sub-types varies considerably. sJIA is a multiorgan disease characterised by arthritic symptoms, fever, transient rash, liver and spleen enlargement. It is distinct from other subtypes and is often resistant to treatment. The overall outcome of the disease is poor with a high risk of long-term functional impairment. Macrophage activation syndrome (MAS) is a severe, life threatening complication to sJIA which affects around 7% of children, which is associated with serious morbidity and sometimes death (Yokota et al., 2010)" (MS, page 20).

"JIA is a relatively rare disease, with an estimated incidence in the UK of 0.1 per 1000 children per year, equivalent to 1000 children diagnosed per year. The prevalence is in the order of 1 per 1000 children, and about 10,000 children in the UK are affected. Approximately 10% of children diagnosed with JIA have systemic disease. Of these patients, those who have had an inadequate response to NSAIDs and corticosteroids and are 2 years of age and older will be eligible for Tocilizumab treatment." (MS, page 21).

2.2 Critique of manufacturer's overview of current service provision

Does the ERG believe that the manufacturer's overview of current service provision is appropriate and relevant to the decision problem under consideration?

The ERG broadly agrees with the manufacturer's description of current service provision. The MS states that there are no specific NICE guidance documents or national protocols for the treatment of sJIA and in addition there are no licensed therapies for the treatment of sJIA.

The British Society for Paediatric and Adolescent Rheumatology recommends treatment for JIA within multidisciplinary teams including paediatric rheumatologist, paediatric rheumatology clinical nurse specialist, ophthalmologist, general practitioner, paediatric physiotherapist, paediatric clinical psychologist, paediatric occupational therapist, podiatrist (Davies et al 2010). Drug therapy for sJIA

typically begins with systemic corticosteroids to treat systemic symptoms. Later in the disease, the systemic features can be mild / absent and at that stage steroid joint injections are often helpful (and

Search strategy for section 5.8, Indirect and mixed treatment comparisons

The MS reported searches of all the required databases: Medline, Medline In-Process, Embase and the Cochrane Library. Medline, In-Process and Embase searches were undertaken using the Datastar host, appropriate date spans, the date of searching and the full search strategies were reported in the MS.[10] Details of the Cochrane Library strategy were absent from the MS, and were requested by the ERG as part of the clarification process.[10] The clarification response[7] from the manufacturer stated the original Cochrane search had not been recorded on 28.3.11. The manufacturer ran additional searches of the Cochrane Library on 13.5.11 in order to provide that search strategy in the clarification response.[8]

The Medline, In-Process and Embase searches were presented as individual Datastar strategies which were clearly structured into population and intervention facets with the addition of a study design filter. The comparator interventions were identified as etanercept, anakinra, adalimumab and infliximab. Methotrexate (MTX) was not included in the indirect comparison searches.

The ERG considered methotrexate an important comparator intervention and was concerned that searches were not undertaken for studies of MTX in sJIA to allow for an indirect comparison of tocilizumab with MTX, therefore the ERG carried out additional searches which are described later in this section.

The same intervention search terms were applied to all searches, and consisted of the comparator drug's generic name in combination with the brand name, limited to the title and abstract fields. In order to make the searches more sensitive, additional synonyms, brand names and the CAS registry number could have been included, together with thesaurus index terms where available. As with the clinical effectiveness searches, all strategies would have benefit from the inclusion of more comprehensive synonyms for the intervention and population.

The Embase search strategy was presented first in the MS, and contained comprehensive variations of the disease terms. The first line of the search contained a potential typographical error which did not appear relevant to the topic i.e. ***(juvenile adj arthritis adj c adj '12').ab.*** This error would have impaired retrieval of records with 'juvenile arthritis' in abstract. The Embase search incorporated an RCT search filter and attempted to remove references to books, conference papers, editorials, letters and reviews from the retrieved results. The ERG noted a few areas of weakness in the RCT filter, the most important of which being the inclusion of '***retracted article'*** as a synonym for randomised controlled trial. The ERG was unclear why this term was included, as it did not appear to relate to any aspect of controlled trials or randomisation. The Embase RCT filter employed appeared to be a pragmatic collection of terms, limited solely to the title and abstract fields. The ERG felt that an objectively derived filter which incorporated relevant Emtree terms would have increased the sensitivity and relevance of the search results. Line 15 of the RCT filter attempted to remove various publication types combined Emtree terms from the results, by means of the Boolean operator 'NOT'. Unfortunately this attempt was not entirely successful as it appeared that line 15 was intended to search the Emtree ***Exp randomised controlled trial***, for example:

(book or conference adj paper or editorial or letter or review).p.t. not (exp adj randomised or randomized) adj controlled adj trial

OvidSP syntax was used, which failed to work in Datastar and resulted incorrect parentheses. The correct Datastar syntax should have applied, e.g.

(book or editorial or letter or review).pt. not Randomized-Controlled-Trial#.DE.

Woo et al 2000.[18] In the absence of the requested data from the manufacturer (individual data for tocilizumab without methotrexate and placebo without methotrexate from the TENDER trial) this was not possible. It should also be noted that data from Woo et al.[18] are probably minimal as most data are reported for children with sJIA and extended oligoarticular arthritis combined; in addition, outcomes from both trials may not be comparable.

The ERG have investigated heterogeneity within and across TENDER and ANAJIS trails. Inclusion criteria are similar for both trials. Table 4.11 presents baseline characteristics for TENDER and ANAJIS.

Table 4.11: Patient characteristics at baseline for TENDER and ANAJIS trials

					ANAJIS	
					Placebo n=12	**Anakinra n=12**
					8(67)	7(58)
					7.5 (3.73)	9.5 (5.19)
					3.2 (1.95)	4.2 (3.33)
					84 (65.74)	66 (64.4)
					57 (27.85)	44 (23.37)
					16 (15.84)	16 (13.12)
					17 (14.91)	16 (14.88)
					57 (29.74)	63 (20.57)
					55 (26.51)	50 (24.39)
					1.44 (0.625)	1.67 (0.845)
					11 (91.6)	8 (66.7)
					11 (91.6)	8 (66.7)